WILEY

HANDBOOK OF Measuring System DESIGN



EDITORS PETER H. SYDENHAM RICHARD THORN

Volume 3

Valve Exhibit 1077
Valve v. Immersion

VOLUME 1

Contributors.

Preface.

Abbreviations and Acronyms.

Introduction.

PART 1. FOUNDATIONS OF MEASURING.

SECTION 1. THE PROCESS OF MEASURING.

1. Sophistication of Measurement and its Body of Knowledge (Peter H. Sydenham).

2. Organization of Instrument Science (Joseph McGhee).

3. Measures and Metrics; Their Application (Peter H. Sydenham)

4. Economic Consideration of Measurement (Peter H. Sydenham).

5. Humans in the Real World (Joseph McGhee).

6. Substructure of Human-Machine Systems (Joseph McGhee).

SECTION 2. MEASURING THEORY AND PHILOSOPHY.

7. Introduction to Measurement Theory and Philosophy (Ludwik Finkelstein).

8. Formal Theory of Measurement (Ludwik Finkelstein).

9. Nature and Properties of Measurement (Ludwik Finkelstein).

10. Extensions of the Representational Theory of Measurement (Ludwik Finkelstein).

11. Measurement Theory in Physical, Social, and Psychological Science (Ludwik Finkelstein).

12. Fuzzy Approaches for Measurement (Eric Benoit, Laurent Foulloy and Gilles Mauris).

13. Signals, Information and Knowledge, and Meaning (Qing Ping Yang).

14. Hierarchical Aspects of Measurement Systems (Joseph McGhee).

15. Typical Measurement Systems Architecture (Joseph McGhee).

SECTION 3. ENVIRONMENTAL FACTORS.

16. Reduction of Influence Factors (Paul P.L. Regtien)

17. EMC and EMI (Kim R. Fowler).

SECTION 4. FEEDBACK IN MEASURING SYSTEMS.

18. Nature and Scope of Closed-loop Systems (Peter H. Sydenham).

19. Dynamic Behavior of Closed-loop Systems (Peter H. Sydenham).

20. Closed-loop Sampled Data Systems (Peter H. Sydenham).

21. Nonlinear Closed-loop Systems (Peter H. Sydenham).

SECTION 5. MESSAGING THEORY.

22. Characteristics and Theory of Knowledge (Luca P. Mari).

23. Principles of Semiotics as Related to Measurement (Luca P. Mari).

24. Principles of Epistemology as Related to Measurement (Timothy Lindsay John Ferris).

SECTION 6. SIGNAL THEORY.

25. Introduction to Signals in Physical Systems (Eugen Georg Woschni).

26. Signal Classification (Eugen Georg Woschni).

27. Signals in the Frequency Domain (Eugen Georg Woschni)).

28. Signals in the Time Domain (Eugen Georg Woschni).

29. Relationship Between Signals in the Time and Frequency Domain (Eugen Georg Woschni).

30. Statistical Signal Representations (Eugen Georg Woschni).

31. Discrete Signal Theory (Eugen Georg Woschni).

32. Geometrical Signal Representation (Eugen Georg Woschni).

33. Coding Theory and its Application to Measurement (Eugen Georg Woschni).

34. Modulation Theory (Eugen Georg Woschni).

SECTION 7. SYSTEMS THEORY.

35. Systems in the Time-Domain (Eugen Georg Woschni).

36. Systems in the Frequency Domain (Eugen Georg Woschni)

37. Relationship Between Systems in the Time and Frequency Domain (Eugen Georg Woschni).

38. Stability Issues (Eugen Georg Woschni).

SECTION 8. SOURCES OF INFORMATION ON MEASUREMENT.

39. Characteristics of Data, Information, Knowledge, and Wisdom (Timothy Lindsay John Ferris).

40. Sources of Information (Peter H. Sydenham).

41. Terminology and Classification of Measurement Systems (Peter H. Sydenham).

42. Information Databases of Relevance to Measurement (Peter H. Sydenham).

PART 2. UNITS, STANDARDS AND CALIBRATION.

SECTION 1. STANDARDS SUPPORTING MEASUREMENT.

43. Units (Brian W. Petley).

44. Types of Paper Standards and their Purpose (Halit Eren).

SECTION 2. CALIBRATION.

45. Calibration Process (Halit Eren).

46. Calibration Interval (Peter H. Sydenham).

47. Internet Calibration (Richard A. Dudley).

PART 3. ERROR AND UNCERTAINTY.

SECTION 1. ERROR AND UNCERTAINTY.

48. Common Sources of Errors in Measurement Systems (Dietrich Hofmann).

49. General Characterization of Systematic and Stochastic Errors (Martin Halaj).

50. Errors in Signal Systems (Eugen Georg Woschni).

51. Errors in Digital Signal Systems (Luca P. Mari).

52. Error Models, Error Budgets and their Calculation (Rudolf Palencár).

53. Calculation and Treatment of Errors (Joseph McGhee).

54. Explanation of Key Error and Uncertainty Concepts and Terms (Luca P. Mari).

55. Uncertainty Determination (Joseph McGhee).

PART 4. MEASURING SYSTEM BEHAVIOR.

SECTION 1. MEASURING SYSTEM SPECIFICATION.

56. Transfer Characteristics of Instrument Stages (Peter H. Sydenham).

57. Static Considerations of General Instrumentation (Peter H. Sydenham).

58. Description of Accuracy: Linearity, and Drift (Peter H. Sydenham).

59. Introduction to the Dynamic Regime of Measurement Systems (Peter H. Sydenham).

60. Zero-order System Dynamics (Peter H. Sydenham).

61. First-order System Dynamics (Peter H. Sydenham).

62. Second-order System Dynamics (Peter H. Sydenham).

VOLUME 2

PART 5. MEASURING SYSTEM DESIGN.

SECTION 1. ENGINEERING A MEASURING SYSTEM.

63. Outline of Systems Thinking (Peter H. Sydenham).

64. Executing A Measuring System Design (Peter H. Sydenham).

65. Life Cycle Concept (Floyd Guyton Patterson Jr.).

66. Phases of System Life Cycle (Kim R. Fowler).

67. Principle of Concept of Operations (ConOps) (Jack Ring).

68. Setting the System Boundaries (Joseph McGhee).

69. Requirements Allocation (Andrew Kusiak and Fang Qin).

SECTION 2. DESIGN METHODOLOGIES.

70. Measuring System Design Methodologies (Ludwik Finkelstein).

71. Modeling Methodology (Peter H. Sydenham).

72. Mathematical Methods of Optimization (Halit Eren).

SECTION 3. ELECTRONIC AND ELECTRICAL REGIME.

73. Overview of Electrical and Electronic Technique (Peter H. Sydenham).

74. Basic Electronic Components (Peter H. Sydenham).

75. Electronic System Building Blocks (Peter H. Sydenham).

76. Electronic Systems Design (Peter H. Sydenham).

77. Limits of Detection in Electronic Systems (Peter H. Sydenham).

78. Embedded Systems (Timothy Wilmshurst).

79. Testing Electronic Systems (Patrick D.T. O'Connor).

SECTION 4. FINE MECHANICAL REGIME.

80. Principles of Fine Mechanics – Kinematic and Elastic Designs (Peter H. Sydenham).

81. Principles of Fine Mechanics – Systems Considerations (Peter H. Sydenham).

82. Kinematical Regime – Members and Linkages (Peter H. Sydenham).

83. Kinematical Regime - Fasteners, Bearings (Peter H. Sydenham

84. 83. Kinematical Regime – Rotary Motion (Peter H. Sydenham

85. Elastic Regime of Design – Design Principles (Peter H. Sydenham).

86. Elastic Regime of Design – Spring Systems (Peter H. Sydenham).

87. Elastic Regime of Design – Plates and Bimorphs (Peter H. Sydenham).

88. Error Sources in Fine Mechanics (Peter H. Sydenham).

SECTION 5. VISIBLE RADIATION REGIME.

89. Optical Materials (Pak L. Chu).

90. Optical Elements (Pak L. Chu).

91. Light Sources and Detectors (Miroslaw Jonasz).

92. Optical Measuring Instruments (Peter H. Sydenham).

93. Testing Optical and Other Radiation Systems (Alan J. Cormier).

SECTION 6. HUMAN FACTORS ENGINEERIN.

94. Human Factors Engineering (Nicholas I. Beagley).

95. Human-Machine Interface (Nicholas I. Beagley).

96. The Domains of Human Factors Integration (Nicholas I. Beagley).

97. Design Methodology (Nicholas I. Beagley).

SECTION 7. QUALITY IN MEASURING SYSTEMS.

98. Reliability and Maintainability (Patrick D.T. O'Connor).

99. Safety Organization (Peter H. Sydenham).

100. Safety Analysis Methods (Peter H. Sydenham).

101. Assessing and Demonstrating Safety (Peter H. Sydenham).

102. Introduction to the Legal Process (Christopher Sweet).

103. Legal Liability Issues for Designers – A Case Study (Christopher Sweet).

PART 6. MODELING MEASURING SYSTEMS.

SECTION 1. MODELING MEASURING SYSTEMS.

104. Models of the Measurement Process (Luca P. Mari).

105. Modeling with LabVIEW™ (Wieslaw Ttaczala).

106. Virtual Instrumentation in Physics (Wieslaw Ttaczala).

PART 7. ELEMENTS: A – SENSORS.

SECTION 1. SENSOR FUNDAMENTALS.

107. Principles of Sensor Science (Joseph McGhee).

108. Transducer Fundamentals (Paul P.L. Regtien).

109. Structure and Energy in Sensor Systems (Joseph McGhee).

110. Signal/Energy Matrix Modeling (Joseph McGhee).

111. Classification of Sensors (Joseph McGhee).

112. Systematic Description of Sensors (Paul P.L. Regtien).

113. Force-feedback Sensors (Barry E. Jones).

SECTION 2. THE SENSING INTERFACE.

114. Models of the Sensor Interface (Qing Ping Yang).

115. Designing the Sensor Interface (Qing Ping Yang).

116. Selection of Sensors (Paul P.L. Regtien).

117. Materials in Measuring Systems (Peter H. Sydenham).

118. Ultrasonic Sensors (Peter J. Lesniewski).

119. Ultrasonic Instrumentation Principles (Lawrence C. Lynnworth).

120. Ultrasonic Instrumentation Design (Lawrence C. Lynnworth).

PART 8. ELEMENTS: B – SIGNAL CONDITIONING.

SECTION 1. ANALOG SIGNAL CONDITIONING.

121. Signals in the Presence of Noise (Richard Burdett).

122. Operational Amplifiers (Joseph McGhee).

123. Instrumentation Amplifiers (Joseph McGhee).

124. Specialized Amplifiers for Measurement Systems (Joseph McGhee).

125. Outline of Purpose of Analog Data Filters (Joseph McGhee).

SECTION 2. ELECTRICAL BRIDGES.

126. Electrical Bridge Circuits – Basic Information (Zygmunt L. Warsza).

127. Unbalanced DC Bridges (Zygmunt L. Warsza).

SECTION 3. AI SIGNAL PROCESSING TECHNIQUES>

128. Name and Scope of AI Techniques (Ajith Abraham).

129. Artificial Neural Networks (Ajith Abraham).

130. Rule-based Expert Systems (Ajith Abraham).

131. Evolution Computation (Ajith Abraham).

VOLUME 3

PART 9. ELEMENTS: C – DATA ACQUISITION AND PROCESSING SYSTEMS.

SECTION 1. DAS COMPONENTS.

132. Data Acquisition Systems (DAS) in General (Gerd Wöstenkühler).

133. Amplifiers and Filters for DAS (Gerd Wöstenkühler).

134. Analog Multiplexers (Gerd Wöstenkühler).

135. Sample-hold Circuits (Gerd Wöstenkühler).

136. Quantizing Theory Relevant to DAS (Gerd Wöstenkühler).

137. Coding for Data Converters (Gerd Wöstenkühler).

138. Sampling Theory Relevant to DAS (Gerd Wöstenkühler).

139. Analog-to-Digital (A/D) Converters (Gerd Wöstenkühler).

140. Integrating Type (A/D) Converters (Gerd Wöstenkühler).

141. Digital-to-Analog (D/A) Converters (Gerd Wöstenkühler).

SECTION 2. DIGITAL SIGNAL PROCESSING (DSP).

142. Z-transforms (Armar Bousbaine).

143. DFT and FFTs (Gerd Wöstenkühler).

144. DSP Chip Sets (Iain Paterson-Stephens).

145. DSP Tools (Iain Paterson-Stephens).

146. Principles of DSP Hardware Design (Iain Paterson-Stephens).

147. Ideal Digital Filters Approximation (Joseph McGhee).

148. General Performance of the Digital Filter (Joseph McGhee).

149. Low-, High-, and Band-pass Digital Filters (Joseph McGhee).

150. Finite Impulse Response (IIR) Digital Filters (Joseph McGhee).

151. Finite Impulse Response (FIR) Digital Filters (Joseph McGhee).

SECTION 3. COMPUTERS IN MEASURING SYSTEMS.

152. Fundamentals of the Stored Program Digital Computer (Joseph McGhee).

153. Single Address Instruction Microcomputer (Joseph McGhee).

154. Internal Operation of the Microprocessor (Joseph McGhee).

155. External Operation of the Microprocessor (Joseph McGhee).

156. Memory Management in the Microprocessor (Joseph McGhee).

157. Data Acceleration in Computers (Joseph McGhee).

158. Microcontroller Systems (Joseph McGhee).

159. Designing and Building Software for Measuring Systems (Joseph E. Kasser).

SECTION 4. INTELLIGENT MEASURING SYSTEMS.

160. Smart Sensor System Features (Peter H. Sydenham).

161. Knowledge-based Systems (Dietrich Hofmann).

PART 10. ELEMENTS: D – MEMS.

SECTION 1. MICRO ELECTRO MECHANICAL SYSTEMS (MEMS).

162. Principles of MEMS (Janusz Bryzek).

163. Uses and Benefits of MEMS (Janusz Bryzek).

164. Principles of MEMS Actuators (Janusz Bryzek).

PART 11. ELEMENTS: E – COMMUNICATION IN MEASURING SYSTEMS.

SECTION 1. DISTRIBUTED AND NETWORKED MEASURING SYSTEMS.

165. Introduction to Networked Instrumentation (Joseph McGhee).

166. Instrument Interconnection (Joseph McGhee).

167. Asynchronous and Synchronous Interface Protocols (Joseph McGhee).

168. RS 232 and EIA/TIA 232 Serial Interface (Joseph McGhee).

169. Voltage and Current Loop Transmission (Joseph McGhee).

170. IEEE-488 Instrumentation Bus (Joseph McGhee).

171. Local Area (LANs) and Wide Area Networks (WANs) (Joseph McGhee).

172. Fieldbus Systems (Halit Eren).

173. Scheduling Systems (Emil Michta).

PART 12. ELEMENTS: F – SIGNALS AND NOISE.

SECTION 1. NOISE AND INTERFERENCE.

174. Typical Signals Arising in Measurement (Eugen Georg Woschni).

175. Comparison of Analog and Digital Signal Handling (Joseph McGhee).

176. Signals and Signal-t0-noise Ratio (Richard Burdett).

177. Grounding and Shielding (Kim R. Fowler).

178. Noise Matching and Preamplifier Selection (Richard Burdett).

179. Input Connections; Grounding and Shielding (Richard Burdett).

SECTION 2. SIGNAL RECOVERY IN THE PRESENCE OF NOISE.

180. Bandwidth Reduction of Baseband DC Signals (Richard Burdett).

181. Amplitude Modulated Signals: The Lock-in Amplifier (Richard Burdett).

182. Boxcar and Signal Averages (Richard Burdett).

183. Correlators in Signal Extraction (Richard Burdett).

184. Photon Counting (Richard Burdett).

185. Pulse Height Discrimination, Ratemeters and Pileup (Richard Burdett).

186. The Family of Signal Recovery Methods (Richard Burdett).

PART 13. COMMON MEASURANDS.

SECTION 1. FLOW MEASUREMENT.

187. Flowmeter Selection and Application (Michael Reader-Harris).

188. Differential Pressure (DP) Flowmeters (Michael Reader-Harris).

189. Basic Principles of Flow Measurement (Richard Thorn).

190. Calibration and Standards in Flow Measurement (Richard Paton).

SECTION 2. DISPLACEMENT AND ANGLE MEASUREMENT.

191. Displacement and Angle Sensors Performance and Selection (Halit Eren).

192. Strain Sensors (Peter H. Sydenham).

193. Specialty Displacement and Angle Sensors (Halit Eren).

194. Large-scale Metrology (Stephen Kyle).

SECTION 3. TEMPERATURE MEASUREMENT.

195. Characteristics of Temperature Measurement (Joseph McGhee).

196. Thermocouple Temperature Sensors (Jacek Kucharski).

197. Metalic Resistance Temperature Detectors (RTDs) (Dietrich Hofmann).

198. Calibration and Standards in Temperature Measurement (D.R. White).

SECTION 4. TIME AND FREQUENCY.

199. Characteristics of Time and Frequency Measurement (Michael A. Lombardi).

200. Calibrations and Standards in Time Measurement (Michael A. Lombardi).

SECTION 5. ELECTRICAL QUANTITIES.

201. Voltage Measurement (Halit Eren).

202. Current Measurement (Halit Eren).

203. Resistance Measurement (Halit Eren).

204. Capacitance and Inductance Measurement (Consolatina Liguori).

SECTION 6. VELOCITY AND ACCELERATION.

205. Theory of Vibration Measurement (Peter H. Sydenham).

206. Practice of Vibration Measurement (Peter H. Sydenham).

207. Acceleration Measurement (Peter H. Sydenham).

208. Amplitude and Velocity Measurement (Peter H. Sydenham).

SECTION 7. CHEMICAL PROPERTIES.

209. Characteristics of Chemical Measurements (Peter H. Sydenham).

210. Optical Transducers for Chemical Measurements (Ashutosh Sharma).

211. Mass Spectrometry (Peter H. Sydenham).

212. Chromatography (Brett Paull).

213. Electrochemical Measurements (David Davey).

PART 14. TEST AND EVALUATION.

SECTION 1. MEASUREMENT TESTING SYSTEMS.

214. Accelerated Testing (Patrick D.T. O'Connor).

215. Automatic Test Systems (Patrick D.T. O'Connor).

216. Test Facilities (Patrick D.T. O'Connor).

217. Instrument Evaluation (Steve Cork).

Subject Index.

1: Sophistication of Measurement and its Body of Knowledge

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Sophistication of Measurement as the Degree of Science 5
2 Measurements and the Body of Knowledge 6
Related Articles 9
References 9

1 SOPHISTICATION OF MEASUREMENT AS THE DEGREE OF SCIENCE

The decision of whether to use existing, or to create new, measuring instruments in the study of a subject comes after measurable variables have been identified. The process is, in the physical sciences, usually considerably easier to realize than in many areas of the empirical sciences. Many stages of prior reasoning precede such a decision: this is not always recognized, especially in engineering. The process can be depicted by the chart given in Figure 1.

Knowledge seeking begins presumably because of certain inquisitive features of man's makeup that stirs up interests in directions that seem to have more relevance than others. The processes involved are complex, and, as yet, not adequately known. Paradoxically, it seems that a great deal of knowledge is used in a very general way from the onset to choose candidate paths of action to follow to gain the knowledge sought.

This process, which involves the cognitive elements of sensation, perception, apperception, advises the knowledge seeker that certain information is more relevant for study than other data. It appears that the biological senses involved provide data input to the brain, coding it with meaning to suit the required task. Two people viewing a plant leaf, for example, see the same object with similar senses, yet both could 'see' quite different attributes. Latent information available has begun to be filtered at this stage.

The assembled data is then sorted and classified according to various kinds of similarities to detect differences. Each group forms a crude measurement standard of comparison for the others. This process can often be continued until advanced knowledge is established without using measuring instruments. Linnaeus (1707–1778) was able to make a major contribution to botany by introducing his binomial classification system (see Figure 2). Darwin's *On Origin of the Species by Means of Natural Selection* of 1859 has been recognized as probably the greatest generalization yet – although gene mapping is taking that over. It was made from vast quantities of data that were all assembled with little use of measuring instruments to enhance man's natural senses.

At some stage, this qualitative form of science can be subjected to increasingly more quantitative methods. Attributes of the various classes became apparent in a way that allows instruments to be applied that give natural sensing, greater sensitivity, and greater power to move from a qualitative mode into the quantitative measurement mode. More detailed knowledge becomes available as measurements produce data that is referred against more adequate, precise, and accurate standards. Thus, it is that physical measuring instruments applied as the degree of science, which is reflected by the degree of quantification used, is improved. This line of reasoning also makes it

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

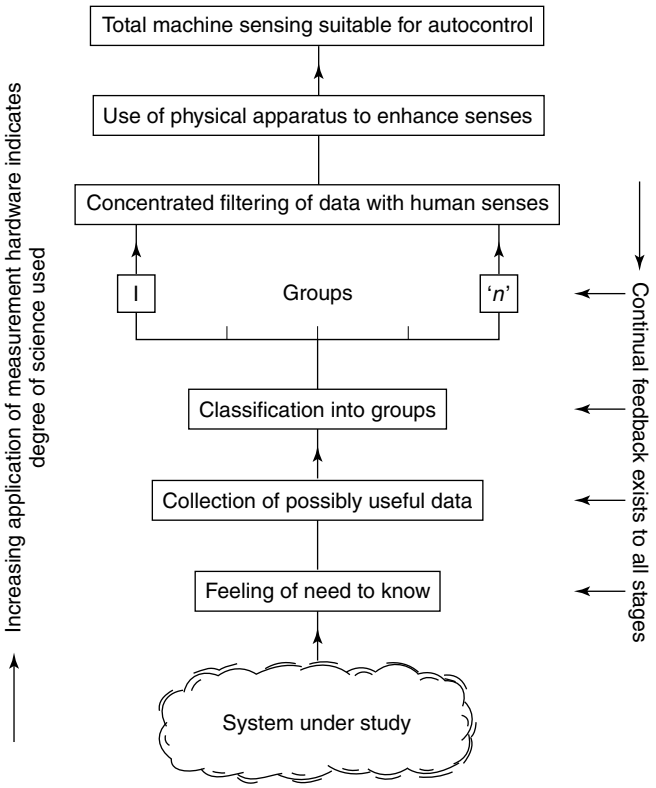


Figure 1. Simplified hierarchy of application of measuring instruments in the study of a problem.



Figure 2. Linneaus resting after a botanical ramble. He devised the binomial classification system now used, reporting it in 'Systema naturae', 1758. (Copyright © Uppsala University.)

vital that appreciation of the qualitative stages preceding proper measurement and the instruments that evolve are understood.

This sentiment is not new as the famous Lord Kelvin quote tells us. The words of Westaway (1937) are in sympathy with this. The concept expressed in Figure 1 is simplified: in practice, the stages at which hardware forms of measuring instruments are used varies widely. In some studies, they are needed at the very beginning.

Finkelstein (1975) sums up the situation in this way:

> Measurement presupposes something to be measured, and measures have to be conceived and sought before they can be found in experience. Both in the historical development and logical structure of scientific knowledge the formulation of a theoretical concept, or construct, which defines a class of entities and the relations among its members, providing a conceptual interpretation of the sensed world, precedes the development of measurement procedures and scales. It is necessary for instance to have some concept of 'degree of hotness' as a theoretical construct, interpreting the multitude of phenomena involving warmth, before one can conceive and construct a thermometer.
>
> As measurement procedures are developed, and knowledge resulting from measurement accumulates, the concept of the measured class becomes clearer and may to a substantial extent become identified with the operational procedures underlying the measurement process.
>
> In some cases the concept of an entity arises from the discovery of mathematical invariances in laws arrived at by measurement, and the entity is best thought of in such mathematical terms, but in general one attempts to arrive at some qualitative conceptual framework for it, if possible.

As the subject matter becomes better known and enables unattended sensing by an observer, as needed for control or monitoring purposes, the use of measuring instruments to enhance human senses may not be appropriate. Hardware sensors then totally replace man's senses.

It is logical, therefore, to expect all of man's endeavors that require measurements to be made (most of them!) to trend steadily toward greater use of measuring instruments. Certainly, time has proven this to be so. But this is also a consequence of man's method of survival on earth. Unlike lower animals, man has the ability to modify his environment to suit his biological structure. He does this usually by the use of technological developments, which rarely operate in the same way as natural equivalents or are made with the same materials. A comparison of natural and man-made vision sensors is given in Figure 3. The knowledge man possesses is being built up of a component about the natural world plus a component about the structures that man has created. Measuring instruments are the means by which man's creations operate and these too are creations of man. The relationship between measurement and knowledge has been explored (Sydenham, 2003).

2 MEASUREMENTS AND THE BODY OF KNOWLEDGE

The sum total of knowledge is termed the *body of knowledge*. As knowledge is a characteristic of man, not of

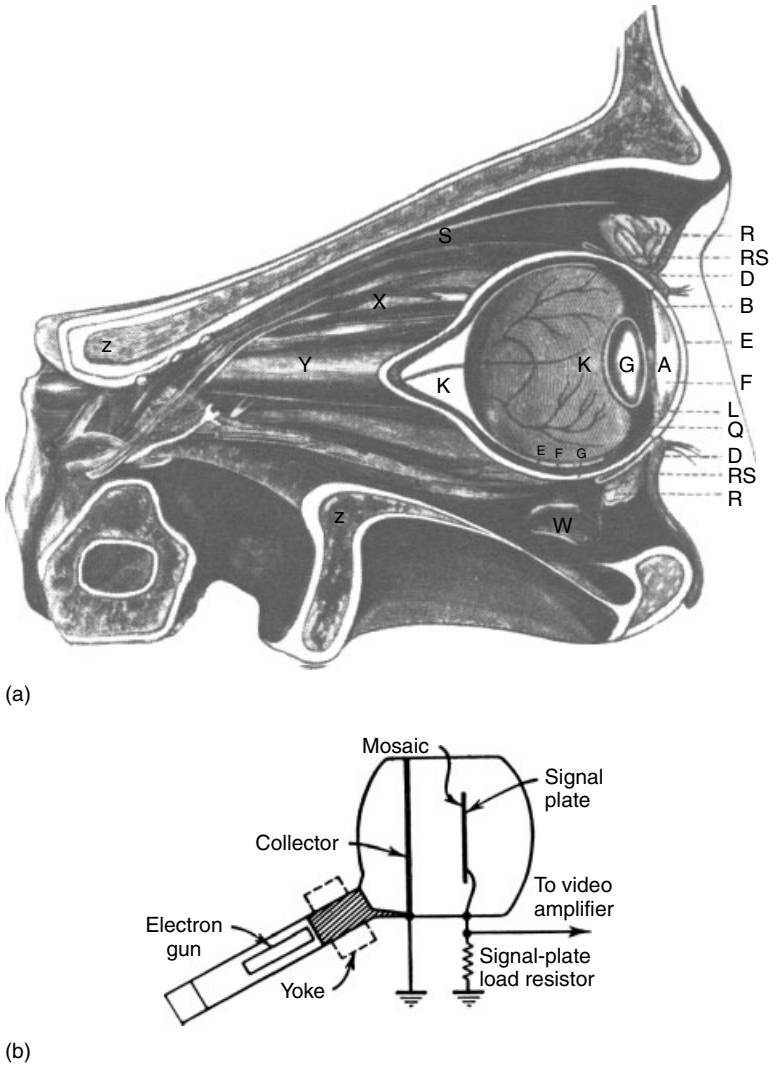


Figure 3. Man's creations generally use different materials and techniques as do natural systems. Here, imaging sensors are contrasted (a) Longitudinal section of eye (Reproduced from *Cyclopaedic Science*, Pepper J.B., (1874), Copyright Frederick Warne) and (b) RCA iconoscope – early form of television camera tube from Kloeffler (1949) (Courtesy RCA Ltd, USA).

existence, it began at zero magnitude and grew with time. No method has yet been devised to measure its magnitude in objective ways but it clearly is enlarged continuously with the passage of time.

It is formed of two groups: that about the natural world and that about the unnatural systems created by man. Man's creation grows, the natural world changes; the extent of the latent information available for conversion into knowledge therefore grows continuously.

As the body of knowledge grew, various workers of the past tried to summarize all that was known. Today, that must be recognized as an almost hopeless task. Collectively, all knowledge must be stored in a manner whereby it is retrievable. The danger of converting latent information back into another form of latent storage via the knowledge conversion state is real; what lies in the literature is not all recoverable in an easier manner than that by which it was first generated!

To retrieve knowledge, it is grouped into convenient classifications. Convenience is a term in which time of action is most important. The memory span of man, especially short term, is very limited, so it has been suggested (Harman, 1973) that major groupings usually total around seven. These in turn are subdivided, giving the various epistemological groups.

Measurements assist in gaining knowledge and knowledge, in turn, assists new forms of measurements to be conducted. A closed-loop mechanism can be observed in the development of measurements; Figure 4 depicts this.

Over the past few decades, the trend toward recognition of the interdisciplinary studies that replaced the specialisms that came to us previously has highlighted the

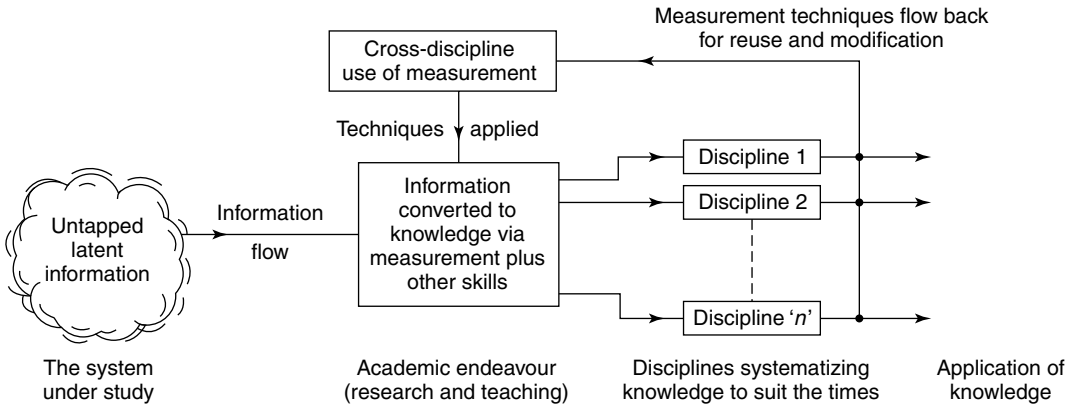


Figure 4. Relationship of measurement principles in ordering the body of knowledge.

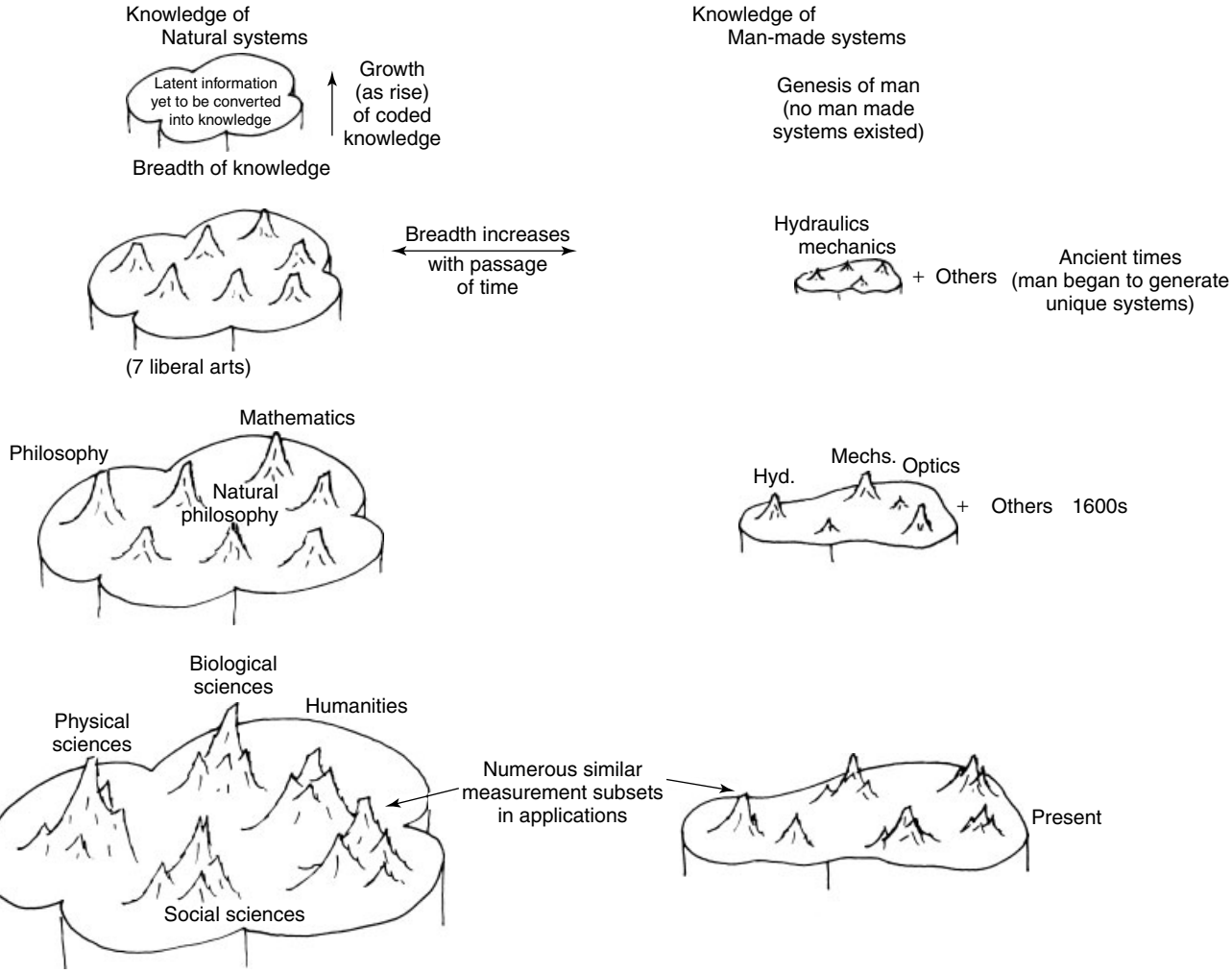


Figure 5. Epistemological mountains in the two plains of human knowledge. Measurement techniques are now duplicated on most contemporary mountains.

fact that not only does such a feedback process exist but it is also often duplicated (a needless waste of effort, therefore) and is often cross-fertilized between epistemological groups.

The Dewey cataloging system gives librarians a set of numerical codes, each having a linguistic description of what subject matter each number represents. Of over 40 000 numerical assignments, some 600 clearly relate to the

measuring process. These are distributed widely over the whole body of knowledge, as classified by that system. Pictorially, this means that most clusters of knowledge possess subclusters concerned with measurement method as depicted in Figure 5.

At present, information scientists – those people that work on the storage, coding, and retrieval of knowledge – consider that the major clusters are changing to reflect the interdisciplinary attitudes. New clusterings are emerging, one which may well be that of the relatively new discipline of measurement science, the pursuit of means to convert latent information into meaningful knowledge by rigorous and objective procedures of philosophy and practice.

RELATED ARTICLES

Article 2, Organization of Instrument Science, Volume 1; **Article 3, Measures and Metrics; Their Application, Volume 1**; **Article 4, Economic Considerations of Measurement, Volume 1**; **Article 5, Humans in the Real World, Volume 1**; **Article 6, Substructure of Human–Machine Systems, Volume 1**.

REFERENCES

Finkelstein, L. (1975) Fundamental Concepts of Measurement: Definition and Scales. *Measurement and Control*, **8**, 105–111.

Harman, G. (1973) *Human Memory and Knowledge*, Greenwood Press, London.

Sydenham, P.H. (2003) Relationship between Measurement, Knowledge and Advancement. *Measurement*, **34**(1), 3–16, Special Issue on Measurement foundations.

Westaway, F.W. (1937) *Scientific Method: Its Philosophical Basis and its Modes of Application*, Hillman-Curl, New York.

2: Organization of Instrument Science

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Definition of Instrument Science 10
2 The Need and Starting Point for Ordering in Instrument Science 11
3 How Instrument Science is Organized 12
4 Orders of Classification 12
Related Articles 14
References 14

1 DEFINITION OF INSTRUMENT SCIENCE

A science is an organized body of knowledge (Finkelstein, 1994). What then is Instrument Science? To answer this question, we must define what an instrument is. When posing the question, '*what is an instrument?*' (McGhee, Henderson and Sankowski, 1986), most people have a visceral feeling for the answer. According to the McGraw-Hill Encyclopaedia of Science and Technology, an instrument is a *SYSTEM, which refines, extends, or supplements* the *HUMAN faculties of sensing, observing, communicating, calculating, controlling, and perceiving*. In other words, instruments are human-made elements embedded within human-machine systems, which help humans to acquire information, by the process of sensing, and to handle data, by performing information handling operations. Using this definition as the key, an implicit use of taxonomy led to the proposal that ordering in instrumentation should involve functional and structural *reticulation* (McGhee, Henderson and Sankowski, 1986). This statement is similar to another definition by Peter Stein (1969) who asserted that *Measurement* combines '*INFORMATION transfer about*' and '*ENERGY transfer from*' a '*process*' using '*SYSTEMS,*' which are made up of '*components or TRANSDUCERS*' forming a '*STRUCTURE or network*'.

A definition that encompasses all of these ideas is given in Figure 1. The systemic nature of measuring instruments demands a holistic approach in design and analysis. It is apparent that the ordering of *information machines* depends upon the holistic relations among specific sensor *structures* performing diverse *functions* within different *energy* domains for the *acquisition, capture, communication*, or *distribution* of *information* in a variety of *signal* forms.

The diagrammatic summary definition given in Figure 1 is based upon the functions performed by measurement systems, the structures that allow them to perform the function, and the energy form from which the information is acquired. It may be regarded as the study of the methods and techniques of extending the human abilities to handle information using information machines. Since information is predominantly carried by signals, measurement is concerned with the acquisition, handling, analysis, and synthesis of signals in measuring instruments. It may also be considered as the measurement analogy of data communications.

To assist with the generalizations that make measurement scientific, it is essential to develop a unified metrological description of every constituent component making up a measurement system. A unified approach allows the evaluation of the metrological characteristics of each element. Thus, the formation and analysis of all contributory factors, and in particular, the measurement errors can be performed (Solopchenko, 1994). Signals, which are acquired using various forms of sensors, are handled using diverse forms of metrological components. These may be conditioners, amplifiers, and filters used in conjunction with suitable multiplexing methods.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

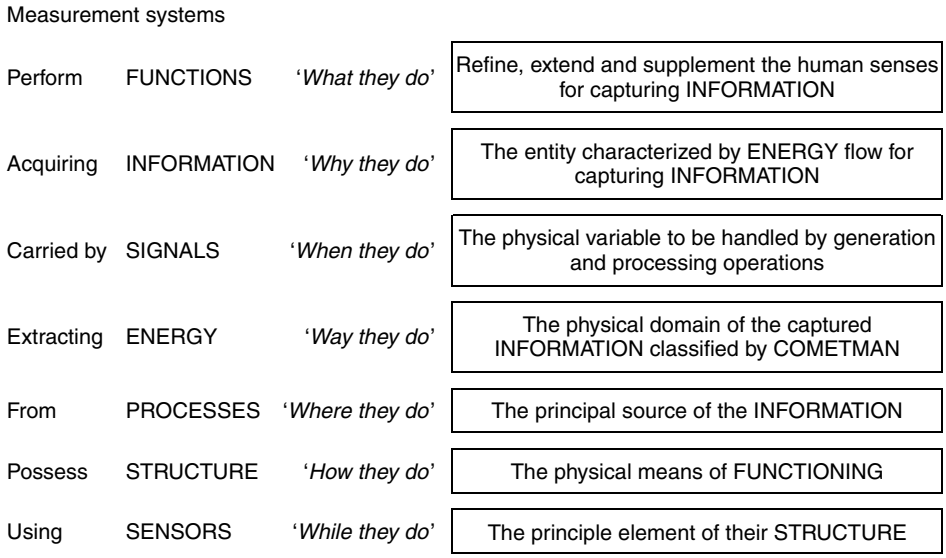


Figure 1. A substantive definition of measurement.

When instruments, which have the primary structure of systems, are viewed from this position, the field of Systems Science and Engineering (M'Pherson, 1980, 1981; Sandquist, 1985), with its related disciplines associated with large-scale systems, must play an important part in their exposition. This systems approach, which possesses holistic or totality features, offers a number of advantages. A principal benefit places instruments within a hierarchy of both systems and machines by structure, function, energy form, and information. McGhee, Henderson and Sankowski (1986) have stated that these aspects are revealed by the methods of reticulation or subdivision. As it happens, reticulation also reveals the places occupied by other types of subsystems within this hierarchy. Thus, advantages are accrued by using this approach in the study of instrumentation. Commencing from this standpoint, the systems approach is essential for the study of instrumentation. Some broad principles of Systems Engineering for instrumentation are adapted for the boundary view of human–machine systems in **Article 68, Setting the System Boundaries, Volume 2**.

2 THE NEED AND STARTING POINT FOR ORDERING IN INSTRUMENT SCIENCE

Every field of scientific activity requires organization or ordering. An essential starting point in the ordering of Instrument Science is the application of a relevant taxonomy (Flint, 1904; Durand, 1899; Broadfield, 1946; Körner, 1970; Knight, 1986; McGhee and Henderson, 1991; McGhee *et al*., 1996; McGhee and Henderson, 1993; Thomson, 1926) using objective methods to ensure that the ordering is justifiable. Such schemes of classification have been compared to nominal scales of measurement using an algebraic formulation (Watanabe, 1996). The following quotation (Knight, 1986) indicates the fundamental importance of classification in all of the applied sciences:

> We are apt to think of classification as a sort of 'natural history stage' through which all sciences pass in their youth before they grow into something handsomer, more mathematical and explanatory... classification is a highly theory-laden activity.... What one thinks one is classifying may make a big difference to the system of classificatory categories one uses.

It is apparent that classification is of basic importance for all activities in the applied sciences.

It has been noted that a taxonomy of Instrument Science will be erroneous if it is based upon its ends (McGhee and Henderson, 1993) as this will only lead to a cataloging of instruments. Indeed, only by organizing the constitution of the topic on the basis of contributory disciplines can Instrument Science be arranged according to its basic nature and inherent characteristics. Considering the nature and scope of the disciplines constituting the taxonomy, analysis, design, and utilization of instruments and instrument systems provides a clear view of the contributory disciplines of Instrument Science (Finkelstein, 1994; Finkelstein and Grattan, 1993, 1994; Measurement, 1994; Sydenham, 1982, 1983; Sydenham and Thorn, 1992) within Instrumentation and Measurement Technology (I&MT).

3 HOW INSTRUMENT SCIENCE IS ORGANIZED

Instrument science must be *holistic* by always using the '*whole-life-whole-system*' approach characterizing the *SYSTEMS ENGINEERING* method (M'Pherson, 1980; McGhee, Henderson and Sankowski, 1986; Sandquist, 1985). Thus, it is seen that instruments and instrument systems perform a diversity of information handling functions allowing the acquisition, capture, communication, processing, and distribution of information about the states of equilibrium and motion of solids, liquids, gases, and their constituent systems using a variety of physical sensing structures in different energy forms. McGhee and Henderson (1991) have suggested that this is the starting point, not only for ordering in Instrument Science but also as the fundamental context for ordering in all of the applied sciences.

The question then arises as to how the science of measurement should be organized into identifiable bodies of knowledge. A method for the organization of knowledge in the biosciences called *Taxonomy* or *Classification Science* provides the answer to this question. This method can be adapted for the organization of measurement. Observation and recording are the embodiment of the scientific method, which is of profound importance in the understanding and utilization of the physical universe and its resources. This aim is achieved through the measurement of the states of equilibrium and motion of solids, liquids, gases, and the systems they constitute (McGhee, Henderson and Sankowski, 1986). Instruments are the means by which these human faculties may be improved and supplemented (Finkelstein, 1994). However, the acquisition of information, or, more generally knowledge, requires some process of ordering or organization. In the case of instrumentation, this ordering of *information machines* depends upon the holistic relations between various instruments. The basic theoretical mechanism, which allows the organization, is the field of taxonomy or classification science. Although this science is well known in the biosciences, it is not so well known, or for that matter understood or applied, in the engineering sciences. This opinion has been expressed on a number of occasions in the references quoted in McGhee, Henderson and Sydenham (1999). It is well worthwhile to provide some basic information on the nature and scope of taxonomy for use in measurement. The systemic nature of instruments implies a holistic approach in their ordering.

Since the time of Plato and Aristotle, many attempts have been made to organize the sciences into hierarchical groupings. A scientific approach for the ordering of science is provided by *TAXONOMY* . Although this science has been used implicitly by bioscientists for centuries (Daly and Linsley, 1970), its intrinsic rules and principles were not studied deeply until the nineteenth-century French philosopher, Durand (De Gros), examined its constitution. Thus, a clear distinction is drawn between the ordered organization of the theory of Taxinomy (its original spelling) itself and its principal applications in a specific field. It has been claimed that the word Taxonomy (from the ancient Greek taxis meaning order) was first used by the seventeenth-century Swiss botanist Augustin Pyrame de Candolle (1778–1841).

What is the nature and scope of taxonomy or classification science? In the view of Durand, the most elementary form of all classification is the series that depends upon the increase or decrease of some *variable* of the scheme of ordering. Hence, any legitimate scheme of instrument classification must ensure that all of its divisions are always determined by one common principle. Instrument classification will thus be erroneous if it is based upon its ends, as this merely leads to a catalog of different kinds of instruments. Rather, instrumentation should always be arranged according to its basic nature, its inherent characteristics, and not upon anything lying outside itself. In other words, the science of classification in instrumentation is not about the sum of the ends of instrumentation but rather about coordinating the science of instruments in such a way as to give it an organized or systematized structure.

4 ORDERS OF CLASSIFICATION

The significant contribution Durand made to the science of taxonomy was the proposal that there are four principle orders or problems of classification. These orders are summarized in Table 1. In the *First Order,* described as *Generality* or *Resemblance*, is embodied what many other theorists of classification have called the 'likeness' of one thing with another thing. The *thing concept* is fundamental to the whole of categorical ordering, not just in bioscience. It is also important in the earth sciences (Von Engelhardt and Zimmermann, 1988) for the classification of minerals, in technology transfer (Zhao and Reisman, 1992), and in KNOWLEDGE ENGINEERING (KE) (Chandrasekaran and Goel, 1988; Gomez and Segami, 1991; Mill and Rada, 1990; Yasdi, 1991). Hence, this concept also has central importance in instrumentation. Likeness, of course, is that *relation* between several concrete things that unites them. Thus, the application of classification by zoologists and botanists in the discrimination between genera and species is a good example of the way in which the problem of generality and resemblance is approached.

In taxonomy, there is an important tendency to group things on the basis of their *Composition* or *Collectivity*. Durand distinguished this as the *Second Order* of taxonomy. While this order is concerned with the relationship of the part to the whole and vice versa, the *Third Order* of

Table 1. A summary of the four orders or problems of taxonomy.

Taxonomy, the science of classification (Putting *things* in a *scientific order*)		
Problem or Order	Definition and Aspects	Comment
Generality or resemblance	1. Concerned with the *likeness* of *separate things* 2. *Likeness* is that *relation* between *things* that *unites* them 3. The *thing concept* is fundamental to all *Categorical Ordering* (*i.e.* Taxonomy)	Also called the *Metaphysical Order* because terms are concerned with *theoretical* or *fictitious* things
Composition or collectivity	1. Concerned with the *relationship* of a *part* of a *thing* to the *whole thing*	All other orders are concerned with the actual things to be classified
Hierarchy	1. Concerned with the *relation* between *heads* or *central members* of *groups* of *things*	Related to the *order* of *composition/collectivity,* especially in the places occupied in each *order* relative to other things of the *same order*
Genealogy or evolution	1. Concerned with the *kinship* of *one thing* with some *other thing*	Hinges upon *notions* of *kinship* by *relationships* of • *ascent* • *descent* • *collaterality.*

taxonomy, called *Hierarchy*, takes account of the relation of rank between the heads or central members of groups of things. In their turn, these are related in the order of composition, but address each concrete thing in the assessment of the place it occupies in each order relative to the other constituents of the same order. Perhaps the most important *Fourth Order* in Durand's theory of taxonomy, especially in bioscience, is that known as *Genealogy* or *Evolution*. This order hinges upon the notions of kinship through the relations involved in the characteristics of *ascent, descent,* and *collaterality*. As with the orders of *Composition* and *Hierarchy, Genealogy* and *Evolution* are also concerned with the actual objects or events that are to be classified. Although there have been minor developments of this theoretical constitution of taxonomy, it is still fair to say that the basis laid by Durand has not been significantly altered. As this theory of taxonomy was formulated in the context of bioscience, it requires modification before being applied to instrumentation.

Another important aspect of taxonomy is the development of a system of nomenclature, which is unambiguous. In bioscience, the binomial nomenclature is due to the eighteenth-century Swedish botanist, Carolus Linnaeus (1707–1778). For example, in plant kingdom classification, the first category of the ordering is called a division. This is followed by subdivision followed by class, order, family, genus, species, and subspecies. It seems logical and convenient to use the same ordering for machine kingdom grouping, although it may cause some controversy. Adapting the basic phenetic and phyletic methods used by bioscientists allows functional and structural grouping in instrumentation. Phenetic discrimination uses similarity and difference in form or physical feature, while phyletic techniques are based upon evolutionary criteria.

A summary of taxonomy for instrumentation (McGhee and Henderson, 1989) points out that it has three objectives and three functions that emphasize its importance. Thus, the three objectives of classification are:

1. the concrete discrimination between different things;
2. the consensus regarding standards for the principles of description;
3. the bringing of order or systematization.

Similarly, the three functions of classification should allow

1. the organization of the means of communication and retrieval of the descriptions used;
2. the acquisition of new information in the extension of descriptions;
3. the highlighting of unifying factors between entities without diminishing the importance of any existing differences.

The materials of taxonomy in Instrument Science are the diverse types of instruments and their operating principles. Assembling the various instrument types is the main activity of classification in Instrument Science because it allows the possibility for further study. The grouping of instruments from the lowest levels of sensors into progressively larger groups so that a hierarchical ordering by function, structure, and energy form, constitute the final ingredients of discrimination and ordering in Instrument Science.

RELATED ARTICLES

Article 1, Sophistication of Measurement and its Body of Knowledge, Volume 1; **Article 6, Substructure of Human–Machine Systems, Volume 1**; **Article 7, Introduction to Measurement Theory and Philosophy, Volume 1**; **Article 14, Hierarchical Aspects of Measurement Systems, Volume 1**; **Article 22, Characteristics and Theory of Knowledge, Volume 1**; **Article 63, Outline of Systems Thinking, Volume 2**; **Article 104, Models of the Measurement Process, Volume 2**; **Article 107, Principles of Sensor Science, Volume 2**.

REFERENCES

Broadfield, A. (1946) *The Philosophy of Classification*, Grafton and Co., London.

Chandrasekaran, B. and Goel, A. (1988) From Numbers to Symbols to Knowledge Structures: Artificial Intelligence Perspective on the Classification Task. *IEEE Transactions on Systems, Man and Cybernetics*, **18**(3), 415.

Daly, H.V. and Linsley, E.G. (1970) Taxonomy, in *Encyclopaedia of the Biological Sciences*, 2nd edn (ed. P. Gray), Van Nostrand Reinhold, New York (p. 920).

Durand (De Gros), J.P. (1899) in *Aperçus de Taxinomie Générale* (ed. F. Alcan), Paris.

Finkelstein, L. (1994) Measurement and Instrumentation Science – An Analytical Review. *Measurement*, **14**(1), 3–14.

Finkelstein, L. and Grattan, K.T.V. (eds) (1993) *State and Advances of Measurement and Instrumentation Science*, *Proc of IMEKO TC1/TC 7 Colloquium*, City University, London.

Finkelstein, L. and Grattan, K.T.V. (1994) *Concise Encyclopaedia of Measurement and Instrumentation*, Pergamon, Oxford.

Flint, R. (1904) *Philosophy as Scientia Scientiarum and A History of Classification of the Sciences*, William Blackwood & Sons, Edinburgh.

Gomez, F. and Segami, C. (1991) Classification Based Reasoning. *IEEE Transactions on Systems, Man and Cybernetics*, **21**(3), 644.

Henderson, I.A. and McGhee, J. (1993) Classical Taxonomy: An Holistic Perspective of Temperature Measuring Systems and Instruments. *Proceedings of IEE-A*, **140**(4), 263.

Knight, D. (1986) Physics and Chemistry in the Modern Era, in *The Physical Sciences Since Antiquity* (ed. R. Harre), Croom Helm, Beckenham.

Körner, S. (1970) *Categorical Frameworks*, Basil Blackwell, Oxford.

McGhee, J. and Henderson, I.A. (1989) Holistic Perception in Measurement and Control: Applying Keys Adapted from Classical Taxonomy. *IFAC Proceedings of Series*, (5), 31.

McGhee, J. and Henderson, I.A. (1991) The Nature and Scope of Taxonomy in Measurement Education. *ACTA IMEKO XII*, **2**, 209.

McGhee, J. and Henderson, I.A. (1993) Current Trends in the Theory and Application of Classification to Instrumentation and Measurement Science, in *State and Advances of Measurement and Instrumentation Science*, *Proc IMEKO TC1/TC7 Colloquium* (eds L. Finkelstein and K.T.V. Grattan), City University, London (p. 32).

McGhee, J., Henderson, I.A. and Sankowski, D. (1986) Functions and Structures in Measurement Systems: A Systems Engineering Context for Instrumentation. *Measurement*, **4**(3), 11–119.

McGhee, J., Henderson, I.A., Kulesza, W. and Korczynski, M.J. (1996) *Scientific Metrology*, ISBN 83-904299-9-3, printed by A.C.G.M. LODART, Lodz.

McGhee, J., Henderson, I.A. and Sydenham, P.H. (1999) Sensor Science–Essentials for Instrumentation and Measurement Technology. *Measurement*, **25**(2), 89–113.

Measurement, **14**(1), (1994) special issue on Measurement and Instrumentation Science.

Mill, H. and Rada, R. (1990) Regularity: Generalising Inheritance to Arbitrary Hierarchies, in *Proceedings of 2nd International Conference on Tools Artificial Intelligence* Washington D.C., (p. 635).

M'Pherson, P.K. (1980) Systems Engineering: An Approach to Whole-System Design. *Radio and Electronic Engineering*, **50**, 545–558.

M'Pherson, P.K. (1981) A Framework for Systems Engineering Design. *Radio and Electronic Engineering*, **51**, 59–93.

Sandquist, G.M. (1985) *Introduction to System Science*, Prentice Hall, Englewood Cliffs, NJ.

Solopchenko, G.N. (1994) Formal Metrological Components of Measuring Systems. *Measurement*, **13**, 1–12.

Stein, P.K. (1969) The Engineering of Measurement Systems. *Journal of Metals*, **21**, 40.

Sydenham, P.H. (ed.) (1982) *Handbook of Measurement Science*, *Vol. 1 Theoretical Fundamentals*, John Wiley & Sons, Chichester.

Sydenham, P.H. (ed.) (1983) *Handbook of Measurement Science*, *Vol. 2 Practice Fundamentals*, John Wiley & Sons, Chichester.

Sydenham, P.H. and Thorn, R. (eds) (1992) *Handbook of Measurement Science*, *Vol. 3 Elements of Change*, John Wiley & Sons, Chichester.

Thomson, A.J. (1926) *Introduction to Science*, Williams & Norgate Ltd, London.

Von Engelhardt, W. and Zimmermann, J. (1988) *Theory of Earth Science*, Cambridge University Press, Cambridge, MA (p. 102).

Watanabe, H. (1996) Theory of Classification of Objects by Predicates. *Measurement*, **18**(1), 59–69.

Yasdi, R. (1991) Learning Classification Rules from Database in Context of Acquisition and Representation. *IEEE Transaction on Knowledge and Data Engineering*, **3**(3), 293.

Zhao, L. and Reisman, A. (1992) Towards meta research on technology transfer. *IEEE Transaction on Engineering Management*, **39**(1), 13–21, 103.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

3: Measures and Metrics; Their Application

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Measures Overview 15
2 The Measurement Situation 15
3 Measures and Metrics 16
4 Terms 17
5 Some Metrics 18
6 Forms of Cognitive Entity 19
7 The Scientific Process 20
8 How to Apply Measures 21
9 The Measures Triangle and its Parameters 21
10 Case Study of the Generation of Measures 22
Related Articles 23
References 23

1 MEASURES OVERVIEW

Measurement is found everywhere; it seems to be a necessary part of human living (Klein 1975; Ellis 1973). It is the process by which we seek to qualify and quantify an issue. It is a key part in the generation of knowledge for that issue. For example, in order to decide if the greenhouse watering system needs to be turned on, the moisture content of the soil is measured, resulting in a number that is compared to a standard value to decide if it is needed.

Measurement is not always well set up. The well-experienced measurement scientist or engineer will easily be able to point to the inefficient way in which much of measurement activity is practiced.

We need to be clear about such questions as:

- What is the purpose of the measurement?
- How does measurement advance the issue in question?
- Is it being done appropriately?
- Is the result expressed appropriately?

2 THE MEASUREMENT SITUATION

Throughout the recorded history of man, there has existed recognition of the connectivity between measurement and the acquisition of knowledge that, in turn, can be related to the advancement of man in general – Sydenham (1979); Bud and Warner (1998).

Measurement can be used to support two kinds of knowledge gathering situations:

- **Controlling a known situation**

A temperature controller in a food storage container uses the measurement value to switch the cooling on and off as needed. Here, the physical process is well understood; the need is to control the flow of cooling as the temperature varies.

- **Investigating a subject under research**

The need here is to glean new knowledge. For example, a theory has been proposed that suggests a relationship between two variables in an illness exists that would suggest a cure. A series of experiments is designed in which measurements are made under controlled conditions to reveal if the relationship holds.

A key statement about the relationship between measurement and knowledge is that of Lord Kelvin. In 1883, in a lecture at the Institution of Civil Engineers, he stated:

> 'In physical science a first essential step in the direction of learning any subject is to find principles of numerical

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

> reckoning and methods for practicably measuring some quality connected to it. I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be.'

This statement is expressing elements of the thinking paradigm known as *reductionism*, the main method by which we gather knowledge in the so-called hard sciences.

Reductionism appears to have come down to us from the contribution of Descartes. In 1693, he stated in his *Discourse on Method* – see Hutchins (1952):

> '.....divide each of the difficulties that I was examining into as many parts as might be possible and necessary, in order to best solve it.'

He suggested that the human mind sorts out its problems and finds solutions by breaking them down into successively smaller elements, until the stage is reached where they are adequately understood.

Descartes suggested four rules for 'properly conducting one's reason':

- Avoid precipitancy and prejudice
- Accept only clear and distinct ideas
- Conduct orderly progression from the simple to the complex
- Complete analysis with nothing omitted.

This is the basis of the measurement methodology outlined in Section 8.

In addition to being used for the 'hard science' physical situation measurements, it is also used in the 'soft science' situations to obtain qualitative knowledge where measurement is vital to such situations – for example, audits of the performance of people and processes.

Studies on the general nature of measurement are available. A few are now selected to show the range of approaches taken.

Finkelstein has covered a large range of fundamental topics. His paper, Finkelstein (1999), is a good summary of how far the ideas have been taken in formal mathematical terms.

Sydenham (1979) is a review of the role of measurement, which attempted to delve into the reasons and processes.

The place of measurement in science is covered by Kariya (1999) and IMEKO (1999). It gives a balanced overview of the hard science involved along with the necessary early stages of idea formulation and expression of what it is about as a process of learning.

Hofmann (1999) makes the link between measurement and practical needs in society.

Yang and Butler (1997) approached the problem of creating a universal framework from the epistemological perspective, suggesting it be modeled as a knowledge-oriented system. They propose that an object-oriented model (Yang and Butler 1998) be used for representing measurement systems.

3 MEASURES AND METRICS

The nature of inquiry used to gather knowledge can be very different across the various disciplines – see Brown, Fauvel and Finnegan (1981). Mathematicians, scientists, engineers, social science, management, and so on, do not have the same belief systems and often use different ways of thinking to solve their problems. They, however, all make measurements of some kind to support their knowledge-development processes.

It often comes as a surprise to the 'hard science' trained scientist or engineer that not all situations can make use of reductionist techniques.

Stumbling blocks for reductionists in accepting the softer sciences and humanities approaches are the:

- apparent lack of sufficient rigor of understanding and expression;
- use of many less familiar terms and words like 'paradigm', 'metaphor', 'holistic', and so on;
- inability to be as precise about ideas as are the laws of physics;
- inability of humanities practitioners to clearly identify the parameters and relationships of their areas of work;
- lack of applicability of the reductionist approach – that surely should be used; after all, it has and still is serving much of science and engineering very well.

The humanities paradigm is known as the *phenomenological approach*. Here, the observer does not metaphorically dismantle, by reductionism, the system of interest to separate its subsystems and then build it up again after changes have been made. Instead, the humanities viewpoint is one of metaphorically getting inside the system of interest, inserting intervention actions to see if current understanding is correct, and the ability to change the system as required. A relevant branch of this is called the *soft system methodology* (SSM), Checkland (1981).

In sharp contrast, reductionism requires all of the system of interest to be first bounded to form a closed system that is then dismantled to be built up again in its new form. The sort of problem that does not lend itself to this paradigm is one in which the boundaries of influence are

unclear, preventing the creation of an adequately closed system model.

There is also another reason why reductionism often fails in the complex systems arena. Success in understanding and problem solving is predicated by the assumption that the solutions for the subsystems resulting from reticulation can all be integrated back into the needed whole. One difficulty is that even slight variations in interface specification of those subsystems parts can have a significant impact on the performance of the whole – to the point where the performance of the new whole differs markedly from expectations.

The reductionist concept for problem solving is not totally accepted – it does have a severe philosophical problem.

A fundamental difficulty is what philosophers call the 'dual body' problem. Behavior of the physical aspect of the human system is well explained by the laws and rules of physics. The human mind, however, seems to behave quite differently. Its behavior defies reduction to formal description and use of the same method of scientific investigation.

Methods of inquiry, and even the scientific process of knowledge discovery, are not taught in most engineering and science courses. A result has been the widening divide between the thinking styles of the Arts/Humanities and the Sciences, existing on the modern university campus.

Many myths about measuring exist – see Sage and Rouse (1999) pg. 584–586 – some are:

- Measurement made with hard quantified measures will lead to the soft issues also being understood – not so; soft systems are different from hard systems and need different approaches in their measurements.
- Measurement is for bean counters and the data cannot be translated into useful improvements – not so, provided it is done well, see Section 8.
- Measurement is about the past and is not relevant to the different future – not so; applications can mature as projects change, by the application of sound and relevant measurement.
- Measurement encourages a box ticking culture – not necessarily so, provided it is done well and not using simple-to-measure, yet nonuseful, data.
- Measurement stifles creativity – not so, as measurement is about knowing about things in an objective manner.
- Measurement thwarts productive human activity – not so, if done appropriately.
- The more the measurement the better the productivity will become – not so, for again it is a matter of devising a good measuring system that truly addresses the requirements.

4 TERMS

Whatever process of measuring is being implemented, a confusing range of terms are used to describe the measures used.

The 'thing which is to be known' within a measuring situation is today, in the engineering world, often called a *measurand*.

A commonly found general term for measures, used extensively in the process performance arena, is the metric (Blanchard and Fabrycky (1998); Sage and Rouse (1999)).

This term is found, where a set of measures (metrics) are established to collectively gain insight into how well the whole process is working. This term is not as frequently used in the physical sciences, for the word 'metric' there is associated with the metric system of units.

Another measures term often used in systems management is the technical performance parameter (TPM) – this is explained later.

Many other terms will be encountered that mean much the same thing – tracking variable or parameter, indicator, index, score board value, and so on.

Measure terms that have specific and different application include:

- measure of effectiveness (MOE)
- measure of performance (MOP)
- system performance parameter (SPP)
- technical performance parameter (TPP).

Where these fit into a hierarchy of measurement is explained in Section 8.

The development and application of truly effective sets of metrics is a skilled task based much on experience in the application area. It is easy to generate the measures for the clearly evident physical measurements such as temperature, speed, and load-carrying capacity. It is often not so easy to decide an effective parameter for more elusive, many to one mapping situations, such as in setting up a measuring system for the quality of a social reform program.

At the single-measurand level, seek to choose the measure with best overall effectiveness. It will not always be obvious; the process involving the measure needs to be understood. For example, in jam making, a rapid change in the pH is a far better indicator of when it is optimally cooked than is the viscosity of the mixture.

Setting up a truly effective metric is not always easy; simple ones are often chosen that, while providing a seemingly comfortable quantitative number, add little to the overall picture being sought. For example, the rate of progress of a software task could be measured as 'lines of code completed in a unit time' compared against the envisaged number of lines used as the norm. This is, however, far too simplistic

as the quality of the code and the number of errors to be subsequently corrected can completely overwhelm the time used to prepare the code for the usable standard.

As the choice and use of metrics is based in considerable experience, a company will often be protective of its metrics database and not release it to the general public for, over time, it develops to have intellectual property value.

In reality, it is the high-level measures that are of real interest, physical variables being but a part of 'many to one' mappings of measures.

5 SOME METRICS

Thousands of metrics exist. A well-organized systems design operation will have a progressively updated database of metrics that has been developed to suit its own kind of industry.

Unfortunately, these tend to not be developed in reusable ways that would permit follow-on projects to extract them from a well-setup library. Also, they are often held in confidence and tied into a project.

They mature as the staff uses them, and for this reason alone the best way to develop effective ones is to ensure they are reused over projects in a controlled manner.

A measure stored in a metric database needs to have the following information recorded:

- Metric/measure name
- Symbol used to represent it
- Acronym used, where applicable
- Synonym usage explanation
- Definition of its purpose
- Brief description of its uses
- Use in multimeasures mapping sets
- Previous projects in which it has been used
- Person who authored the entry
- Level of confidentiality assigned
- Authorizing person
- Persons who accessed it in past use.

With so much to set up to ensure traceability, soundness, and uniqueness, it is not surprising that good metric databases are not readily available.

The following short collection of metrics is a motivational starting point.

5.1 Physical measurands

- Velocity
  - Time lapsed
  - Mass
  - Force
  - Temperature
  - Viscosity
  - Tensile strength
  - Strain

and so on.

5.2 General systems use

- Time to market
  - Time to completion
  - Number of items produced
  - Sales made
  - Sales returns
  - Defects rate
  - Repair time
  - Mean time between failure (MTBF)

and so on.

5.3 Customer responsiveness

The following are from Sage and Rouse (1999), pg 569. These require many-to-one measurement mappings to arrive at a measured quantity – see Section 8.

- Product features added
- Product quality
- Customer satisfaction
- Speed of response to customers
- Market expansion
- Product uniqueness
- Listening to customers
- Customer visits
- Sales improvements
- Innovation
- Organizational acceptance to customer evolution

5.4 Innovation measurement

Some of the lists provided in Sage and Rouse (1999), pg 570 are now given:

- Number of innovative small parts
- Service innovations
- Number of pilots and prototypes
- Number of benchmarked ideas adopted
- Measures of word-of-mouth marketing
- Number of innovation awards.

5.5 Software development

- Lines of code
  - Rate of completion of lines of code
  - Efficacy of coder
  - Error rate per 1000 lines of code
  - Recursion time
  - CPU needed
  - Speed of execution of standard benchmark operation
  - Latency time
  - Number of branches
  - Compilation time
  - Reset time
  - Cyclometric complexity
  - Level of cohesion
  - Level of coupling

and so on.

5.6 Defence systems

While extracted from defence material, Hoivik (1999), these may also be relevant to civil projects and situations.

- Quantity of x
- Quality of x
- Coverage of x
- Survivability
- Lethality
- Sea, air, and land worthiness
- Warhead size
- Speed
- Range
- Altitude of operation
- Evaluability
- Weight
- Power
- Computer throughput
- Memory size
- Cooling capacity
- Target location accuracy
- Reaction time
- Receiver sensitivity
- Ranging accuracy
- Range
- Hardness to damage
- Damage tolerance
- Drift rate of guidance unit
- Radiation hardness
- Engine power
- Rate of turn
- Climb rate
- Payload
- Subsystem x weight
- Number of crew needed
- Firing rate

and so on.

6 FORMS OF COGNITIVE ENTITY

When measuring, it is important to differentiate between the terms data, information, knowledge, and wisdom when used in relation to knowledge gathering via measurement. This issue is addressed in more depth in **Article 13, Signals, Information and Knowledge, and Meaning, Volume 1**; **Article 23, Principles of Semiotics as Related to Measurement, Volume 1**; **Article 24, Principles of Epistemology as Related to Measurement, Volume 1**; and **Article 39, Characteristics of Data, Information, Knowledge, and Wisdom, Volume 1**.

It is useful at this stage to classify the four levels of the development of a cognitive entity. Using the Oxford Universal Dictionary, 1968, we get the following definitions:

- Data:
  'a thing given or granted: something known or assumed as fact and made the basis of reasoning or calculation.'
- Information:
  'the action of informing' stemming from
  'to put into form or shape.'
- Knowledge:
  'the fact of knowing a thing, state, etc.' stemming from 'to recognize or identify.' Also, 'an organized body of information'.
- Wisdom:
  'the quality or character of being wise' stemming from 'having or exercising sound judgment or discernment.'

These give some useful clues about the terms, but we really need an explanation that is better related to measurement systems. Here, follow the author's definitions Sydenham (1986).

Data: Raw symbols that are obtained from a measurement system and that have no assignment of meaning associated with them. They are just simply numbers, letters, ikons, cuneiform stabs in clay, and so on. An example is the symbol set of '10'.

Information: This is data that has associated, either tagged with it or held elsewhere, a small amount of cognitive material that gives it a certain meaning. Reduction of 'raw data' into 'engineering units' is an example. An example is '10 m'. The assignment of a tag that has a cognitive meaning; here, the distance unit 'm' for the

unit of length, the meter, creates useful information from the number symbol.

Knowledge: This is sets of information put into a context of a particular use. Representational information is organized into a coherent model structure. As with 'beauty', what constitutes knowledge is in the mind of the beholder.

It possesses specificity of application. For example, the raw data from a strain gauge on a wing of an aircraft for a given location and time, and with known units, constitutes a segment of knowledge.

Wisdom: This is a higher level of cognition than knowledge. It is a set of knowledge components having associations between entities. For example, the pattern of strain gauge readings across the aircraft wing may have peculiarities that suggest, to the expert mind, that it is in an unsafe state.

A level is reserved for the highest level of, as yet unfathomable, intelligence.

These entities form the intelligence tree shown in Figure 1. It is clear how measures lead to an increase in wisdom.

Fashions in the use of terms change. Overall, what used to be called *information* often tends to be called *knowledge* now. The fact is the various cognitive entities have yet to be consistently used.

7 THE SCIENTIFIC PROCESS

An understanding is needed of how quantifying measurement can contribute to increasing the available knowledge on a topic. This is explained by the scientific process used as the basis of reductionist thinking.

This stems from as early as the sixteenth century and has gradually become the norm. In 1931, Bertrand Russell published his understanding of the basic process steps of the scientific method.

'In arriving at a scientific law there are three stages:

- The first consists in observing the significant facts
- The second in arriving at an hypothesis, which if it is true would account for these facts
- The third deducing from this hypothesis consequences which can be tested by observation.'

The scientific method relies on:

- reducing the complexity of the variety of the real world to a manageable state;

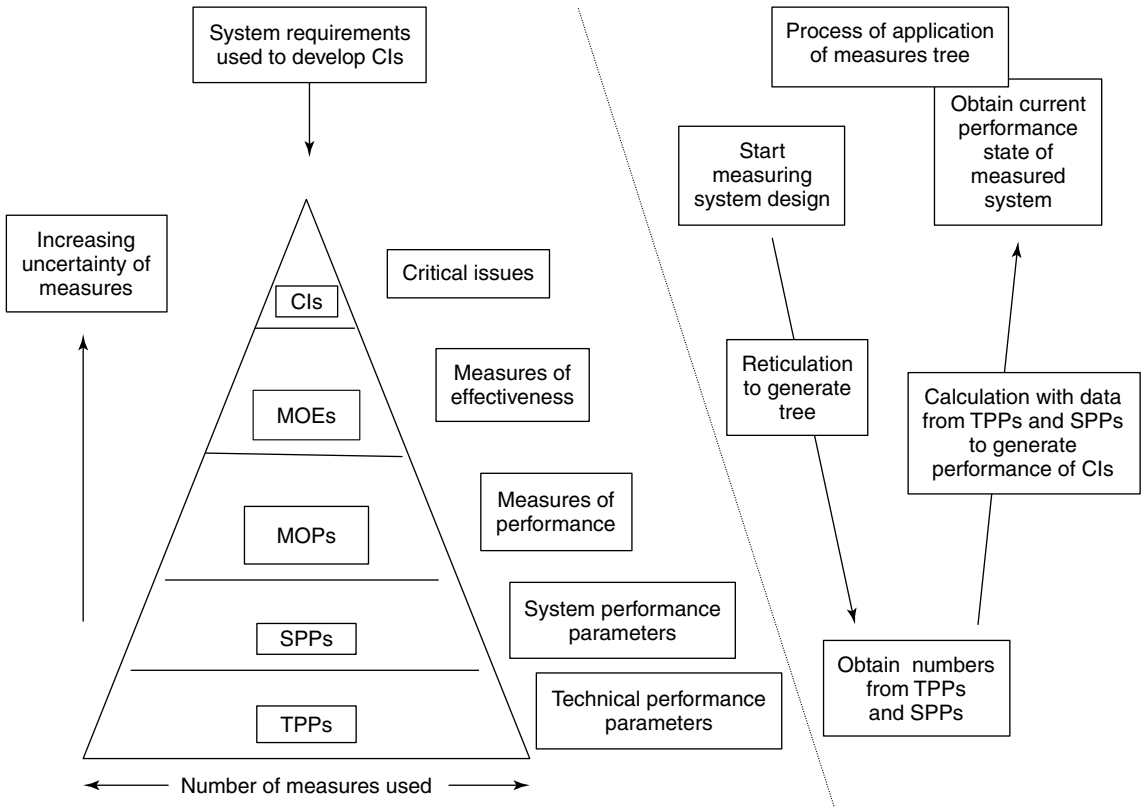


Figure 1. Intelligence tree shows relativity of the various cognitive variables and relationship to measures triangle.

Table 1. Stages of the measurement process and the role of measurement in its execution.

Generalized scientific method	Role of measurement theory and practice
Develop hypothesis	
1. Identify question/problem	1. Develop test objectives
2. Formulate hypothesis	2. Estimate performance
Experiment	
3. Plan the experiment	3. Develop test method
4. Conduct the experiment	4. Collect test data
5. Analyze the results	5. Calculate the measures
Verify hypothesis	
6. Check the hypothesis	6. Compare results
7. Refine the hypothesis	7. Rerun tests or extrapolate

- performing analysis or experimentation on simple models of the world to examine a hypothesis;
- validating a hypothesis by looking repeatedly to see if it can be disproved – the 'null hypothesis' basis. It is actually not achieved by showing it to be always true, as is commonly understood (infinite testing needed there!);
- building knowledge, therefore, by eventually refuting the hypotheses and forming an improved one.

The scientific process of inquiry and its stages are summarized as Table 1. Alongside are given the various functions of measurement in that process.

Areas of measurement are needed to undertake all stages of this knowledge-gathering activity. Measurement is, therefore, a key part in its application. Poorly undertaken measurement can lead to incorrect knowledge, or more usually the case, to less precise knowledge, possibly giving rise to misinformation or negative knowledge.

The process acquires new data from measurements made, and the observer uses that data to draw conclusions about the hypothesis being developed by evaluating the data in the context of the hypothesis.

So far, we have discussed the role of measurement in the scientific process. It is an easy step to see that this process is applicable to any measurement situation itself for a measurement activity is an experiment to see what you have. This is the time to review how that data flows into evaluation of the hypothesis.

8 HOW TO APPLY MEASURES

We need to ask a fundamental question. What is the holistic purpose of making a measurement?

In the closest inward looking boundary, it is to satisfy the need of the person requesting the test. This is, however, far too restricted a horizon to take because that test is being done to integrate into a much larger problem situation.

The scientifically executed process is the only way by which measures are obtained that are as objective as possible. The physical experiment performed in measuring is the only way to obtain verified data on the physical world.

A single measurement entity is being measured as part of a large array of measurements needed for evaluation purposes of a system of some kind. Examples might well be to assess the airworthiness of a new aircraft or to see if a medical intensive monitor unit is operating within all critical performance parameters.

The above sample lists of metrics show that for a project numerous things can be measured. The question needing an efficient solution is how can one set up an optimal measuring system when time, access, and cost, usually, severely limit the number of measurements that can be made.

9 THE MEASURES TRIANGLE AND ITS PARAMETERS

What is needed is a plan to set up and use many scientifically executed physical measurements that are integrated, in a traceable manner, to form decisions that map into a few high-level measures about the overall system.

This leads to the concept of the measures triangle. Figure 2 shows the various levels and types of measures that form this measures treelike diagram.

To set up a system's measurement plan, the first thing to do is to identify the critical issues (CI) from the system requirements documentation. CIs are those high-level issues that, will make the development fail if not achieved.

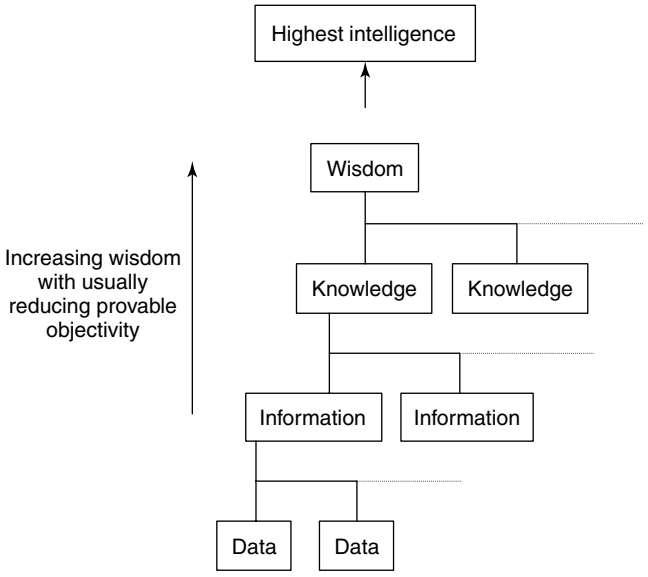


Figure 2. Measures triangle and its levels of measures.

Each CI is then broken down to obtain its measures of effectiveness (MOE). These are expressed in terms of what is to be achieved, for example, the requirement states that 'the customers must be satisfied' so that MOE needs measures of customer satisfaction to be set up. That it may not be immediately obvious how to measure it is not an issue at this stage of the reticulation. One should not start from what can be measured, but from what should be measured – a commonly ignored requirement!

The MOEs, in turn, reticulate down to give measures of performance (MOP). These break down the MOE into the MOP that, when combined, lead to the MOE value. Customer satisfaction could be measured in terms of the return rate of customers, from a direct survey of them using a written survey instrument, or from use of a video camera that records their demeanor as they pay for the goods. MOEs and MOPs cannot be measured directly.

This in turn gives the number of returns per customer; consolidated survey results grade; and an image interpretation as the system performance parameter SPPs to measure.

The final level is the technical performance parameter (TPP) wherein physical measurements are possible. SPP measures will often go directly to the TPP level as is the case here for return rate: a counting exercise. The paper survey will, however, not necessarily need TPP measures as it will be done by discussion and box ticking in a form. The paper survey is already able to yield some SPP values. The customer demeanor grade has to be determined from image processing set up to measure various indices. This needs a physical measuring instrument to obtain a set of TPPs that map into the SPP of customer demeanor.

In this way, from each CI, a tree of measures is generated. A measuring system design has been generated. Of these layers, only the TPP is hard data obtained by use of the defendable scientific process. The measures on the other layers invariably contain some degree of subjective decision making.

Once all of the TPP and SPP data has been obtained, the values are fed upward inside the tree to give the various other measures their value, by direct calculation up the tree. In this way, traceability results for all measurements and only essential measurements are performed.

When the first draft of the system is available and representative figures can be inserted for each measure, judgment can be exercised to see which measures are not significant; these can be omitted to reduce the measurement effort needed.

As the calculations rise up the trees, the integrated knowledge becomes less accurate in its nominal value and will have increasing uncertainty. This situation cannot be avoided unless all decisions are made entirely with physical measurements and with nonfuzzy logical combination. This ideal of perfect measurement appears to be a goal but it, usually, would be too costly, would take too long to organize, and more often than not, would not be feasible to build.

At any point in the tree a nodal value can be extracted, and its measure calculated from data below it. The widely used technical performance measure TPM is, in effect, such an extraction, but it has little recognition of where it fits into the traceable hierarchy from data measurement to proof of requirement satisfaction, for it is usually generated from experience without use of a measures tree.

A more detailed explanation of test and evaluation of large systems with many variables is provided in Sydenham (2004).

10 CASE STUDY OF THE GENERATION OF MEASURES

A simple example of how a measures tree is developed from the CI is the generation of a suitable measuring system for automatically monitoring the quality of loaves of bread coming from a bread baking line. This results in a fully physical measurement situation.

Figure 3 shows a set of reticulated measures that allow the appropriate TPPs (the actual measurements of physical variables) to be measured and their values integrated into the result that each loaf is acceptable, having met all requirements that are extracted as CI. If the integrated measure for a loaf is not acceptable, then the out-of-limit measurements will show where action is needed for correction.

A measures tree can be set up for a measuring system engineering development project. As the development progresses, the respective measures are monitored and used, with the best estimates of the time, to yield the overall value of merit for progress of the project.

This methodology can become very complex in large systems developments for the following reasons:

- CIs can exceed the 100 to 1000s;
- reticulation needs time to decide good breakdown sets of variables;
- measurements cannot always be made in a timely manner;
- uncertainty of measurements is variable;
- nature of measure, such as statistical distribution, are different, making combination problematic;
- measures are often taken from similar prior tests and need careful application;
- required measurements cannot be made because of a reduced budget, too little time, lack of test object, or lack of a test facility;

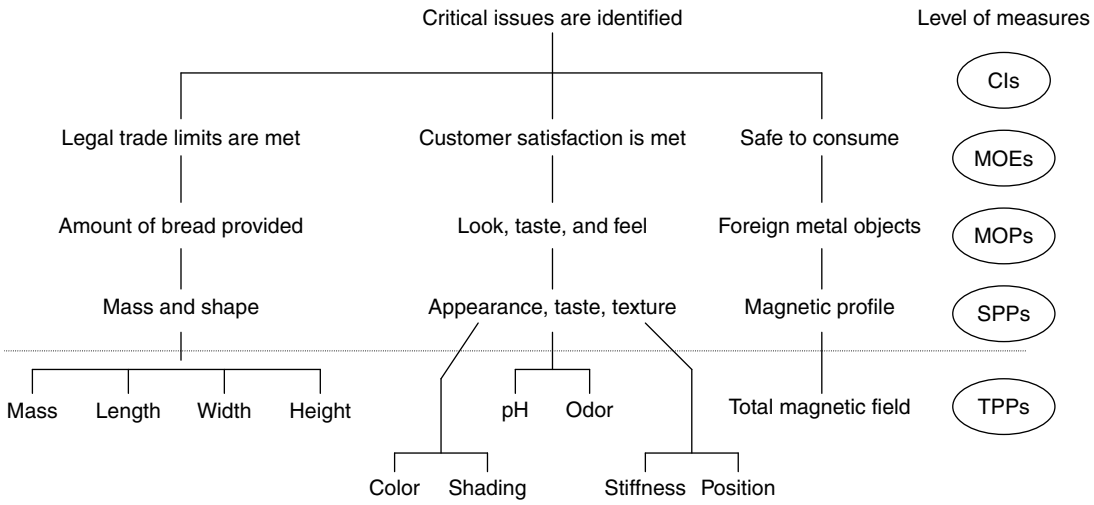


Figure 3. Development of a measuring system for loaves of bread from an automatic bakery.

- prior history, inherited data, and poorly defined processes are forced into the program.

For these reasons, many large programs do not get well-organized measures, with subsequent lack of confidence and timeliness in delivery.

Small programs need to tailor the measures tree concept to suit their budget. An example is Pratt (2004).

RELATED ARTICLES

Article 4, Economic Considerations of Measurement, Volume 1; **Article 10, Extensions of the Representational Theory of Measurement, Volume 1**; **Article 11, Measurement Theory in Physical, Social, and Psychological Science, Volume 1**; **Article 14, Hierarchical Aspects of Measurement Systems, Volume 1**; **Article 15, Typical Measurement Systems Architectures, Volume 1**; **Article 39, Characteristics of Data, Information, Knowledge, and Wisdom, Volume 1**.

REFERENCES

Blanchard, B.S. and Fabrycky, W.J. (1998) *Systems Engineering and Analysis*, Prentice Hall, New Jersey.

Brown, S., Fauvel, J. and Finnegan, R. (1981) *Conceptions of Inquiry*, Methuen and Open University Press, London.

Bud R. and Warner, D.J. (eds) (1998) *Instruments of Science: An Historical Encyclopedia*, Garland, London.

Checkland, P. (1981) *Systems Thinking, Systems Practice*, Wiley, Chichester.

Ellis, K. (1973) *Man and Measurement*, Priory Press, London.

Finkelstein, L. (1999) Foundational Problems of Measurement Science, in *Proceedings of. International Workshop on Advances of Measurement Science*, IMEKO TC-7, Kyoto, (pp. 23–31).

Hofmann, D. (1999) The Role of Measurement for Innovation and Society, in *Proceedings of International Workshop on Advances of Measurement Science*, IMEKO TC-7, Kyoto, (pp. 101–109).

Hoivik, T. (1999) *Advances in Test and Evaluation, Short Course*, Defence Engineering Group UCL, London.

Hutchins, R.W. (ed.) (1952) *Great Books of the Western World – 31 Descartes and Spinoza*, University of Chicago, Chicago, IL.

IMEKO. (1999) All papers, in *Proceedings of International Workshop on Advances of Measurement Science*, IMEKO TC-7, Kyoto, (pp. 1–235).

Kariya, K. (1999) Construction of Measurement Science as Basis of Empirical Sciences, in *Proceedings of International Workshop on Advances of Measurement Science*, IMEKO TC-7, Kyoto, (pp. 67–98).

Klein, H.A. (1975) *The World of Measurements*, Allen and Unwin, London.

Pratt R.W. (2004) *Framework for Issue Management*, MEng thesis, University of South Australia.

Sage, A.P. and Rouse, W.B. (1999) *Handbook of Systems Engineering and Management*, Wiley, New York.

Sydenham, P.H. (1979) *Measuring Instruments: Tools of Knowledge and Control*, Peter Peregrinus and Science Museum, Stevenage.

Sydenham, P.H. (1986) Structured Understanding of the Measurement Process Pt.2: Development and implementation of a Measurement Process Algorithm. *Measurement*, **3**, 161–168.

Sydenham, P.H. (2004) *Systems Approach To Engineering*, Artech House, Boston, MA.

Yang, Q. and Butler, C. (1997) On the Framework of Measurement Science, in *Proceedings of XIV IMEKO World Congress*, Vol. 5, Tampere, (pp. 101–106).

Yang, Q. and Butler, C. (1998) An Object-oriented Model of Measurement Systems. *IEEE Transactions on Instrumentation and Measurement*, **47**(1), 104–107.

4: Economic Considerations of Measurement

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Cost Issues 24
2 The Cost of an Instrument 24
3 Analyzing the Cost of Making a Measurement 26
4 Calculating the Cost of Measurement 28
5 Balancing the Cost-benefit Factors 28
6 The National Measurement System Matrix Approach 29
7 Guidance to Management on the Cost-benefit of Measurement 29
Related Articles 32
References 32

1 COST ISSUES

Measurements are made using physically existing equipment. The instruments must be created from the basic materials and applied using the labor of people. Material and labor resources have never been unlimited or freely available to a measurement user. Priorities must be decided in situations appearing to need measurement equipment.

A measuring instrument does not stand alone as far as its costs are concerned. The cost of measurement should not be assessed by looking only at the purchase or manufacturing cost of the apparatus used. Here, attention is drawn to the real cost of the individual measurement, for too little attention is often given to this aspect of measuring instruments.

To provide a balanced economic statement, it is necessary to consider the worth of a measurement in terms of the cost-to-benefit ratio for making the measurement.

The cost of measurements and their worth has not been a subject that many writers have felt they wanted to openly address. In general, measurement is seen as a low value item (around 2–8%) in a much larger overall budget of a system and thus needs little attention. It is considered by many that making a measurement is straightforward and simple to set up using available systems off the shop shelf.

Costs to produce and sell a measuring system are, however, very much the concern of the manufacturer who needs to set a realistic selling price, and to a nation, for it is a significant factor in a national economy due to the high gearing effect of good measurement practice.

This account begins first by looking at the cost of the apparatus, broadening outward through the cost of making a measurement, and then to the cost of measurement to a nation.

2 THE COST OF AN INSTRUMENT

The cost of finally making a measurement is formed from several subcomponent costs:

- C_i – cost to investigate the measurement need by the customer
- C_m – cost to research need, design, manufacture and market the measurement system hardware and software
- C_s – cost to set up the measurement in its application ready for use
- C_m – cost of maintenance and spares holdings
- C_a – cost to apply the measurement.

Costs to decide what is needed and to locate and purchase the instrument system C_i are often absorbed into other

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

project budget lines and thus can be transparent. They, however, can be considerable. Paradoxically, the less expensive the system needed, the higher the relative costs can be for locating a suitable unit; sellers will often assist with this cost where the price tag is large.

The cost C_m of designing and making the instrument system is built up from many parts, such as researching the need, design development, embellishment, paper work about specifications, use and service, packing, delivery, and many more contingencies. Figure 1 depicts the factors involved in costing the making of an instrument. Manufacturers well understand the issues and costs involved.

In classical antiquity, many of these cost factors did not arise. How to use the instrument was usually self-evident. Rarely were more than one of exactly the same model made; often the maker was the user, so he or she did not need to be concerned with records and sales. As the instruments were made by craft artisans, they usually had more decorative embellishment than was needed to be adequately functional.

The time taken to make a piece of measuring equipment has changed over time. It was not uncommon for instrument makers up to the seventeenth century to spend not months but years on manufacture. Harrison produced his four prize-winning chronometers of the eighteenth century as the result of devoting most of his life to the task. Originally, instrument making was more akin to the production of fine arts products of today than to the fierce open competition of an industrial market.

The introduction of production manufacturing methods changed this situation greatly. Markets widened, production run numbers rose, and cheaper means of producing the desired performance were realized.

The measuring instrument manufacturing market game has the feature that its players seem to be doing their best to work themselves out of a job! As soon as a significant new principle or technique is discovered that could cause major sales, the manufacturers move to produce it for sale first at a relatively high price. Demand grows with reduced price, so means are sought to stimulate demand, leading to larger production runs sizes. This tactic, if successful, drops the price, which saturates the market; the manufacturer must again look for another product.

Some makers will make adequate profits in this cycle and will return again to play another round. Many, however, do not get their timing and marketing methods right. They end up with an abundance of out-of-date products that cannot be sold at even the cost-to-produce price. Companies can go to the wall if this happens too often. Many an instrument firm has poured funds into developing a certain class of

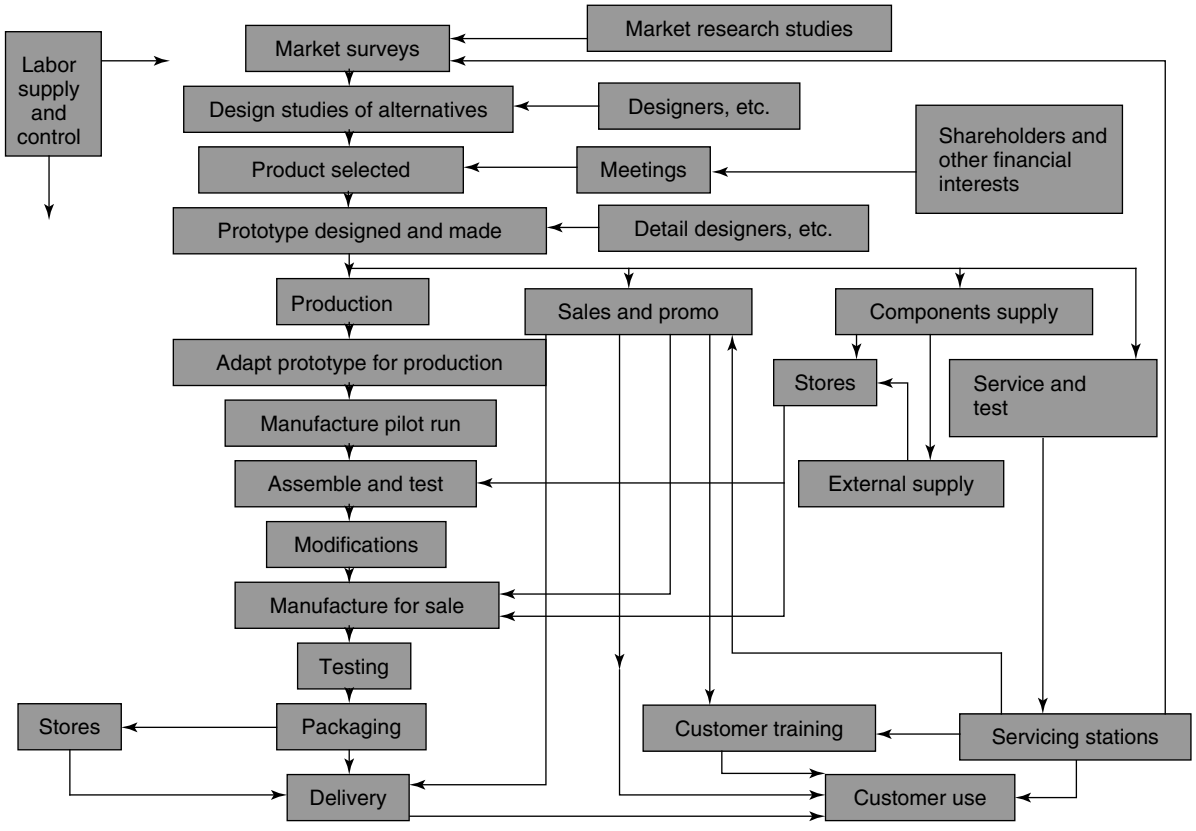


Figure 1. Factors involved in costing the manufacture of an instrument.

instrument only to have other makers systems sell better than theirs.

The instrument market has the following specific features:

- Production volume is small to medium, rarely large.
- Cycle time to develop is only a matter of months or so.
- Costs to produce have fallen dramatically as development by the whole group of this industry increased.
- Useful life of the physical aspect of a product is comparatively long.
- Information Technology (IT) content changes rapidly but the basic principle remains reasonably static.
- Purchasers expect ultimate reliability for minimal cost.
- Applications are not well worked out.
- Users expect 'fit and forget' to apply despite the low price paid.
- Users expect spares to be available for decades.
- Companies making instruments and measuring systems change hands a great deal.

Furthermore, as the per-unit production cost is lowered to make the system more affordable, the reliability of the components, and thus that of the system, are often sacrificed.

Another feature that the instrument maker has to contend with is that systems designers consider it reasonable and normal to call up specials when specifying the measuring apparatus required. One-offs are product lines that the firms must produce to maintain customer faith yet they often cost far more than the customer thinks should be paid. People have become used to the idea that their electric toaster works as the designers suggest it should, yet they would not always agree to make the same degree of compromise in using the instrument makers suggestions of what are reasonable basic units to work with.

Just as the knowledge about measuring systems has to be restructured for future, more efficient use, so must the range and number of marketed instrument variations be restructured.

It is generally acknowledged and understood that the cost of the instrument itself can be assessed from a long list of the production component costs, as shown in Figure 1. Management must budget for a manufacturing price to decide sales potential and profit margins. The same, however, cannot be said of users when they assess the cost of a measurement to their organization or enterprise. The truth is rather startling: the real cost of a measuring instrument is usually many times the purchase price of the apparatus.

3 ANALYZING THE COST OF MAKING A MEASUREMENT

Assessment of the real cost of a measuring instrument should be on the basis of considering what costs would have been avoided, or incurred through lost benefit, if it had never been considered – the economists call this the 'opportunity cost'. Figure 2 shows the total costs that comprise the cost of making a measurement.

To begin the costing process, it is necessary to first include the cost of the time spent in deciding what has to be measured (C_i). Too often this stage is seen as barely existing because the measurand, the parameter to be measured, appears obvious. It is the writer's experience that many users give too little emphasis to this very basic step, ending up measuring the wrong entities.

Having decided the variable(s) concerned, it is then necessary for the designer to devote time to generating sound specifications for the method selected. For example, assume the task has been defined as one of measuring the temperature of a product because that seems to indicate when it is ready to be transferred to another process – as in jam making. Is temperature really the best parameter to monitor? If so, what must be the precision, accuracy, absolute value, shape of sensor, and so on? To ascertain when jam is cooked sufficiently, it is better to measure the pH, and not the temperature! Making these requirement decisions incurs time and monetary costs.

Consider next that the specific instrument principle and the specifications have been chosen. The measurement task has still barely begun, for it is then necessary to locate trade information of available products so that a specific purchase can be organized. Tenders are called – or firms contacted – to obtain quotations, data sheets, delivery information, installation details, and so forth. At this juncture in costing of the real cost of a measurement, one must not forget to also include the time spent in preparing budget submissions and filing the records of this investigative work.

Eventually, the stage is reached where a purchasing decision can be made. That, too, takes up more time by skilled staff. It is unlikely that these above mentioned stages can be passed through in less than a minimum of a few hours of effort for even the simplest of measurements. Extensive measurement systems take months, sometimes many-years, of work to reach this stage.

The instrument system has then been selected and the order placed for supply; the cost of the ordering procedure should be included. Delivery is then awaited. If all proceeds perfectly, the instrument arrives on time without the need to make enquiries about late delivery. More often than not it is necessary to chase delivery, adding more cost.

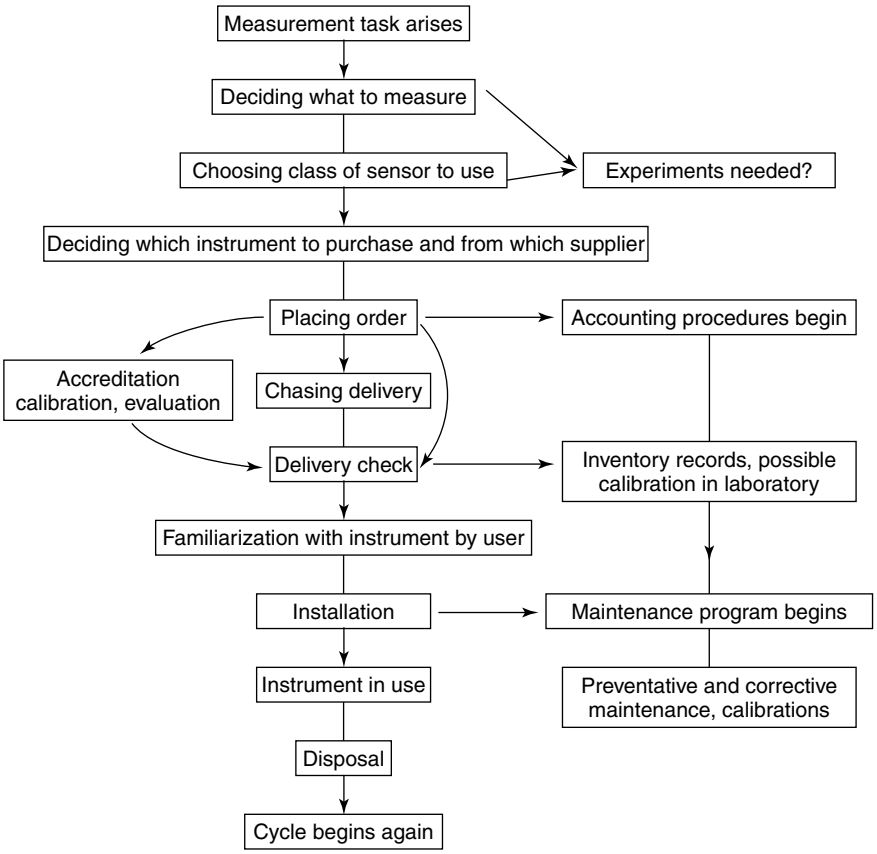


Figure 2. The cost of a measurement far exceeds the purchase price of the instrument.

Upon receipt of the system, the wise purchaser will carry out an immediate overview delivery check to ascertain that it is undamaged, is generally within specifications, contains the necessary documentation (which might be yet another cost to consider), and that it will suit the purpose intended. Soon after that a more extensive test is needed; if performed properly, this can take many hours and will need the use of expensive test equipment, therein also placing a cost into the overall account.

If things are found to be wrong at delivery time, additional effort will have to be devoted to getting it corrected through the purchase agreement or perhaps, which is often quicker and perhaps cheaper, having it repaired in-house.

In parallel with this acceptance stage is the need for the user to become familiar with the instrument operation. It may need attendance by staff at a training course, or be simply a case of reading the operating instructions. Whichever it is, still more time must go onto the costing sheet. The resources expended on this should not be underestimated, for a complex piece of measuring equipment may require the user to spend much unproductive time learning to apply it. Errors in its use in the early stages will also give rise to costs that have to be met somewhere.

Installation is then needed. This requires the organization of varied skills. It may necessitate inventory records to be created along with an instrument dossier and a spare parts reserve to be set up. Plant drawings usually need to be updated with instructions written and issued to operators.

Costs of the measurement continue to mount. To ensure that the instrument is performing properly, it will be necessary for it to have a service program, a calibration program, and a maintenance program in operation, as well as a possible method of monitoring the measuring system if its role is in a key area. To remove an instrument, take it to a calibration laboratory, carry out a calibration, and return it to duty will take a minimum of another two or three hours of work, but it can even take many days!

Too often the decision is made to purchase the instrument offered at the lowest price. This is oversimplifying the purchase decision; the fact remains that most instruments will need some unpredictable service during their useful life. The cost of this kind of service is high, for it usually occurs when labor repair costs are at their highest penalty rate. Furthermore, the emergency technician may not be familiar with the system and has to learn at the owner's expense.

Eventually, the measuring system has provided a useful and satisfactory life; its costs to the user go on. At some

stage, effort has to be expended to establish a replacement schedule and the choice of its replacement, and these can be complex issues. The model of instrument used originally will often have been superseded by something that may not be quite the same in some aspects that are critical to the application; it may not be possible to get a replacement at all.

Finally, the last of the costs associated with the measuring system parts is its disposal. It may be possible to recoup some of the aforementioned costs through sale of the instrument, but the opportunity cost of a well used, now superseded instrument is more likely to be a debit, not a credit. This cost often does not arise until years after installation, for most large complexes tend to store old instruments for many years before they are finally discarded, finding their way into a museum collection or being dismantled for use as pieces of some other plant.

4 CALCULATING THE COST OF MEASUREMENT

Any genuine attempt to estimate the cost of a measurement must include all of the above factors. The overall cost issue to be studied is one of establishing what resources were tied up by the existence of the measuring system that might have been used elsewhere if it had not been considered. A detailed costing on the above-given basis can lead to a satisfactorily quantitative assessment of which measuring system to purchase and in which way to use it to make the measurement needed.

All of the above costs are concerned only with the choice, purchase, installation, and maintenance of the correct measuring system. The cost of the whole measurement can continue to rise if the right decisions have not been made.

Much of measurement conducted is, unfortunately, not entirely productive. A wrong measurement has often been the cause of disastrous losses to the enterprise. A serious malfunction in the commissioning stage of a new ship can cost the builder thousands of dollars per hour. Incorrect operation of a weapon of war has a cost that goes without saying.

5 BALANCING THE COST-BENEFIT FACTORS

The total of all of the above costs has to be weighed against the benefits of making the measurement when it is expressed in monetary terms. Whereas it is reasonably clear and straightforward to obtain the cost of making the measurement, it is not so clear how to assess the value of the benefits of making it.

This can be done by conducting a risk analysis, for example along the lines of developing a safety case – *see* **Article 99, Safety Organization, Volume 2**; **Article 100, Safety Analysis Methods, Volume 2**; and **Article 101, Assessing and Demonstrating Safety, Volume 2**. Some simple examples should assist one to understand what can be at stake.

A large blower was used to aerate sludge in a mineral separation process. The 100-kW blower had a temperature sensor installed to shut it down when its internal lobes overheated, for that could cause it to seize and burn out the motor. The cost of the blower unit was around $200 000 installed, and the inexpensive and somewhat crude temperature sensor used was a mere $200. As might reasonably be expected, the blower was eventually seriously damaged when it overheated at a time when the sensor had not operated properly. Lost production and repairs cost as much as a new blower unit. In this case, the outlay of a mere 0.1% of the purchase value on protection seems to be too low. Lost benefit can be expected to be one of the higher costs.

A second example is found in the factory of a permanent magnet manufacturer who had taken over an old plant. These magnets are cast from a molten NiCo alloy that melts at a temperature higher than that of steel. The actual pouring temperature and subsequent annealing process are crucial to the final field strength of the magnet.

The whole manufacturing process, with its casting, several annealing and curing steps, takes some 10 days before it is learned from tests how well the magnets will perform. Along with the process, several measurements are made of critical parameters. At the end of the process there existed a final inspection station used to test and certify the quality of each magnet.

First, the process relied on a very old uncalibrated infrared radiometer and a set of record cards that had been inherited, giving the different temperatures needed (as indicated by that radiometer) for the various shapes of magnet cast. As no independent calibration of the radiometer existed, if it should fail, it would be virtually impossible to decide what temperatures to use in future.

Next, the final inspection area that was set up inside the dirty foundry area was a dark mess. The force measuring system for the magnet test was red with rust and its measuring instrument dials were hardly readable. This was the station audited by the customer! It was rejected, as it was not a suitable testing process for them to accept; they then did their own tests with a higher rejection ratio.

The manufacturer decided he was unable to make the investment in improvements and eventually was forced to close down.

A third example situation was that of a temperature controller installed in a greenhouse used for plant research in a

university. After 5 years of apparently stable temperatures in the greenhouse, the thermocouple controller system was overhauled.

It was found that the thermocouple circuitry was not in the control regime expected, owing to an incorrect change to the circuit card to hopefully reduce the indication span. This had reset the recorder's chart zero point to where it was expected to be, but in doing so the changes had increased the chart span to 400 °C, full scale, instead of 40 °C. It seemed to behave as expected but it was then seen to be recording the temperature of the control room, not that of the greenhouse. The cost of this incorrect measurement would be hard to assess but it is clear that the research findings for the previous 5 years were subject to much uncertainty.

The benefits of a measurement need to be carefully investigated using an objective standpoint and methodology that leads to monetary values. These can then be weighed against the monetary cost to make the measurements.

6 THE NATIONAL MEASUREMENT SYSTEM MATRIX APPROACH

The cost of measurement also can be of importance at the whole country level; measurement has been identified as an important factor in a national economy. This discussion now considers the cost to a nation of making measurements.

Studies of the cost of national measurements were given in-depth consideration in the 1970s. Prior to that, costing was certainly performed on the annual budgets of such institutions as the national measurement laboratories and similar identifiable groups, but little attempt was made to go beyond this small section of the total national cost devoted to measurement.

The term *National Measurement System* (NMS) came to the fore when the then National Bureau of Standards (NBS), USA, (now the National Institute of Science and Technology (NIST)) conducted a major nation-wide study of the value of the United States' NMS.

Development of their study can be followed in the National Conference of Calibration Laboratories (NCSL) Newsletters from 1972, through to 1978. Minck (1977) defined the NMS as 'consisting of all national activities and mechanisms that provide physical measurement data to allow the creation of the objective, quantitative knowledge required by our society'.

Previously, Huntoon (1967) had developed the concept of the NMS in a paper. The NBS study reached the final report stage in 1975 after 151 separate studies had been made. The report contained 27 documents. Two key readings are NBSIR 75–925 'Final summary report study of the National Measurement System, 1972–1975' and NBSIR 75–949 'Structure and function of the National measurement system'.

The report included a major impact matrix (Figure 3), showing the measurement relationships existing for 25 groups of societal needs. It graphically portrays the dynamics and the degree of such interactions and, in doing so, well illustrates the great complexity and widespread use of measurements in our modern lifestyle.

This NBS study was catalytic; many countries now maintain NMS activities. Owing to the importance of the NMS to national economic welfare, they are now usually cared for by departments of trade or industry.

Information in the UK system is found on DTI (2003). The Australian one is explained in NSC (2003). Taiwan's is discussed in NML Taiwan (2003). Others are to be found from a web search.

Interestingly, the organization that began the NMS movement, NIST, does not list any NMS activity, as such, on its current information service, NIST (2003).

7 GUIDANCE TO MANAGEMENT ON THE COST-BENEFIT OF MEASUREMENT

Very little has been done to find and develop means by which management can be guided in budget decisions concerned with the use of new, or existing, measuring equipment. Easy to reveal are the simple benefits of money saved by implementing a better-controlled process, or by labor costs saved by the use of a more advanced measuring technique. More difficult to quantify in an adequate manner, but often equally as important, are such factors as

- general improvement of the standard of a product resulting in increased sales;
- increased reliability of a product;
- less wastage in manufacture;
- potential to produce a product that was otherwise impossible to make;
- improved performance of the product;
- improved working relationships between those groups of a nation that impact together in an area of measurement; and
- increased confidence by others in a nation's products, with subsequently improved exports.

At present, measurers are not yet able to measure the worth of measurements to the same degree as they can assess the cost of measurement. Perhaps this is the reason many measurement budget decisions generally do not allocate finance to measuring equipment in proportion to

DIRECT MEASUREMENTS TRANSACTIONS MATRIX FOR NATIONAL SYSTEM OF PHYSICAL MEASUREMENTS (March 1976) — USES / SUPPLIERS	KNOWLEDGE COMMUNITY (Science, Education, Prof. Soc. & Publ.) 1	INTERNATIONAL METROLOGICAL ORGANIZATIONS 2	DOCUMENTARY STANDARDS ORGANIZATIONS 3	INSTRUMENTATION INDUSTRY (SIC Major Gp 38) 4	NBS 5	OTHER U.S. NATIONAL STANDARDS AUTHORITIES 6	STATE & LOCAL OFFICES OF WEIGHTS & MEASURES (OWM's) 7	STANDARDS & TESTING LABORATORIES AND SERVICES 8	REGULATORY AGENCIES (excl. OWM's) 9	DEPARTMENT OF DEFENSE (excl. Stds. Labs) 10	CIVILIAN FEDERAL GOVT AGENCIES (exc. Stds Labs & Reg. Ag.) 11	STATE & LOCAL GOV'T AGENCIES (exc. OWM's Reg. Ag.) 12	INDUSTRIAL TRADE ASSOCIATIONS 13	AGRICULTURE, FORESTRY FISHING; MINING (SIC Div. A & B) 14	CONSTRUCTION (SIC Div. C) 15	FOOD/TEXTILE/LBR/ PAPER/LEATHER/ETC. (SIC 20–26, 31) 16	CHEM/PETROL/RUBBER/ STONE/CLAY/GLASS... (SIC 28–30, 32) 17	PRIMARY & FAB. METAL PRODUCTS (SIC 33–34, 391) 18	MACHINERY EXCEPT ELECTRICAL (SIC Major Gp 35) 19	ELECTRIC AND ELECTRONIC EQPMT (SIC Major Gp 36) 20	TRANSPORTATION EQUIPMENT (SIC Major Gp 37) 21	TRANSPORTATION & PUBLIC UTILITIES (SIC Div. E) 22	TRADE/INS/FIN/REAL EST/PERS SVCS/PRINT (SIC F-H, ba l, 27) 23	HEALTH SERVICES (SIC Major Gp 80) 24	GENERAL PUBLIC 25
1 KNOWLEDGE COMMUNITY (Science, Education, Prof. Soc. & Publ.)	4 1 6 2	3 1 2 1	3 1 3 1	2 1 3 1	4 1 4 2	1	1	2 1 2 1	3 1 2 2	3 1 3 1	3 2 3 2	2 1 2 1	2 2 1 2	1	1	2 2 1 1	3 1 2 2	3 1 2 1	2 1 2 1	3 2 3 1	3 1 3 1	2 2 2 2	3 1 1 1	2 1 1 1	2 2 1 1
2 INTERNATIONAL METROLOGICAL ORGANIZATIONS	2 1 1 1	3 1 3 1	3 1 2 2	2 2 1 1	3 1 3 2	1		1	1	1	1			1						1			1			
3 DOCUMENTARY STANDARDIZATION ORGANIZATIONS	3 2 1 2	3 1 2 2	3 2 3 1	3 1 3 1	3 1 3 2		3 1 1 1	3 1 2 1	3 2 2 2	3 1 2 1	2 1 1 1	1	3 1 2 2	1	2	2	2 1 1 2	2	2 1 2 1	2 1 2 1	2 1 2 1	3 1 2 2	1	2	1
4 INSTRUMENTATION INDUSTRY (SIC Major Gp 38)	2 1 2 1	2 1 1 1	2 1 2 2	4 1 5 2	3 1 3 2	1	1	3 1 3 2	3 1 2 2	3 1 5 2	3 1 4 1	3	2 1 2 1	3 1 3 2	4	3 1 3 2	3 1 3 1	3 1 4 1	3 1 5 1	3 1 5 2	3 1 5 1	3 1 5 2	3 1 5 1	3 4 2 2	3 1 4 2
5 NBS	3 1 3 1	3 1 3 2	3 1 3 2	3 1 3 2	3 1 4 1	3 1 1	4 1 2 2	3 1 3 1	3 2 2 1	3 3	3 1 3 2	2 1 1 1	2 2 1 2	4 1 2 1	1	2 1 2 1	2 2 2 2	4 2 2 2	2 1 2 1	3 1 2 1	4 1 2 1	4 2 2 2	4 3 2 2	4 3 2 2	2 1 1 1
6 OTHER U.S. NATIONAL STANDARDS AUTHORITIES	1	1		1	3 1 1 1	3 1 1 1		1	1	2	1	1	1	1	1							1	1		
7 STATE & LOCAL OFFICES OF WEIGHTS & MEASURES (OWM's)		1	1	3 1 1 R	4 1 4 2 1 R		4 2	1	1			3 1 1 1	1	1		3 1 2 2	1		3 1 2 1		1	2	4 1 3 1		4 1 1 1
8 STANDARDS & TESTING LABORATORIES AND SERVICES	2 1 1 1		2 1 2 1	2 1 3 1 R	2 1 2 2 R			3 1 3 1	2 1 2 2	3 1 3 1	3 1 2 1	1	3 1 2 1	1	1	1	2 1 2 1	3 1 2 1	3 1 2 1	3 1 3 1	3 1 3 1	3 2 2 2	3 1 1 1	3 2 2 2	
9 REGULATORY AGENCIES (excl. OWM's)	3 1 2 1		3 1 2 2	3 1 2 2 R	3 2 2 2 R			3 1 2 2 R	3 2 3 1	2 1 1 1 R	3 1 2 2 R	1 R	2 1 2 2 R	4 2 2 1 R	4 1 2 2 R	3 2 2 1 R	3 1 2 1 R	3 1 2 1 R	4 1 2 1 R	3 2 2 1 R	3 1 2 2 R	3 1 2 2 R	3 1 2 2 R	4 2 3 2 R	3 1 2 2 R
10 DEPARTMENT OF DEFENSE (excl. Stds. Labs)	3 2 2 2		3 1 2 1	3 1 3 2 R	2 1 3 1 R	1		3 1 3 1 R	1	3 1 7 1	3 1 3 1		2 1 1 1				2 1 1 1 R	3 1 2 1 R	4 1 2 2 R	3 1 4 2 R	3 1 4 2 R	2 1 2 2 R	2 2 R	1 R	
11 CIVILIAN FEDERAL GOV'T AGENCIES (excl. Stds. Labs & Reg. Ag.)	3 1 2 1		2 1 2 2	2 1 2 2	3 1 3 1	1		2 2 2 1	2 1 2 2	2 1 1 1	3 1 7 2	2 1 1 1	2 1 1 2	2	1	1	2 2 2 2	3 1 2 1	3 1 1 1	3 1 2 2	3 1 2 1	4 1 4 1	1	2	2 3
12 STATE & LOCAL GOVERNMENT AGENCIES (exc. OWM's & Req. Ag.)	1		1	1 R	2 1 1 1 R	1	1 R	1 R	1	1	1	3 2 6 2	1	1	2	1	1	1	1		1	1	1	1	1	1
13 INDUSTRIAL TRADE ASSOCIATIONS	2 1 1 1	1	2 1 3 2	2 1 2 2	2 1 2 2		1	3 1 1 1	2 1 2 2	2 1 1 1	2 1 1 2	2 2 2 2	2 1 2 1	1	1	1	2 1 1 1	1	2	2 2 1 2 1	2 1 1 1	2 1 1 1	1	1	1
14 AGRICULTURE, FORESTRY FISHING; MINING (SIC Div. A & B)	1		2 1 1	1 R	2 1 1 R		1 R	1 R	1	1	2	1	1	2 5	1	2	2	2	1	1	1	2	1		1
15 CONSTRUCTION (SIC Div. C)	1		1	1 R	1 R		1 R	1 R	1	1	1	1	1	1	6			1 R	1	1	1	1	1		1
16 FOOD/TOB/TEXTILE/ APPAREL/LBR/FURN/PAPER/ LEATHER (SIC 20–26, 31)	1		3 1 1 1	3 2 1 2 R	3 2 1 2 R		2 R	2 1 R	3 1 2	1	1	1	3 2 1 2	1	1	2 1 6 1		1 R	1 R	1 R		1 R	1		2
17 CHEM/PETROL/RUBBER/ PLASTICS/STONE/CLAY/ GLASS (SIC 28–30, 32)	2 1 2 1	1	2 1 2 2	2 1 2 1	3 2 2 2		1	2 1 1 1	2 2 2 2	2 1 1 1	2 2 2 2	1	2 1 1 1	2	1	1	2 1 6 1	1	1	1	1	1	1	1	1

Figure 3. Direct measurements matrix from the USA foundational NMS study.

18 PRIMARY & FAB. METAL PRODUCTS (SIC 33-34, 391)	2 1 1		3 2	3 2 1 R	3 1 2 R		1	3 1 R	2	3 2 1 1	2	1	2	2	1	1	1	3 1 6	2 1 1	1	3 2	2	1	1	1
19 MACHINERY, EXCEPT ELECTRICAL (SIC Major Gp 35)	2 1 1	1	3 1 2	3 1 2 1 R	3 2 R		1	2 1	1	3 1 1 1	3 1 1	1	2	2 2	1	1	1	2	3 1 6	2 1	3 2	2	2	1	2
20 ELECTRIC AND ELECTRONIC EQPMT (SIC Major Gp 36)	3 1 2 2		3 1 2 2	3 1 3 2	3 1 2 2			3 1 2 2	3 1 1 2	3 1 3 1	3 1 2 2	1	3 1 2 1	1	1	1	1	2	3 1 6 3	3 2 2 3	2	1	2	2 2 2 2
21 TRANSPORTATION EQUIPMENT (SIC Major Gp 37)	2 1 2 1		2 1 2 1	3 1 2 1 R	2 1 2 2 R			2 1 2 1 R	2 1 2 2	2 1 2 2	2 1 2 1	2 1 1 1	3 1 1	1	1	1	2 1 1 1	2 1 1 1	2 1 1	3 2 1 2	3 1 6 1	3 2 1	1		1
22 TRANSPORTATION & PUBLIC UTILITIES (SIC Div. E)	2 1 2 1		2 1 2 2	2 1 2 1 R	2 1 2 2 R	3 1 1 2	1	2 1 2 1 R	3 2 2 2	2	2 1 2	1	2 2	2	1	1	1	1	1	1	2 1 2 1	3 1 6 1	2	1	3 1 4 1
23 TRADE/INS/FIN/REAL EST/PERS SVCS/PRINT-PUB (SIC F-H, Bal. I, 27)	1		1	2 R	3 1 1 R		2 R	1 R	2	1	1	1	1	1	1	2	2	2	1	1	1	2	7	1	6
24 HEALTH SERVICES (SIC Major Gp 80)	1		1	2 R	4 2 2 3 R			1 R	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	5	2
25 GENERAL PUBLIC			1 R	2 R	2 1 1 R		1 R		2 1 1 2 R	1	2 R	1 R	1 R	1 R	1 R	1 R	1 R	1 R	1 R	1 R	1 R	1 R	2 R	2 R	1 R	2 1 7 1

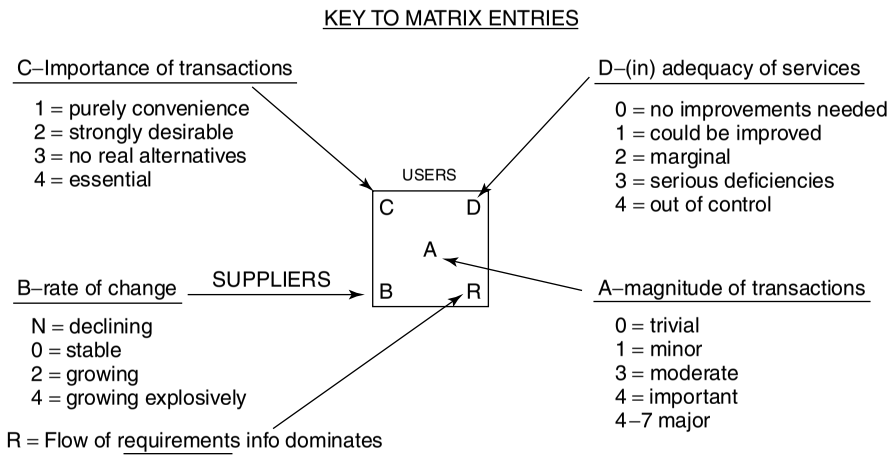


Figure 3. (*Continued*).

the money expended on the whole system. Comparatively little money goes into calibration, education and research of the basic fundamentals of measurement using scientific principles and its technology. 'Fit and forget' measurement technology might be the desire of management but, as with all uses of technology, it is not a reality. Measuring instruments are no exception and they are certainly not perfect.

Throughout history, there have been occasions when the governing power of nations have considered that certain measurements or measuring instruments would be of great benefit to the nation.

In 1598, the Spanish monarch, Philip III, offered a large sum of money to any person who could produce a timekeeper that would reduce navigational inaccuracies across the Atlantic Ocean regions. In England, Harrison was awarded £20 000 from the British Government who had offered this huge reward, in 1714, for a similar reason.

On another occasion, James Cook took the voyage to southern regions of the globe to measure the transit of the planet Venus in June 1769. This too was concerned with improving navigational accuracy. Here, it was the Royal Society of London that induced the King and the Admiralty to expend a considerable sum of money conducting this one measurement.

In summary, it is relatively easy to make a case for spending money on measurement for clear-cut scientific, knowledge-seeking reasons. The long-term less quantifiable benefits, however, are often seen in their historical perspective as being more significant than the originally argued case; these are the benefits that management would like to be able to cost in money terms.

RELATED ARTICLES

Article 3, Measures and Metrics; Their Application, Volume 1; **Article 5, Humans in the Real World, Volume 1**; **Article 64, Executing A Measuring System Design, Volume 2**; **Article 65, Life Cycle Concept, Volume 2**.

REFERENCES

DTI (2003) UK National Measurement System. URL: http://www.dti.gov.uk/nms/.

Huntoon, R.D. (1967) Concept of a National Measurement System. *Science*, **158**, 67–71.

Minck, J. (1977) The National Measurement System – The Hidden Giant. *NCSL Newsletter*, **17**, 3–5.

NIST (2003) National Institute of Science and Technology, USA. URL: http://www.nist.gov.

NMLTaiwan (2003) National Measurement Laboratory, Taiwan. URL: http://www.nml.org.tw/en/NMS/nms_1.html.

NSC (2003) National Standards Commission, Sydney, Australia. URL: http://www.nsc.gov.au/PAGES/Nms/nms.html.

5: Humans in the Real World

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 General Remarks on Humans, Machines, and Interfaces 33
2 The Overall Human, Machine, and Environment Model 33
3 Humans and Technical Instruments – an Evaluation and Comparison 34
4 Human and Information Machine Model – The Crucial and Central Measurement Link 36
5 The Two Problems of Measurement 36
6 Direct and Indirect Measurement 36
7 Inferential Measurement 37
Related Articles 37
References 37

1 GENERAL REMARKS ON HUMANS, MACHINES, AND INTERFACES

From the earliest times, humankind has attempted to control its environment. This compulsion has required the extension of the human abilities to sense the environment and so to perform useful work upon its materials and substances. Human-made implements and artifacts are the means by which these faculties are extended. A generic name, which includes all of these items, is embodied in the concept of *machines*. There are two main kinds of machines. The first, which extends the human ability to do useful work, may be referred to as *energy handling* or *energy transforming machines*. A second group of machines to extend human sensing, given much emphasis by the present revolution is *information handling*, constituting what may be called *information transforming machines*. Information handling machines may also be referred to as *information machines* in short or also as *technical instrumentation*.

2 THE OVERALL HUMAN, MACHINE, AND ENVIRONMENT MODEL

The important interfaces between humans, machines, and their environment are illustrated in Figure 1 (McGhee *et al*., 1996, 1998). The human biosystem is a complex combination of muscles and senses. Muscular activity is associated with effort and effort with work. The practical limitations on the ability of human muscle power to work have led to the extension of almost all humans working in complex industrial societies using machines that amplify the human ability to gain mechanical advantage. Simple examples of this are the spades, pickaxes, and spanners used in such diverse activities as cultivation, road repair, and motor mechanics. More sophisticated examples used in the same activities, which allow orders of magnitude increase in the mechanical advantage, are tractors, mechanical diggers, and pneumatic spanners. The impact that these *energy handling* machines have had upon human activity is mirrored in those machines allowing extension of the human abilities of sensing and perception. For example, although humans can sense hotness relations through the tactile and visual senses, they require extension when the objects under observation are either very hot or very cold. Thus, the need arose for inanimate implements to extend this human sensing ability. So it is with the other human senses as well.

Because of the inextricable link between humans and their environment, and between humans and machines, the ordering of these interfaces helps to explain the links given in Figure 1. Firstly, there is always a direct link between

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

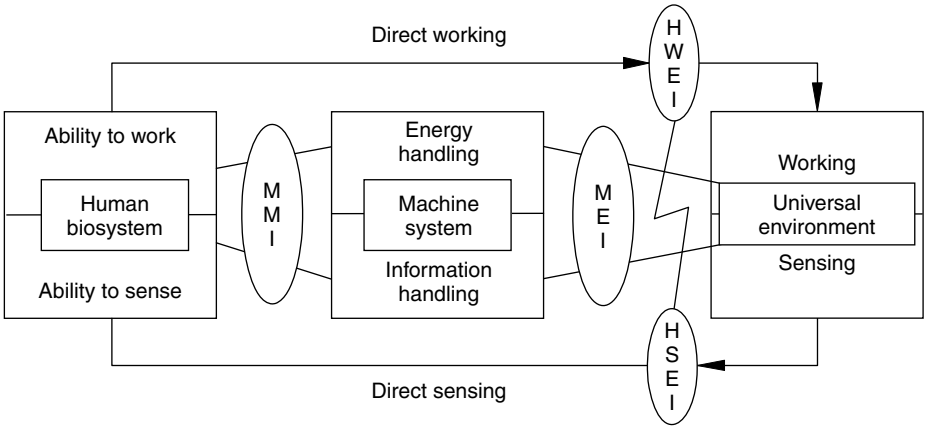


Figure 1. Interfaces between humans, machines, and the universal environment.

humans and the universal environment. This interface has two component parts. One of these, which may be called the *human-working environment interface*, or (HWEI), explains the activities undertaken by humans in performing useful work as part of the process of providing sufficient resources for survival. Related to this, to a greater or lesser extent, is an interface called the *human-sensing environment interface*, or HSEI. This embodies the interacting link between the human senses, which interface with the universal environment for the purposes of gathering information. From the beginning of the machine era to the present day and beyond, an interface has existed between humans and machines. Known as the *human–machine interface*, or MMI, it provides the basis for humans to use machines to extend their abilities to perform useful work by acquiring some form of mechanical advantage, as well as supplementing the human abilities of sensing and perception. The third interface, allowing machines to interact with the universal environment, is called the *machine environment interface*, or MEI.

These interfaces between humans, machines, and their environment give a clear picture of the primary links between them. It also indicates that instruments should be studied, analyzed, and designed on the basis of how humans interact with them and how they interact with the environment within which they extend the human abilities to handle information.

3 HUMANS AND TECHNICAL INSTRUMENTS – AN EVALUATION AND COMPARISON

It is instructive to compare the human sensing system with human-made technical sensors. This comparison is summarized in Figures 2 and 3 (McGhee, Henderson and Sydenham, 1999), which are freely adapted from Hofmann (1982). The five human senses of seeing, hearing, tasting, smelling, and feeling, which involve some 122 million sensing receptors, can be classified by the energy form of the observed variable using the COMETMAN acronym. The origin of this acronym and the classification of energy are given in **Article 109, Structure and Energy in Sensor Systems, Volume 2**. From the information processing point of view, humans can process information at a rate of around 100 bps with an unconscious information transfer rate of around 1 Gbps from some 4 million nerve strings linked through the human nervous system to the cortex or brain stem.

Figure 2 shows that a biological sensing system is characterized by

- an abundance of sensors also called *receptors*,
- a very large number of message channels,
- a parallel processing capacity,
- a complex interacting structure with fusion of sensors,
- an ability to learn and adapt.

In the case of technical sensor systems, it is their clear function to extend and enhance the human senses. Although they execute this intention in most cases, Figure 3 shows that they are nevertheless characterized by

- a limited range of basic sensors,
- a small number of message channels,
- a series processing of information,
- a transparent physical structure, which is neither complex nor interacting, nor is there extensive fusion of sensors,
- an as yet underdeveloped ability to learn and adapt.

It is fair to say that the advent of large-scale integration of both analog and digital processing circuits, accompanied by the use of microprocessors with large memory capacity,

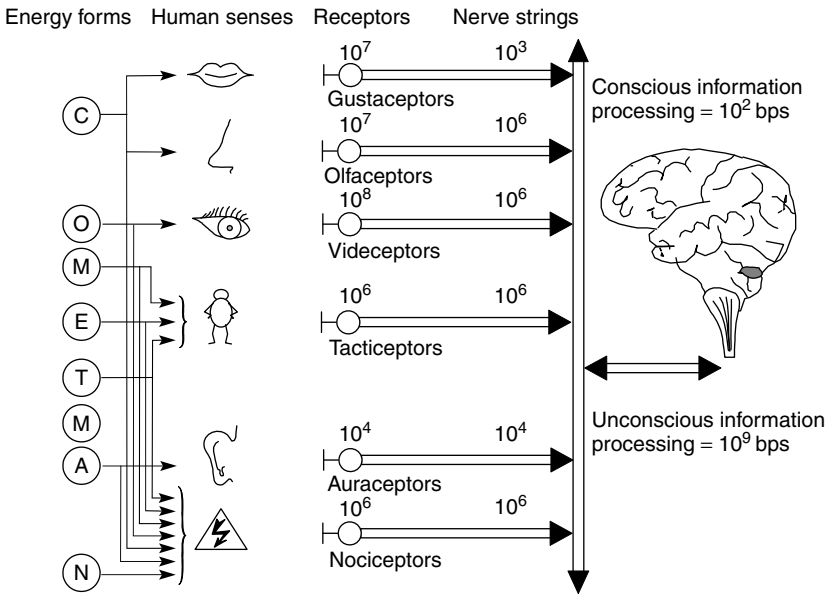


Figure 2. The human biological sensing system including sight, hearing, tasting, smelling, touching, and pain.

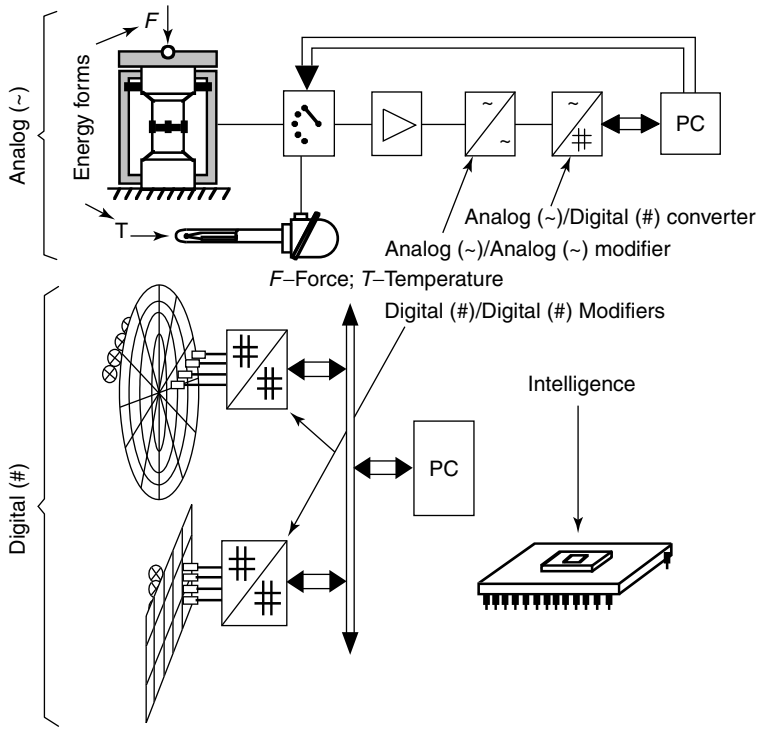


Figure 3. Technical sensor systems.

has led to tremendous changes in *technical instrumentation*. Even so, there is broad scope for further development.

Figure 3 shows that almost all of the measuring system is actually handling data, even in the strictly analog parts of the instrument such as the sensor, where an information-bearing signal is converted into another signal in a different energy domain. Hence, measurement technology should perhaps be renamed as *data measurement* or DM. In contemporary measurement, because of the influence of computers, there is an inexorable trend toward the use of a minimum number of analog components. For this reason, it seems appropriate to view measurement technology as data measurement. Data measurement is to measurement as data communication is to communication.

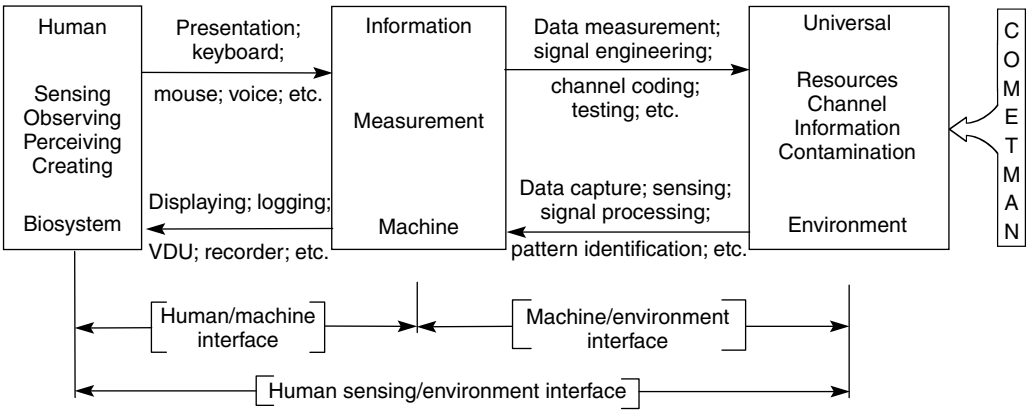


Figure 4. Block diagram of human and measurement information machine in a universal environment.

4 HUMAN AND INFORMATION MACHINE MODEL – THE CRUCIAL AND CENTRAL MEASUREMENT LINK

Information technology, or IT, has produced radical changes through the pervasive influence of the personal computer, or PC. Each increase in the system clock frequency and corresponding improvement in the information machine input/output, or I/O cards opens up endless measurement applications.

From the functional point of view, the relationship between the information machine aspects of Figure 1 are given in more detail in Figure 4 (McGhee, Henderson and Sydenham, 1999). The human faculties of sensing, observing, perceiving, and creating are visualized as being extended using a measuring information machine. Information flows from the human through the MMI, which consists of the PC keyboard, the human voice, the mouse, and so on. In the reverse direction, information is communicated to the human using various types of displays, Visual Display Units (VDUs) as well as data loggers and recorders. The measuring information machine enhances the human faculties by assisting in the generation of suitable coded interrogating signals for testing purposes. This process is one of the aspects of *data measurement*. Technical sensors, of the same general diversity as in Figure 3, enhance the human abilities to capture information associated with the flow of energy in the universal environment under observation.

5 THE TWO PROBLEMS OF MEASUREMENT

It is important to notice the distinction between the two main types of measurement in the above model of measurement. Measurement in technical science has two main purposes. The first purpose is to acquire data about the static or equilibrium properties of real objects or events. The second is to acquire data about the dynamic behavior of these real objects or events. These may be referred to as *direct* and *inferential* measurement respectively (Finkelstein, 1994; McGhee, Henderson and Sydenham, 1999). They may be referred to as the two problems in measurement. Both require the use of sensors to measure the states of equilibrium and motion of solids, liquids, gases, and the systems they constitute. A primary goal in both these measurement processes is to take account of errors, their sources, and their quantification (Solopchenko, 1994; McGhee, Henderson and Sydenham, 1999). The first measurement problem is to acquire data regarding the static or equilibrium properties of some real object or event. Acquiring data about the dynamic behavior of these objects or events is the second problem. Thus, real systems, objects, or components may be measured from two different points of view known as *direct measurement* and *inferential measurement*.

6 DIRECT AND INDIRECT MEASUREMENT

Direct measurement *per se* involves the process of directly *sensing* those active observable quantities characterizing the flow of energy and information (Finkelstein, 1994). Figure 3 (McGhee, Henderson and Sydenham, 1999) shows a typical structure for direct measurement using only two analog sensors and two digital sensors for simplicity. A load cell is used to measure an applied force F, while an electrical temperature sensor is used to measure a temperature, T. Optical methods for digitally measuring linear and angular positions are also shown. The instrumentation also includes a diagrammatic representation of multiplexing, amplification, and conversion under the control of a PC.

Indirect measurement uses directly measured quantities to compute the value of another, ostensibly nonmeasurable variable. The measuring instrument captures a signal carrying information about the flow of energy or matter associated with the system or event. Signal processing is used in both cases to extract the information.

7 INFERENTIAL MEASUREMENT

Figure 5 gives a block diagram of the inferential measurement problem. All systems may possess any one, or a combination of the characteristics associated with storage, transformation, transmission, or dissipation of energy. These system characteristics determine the dynamic behavior of the process whose dynamic properties can only be verified by measurement with the help of interrogative testing, also referred to as *inferential measurement* or *identification.* The dynamic characteristics of a system can be described by models. A model of a physical process is a representation of its style, form, or pattern. Hence, inferential measurement or identification consists of constructing a model of a dynamic system. Modeling is based upon input and output observations or measurements, as well as on prior knowledge. Dynamic models have significance in many engineering and nonengineering systems. Frequently, they play a vital role in the proper analysis and design necessary for the efficient operation of the modeled system. Models are used in physics, biology, astronomy, engineering, economics, sociology, psychology, medicine, ecology, agriculture, as well as others. A boundary view of the identification problem is given in **Article 68, Setting the System Boundaries, Volume 2**. Modeling and model building as an essential ingredient of sensor science is justified in **Article 107, Principles of Sensor Science, Volume 2**. Two-space, three-space, and tetrahedron modeling of sensors is described in **Article 109, Structure and Energy in Sensor Systems, Volume 2**. Signal/Energy matrix modeling of sensors is considered in **Article 110, Signal/Energy Matrix Modeling, Volume 2**.

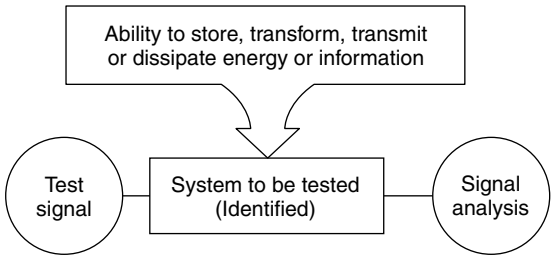


Figure 5. A block diagram illustrating inferential measurement.

Humans have a compelling urge to sense and test. Classifying and synthesizing signals is a very important activity in inferential measurement. In its engineering context, measurement can lead to improved design, since physical processes become better known. This knowledge also leads to more efficient energy utilization and product quality by allowing effective control.

RELATED ARTICLES

Article 2, Organization of Instrument Science, Volume 1; **Article 6, Substructure of Human–Machine Systems, Volume 1**; **Article 95, Human–Machine Interface, Volume 2**.

REFERENCES

Finkelstein, L. (1994) Measurement and Instrumentation Science – An Analytical Review. *Measurement*, **14**(1), 3–14.

Hofmann, D. (1982) Automatic Testing with Intelligent Sensor Systems – Measurement or Classification, in *ACTA IMEKO IX* (p. 545).

McGhee, J., Henderson, I.A. and Sydenham, P.H. (1999) Sensor Science – Essentials for Instrumentation and Measurement Technology. *Measurement*, **25**, 89–113.

McGhee, J., Henderson, I.A., Korczynski, M.J. and Kulesza, W. (1998) The Sensor Effect Tetrahedron: An Extended Transducer Space. *Measurement*, **24**, 217–236.

McGhee, J., Henderson, I.A., Kulesza, W. and Korczynski, M.J. (1996) *Scientific Metrology*, ISBN 83-904299-9-3, printed by A.C.G.M. LODART, Lodz.

Solopchenko, G.N. (1994) Formal Metrological Components of Measuring Systems. *Measurement*, **13**, 1–12.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

6: Substructure of Human–Machine Systems

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Boundary Perspective of Human–Machine Systems 38
2 Resources 38
3 Information 38
4 Contaminants or Influence Effects 39
5 The Information/Energy Tetrahedron 40
6 Substructure of Human–Machine Systems 41
Related Articles 41
References 41

1 BOUNDARY PERSPECTIVE OF HUMAN–MACHINE SYSTEMS

As defined in **Article 5, Humans in the Real World, Volume 1**, instruments are human-made systems. Between men and machines there is an ever-present relation, whose manifestation alone does not constitute a system. Systems are understood to have an opaqueness and obscurity of structure, characteristic of complexity. Although complexity is a necessary property, its possession is insufficient for an object to be classified as a system. The object must also have a purposeful function. Hence, a system is a purposefully directed labyrinth of complex operations. It will be seen later that a system also possesses recursive properties. Measurement occurs within the context of systems (**Article 5, Humans in the Real World, Volume 1**). An exemplary aspect of this systems approach is the totality of the considerations, which are taken into account. Thus, systems are examined on the basis of the bilateral influences and effects, which occur between them and the wider system within which they operate.

The primary functions of instruments inextricably link them with humans (**Article 5, Humans in the Real World, Volume 1**). A human–machine system is the context within which men and machines are united in the execution or pursuit of aims. Hence, human–machine systems provide an appropriate starting point for the introduction of instrumentation. These systems do not exist in isolation. The environment, within which they function, regards them as subsystems of itself, the wider system. Hence, the bilateral impacts of this linking must be considered. A boundary perspective of these systems, shown in Figure 1 (M'Pherson, 1980; McGhee, Henderson and Sankowski, 1986), allows a distinction between the three groups of inputs and outputs.

Inputs and outputs in human–machine systems may be grouped as *Resources*, *Information*, or *Contamination*.

2 RESOURCES

Manpower, Money, Machines, and Materials are often called the 4M's of systems. *Manpower* is remunerated by a *Money* resource, which is also used to purchase *Machines* and *Materials*. Actuators and tools (forms of *energy handlers or transformers*) are machines that supplement the human capacity to do work. Those that extend the human faculties of sensing and perception, belong to the general class of *information machines*. Raw materials support production at all stages and levels. The transformation of input resources into consumable output resources is an aim of human–machine systems. Electrical power is the most common form of resource for modern instrumentation.

3 INFORMATION

This could be considered as the most important in the measurement context. There are three main classes of

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

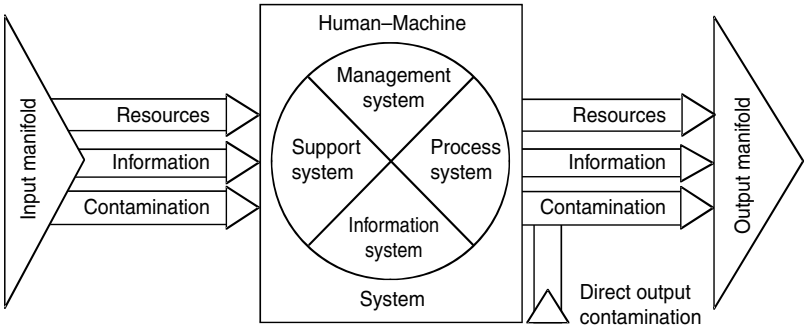


Figure 1. Boundary perspective of a human-machine system.

information. *Subjective information* is the most nebulous since it is embodied in those human perceptions resulting from intuitions, opinions, and feelings. As has been stated above, technical instruments extend the human faculties of sensing and perceiving and so enable humans to acquire *Objective information* about the physical universe.

The theories of Shannon and Weaver (1972), which permit the treatment of information handling processes with a higher level of abstraction, provide the means of quantifying objective information. In this theory, there are two coding theorems, which, although developed for communications instrumentation systems, are also important for the other instrumentation systems. One theorem states the important principles for an economical, coded representation of source data, that is, each source alphabet should contain the minimum number of alphabet symbols. Although this alphabet should be used as economically as possible in the formation of code words, the code words must still be uniquely and instantaneously decodable, thus ensuring no ambiguity. The other theorem refers to the coding of source coded signals for transmission through interference generating channels. A pretransmission coding of these signals should be such that their statistics optimally match the statistics of the communications instrumentation channel, so allowing high fidelity information transmission. Hence, two levels of coding are evident. The limits, beyond which it is impossible to advance, are usually referred to as *Fundamental Limits* in all physical processes (IEEE, 1981).

Fuzzy information, which was first formulated by Lotfi Zadeh (1984) in the 1960s, provides a means of quantifying the possible meanings for the same group of symbols. It may be concluded that fuzzy events should be treated as independent, 'fuzzy Shannon sources'. The average information of these 'fuzzy Shannon sources' has one component associated with the randomness of the two events in a binary alphabet. The other specifies the average information resulting from the fuzziness of the fuzzy set relative to its ordinary or binary set.

4 CONTAMINANTS OR INFLUENCE EFFECTS

Contamination is properly regarded as implying impurity and interference. Three types of contaminants (or influence effects) are (i) *disruptions*, which could have drastic or catastrophic effects due to natural disasters, for example, climate, (ii) *social factors* such as wars, strikes, or prohibitions and (iii) *failure of resources* due to impurities and effluents from various sources.

Bilateral Power Flows occur along the interconnections that link a human–machine system with its environment. Apart from feedstock-power flows, which are essential for correct operation of the system, other types of power flow can cause interference within the system resulting in poorer system performance. This type of contamination may be referred to as a *disturbance power flow*. In most situations, as mentioned above, a limitation, called a *fundamental limit*, exists. These are of interest in all systems. As they are of special interest in information theory, they are of interest in all information handling processes. If the information handling operation is measurement, then *fundamental measurement limits* exist due to the equipartition of energy, which is the most common source of *fundamental fluctuations*. McGhee, Henderson and Sankowski (1986) give a list of references.

Components that are used to build systems, are not immortal. The aging of the physical elements gives rise to a change in their material properties, which can affect performance. If burnout occurs, then this type of failure may cause a disruption. Parameters of the system, which are instrumental in determining the performance of the system, give the system '*parameter-sensitivity*'. *Sensitivity analysis* (Frank, 1978; Lojek, 1982) provides a means of assessing the effects of all impacts and parameters upon the performance of systems. In general, all of these sources of contamination must be considered in instrument systems. The boundary perspective of human–machine systems provides a basis for highlighting all effects impacting upon systems. Hence, it indicates the importance of a detailed

consideration of all causes and effects at the boundaries of any system (Sandquist, 1985).

5 THE INFORMATION/ENERGY TETRAHEDRON

It is also possible to visualize the range of possible information and energy forms for the groups of inputs discussed above. A way of providing this view is shown in Figure 2 (McGhee *et al*., 1996, 1998) using the COMETMAN acronym. This acronym is one of the ways of classifying energy types described in much fuller detail in **Article 109, Structure and Energy in Sensor Systems, Volume 2**.

A spatial representation for the boundary inputs and outputs of a human–machine system in its most general form requires a multidimensional space. As there is a total of three groups of inputs and three groups of outputs, the representation becomes too complex for graphical illustration. Each of the input and output groups may be specified as information bearing signals or energy flowing forms. Figure 1 greatly simplifies the possibilities by visualizing the output energy/signal form for an information machine, or more specifically a measurement system. In such a system, the main output is an information output, which also corresponds to the resource output. Although contamination is also produced at the output, it is not necessary to consider it for the present purposes.

Along the base plane of the spatial representation introduce a topological grouping consisting of three axes. Each of these axes is taken as representing one of the three input groups with their COMETMAN classification. This organization of inputs on the base plane may be regarded as *the input triangle*. Such a name is appropriate, as illustrated by the *input triangle* given by the dotted lines in Figure 2. Here the support energy form is electrical, the contamination energy/information is chemical, and the measurand energy form is mechanical using a sensor like a resistance strain gauge. The chemical contamination could be responsible for parasitic electrochemical potentials at the contacts in the electrical circuit. Each point on the *input triangle* axes may be joined to an energy/information form on the output axis. When the output is in the form of an information bearing electrical signal, the dash-dot line combinations shown in Figure 2 allow the link between all three input forms and the output form to be clearly seen. The geometrical shape, which is outlined by the dotted and the dash-dot lines, has one base plane surface and three inclined surfaces. This represents a tetrahedron or four-faced object. This way of representing the boundary impacts on systems may be called the *information/energy tetrahedron* when used in the context of systems. When used in the context of sensors it may be called the *sensor effect tetrahedron*.

Illustration of the manner in which the contamination inputs influence the output may be extended to cover contamination from other energy forms. The dot-dot-dash lines in Figure 2 show the case when the contaminating energy form is due to the influence of temperature on the wire of the strain gauge. The deep perception allowed by the *information/energy tetrahedron* given in Figure 2, which is

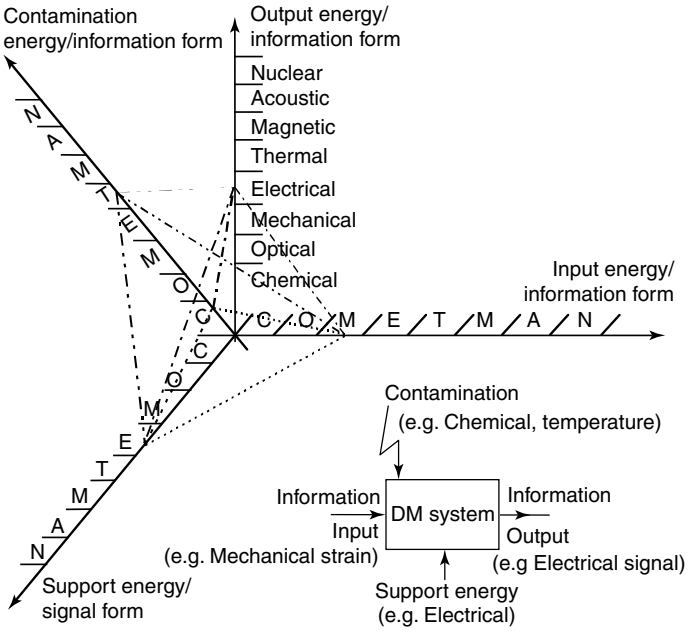


Figure 2. The information and energy tetrahedron.

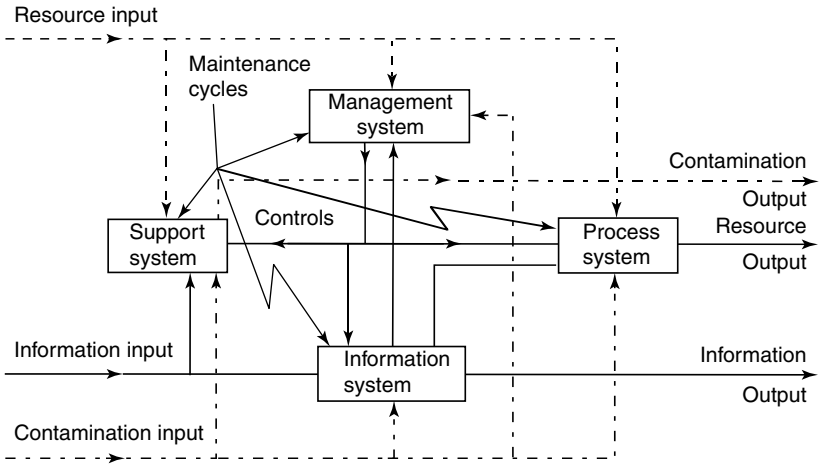


Figure 3. Interconnecting substructures in a human-machine system.

fairly straightforward, is still realistic and comprehensive. The use of the information/energy tetrahedron in sensor systems is given a fuller treatment in **Article 109, Structure and Energy in Sensor Systems, Volume 2**.

6 SUBSTRUCTURE OF HUMAN–MACHINE SYSTEMS

Human–machine systems may be divided into the four main subsystems that are indicated in Figure 2 and described by McGhee, Henderson and Sankowski (1986), M'Pherson (1980) and McGhee *et al*. (1996). The interconnection between these subsystems is shown in Figure 3. A *Process system*, which can produce a useful resource output, provides the means by which the aim of the human–machine system is realized. Controls and instructions, generated by a *Management system*, with the aid of an *Information system*, are required for the proper coordination and execution of all system functions at all levels. The *Information system* is responsible for acquiring and processing all of the information required to permit effective operation of the complete system, through the management system. Supplying feedstock and energy is the responsibility of a *Support system*, which also provides the human–machine system with its functional means of repair and survival. Survivability means that the system will continue to operate in a hostile environment and will complete its mission or operation. Survivability in the context of microprocessor based real-time operating systems is important. An ability for structural modification may be required. Reliability analysis provides a method of assessing the probability that a system will complete its mission successfully.

RELATED ARTICLES

Article 2, Organization of Instrument Science, Volume 1; **Article 5, Humans in the Real World, Volume 1**; **Article 16, Reduction of Influence Factors, Volume 1**; **Article 95, Human–Machine Interface, Volume 2**.

REFERENCES

Frank, P.M. (1978) *An Introduction to System Sensitivity Theory*, Academic Press, London.

IEEE (1981) Fundamental Limits in Electrical Engineering. Special edition of *Proceedings of IEEE*, **69**(2).

Lojek, B. (1982) Sensitivity Analysis of Non-Linear Circuits. *IEE Proceedings*, **129G**, 85–88.

McGhee, J., Henderson, I.A. and Sankowski, D. (1986) Functions and Structures in Measurement Systems: A Systems Engineering Context for Instrumentation. *Measurement*, **4**(3), 111–119.

McGhee, J., Henderson, I.A., Kulesza, W. and Korczynski, M.J. (1996) *Scientific Metrology*, ISBN 83-904299-9-3, printed by A.C.G.M. LODART, Lodz.

McGhee, J., Henderson, I.A., Kulesza, W. and Korczynski, M.J. (1998) The Sensor Effect Tetrahedron: An Extended Transducer Space. *Measurement*, **24**, 217–236.

M'Pherson, P.K. (1980) Systems Engineering: An Approach to Whole-System Design. *Radio and Electronic Engineering*, **50**, 545–558.

Sandquist, G.M. (1985) *Introduction to System Science*, Prentice-Hall, Englewood Cliffs, NJ.

Shannon, C.E. and Weaver, W. (1972) *A Mathematical Theory of Communication*, University of Illinois Press, Illinois.

Zadeh, L.A. (1984) Making Computers Think Like People. *IEEE Spectrum*, **12**, 26–32.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

7: Introduction to Measurement Theory and Philosophy

Ludwik Finkelstein
City University, London, UK

1 The Basic Concepts 45
2 Outline of the Historical Development of Measurement Theory 46
3 Nature and Properties of Measurement 46
References 47
Further Reading 47

1 THE BASIC CONCEPTS

Measurement can be defined in the wide sense as a process of empirical, objective assignment of symbols to attributes of objects and events of the real world in such a way as to represent them, or to describe them.

Description, or representation, means that when a symbol, or measure, is assigned by measurement to the property of an object, and other symbols are assigned by the same process to other manifestations of the property, then the relations between the symbols or measures imply, and are implied, by empirical relations between the property manifestations.

Objective process in the definition of measurement means that the symbols assigned to a property manifestation by measurement must, within the limits of uncertainty, be independent of the observer.

An empirical process in the definition of measurement presented above means, first, that it must be the result of observation and not, for example, of a thought experiment. Further, the concept of the property measured must be based on empirically determinable relations and not, say, on convention.

This wide definition of measurement is often disputed by those who consider the paradigm of measurement in the physical sciences as normative, or, at least, they require measurement to be a numerical representation in which reflects an order.

For this reason, it is convenient to distinguish between strongly and weakly defined measurements. Strongly defined measurement is defined as a class of widely defined measurement, which follows the paradigm of the physical sciences. In particular, it has precisely defined empirical operations, representation by numbers, and well-formed theories for broad domains of knowledge.

Measurement that constitutes representation by symbols of properties of entities of the real world, based on an objective empirical process, but lacks some, or all, of the above distinctive characteristics of strong measurement, may be termed weakly defined (Finkelstein, 2003).

It remains to explain the terms *theory of measurement* and *philosophy of measurement* as used in this introduction.

Measurement theory is the study of the logical and epistemological basis of representation of property manifestations by symbols.

Measurement philosophy is the study of the basic concepts underlying measurement, embracing measurement, as defined above, and concerning the relation of measurement to language, reasoning, knowledge, and the like.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

2 OUTLINE OF THE HISTORICAL DEVELOPMENT OF MEASUREMENT THEORY

The practical pursuit of measurement arose in the earliest stages of the material culture of man. However, it was based on an intuitive understanding of the process.

The ancient Greeks were the first to investigate the philosophical foundations of measurement. The schools of Pythagoras, Plato, and Aristotle were concerned with the nature of number, quantity, and their relation to the real world.

The Middle Ages saw much scholarly study of the theory of measurement, though scholars were not concerned with the application of measurement to observation of the physical world.

The rise of modern science, from Galileo, through Newton to Lagrange and Maxwell, put mathematics at the heart of physics, applied measurement to physical observation, and developed comprehensive mathematical theories of domains of physics. These developments were, however, not accompanied by analysis of the fundamental concepts of measurement.

Helmholtz (1887) laid the true foundations of the modern theory of measurement in a thorough logical analysis of the epistemology of counting and measuring. Hölder (1901) further developed this.

Campbell (1920) provided a lucid and thorough analysis of the fundamental basis of the measurement of physical quantities and his theory became generally accepted.

The classical theory of measurement from Helmholtz to Campbell was concerned with physical measurements. The theory was inapplicable to the social and psychological sciences. A report of a committee of the British Association for the Advancement of Science, published in 1940, which considered quantitative methods, rejected the possibility of psychological measurements.

These rigid positions of the classical theory of measurement were broken down by work in the social and psychological sciences. The concern of the social sciences with the concept of utility led to an axiomatic theory of utility that supported development of a new measurement theory. In psychology, Stevens (1951) carried out much fundamental work on developing an appropriate analysis of the nature of measurement.

The proceedings of a conference in the United States in 1959 (Churchman and Ratoosh (1959)) presented a review of the classical approaches to measurement as extended to needs of the social and behavioral sciences.

Modern formal measurement theory may be said to be based on the work by Tarski (1954) on relational systems and model theory, broadly describable as concerned with the representation by symbols of a formal language.

Suppes and Zinnes (1965) provided an early, clear exposition of the theory, in the development of which Suppes has been one of the key workers. The representational theory of measurement was then developed and well documented in a group of important books published between 1971 and 1990. The principal works are those of Pfanzagl (1968), a very detailed and thorough account of the representational foundations of measurement in three volumes by Krantz *et al*. (1971, 1990) and the important treatises of Roberts (1979) and of Narens (1985)

The representational theory has now been accepted widely in the social and behavioral sciences. However, while it embraces physical measurement, the theory has not found a recognized place in the theory and practice of physical sciences.

Modern measurement theory has been debated in the measurement and instrumentation technology community since the 1970s.

Representational measurement theory has not progressed significantly in recent years, as it seems to have proved adequate for the purposes for which it has been developed.

However, the rapid development of symbolic computation and soft computing has offered prospects for the further development and extension of general symbolic representation.

3 NATURE AND PROPERTIES OF MEASUREMENT

The properties of measurement arising from its definition as a process of empirical, objective, assignment of symbols to properties of objects or events of the real world in such a way as to describe them will now be analyzed and discussed. They are an explanation of the primacy of measurement in science.

Measurement provides an objective description of the measurand. It is thus invariant in rational discourse. The description is not merely a matter of opinion or feeling.

Measurement is based on a well-defined empirical process of observation. It is thus a basis of justified, true belief; in other words it is the basis of true knowledge.

Measurement is not naming. It provides descriptions of relations of the property manifestation measures to other manifestations of the same property. The value of a measurement process depends upon the richness of the relations it can represent.

Measures are descriptions of great conciseness. A single number tells us what it would take many words to express.

Measurement gives, further, a description that is precise, pinpointing by a single number a particular entity, where a verbal description indicates a range of similar but differing things.

Measurement is description by a well-defined symbolism. A measure of a property gives us an ability to express facts and conventions about it in a formal symbolic language. Without the convenient notation of such a language, the complex chains of induction and deduction by which we describe and explain the universe would be too cumbersome to express.

It follows from what has been said that description by symbols is not good in itself. The only value of measurement lies in the use to which the information is put. Science is not just the amassing of numerical data; it depends upon the way in which the data are interpreted, analyzed, and organized.

Finally, measurement describes measurands by symbols, which can be realized as signals. They can be acquired, processed, and effectuated by information machines.

REFERENCES

Campbell, N.R. (1920) *Physics: The Elements*, Cambridge University Press, Cambridge; reprinted 1957, *Foundations of Science*, Dover Press, New York.

Churchman, C.W. and Ratoosh, P (eds) (1959) *Basic Concepts of Measurements*, Cambridge University Press, Cambridge.

Finkelstein, L. (2003) Widely, Strongly and Weakly Defined Measurement. *Measurement*, **34**, 39–48.

Helmholtz, H.V. (1887) *Zählen and Messen Erkenntis – Theoretisch Betrachet*, Philosophische Aufsaetze Eduard Zeller Gewidmet, Leipzig; translated by C.L. Bryan (1930) *Counting and Measuring*, Van Nostrand, New York.

Hölder, O. (1901) Die Axiome der Quantität und die Lehre vom Mass. *Berichte ueber die Verhandlungen der koeniglich Saechsichen Geselschaft der Wissenschaften zu Leipzig, Math-Phys Klasse*, **53**, 1–64.

Krantz, D.R., Luce, R.D., Suppes, P. and Tversky, A. (1971) *Foundations of Measurement*, Vol. 1: 1971; Vols. 2, 3: 1990, Academic Press, New York.

Narens, L. (1985) *Abstract Measurement Theory*, MIT Press, Cambridge, MA, London.

Pfanzagl, J. (1968) *Theory of Measurement*, Physica Verlag, Würzburg, Vienna.

Roberts, F.S. (1979) *Measurement Theory with Applications to Decision Making, Utility and the Social Sciences*, Addison-Wesley, Reading, MA.

Stevens, S.S. (1951) *Handbook of Experimental Psychology*, Wiley, Chichester.

Suppes, P. and Zinnes, J.L. (1965) Basic Measurement Theory, in *Handbook of Mathematical Psychology* (eds R.D. Luce, R.R. Bush and E. Galanter), Wiley, New York.

Tarski, A. (1954) Contributions to Theory of Models. *Indagationes Mathematicae*, **16**, 572–588.

FURTHER READING

Finkelstein, L. (2000) Foundational Problems of Measurement, in *Measurement Science – A Discussion* (eds K. Kariya and L. Finkelstein), Ohmsha Press, Amsterdam (pp. 13–21).

8: Formal Theory of Measurement

Ludwik Finkelstein
City University, London, UK

1 Philosophical and Mathematical Basis of the Formal Theory of Measurement 48
2 The Concept of Quality 48
3 Representation 49
4 Some Direct Scales of Measurement 50
5 Uniqueness: Scale Classification and Meaningfulness 50
6 Uncertainty 51
References 51

1 PHILOSOPHICAL AND MATHEMATICAL BASIS OF THE FORMAL THEORY OF MEASUREMENT

The formal theory of measurement has its origins in the needs of psychology and decision sciences. However, its form has been influenced by developments in philosophy.

The approach of the theory is in its essence based on two philosophical movements: logical positivism and the philosophy of logic.

Logical positivism was an important philosophical movement initiated in the so-called Vienna Circle of philosophers in the early twenties of the twentieth century. It sought to base all knowledge on empirical observation and logical deductions from them. Central to it was the verification principle: the notion that the meaningfulness of statements lies in the specification of the empirical steps required to verify them. In that view, data obtained by empirical observation are foundational and knowledge is built bottom-up from them. Closely allied to it is the view of operationalism in the philosophy of science, which defines concepts in terms of precisely defined empirical operations. While those views have run their philosophical course by the sixties of the twentieth century, they remain influential in scientific thinking.

The other movement on which the formal theory of measurement is founded is the development of modern logic. From the work of Boole in the mid-nineteenth century, it developed mathematical formalisms, such as the notions of symbolic representation, sets, relations, mappings, and the like to represent valid reasoning. It culminated, with respect to measurement, in the work of Tarski (1954), on model theory, concerned with the representation of extralinguistic systems by symbolic systems.

The theory of measurement defines the concept of measurand quality as an empirical relational system based on empirically observed data and relations. It maps it into a symbol system based on the notions of sets and relations. It considers the representation of the quality system by the symbol system and its uniqueness. The foundation of the underlying concepts of measurement on empirical observation gives this theory of measurement a strong validity. The expression of the concepts and reasoning in the formalism of mathematical logic gives it its rigor.

The formal theory of measurement is presented here concisely. Detailed presentations are given in Krantz *et al*. (1971, 1990), Roberts (1979). A wider discussion is given in Finkelstein (2000).

2 THE CONCEPT OF QUALITY

Measurement presupposes something to be measured. Both in the historical development and logical structure of scientific knowledge, the formulation of a theoretical concept

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

or construct, which defines a quality, precedes the development of measurement procedures and scales.

The basic notion is that of a manifestation of a quality, an abstract, a single-sensed aspect of an object or event, such as, for example, the smell of a substance. Observation of the real world leads to the identification of empirical relations among these single manifestations. Examples of such relations are similarity and difference. As a result, the concept of a quality is formed as an objective rule for the classification of a collection of empirically observable aspects of objects into a single set, together with the family of objective empirical relations on that set. The resulting relational system is a quality, and each single member of the set is termed a manifestation of the quality.

We can thus see that there is a difficulty in the measurement of such qualities as beauty. The existence and meaningful use of the word beauty indicates the usefulness of the concept. However, there is not an objective rule for classifying some aspect of observable objects as manifestations of beauty. Similarly, there are no objective empirical relations such as indistinguishability or precedence, in respect to beauty. The basis for measurement of beauty is thus absent from the outset.

When there exists a clearly defined quality, as a set into which its manifestations can be objectively and empirically classified, together with a set of empirical relations, then we can always find some symbolical relational set by which it can be represented.

Consider some quality and let q_i represent an individual manifestation of the quality, so that we can define a set of all possible manifestations as

$$Q = \{q_1 \ldots\} \tag{1}$$

Let there be on Q a family R of empirical relations R_i

$$R = \{R_1, \ldots, R_n\} \tag{2}$$

Then, the quality is represented by an empirical relational system

$$Q = \langle Q, R \rangle \tag{3}$$

Finally, let us explain the concept of an empirical quantity. If there is an order relation in the quality relational system, enabling us to order quality manifestations in a way that has formal similarity to the relations equal, greater, and less, then the quality is termed a quantity.

In some cases, the concept of a quality arises from invariances in numerical laws arrived at by measurement. It will be considered in connection with indirect measurement.

It is usual that once a scale of measurement is established for a quality, the concept of the quality is altered to coincide with the scale of measurement. The danger is that the adoption in science of a well defined and restricted meaning for a quality like intelligence may deprive us of useful insight that the common natural language use of the word gives.

3 REPRESENTATION

Consider some quality represented by a relational system

$$Q = \langle Q, R \rangle$$

Consider a numerical relational system defined as follows. Let N represent a class of numbers

$$N = \{n_1 \ldots\} \tag{4}$$

Let there be on N a family P of relations

$$P = \{P_1, \ldots, P_n\} \tag{5}$$

Then,

$$N = \langle N, P \rangle \tag{6}$$

represents a numerical relational system.

Commonly, N is just the real number line.

The representation condition requires that measurement be the establishment of a correspondence between quality manifestations and numbers in such a way that the relations between the referent property manifestations imply and are implied by the relations between their images in the number set. Formally, measurement is defined as an objective empirical operation M

$$M: Q \longrightarrow N \tag{7}$$

so that

$$n_i = M(q_i) \tag{8}$$

such that $Q = \langle Q, R \rangle$ is mapped homomorphically into (onto) $N = \langle N, P \rangle$.

The above homomorphism is the representation condition.

Firstly, it implies that if q_i is related to q_j by an empirical relation R_k, that is $R_k(q_i, q_j)$, and P_k is the numerical relation corresponding to R_k, then $R_k(q_i, q_j)$ implies and is implied by $P_k(n_i, n_j)$.

Measurement is a homomorphism rather than an isomorphism because M is not one-to-one, it maps separate

but indistinguishable property manifestations to the same number.

$$S = \langle Q, N, \mathrm{M} \rangle \tag{9}$$

constitutes a scale of measurement for Q.
$n_j = M(q_j)$, the image of q_j in N under M is called the measure of q_j on scale S.

4 SOME DIRECT SCALES OF MEASUREMENT

An attempt will now be made to analyze some qualities as empirical relational systems and to explain the logical basis of deriving a scale of measurement for them. Extensive measurement is the basis of measurement in the strict sense and will be considered in detail.

The extensive scales of physical measurement are based on establishing for the quality Q of empirical objects, for which a scale is to be determined, an empirical ordering with respect to Q of the class Ω of all objects possessing elements of Q, together with an operation o of combining the objects, elements of, which has, with respect to Q, the formal properties of addition. Such scales are known as extensive.

The above will now be stated more formally.

The basis of a scale of measurement of Q is the definition of the set Q.

Secondly, there must be an operational procedure that establishes on the set of objects Ω possessing Q an empirical equivalence relation $\sim$ and a transitive empirical relation $\prec$ with respect to Q such that $\langle Q, \sim, \prec \rangle$ is an order system.

Finally, consider objects $\omega_1, \omega_2, \omega_3, \omega_4 \in \Omega$ exhibiting property manifestations $q_1, q_2, q_3, q_4 \in Q$.

For an extensive measurement scale, there must be an operation $\circ$ for combining ω_1 and ω_2 with respect to q_1 and q_2, which we shall denote by $q_1 \circ q_2$, with the same formal properties as addition.

For all $q \in Q$,

1. $q_1 \circ q_2 \in Q$
2. $q_1 \prec q_1 \circ q_2$
3. $q_1 \circ q_2 \sim q_2 \circ q_1$
4. $q_1 \circ (q_2 \circ q_3) \sim (q_1 \circ q_2) \circ q_3$
5. if $q_2 \sim q_3$, then $q_1 \circ q_2 \sim q_1 \circ q_3$
   if $q_3 \sim q_2$ then $q_1 \circ q_3 \sim q_1 \circ q_2$
6. if $q_1, q_2, q_3, \ldots$ bear to each other the relation $\sim$, and $q_1 \prec q_1'$, then there is a number n such that $q_1' \prec q_1 \circ q_2 \circ q_3 \ldots \circ q_n$

With these definitions, the empirical relation system $\langle Q, \sim, \prec, \circ \rangle$ has a structure with the same properties as the numerical relation system $\langle \mathrm{Re}, =, <, + \rangle$.

With an empirical ordering operation and an 'additive combination' thus established, one proceeds to the setting up of a scale. A single object with $s_1 \in Q$ is chosen as standard, assigned the number 1, and chosen as the unit of the scale. Another object with the equivalent property is then found and combined by additive combination to produce the standard 2. This is repeated to produce an extended series of integral standards. Fractional standards are generated by combining two equivalent standards, which when combined are equivalent to s_1, and assigning them the number 1/2, and so on. A property empirically observed to be equivalent to s_I is assigned the measure i.

Consider now a matching scale. Such a scale is based on the establishment on the set of quality manifestations Q of an empirical indifference relation $\sim$.

Given $\langle Q, \sim \rangle$, a set of differing elements $s_i \in Q$ are selected to form a standard set $S = \{s_1, s_2, \ldots, s_n\}$. Numbers (or other symbols) $n_i \in N$ are then assigned to each $s_i \in S$, the same number n_i not being assigned to two differing elements. The fundamental measurement operation M of the scale consists of an empirical operation in which measurands $q_i \in Q$ are compared with members of the standard set S. If $q_i \sim s_i$, then it is assigned the number n_i. An example of this form of scale is a color code in which the relation 'matches' constitutes the empirical indifference relation.

Consider now ranking scales. In such a scale, an empirical order system $\langle Q, \sim, \prec \rangle$ is established on Q. A set of differing standard objects having $s_i \in Q$ is then selected and arranged in an ordered standard series $S = \langle s_1, s_2, \ldots, s_n \rangle$ according to $\langle Q, \sim, \prec \rangle$. Numerals are assigned to each s_i say i, in such a way that the order of numerals corresponds to the order in S, of standards to which they are assigned. Any $q_i \in Q$ can then be compared with the elements of S in the same way as in nominal measurement. If q_i bears the relation $\sim$ to any $s_i \in S$, it is assigned the numeral i. If an entity is not equivalent to any $s_i \in S$, one can determine between which two standard elements it lies in the empirical order system.

The best example of a ranking scale of measurement is the Mohs scale of hardness of minerals in which each material successively only scratches those higher than it in the sequence.

5 UNIQUENESS: SCALE CLASSIFICATION AND MEANINGFULNESS

The requirement that the fundamental measurement procedure of a scale should map the empirical relational system Q homomorphically into the numerical relational system

N does not determine the mapping uniquely. There is an element of arbitrary choice in the setting up of scales of measurement. In the case of scales based on additive combination, for instance, the choice of the unit standard is arbitrary. In the case of a ranking scale of measurement, the actual numbers assigned to the standards are arbitrary subject only to the requirement that they should be in the required order.

The requirement of homomorphism thus defines a class of scales that may be called equivalent. The class of transformations that transform one member of a class of equivalent scales into another is called the class of admissible transformations. The conditions that admissible transformations must satisfy are known as the uniqueness conditions.

We can classify scales by the classes of transformations admissible for them. Let m be numbers representing measures on a scale and let m' be the corresponding numbers on the transformed scale: $m' = F(m)$. The generally accepted classification of scales is as follows:

If F() is any 1–1 substitution, the scale is nominal.
If F() is any monotonic increasing function, the scale is ordinal.
If $F() = am + b$, $a > 0$, the scale is interval.
If $F() = am$, $a > 0$, the scale is ratio.

The problem of the meaningfulness of statements made about a quality in terms of its measures is important. Such a statement is meaningful if its truth is unchanged by admissible transformations of the scales of measurement; in other words, if it reflects the empirical relational system on which the scale is based and not just the arbitrary conventions of the scale.

Thus, as a very simple example, it is meaningful to speak of the ratio of two masses, since that ratio is invariant with respect to changes of the unit of mass. It is not meaningful to speak of the ratio of two hardnesses measured on the Mohs scale, since that ratio would be changed by a monotonic increasing transformation of the scale.

Another view of meaningfulness that can be taken is that only such statements involving measures are meaningful, which can be logically traced to the empirical operations on which the measurement is formed.

6 UNCERTAINTY

The above discussion of measurement has been given in terms of deterministic relations and mappings. However, all empirical observations are accompanied by scatter and bias.

The basic empirical observation process is unable to establish exactly the required relations such as indifference and order, on which the scale of measurement is based. Thus, for example, the empirical indifference relation is assumed to be an equivalence relation, which is reflexive, symmetric, and transitive. In practice, there is a least detectable difference, which means that $a_1 \sim a_2$ does not exclude $a_1 \prec a_2$, or $a_1 \succ a_2$.

Thus, there is uncertainty in the empirical relational system on which the representational measurement is based.

The problem can be dealt with using the concept of probabilistic relational systems, or employing fuzzy logic representation.

REFERENCES

Finkelstein, L. (2000) Foundational Problems of Measurement, in *Measurement Science – A Discussion* (eds K. Kariya and L. Finkelstein), Ohmsha Press, Amsterdam (pp. 13–21).

Krantz, D.R., Luce, R.D., Suppes, P. and Tversky, A. (Vol. 1: 1971; Vols. 2, 3: 1990) *Foundations of Measurement*, Academic Press, New York.

Roberts, F.S. (1979) *Measurement Theory with Applications to Decision Making, Utility and the Social Sciences*, Addison-Wesley, Reading, MA.

Tarski, A. (1954) Contributions to Theory of Models. *Indagationes Mathematicae*, **16**, 572–88.

9: Nature and Properties of Measurement

Ludwik Finkelstein
City University, London, UK

1 Measurement and Other Forms of Symbolic Representation 52
2 Special Properties of Measurement as a Form of Symbolic Representation 53
3 Measurement and Information 53
4 Measurement and Knowledge 54
Reference 54

1 MEASUREMENT AND OTHER FORMS OF SYMBOLIC REPRESENTATION

Measurement is only one form of representation of entities by symbols. It is closely related to other forms of symbolization.

The formal representational theory of measurement based on model theory, outlined here, can be extended to the more general case of representation by symbol systems (Finkelstein, 2000).

A symbol will be defined here as an object or event, which has a defined relation to some entity, for the purpose of eliciting a response appropriate to that entity in its absence.

Let Q be some set of entities and let R be some set of relations on Q constituting a relational system:

$$Q = \langle \mathrm{Q}, \times R \rangle \tag{1}$$

Note that Q may now be a set of any objects, events, abstract entities, and so on, and R need not be empirical.

Now, let Z be a set of objects or events to be used as symbols and let P be a set of relations defined or existing on Z to constitute a symbol relational system:

$$Z = \langle \mathrm{Z}P \rangle \tag{2}$$

With F a mapping from R onto P, and M a mapping from Q into (onto) Z, such that M and F map homomorphically into (onto) Z, then,

$$C = \langle Q, Z, \mathrm{M}, \mathrm{F} \rangle \tag{3}$$

Let $\mathrm{z_i} = \mathrm{M}(\mathrm{q_i})$ be the image of $\mathrm{q_i}$ in Z under M. Then $\mathrm{z_i}$ is termed the symbol of (or for) $\mathrm{q_i}$ under C, and $\mathrm{q_i}$ is termed a *referent* or meaning of $\mathrm{z_i}$ under C. C is termed the *symbolism* or *code*.

We may extend the concept of general symbolization to the concept of language.

A language L may be defined as a *subset of all finite strings*, that is, concatenations of elements of an alphabet or set of symbols, A. In the context of the symbolization C, the alphabet consists of the elements of Z and P, supplemented by function symbols.

A grammar, or syntax, G, is a set of rules, which, operating on the alphabet A, can generate all the strings of L. An alphabet and grammar constitute a compact description of a language.

All the considerations of the formal theory of measurement can be generalized to the symbolization of any relational systems by any general symbol system and descriptive language.

Formalisms such as symbolic logic and the like enable general symbolic descriptions to be effectively handled. The capability of computers to manipulate general symbols has made general symbolic representation technically important.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

2 SPECIAL PROPERTIES OF MEASUREMENT AS A FORM OF SYMBOLIC REPRESENTATION

Measurement, widely defined, can then be viewed as a special form of general symbolic representation. It is, therefore, useful to review those properties of measurement that distinguish it from other forms of symbolic description. It is useful to note here again the distinction between strongly defined measurement that is based on representation in the language of mathematics and weakly defined measurement that may employ representation by more general symbolism.

Measurement is not mere naming, but provides information about the relation of a particular manifestation of a quality and other manifestations of the same quality.

Measurement is an objective description. The symbol assigned to an entity, given the representation relation, is independent of the observer, within the limits of uncertainty of the representation relation. The objectivity of measurement makes measures invariant and indisputable in logical discourse.

Measurement is an empirical process. This means first that it must be the result of observation and not, for example, of a thought experiment. Further, the concept of the property measured must be based on empirically determinable relations and not, say, on convention. The truth of descriptions based on measurement is thus empirically justifiable.

The utility of symbolic descriptors depends upon their conciseness and the compatibility with further processing. Strongly defined measurement yields measures that are numbers, and may be manipulated mathematically. Weakly defined measurement may yield measures that are more general symbols. Their utility depends upon their compatibility with further processing. While mathematical processing of data is particularly powerful, there are effective ways of processing other symbolic descriptions.

As stated above, measurement is not mere naming. It is the representation of a measurand relational system in the real world, by a symbolic relational system. The utility of measurement depends upon the richness of the measurand relational system and the power of the symbolic relational system. In strongly defined measurement, the measurand relational system is embedded, in general, in a well formed and complete theory of a domain. The symbolization in strongly defined measurement is in the language of mathematics, with powerful capabilities of manipulating the descriptions. In weakly defined measurements, the symbolization of the domain may not be complete. Further, the relations representable in the symbolization may be few, so that the measurement is not powerful. Thus, in nominal scales, it is only possible to represent the indistinguishability, or difference of two entities, and no other relations between them.

3 MEASUREMENT AND INFORMATION

If z_i is a symbol of or for q_i under C, then information about q_i given z_i can be denoted by $J(q_i|z_i)$, where

$$J(q_i|z_i) = \langle z_i, C \rangle \tag{4}$$

A measure provides information about the measurand.

In information theory, we consider an information transmission channel that transforms symbols x_i, elements of a symbol set X, acting as inputs into symbols y_i, elements of a symbol set Y, constituting outputs. The transformation is in general many to many.

The quantity of information I about an input x_k provided by the occurrence of an output y_i is given by

$$I(x_k; y_i) = \log \left[\frac{P(x_k/y_i)}{P(x_k)} \right] \tag{5}$$

where $P(\,)$ represents probability. The base of the logarithm defines the unit of the scale.

The definition of quantity of information given by information theory and the definition of information based on symbolic representation theory are consistent, and indeed similar. The information theoretic definition presupposes the concepts of representation by symbols. In both, information is knowledge about an entity provided by an image of the entity under a mapping. Information theory thus deals with a restricted class of problems.

Measurement is then an information process. Measuring instruments are information machines.

Information machines are machines or systems of machines that have as their function the acquisition, processing, outputting, and effectuation of information. They operate by transforming input symbols into output symbols by defined transformations.

The signals that carry information in measuring instruments may be viewed as sequences of symbols. Regarding digital data as a sequence of symbols is intuitively obvious. Analog signals are not so obvious. However, we may view a continuous analog signal $s(t)$ as a sequence of Dirac unit impulses. The elements of the sequence are then symbols.

With regard to symbol processing in information machines, we may regard the processing as a prescribed mapping

$$T: Z_i \longrightarrow Z_o$$

from an input symbol set Z_i (symbolism C_i) into or onto an output symbol set Z_o (symbolism C_o).

More generally, we may view the transformation as a prescribed mapping of sentence in an input language L_i into sentences in an output language L_o.

$$T: L_i \longrightarrow L_o$$

The treatment of measurement as an information process and of measuring instruments as information machines is the basis of systematic approaches to measurement and instrumentation science and technology.

4 MEASUREMENT AND KNOWLEDGE

The result of a measurement constitutes knowledge.

The classical view is that knowledge consists of a set of propositions that are true, that a subject believes to be true, and that the subject is justified in believing to be true. The main problem in this view of knowledge is the nature of the justification of the belief.

There are theories that provide a basis of justification. Two significant theories in the present context are based on causality and reliability. According to causal theories, knowledge consists of true belief that bears a true relation to the fact in question. According to reliability theories, knowledge requires that it be acquired by a reliable process or method.

A measurement result is a statement in the form

$$z_k = \langle z_l, S \rangle \tag{6}$$

where z_k is the symbolic designation of the measurand, z_l is the symbol assigned by or to the measurand by the measurement process, and S is the scale on which the measurement is carried out.

Knowledge obtained by measurement satisfies the conditions of valid knowledge, meeting the requirements of the theories discussed above. The fact that measurement is a precisely defined empirical and objective process ensures that it bears a true relation to the fact in question. These properties also make data that is based on measurement reliable.

We may distinguish between declarative and procedural knowledge.

Declarative knowledge is knowledge about 'that', and has been discussed above.

Procedural knowledge is knowledge about the 'how'. It is essentially expressed as transformational rules for declarative knowledge.

Knowledge based on measurement is essentially declarative knowledge. However, the implementation of the measurement process is based on procedural knowledge.

The application of knowledge takes, in general, the form of a problem solving or design process.

In such a process, the basic operations are an analysis of requirements, leading to a value model or objective function, possibly multidimensional. Following the generation of candidate solution concepts, which are commonly knowledge-based symbolic models, the candidates are analyzed in terms of the value model. On the basis of an analysis of the thus derived values of the solutions, one or more of the candidates are chosen by a decision process to be accepted for further development, or implementation.

The value model is based on a measurement-like process of assigning numbers to attributes. The value model is however subjective, reflecting the view of the decision-maker. It is not measurement.

REFERENCE

Finkelstein, L. (2000) Foundational Problems of Measurement, in *Measurement Science – A Discussion* (K. Kariya and L. Finkelstein), Ohmsha Press, Amsterdam, New York (pp. 13–21).

10: Extensions of the Representational Theory of Measurement

Ludwik Finkelstein
City University, London, UK

1 Development Points of the Representational Theory of Measurement 55
2 Indirect Measurement Scales 55
3 Measurement Based on Theory 56
4 Symbolizations that are not Measurement 57
Reference 57

1 DEVELOPMENT POINTS OF THE REPRESENTATIONAL THEORY OF MEASUREMENT

The formal representational theory of measurement is based on direct measurement; the representation of particular attribute manifestations on the basis of an objective empirical process that involves no theory either in the formation of the concept of the measurand quality, or in its representational process.

The formal representational theory accords with the philosophical perspectives of logical positivism and operationalism, which base all valid knowledge on empirical observations and logical deductions from them. While both these points of view have received much philosophical criticism, they remain an important basis for the approach of the sciences.

Representational theory provides a sound basis for the foundations of measurement in all its aspects. It nevertheless requires extensions to deal satisfactorily with a number of problems.

The first is the problem of indirect measurement. The exposition of the representational theory that is outlined here – *see* **Article 7, Introduction to Measurement Theory and Philosophy, Volume 1**; **Article 8, Formal Theory of Measurement, Volume 1**; and **Article 9, Nature and Properties of Measurement, Volume 1** – considers measurement scales formed by direct mapping from a quality relational system to a numerical relational system. Frequently, however, scales of measurement for qualities are constructed indirectly through a relation of the quality to be measured and other qualities, for which measurement scales have been defined.

The second problem is that of establishing measurement scales for domains of knowledge for which there are complete, well validated, and grounded theories. The basic representational theory does not rely on any theories concerning the measurand. However, the establishment of measurement scales in domains for which valid theories exist is generally based on the underlying theory.

Finally, we have the problem of qualities that have a number of components, or dimensions. The component qualities may be directly measurable in the sense considered above; however, the measurand quality may not. The formation of measures based on such multidimensional qualities is an important practical problem.

The three problems outlined above will be considered in terms of the representational theory.

A more extended discussion is provided in Finkelstein (2000) and the literature cited therein.

2 INDIRECT MEASUREMENT SCALES

In direct measurement, scales are formed by direct mapping from a quality relational system to a numerical relational system. Frequently, however, scales of measurement for

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

qualities are constructed indirectly through a relation of the quality to be measured and other qualities for which measurement scales have been defined. The reason is, generally, the impossibility of setting up a satisfactory measurement scale directly.

In its simplest form, consider a case in which every object that manifests the quality to be measured exhibits a set of other qualities that are measurable. Then, to each manifestation of the measurand quality, there corresponds a set of measures of the associated qualities. These associated or component measures can be arranged in an ordered array. If manifestations of the measurand quality have identical arrays of component measures, and if and only if they are indistinguishable, then the array of component measures characterizes the measurand.

If we can combine the various component measures or, in other words, map them into a single number so that numbers assigned to the quality manifestations by this process imply and are implied by empirical relations between the quality manifestations, then this sets up an indirect scale of measurement.

Consider as an example the scale of measurement of density of homogeneous bodies. Each such body possesses mass m and volume v (where m and v are assumed to be measures on already defined scales). It is an empirically established law that objects of the same material, and hence conceptually of the same density, have the same ratio (m/v). When different materials are ordered according to our concept of density, they are also ordered according to the respective ratio m/v. Hence, a scale of measurement of density is based on the ratio of mass to volume.

A few observations should be made here. The function m/v that maps measures of mass and volume into measures of density is not unique in its order preserving properties with respect to density. For instance, $(m/v)^2$ would be an equally valid derived measure of density. The form of m/v is chosen to result in the greatest simplicity of mathematical relations involving density. The properties of the function m/v are an idealization of real observations.

3 MEASUREMENT BASED ON THEORY

Consider then the establishment of concepts of quantities and of measurement scales in domains of knowledge of the world, for which there are complete, well validated, and grounded theories.

Physics is one such domain. It is the paradigm for the strong definition of measurement. It is proposed to consider measurement in domains for which there are good theories in terms of the processes of establishing measurement in physics.

The classical theory of measurement was developed to give an account of measurement in the physical sciences. In terms of this classical theory, measurement in the physical sciences is based on the establishment of direct extensive scales of measurement for a number of physical quantities as described here. These quantities are used as the base of a system. Scales for other physical quantities are obtained as derived scales, that is, indirect scales in terms of the base quantities, in the form of multiplicative monomial functions of the base quantities.

This theoretical view does not account for the way in which measurement units and scales for physical quantities are established. For example, the unit of length in the SI system is now defined in terms of the velocity of light in vacua defined as standard. The scale of temperature is defined in terms of thermodynamic theory. Thus, the paradigm of measurement that involves no theory either in the formation of the concept of the measurand quality, or in its representational process is inappropriate.

The situation of physics is that it consists of a number of axiomatized theories such as Euclidean geometry, classical mechanics, thermodynamics, electromagnetism, and so on. The scales of measurement of classical physics are based on the acceptance of these theories as representations of the real world and defining the units on that basis rather than on individual axiomatization and establishment of scales of particular physical quantities.

It is now proposed to outline this formally in terms of the processes and notation of formal representational theory.

It is proposed here to define the term theory in the formal terms of model theory. A theory T is then defined as a set of sentences of the language L. It may be possible to present T as an axiomatic system in which all sentences of the theory are clearly derived from a set of axioms that is considered to be true. A fully formalized axiomatic system consists of

1. primitive symbols, (Z), in the notation adopted above;
2. rules for the formation of well-formed expressions (G);
3. axioms;
4. rules of inference;
5. a code C establishing the definition of the symbols and the interpretation of symbols and sentences.

In the strongly defined measurement of physics, L is the language of mathematics. This establishes the rules for the formation of well-formed expressions and the rules of inference. The symbols Z are numbers.

Measurement scales establish a code C constituting the definition of the symbols and the interpretation of symbols and sentences.

4 SYMBOLIZATIONS THAT ARE NOT MEASUREMENT

The concepts of representational theory can be extended to assignments of numbers, or other symbols, to properties of objects in such a way as to describe them, but where the symbolic descriptions are not measurements.

One example is classification of industries by a standard industrial classification. Assignments, on such a classification, of numbers to enterprises are objective and based on an empirical observation. They represent the relation of the enterprise to others symbolized using the same classification. However, they are not measurement to the extent that the system of classification is not based on an objective empirical system of relations of industries.

An important class of the descriptive assignment of numbers, the measurement status of which is problematic, arises in educational testing. Marks in examinations may be objective, and are based on an empirical process, but it is problematic what they measure, other than the performance in a particular test. It is doubtful whether, when marks are treated as measures on a ratio scale, they are not, in fact, measures on an ordinal scale. This affects the meaningfulness of statistics on marks, such as calculations of averages and the like. The conflation of marks, such as the calculation of weighted sums of marks, contains an element of subjectivity in the conflation scheme, which probably disqualifies such conflated marks from being considered measurements.

As a final example of descriptive numerical assignments that have doubtful measurement status, it is important to consider utility evaluation in formal decision making.

Problems of decision sciences were an important stimulus to the development of wider concepts of measurement. Formal concepts of utility, with considerable power, have been developed. However, it appears, that utility, however strongly defined, is only a description of the subjective judgment of the decision maker. It is not a measure of any objectively defined characteristic of the object evaluated.

REFERENCE

Finkelstein, L. (2000) Foundational Problems of Measurement, in *Measurement Science – A Discussion* (eds K. Kariya and L. Finkelstein), Ohmsha Press, Amsterdam, New York (pp. 13–21).

11: Measurement Theory in Physical, Social, and Psychological Science

Ludwik Finkelstein
City University, London, UK

1 Measurement Theory in Physical Science 58
2 Measurement Theory in Social and Psychological Science 58
References 59

1 MEASUREMENT THEORY IN PHYSICAL SCIENCE

Measurement in physical science is generally strongly defined. It is based on the mapping of empirically observed relational systems in a physical domain into a symbolical relational system of mathematics (*see* **Article 7, Introduction to Measurement Theory and Philosophy, Volume 1**; **Article 8, Formal Theory of Measurement, Volume 1**; and **Article 9, Nature and Properties of Measurement, Volume 1**).

In strongly defined measurement

1. definition of the quantities is based on the theory;
2. empirical relational system is also defined by the theory;
3. symbolic relational system is rich;
4. theory for the domain is complete.

However, there are problems in physical measurement that present theoretical problems: they arise in measurement of properties for which there is no adequate theory and also in quantum and relativistic physics.

In technological measurements of physical properties of objects or materials, there often arises the requirement to characterize properties for which there is no adequate theory. For example, in testing of food we may need to describe empirically and objectively the ease with which a fat can be spread on a surface. In mining practice, there arises the need to characterize the hardness of a stratum that has to be worked upon. In such problems, the measurement scales formulated are weakly defined. There is, however, always the endeavor to investigate the relevant materials and phenomena in order to model them in terms of established theory and to produce strictly defined measurement scales.

Special problems arise in quantum and relativistic physics, which cannot be handled by the theory of measurement presented here. In quantum physics, the interaction between observer and observed system imposes a limit on the certainty of the joint measurement of the attributes of the system such as position and momentum. This imposes a fundamental difficulty on the measurement theory. The theory of relativity also has a large impact on the theory of measurement in the physical sciences. In terms of the view presented here it attempts to represent reality using a theory different from those used in classical measurement as, for example, in classical mechanics. The rejection of the concept of simultaneity and an upper limit to velocity are examples of such differences.

There are thus fundamental developments in the application of measurement theory to the physical sciences to be undertaken.

A more extended discussion of the problems is given in Finkelstein (2000, 2003).

2 MEASUREMENT THEORY IN SOCIAL AND PSYCHOLOGICAL SCIENCE

It is to be noted that it is the requirements of social and psychological sciences, which have led to the replacement

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

of the restrictive classical theory of measurement by the more comprehensive modern wide definition.

The social and behavioral sciences are very much concerned with such attributes or qualities as utility, standard of living, alienation, intelligence, and the like. The first problem in attempting to measure them is the difficulty in establishing an adequate objective concept of these qualities based on empirical relations. The conceptual framework is often absent.

When a scale of measurement for a quality such as 'standard of living' is formed, there remains a fundamental problem of establishing that the measure and concept correspond. For example, index figures that are often adopted for the purpose of measuring standard of living are disputed by those whom they do not suit, as not measuring what they consider to be the meaning of standard of living.

The empirical operations involved in establishing scales of measurement in social and behavioral judgment, commonly involve responses by human observers. These are, for example, required to judge whether two stimuli, such as light intensities, pitch of sound, and so on, are indistinguishable. As another instance, they are required to give an ordinal ranking to a number of alternatives. The data thus derived may be sufficiently consistent for a population of observers to consider their objective; they are nevertheless subject to considerable random scatter.

The scales of measurement, the social and behavioral sciences being frequently based on determination of equivalence and order only, are then only nominal or ordinal.

Commonly, some conceptual quality such as the 'alienation' of a work force in terms of measurable quantities, such as worker-days lost through disputes and absenteeism is difficult to establish. The difficulty again is the establishment of agreement between the concept of the quality and the measures adopted.

There are no wholly axiomatized theories in the social and behavioral sciences, which correspond to, say, classical mechanics or thermodynamics.

Measurement in social and psychological science is thus generally weakly defined. Such measurement has one or more of the following features:

1. It is based on a concept of the quality that is often fuzzily defined – *see* **Article 12, Fuzzy Approaches for Measurement, Volume 1**.
2. There is significant uncertainty in the empirical relational system that it represents.
3. The symbolic relational system has limited relations defined on it.
4. There is no adequate theory relating the measurement to other measurements in the same domain.

Finally, it may be stated that in these sciences it is by no means universally agreed that the clear formation of concepts in terms of empirical observation, is possible or desirable. Nor is there agreement that the search for data, through measurement, advances knowledge and understanding. The opponents of quantification would say that human nature and behavior are too variable to enable the methodologies of the physical sciences to be applicable to them.

A more extended discussion of the problems is given in Finkelstein (2000, 2003).

REFERENCES

Finkelstein, L. (2000) Foundational Problems of Measurement, in *Measurement Science – a Discussion* (eds K. Kariya and L. Finkelstein), Ohmsha Press, Amsterdam, New York (pp. 13–21).

Finkelstein, L. (2003) Widely, Strongly and Weakly Defined Measurement. *Measurement*, **34**, 39–48.

12: Fuzzy Approaches for Measurement

Eric Benoit, Laurent Foulloy, Gilles Mauris
Université de Savoie, Annecy, France

1 Definition of Fuzzy Sensors 60
2 Numerical Fuzzy Sensors 61
3 Linguistic Fuzzy Sensors 61
4 Fuzzy Nominal Scales 62
5 Different Types of Fuzzy Blocks 63
6 Rule-based Fusion of Information 63
7 Fuzzy Classification Approaches 64
8 Interpolation-based Fusion of Information 65
References 66

1 DEFINITION OF FUZZY SENSORS

Introduced by Zadeh in 1965 the fuzzy subset theory is now a common tool in main domains. The basis of this theory and its symbology is first defined. A fuzzy subset A of a set X is a set that can contain partially some members of X. Any member x of X has a membership degree to the set A. This membership degree is denoted $\mu_A(x)$, and μ_A is the membership function of the fuzzy subset A.

The main objective of a sensor is to convert the physical state of an entity into an information entity (Mari, 2001). Fuzzy sensors are defined as sensors producing a fuzzy subset as their information entity. The idea of using fuzzy subsets in process control systems arose from Chang and Zadeh's (1972) study of fuzzy mappings.

This section provides basic definitions of the three general stages involved in fuzzy information systems (Foulloy and Galichet, 1995).

Fuzzification is an interface that produces a fuzzy subset of information from the measurement; it is a mapping from the set of measurements X to a set of fuzzy subsets associated with the input, denoted $F(Y)$.

The inference resulting is an interface that produces a new fuzzy subset from the result of the fuzzification, using, for example, a set of rules. The result of the inference is a fuzzy subset associated with the output. It can be represented by a mapping from the set $F(Y)$, associated with the input, to a set of fuzzy subsets associated with the output and denoted $F(Z)$.

Defuzzification is an interface that produces a crisp output from the result of the inference, that is, it is a mapping from the set $F(Z)$, associated with the output, to the set U.

The set Y related to the input can either be the set of numerical values, that is, X, or the set of linguistic values, that is, LX. In the same way, the set Z related to the output can either be the set of numerical values, that is, U, or the set of linguistic values, that is, LU. Therefore, the previous definitions can be used to provide an efficient classification, that is, a typology of fuzzy information handling in four classes as shown in Figure 1.

From the preceding typology, three types of components can be introduced:

- Fuzzy sensors provide a representation of the measurements by means of fuzzy subsets; they perform the fuzzification.
- Fuzzy actuators are components that are able to act on the physical world depending on the fuzzy subset they receive; they perform the defuzzification.
- Fuzzy inference components produce new fuzzy subsets from fuzzy subsets they have received, thus performing fuzzy reasoning.

This article focuses on linguistic fuzzy sensors that deal with linguistic fuzzy subsets and on numerical fuzzy sensors

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

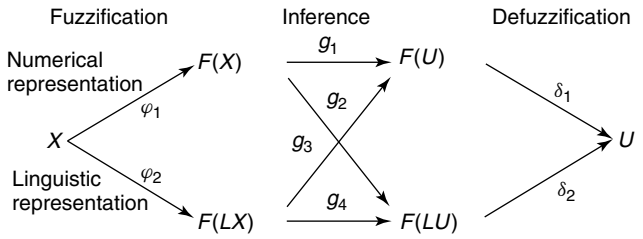


Figure 1. Typology of fuzzy processing in measurement and control systems.

that deal with numerical fuzzy subsets. Linguistic fuzzy sensors can be represented directly by means of the φ_2 mapping (see Figure 1). They can also be represented indirectly by means of the composition of several mappings as long as the resulting information is a fuzzy subset of linguistic terms.

In many situations, information sought is not available from a single entity, but from many different complementary entities. In measurement theory, this case corresponds to the so-called conjoint measurement. Here, it will be called the *aggregation of complementary information*, and is presented in Sections 4 to 8.

2 NUMERICAL FUZZY SENSORS

The basic idea developed in the design of numerical fuzzy sensors is to take into account and manipulate the measurement uncertainties. Mauris, Lasserre and Foulloy (2001) propose a theoretical well-founded approach, compatible with the ISO Guide on the expression of uncertainty, for building a fuzzy subset that is interpreted as a possibility distribution that represents the uncertainty associated with a measurement.

According to the ISO Guide prepared jointly by the International Organization for Metrology; BIPM (Bureau International des Poids et Mesures); IEC (International Electrotechnical Committee); ISO (International Organization for Standardization); and the OIML (International Organization of Legal Metrology), the expression of measurement uncertainty must satisfy certain requirements in order to be widely used at the practical level. The Guide recommends use of a parametric representation of the measurement uncertainty that

- characterizes the dispersion of the observed values; for example, the standard deviation or a half-width of an interval at a given level of confidence,
- can provide an interval of confidence, which contains an important proportion of the observed values,
- can be easily propagated in further processing.

Thus, the ISO Guide proposed to characterize the measurement result by providing the best estimation of the measurand (i.e. in general, the mean value) and the standard deviation. In fact, it simplifies the probability approach by considering only the first two moments (mean value and variance) of the probability distribution.

Nevertheless, the Guide distinguishes two categories of uncertainty according to the method used to estimate their numerical values: those that are evaluated by statistical methods called type A, and those that are evaluated by other means called type B. Note that for the latter, probability approaches are not more natural than the interval approach, when the uncertainty expression comes from experts (Reznik and Dabke, 1999).

In this article, we consider the fuzzy and possibility theories as an alternative to the conventional probabilistic approach (Kendall and Stuart, 1977), and as an extension of the interval calculus (Moore, 1966). Many methods, more or less justified, are proposed in the fuzzy literature to represent measurement uncertainty (Geer and Klir, 1992; Ree and Jager, 1993). The basis of this approach is to represent measurement uncertainty by a fuzzy subset (interpreted as a possibility distribution) built to be compatible with the ISO Guide.

The foundation of this approach is the notion of confidence interval. Let us first recall this notion: let p be a probability distribution, and let it be a 'one point' estimation (for example, the mean value of the probability distribution) of the 'real' value.

An interval is defined around the 'one point' estimation and corresponds to the probability for this interval to contain the 'real' value. Such an interval $I_{1-\alpha}$ is called a confidence interval of confidence level $(1 - \alpha)$ (95%, 99% for example); it is the risk level, that is, the probability for the real value to be outside the interval.

Mauris identified each confidence interval of level $1 - \alpha$, with each α-cut of a fuzzy subset F, which thus gathers the whole set of confidence intervals in its membership function. Figure 2 represents this situation.

3 LINGUISTIC FUZZY SENSORS

Where numerical values are commonly used to represent measurement results some applications would better manipulate symbolic values or linguistic terms. For example, a rule-based decision system uses linguistic terms in its rules. The advantage of using linguistic terms instead of numerical ones is to reduce the number of symbols involved. Zadeh (1971) subsequently introduced the concepts of fuzzy meaning and the descriptor set that are extensively used in fuzzy sensors designs. They provide

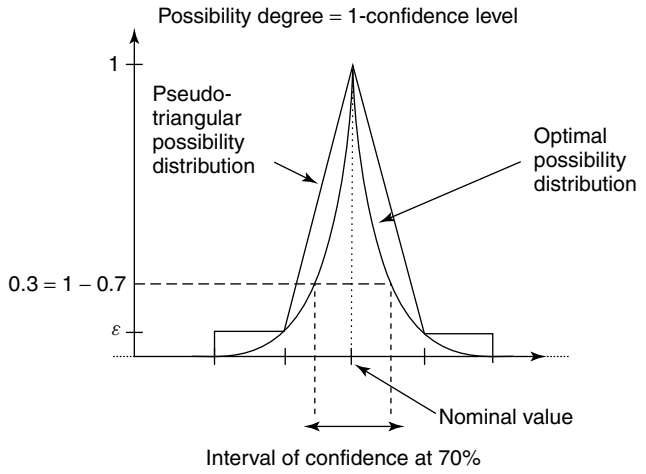


Figure 2. A numerical information entities as a possibility distribution.

a simple means for representing measurement results by a fuzzy subset of linguistic terms. The conversion from numeric to linguistic representation presented in the following is called the *linguistic fuzzy description* or, more simply, the *fuzzy description*.

Let X be the universe of linguistic discourse associated with the measurement of a particular physical quantity. In order to linguistically characterize any measurement over X, let W be a set of linguistic terms, representative of the physical phenomenon. Denote R the relation defined on $X \times W$ that formalizes the link between items of the universe of discourse and the linguistic terms of the lexical universe.

Two mappings can be defined from this relation. The first one is a one-to-one mapping m, called the *meaning* of a linguistic term. It associates any term w of W with a subset of X,

$$\forall w \in W, m(w) = \{x \in X | xRw\} \tag{1}$$

The other one is a one-to-one mapping d, called the *description* of a measurement, which associates any measurement x of X with a subset of W.

$$\forall x \in X, d(x) = \{w \in W | xRw\} \tag{2}$$

Since the two mappings are defined from the relation, we have obviously the following equivalencies:

$$xRw \iff x \in m(w) \iff w \in d(x) \tag{3}$$

Consider now that the relation R is a fuzzy relation on $X \times W$. It is characterized by its membership function denoted μ_R. The grade of membership $\mu_R(x, w)$ characterizes the strength of the link between x and w. Then, the previous

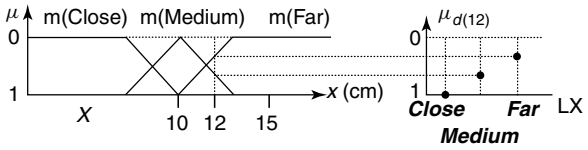


Figure 3. Fuzzy meaning and fuzzy description mechanism.

concepts can be extended to the fuzzy case. It leads to the fuzzy meaning of a term and the fuzzy description of an object state, respectively defined by

$$\forall w \in W, \ \forall x \in X, \mu_{m(w)}(\mathrm{x}) = \mu_R(x, w) \tag{4}$$

and,

$$\forall w \in W, \ \forall x \in X, \mu_{d(x)}(w) = \mu_R(x, w) \tag{5}$$

Equivalencies given in (3) become equalities between grades of membership as follows:

$$\forall w \in W, \ \forall x \in X, \mu_R(x, w) = \mu_{d(x)}(w) = \mu_{m(w)}(\mathrm{x}) \tag{6}$$

Figure 3 shows the link between the fuzzy meaning of each word of the set $LX = \{Close, Medium, Far\}$ and the fuzzy description of a numerical value.

4 FUZZY NOMINAL SCALES

Scales have been introduced to model the link between physical states and information entities created by the measurement. Finkelstein (1975) proposed to define any scale by the following symbolism:

$C = \langle X, S, M, R_X, R_S, F \rangle$, where

- X refers to a set of object states and R_X is a set of relations on X,
- S refers to a set of information entities, and R_S is a set of relations on S,
- M, called the representation, is a mapping from X to S,
- F is a one-to-one mapping with domain R_X and range R_S.

Knowing what kind of scale is used for measurement is very important. This knowledge defines the set of relations and operators allowed for the manipulation of information entities.

In the case of nominal scales, R_S contains only the equality relation on S and R_X contains an equivalence relation on X, denoted $\sim$. The mapping F is then simply defined by $F = \{(\sim, =)\}$. The representation respects the following equation.

$$\forall x_i \in X, \ \forall x_j \in X, x_i \sim x_j \Leftrightarrow \mathrm{M}(x_i) = \mathrm{M}(x_j) \tag{7}$$

The equivalence relation $\sim$ defines a partition of the set of object states and M associates each item of the partition with information entity. The general symbolism for a nominal scale is

$$C = \langle X, S, M, \{\sim\}, \{=\}, \{(\sim, =)\}\rangle \tag{8}$$

It was shown by Benoit and Foulloy (2003) that the fuzzy description presented in Section 3 can be interpreted as a nominal scale with the following symbolism. This scale is named the fuzzy nominal scale.

$$C = \langle X, F(W), d, \{\sim\}, \{\equiv\}, \{(\sim, \equiv)\}\rangle \tag{9}$$

where d is the fuzzy description, and $\sim$ is a fuzzy equivalence relation on X, and $F(W)$ is the set of fuzzy subsets of W.

It was also shown that a proximity relation between lexical entities can be deduced from the preceding symbolism, which means that

- a fuzzy nominal scale is not only a nominal scale; in addition to equivalence relation, proximity relation and associated pretopological operators can be used on information entities;
- the introduction of fuzziness in measurement can be made only if a proximity relation really exists between physical states.

Therefore, fuzzy nominal scales can be useful for taking into account the fuzziness of a humanlike linguistic representation, but they must be used carefully.

5 DIFFERENT TYPES OF FUZZY BLOCKS

Using the preceding approaches, it is now possible to define a set of function blocks that can be merged to create complex systems. The classification of the function blocks is based on the nature of the fuzzy subsets and the nature of the function. Fuzzy subsets can be numerical fuzzy subsets (NFS) or linguistic fuzzy subsets (LFS). Function blocks can perform the fuzzification, the inference or the defuzzification. This classification, issued from the typology presented in Section 1, distinguishes 8 fuzzy function blocks.

- Numerical fuzzification: it has a numerical entity as input, that is, a number or a vector, and an NFS as output.
- Linguistic fuzzification: it has a numerical entity as input, and an LFS as output. It performs a fuzzy description of the numerical input.

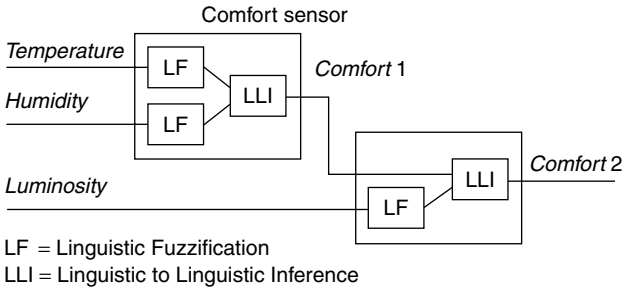


Figure 4. Fuzzy sensor made of fuzzy blocs.

- Numerical inference: input and output are NFS.
- Numerical to linguistic inference: input is an NFS and output is an LFS.
- Linguistic to numerical inference: Input is an LFS and output is an NFS.
- Linguistic inference: input and output are LFS.
- Numerical defuzzification: it has an NFS as input and a numerical entity as output.
- Linguistic defuzzification: it has an LFS as input and a numerical entity as output.

A measurement system can be created with a graph of fuzzy blocks respecting the basic constraint that each end of the connectors must have the same nature and the universe of discourse. Linguistic entities must be fuzzy subsets of the same lexical set, and numeric entities must be defined on the same numerical set. Figure 4 shows how this applies to an example system.

6 RULE-BASED FUSION OF INFORMATION

Some inference blocks use linguistic fuzzy subsets for their inputs and their outputs. As the fuzzy nominal scale concept is too young to be applied at this level, the linguistic fuzzification is usually considered as a simple nominal scale. The only relation that links linguistic entities is an equivalence relation. With an equivalence relation, the fusion of information can be made with a set of rules. Generally, most of the rules are expressed in the following manner:

$$\text{If } X_1 \text{ is } L_1 \text{ then } Y \text{ is } a$$

for example, if the size (of a shoe) is *large* then the price (of a shoe) is *high*.

In fact, the rule represents a correspondence between the variable X_1 (e.g. the size) and the variable Y (e.g. the price). The purpose of this sort of metaimplication is to infer information about Y from information about X_1.

As proposed by Zadeh (1975), the if ... then ... metaimplication could be viewed as a fuzzy relation between X_1 and Y represented by the fuzzy graph $\mu_{\Gamma 1}(X_1, Y)$. Then the image F of a fuzzy subset E of the variable X_1 is obtained by the combination/projection principle (also referred to as generalized modus ponens) according to

$$\mu_F(Y) = \mu_{f(E)}(Y) = \perp_{X_1}(\mu_E(X_1) T \mu_{\Gamma 1}(X_1, Y)) \quad (10)$$

where T is a norm and $\perp$ a conorm.

To compute $\mu_F(Y)$, we must know $\mu_{\Gamma 1}(X_1, Y)$ and $\mu_E(X_1)$.

The problem is first how to build $\mu_{\Gamma 1}$ and secondly how to define $\mu_E(X_1)$.

The most natural representation of the rules is to consider that the rules relate the two linguistic terms L_1 and a, that is, to consider that $\boldsymbol{L}_1$ and $\boldsymbol{a}$ are elements of the lexical sets LX_1 and LY and that $\Gamma 1$ is a fuzzy graph on the Cartesian product $LX_1 \times LY$.

In this case, the expression is transformed for every $\boldsymbol{a}$ belonging to LY into

$$\mu_F(a) = \mu_{f(E)}(a) = \perp_{L_1 \in LX_1}(\mu_E(L_1) T \mu_{\Gamma 1}(L_1, a)) \quad (11)$$

with $\mu_{\Gamma 1}(L_1, a)$ representing the degree of validity of the rule and $\mu_E(L_1)$ a linguistic description of the variable X_1.

Now, if we consider n rules, we have n elementary graphs Γ_i's. If we consider that every rule provides a relation, a unique graph Γ that represents all the rules could easily be built by taking $\Gamma = U_i \Gamma_i$. If we consider more complex rules of the form

$$\text{if } X_1 \text{ is } L_1 \text{ and } X_2 \text{ is } L_2 \text{ then } Y \text{ is } a,$$

we then consider X_1, X_2, Y as elements of LX_1, LX_2 and LY, and Γ is defined on the Cartesian product $LX_1 \times LX_2 \times LY$.

The expression becomes

$$\mu_F(a) = \perp_{(L_1, L_2) \in LX_1 \times LX_2}(\mu_{(E,E')}(L_1, L_2) T_1 \mu_\Gamma (L_1, L_2, a)) \quad (12)$$

It remains then to define $\mu_{(E,E')}(L_1, L_2)$.

If we make the assumption that the two variables are independent $\mu_{(E,E')}(L_1, L_2)$ is decomposable and we can write

$$\mu_{(E,E')}(L_1, L_2) = \mu_E(L_1) T_2 \quad \mu_{E'}(L_2) \quad (13)$$

(with T_2 an operator of intersection, i.e. a norm).

If we have two numeric measurements x_1 and x_2, then

$$\mu_{(E,E')}(L_1, L_2) = \mu_{i(x_1,x_2)}(L_1, L_2) = \mu_{i(x_1)}(L_1) \; T_2 \; \mu_{i(x_2)}(L_2) \quad (14)$$

Computation of (12) leads to obtaining the description of (x_1, x_2) on LY, that is, $\mu_{i(x_1,x_2)}(a)$ for every a belonging to LY.

In fact, we have built a new fuzzy sensor that is an aggregation of the two primary fuzzy sensors.

An interesting problem is the determination of the operators $\perp$ and T. This choice is implied by the structure we will obtain for the aggregated information and by the properties of the graph Γ. Such studies have been performed in the case in which Γ is a crisp graph (Benoit, Mauris and Foulloy, 1995; Mauris, Benoit and Foulloy, 1996).

7 FUZZY CLASSIFICATION APPROACHES

Another approach for fuzzy measurement is to consider fuzzy sensors as fuzzy classifiers. This approach can be chosen when the initial knowledge about the measurement is a set of examples instead of an expert set of knowledge.

The idea of similarity, naturally associated with fuzzy sets, has been used mainly in clustering analysis. Indeed, it may be more realistic for describing a data set to look for fuzzy clusters when some clusters are not well separated.

Since Dunn (1973) and Bezdek (1981), the theory of fuzzy sets has been introduced into several clustering techniques. In particular, the fuzzy c-means (FCM) and the fuzzy k-nearest neighbors algorithms (KNN), which are the generalization of the crisp ones, have been widely used. More recently, some authors (Grabisch and Nicolas, 1994), have used the fuzzy integral in classification problems. Introduced by Sugeno (1974), the fuzzy integral has been mostly used as an aggregation operator in problems such as decision making. Bezdek and Pal (1992) contains most of the seminal works on the use of fuzzy sets theory in the field of pattern recognition and classification.

The FCM clustering algorithm is the fuzzy equivalent of the nearest mean hard clustering algorithm (Duda and Hart, 1973). Data are supposed to be structured into n vectors whose dimension is $p : X_j$; $j = 1,2, \ldots, n$; each vector characterizes an object described with p attributes. We assume here that the number of clusters is known. For this preselected number c, the FCM algorithm produces c vectors, which represent the cluster centers, and for each data point c membership values, which measure the similarity of the data points to each of the cluster centers.

Let u_{ij} be the membership value of the vector X_j to the cluster i describing how close X_j is to this cluster's center C_i. The classification is obtained by minimizing the following objective function with respect to the memberships $\{u_{ij}\}$ and the cluster centers $\{C_i\}$:

$$\sum_{j=1}^{n}\sum_{i=1}^{c}(u_{ij})^m d^2(Xj, Ci)$$

where d is a distance in the space of data. The value of the fuzzy index m tunes the degree of fuzziness of the clustering. The membership values indicate how well the point has been classified. When the input is close to a particular center, the membership value to the corresponding class is close to one.

A fuzzy k-nearest neighbors algorithm was proposed by Keller, Gray and Givens (1985). The conventional KNN classification method assigns each input to one of the possible classes. First the k-nearest neighbors are found. Then the input is assigned to the class that includes the majority of the neighbors. In the fuzzy KNN classifier, the second step consists in assigning to the input a membership degree to each class. Let x be an input, $u_i(x)$ its membership to the ith class and w_{ij} the one of the jth neighbor z_j, then

$$u_i = \frac{\sum_{j=1}^{k} w_{ij}\left(\frac{1}{\|x - z_j\|}\right)^{2/(m-1)}}{\sum_{j=1}^{k}\left(\frac{1}{\|x - z_j\|}\right)^{2/(m-1)}} \qquad (15)$$

where m is an adjustable parameter that tunes the weighting effect of the distance.

Since the introduction of the Sugeno integral, different families of fuzzy integrals have been defined. We only present here the Choquet integral that we use as an aggregation operator for multiattribute classification (Grabisch and Sugeno, 1992). Let f be a function from X to [0,1], and μ a fuzzy measure on X. Assume the elements $\{x_i\}$ are ordered such that we have $f(x_1) \leq f(x_2) \leq \cdots \leq f(x_N)$. The Choquet integral with respect to μ is defined by

$$\int f d\mu = \sum_{i=1}^{N}(f(x_i) - f(x_i - 1))\mu(A_i) \qquad (16)$$

where $A_i = \{x_i, \ldots, x_N\}$ and $f(x_0) = 0$ by convention.

The $\{x_i\}$ represent the attributes that are taken into account for the classification. The $\{f(x_i)\}$ represent the individual degrees of matching with each of these features. For each class, we define one or more prototypes and a fuzzy measure on the set of subsets of attributes. This measure gives the importance of each attribute, and every subset of attributes from the point of view of this class in the classification process. This allows us to take into account the redundancy in the information collected or the dependence between two or more features. Naturally, the features, or subsets of features, whose contribution to the identification of a given class is the greatest are assigned the greatest measures with respect to this class.

8 INTERPOLATION-BASED FUSION OF INFORMATION

An alternate solution to performing a clustering is to create the fuzzy meanings with simple interpolation techniques. In this article, we consider initial knowledge about measurements. This knowledge is materialized by the meaning of lexical entities on a small subset V of the measurements set X.

Then we have to extend the meaning of each symbol on all the measurements set. In order to obtain a fuzzy partition, the measurement set is cut into n-simplexes. An n-simplex in an n-dimensional space is a polyhedra with $n + 1$ vertices. For example, a 2-simplex is a triangle and a 3-simplex is a tetrahedron. Then meanings are defined on each n-simplex respecting the fuzzy partition constraint.

First, the measurements set is partitioned into n-simplexes using Delaunay's triangulation. The points used to perform the triangulation are the elements of the set V. As any triangulation can be used to cut the measurements set, it must be chosen according to constraints associated with the measurements. The Delaunay triangulation is preferred when no constraint can be found (Benoit, Mauris and Foulloy, 1994).

The membership function of the meaning of each word is defined on each n-simplex by a multilinear interpolation. Suppose the restriction on an n-simplex of the membership function of the meaning of a word $\boldsymbol{s}$ is

$$\mu_{m(s)}(x_1, x_2, \ldots, x_n) = a_1x_1 + a_2x_2 + \cdots + a_nx_n + a_{n+1} \qquad (17)$$

The value of this function is known for the $n + 1$ vertices of the n-simplex, therefore the $n + 1$ factors ai can be calculated by the following equation system:

$$A = M^{-1}B \quad M = \begin{vmatrix} x_{1,1} & \ldots & x_{1,n} & 1 \\ \ldots & \ldots & \ldots & \ldots \\ x_{n+1,1} & \ldots & x_{n+1,n} & 1 \end{vmatrix}$$

$$A = \begin{vmatrix} a_1 \\ \ldots \\ a_{n+1} \end{vmatrix} \quad B = \begin{vmatrix} \mu_{m(s)}(v_1) \\ \ldots \\ \mu_{m(s)}(v_{n+1}) \end{vmatrix} \qquad (18)$$

The ith vertex of the n-simplex is denoted v^i, and its jth component is denoted $x_{i,j}$.

Once this process is performed on each n-simplex for each word, we obtain a fuzzy nominal scale defined on X. This scale is an extension on X of the fuzzy nominal scale on V.

With this method, the knowledge needed for the configuration of the sensor is very compact. It can be acquired by means of a system called a *teacher*, which can be a human or an expert system. During the configuration phase, the teacher and the sensor analyze the same phenomenon, and the teacher gives its description to the sensor. The sensor increases its knowledge with its measurement associated with the teacher description. Then it can build the fuzzy nominal scale. It should be noted that the description given by the teacher must be a fuzzy subset of words, which verifies the fuzzy partition condition. If the teacher is a human, then its description is generally a crisp set that contains only one symbol. Figure 5 gives the teacher's knowledge about colors that is expressed in a two-dimensional chrominance space and the fuzzy meaning of the color Red.

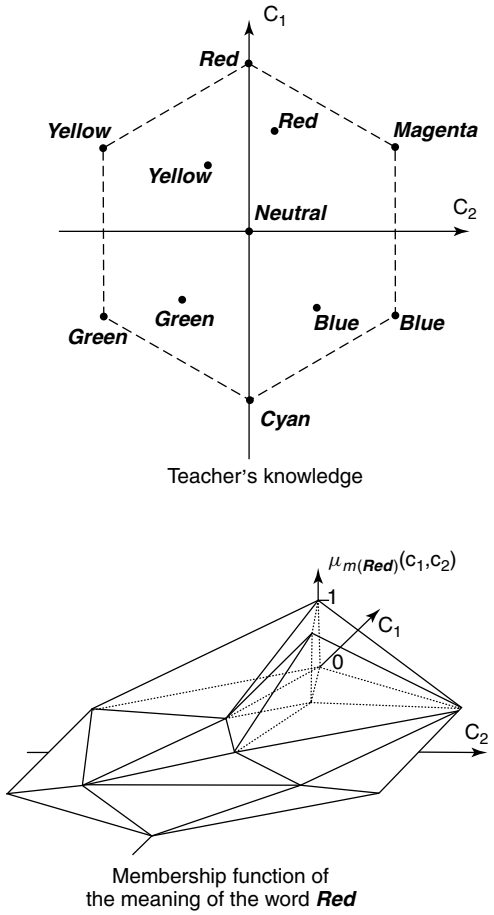


Figure 5. Teacher's knowledge and fuzzy meaning of the color red.

REFERENCES

Benoit, E., Mauris, G. and Foulloy, L. (1994) A Fuzzy Colour Sensor, in *Proceedings of XIII IMEKO World Congress*, 5–9 September, Torino (pp. 1015–1020).

Benoit, E. and Foulloy, L. (2003) Towards Fuzzy Nominal Scales. *Measurement*, **34**(1), 49–55.

Benoit, E., Mauris, G. and Foulloy, L. (1995) Fuzzy Sensor Aggregation: Application to Comfort Measurement, in *Fuzzy Logic and Soft Computing* (eds B. Bouchon-Meunier, R. Yager and L. Zadeh), World Scientific, London (pp. 219–227).

Bezdek, J.C. (1981) *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum, New York.

Bezdek, J.C. and Pal, S.K. (eds) (1992) *Fuzzy Models for Pattern Recognition Methods that Search for Structure in Data*, IEEE Press, New York.

Chang, S.S. and Zadeh, L.A. (1972) On Fuzzy Mapping and Control. *IEEE Transactions on Systems, Man and Cybernetics*, **SMC-2**(1), 30–34.

Duda, R.O. and Hart, P.E. (1973) *Pattern Classification and Scene Analysis*, Wiley, New York.

Dunn, J.C. (1973) A Fuzzy Relative of the ISODATA Process and its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics*, **3**(3), 32–57.

Finkelstein, L. (1975) Representation by Symbol Systems as an Extension of the Concept of Measurement. *Kybernetes*, **4**, 215–223.

Foulloy, L. and Galichet, S. (1995) Typology of Fuzzy Controllers, in *Theoretical Aspects of Fuzzy Control* (eds H.T. Nguyen, M. Sugeno, R. Tang and R. Yager), Wiley, New York.

Geer, J.F. and Klir, G.J. (1992) A Mathematical Analysis of Information-Preserving Transformations Between Probabilistic and Possibilistic Formulations of Uncertainty. *International Journal of General Systems*, **20**(2), 143–176.

Grabisch, M. and Nicolas, J.M. (1994) Classification by Fuzzy Integral: Performance and Tests. *Fuzzy Sets and Systems*, **65**, 255–271.

Grabisch, M. and Sugeno, M. (1992) *Multi-Attribute Classification Using Fuzzy Integral*, 1st IEEE International Conference on Fuzzy Systems, 8–12 March, San Diego, CA (pp. 47–54).

Keller, J.M., Gray, M.R. and Givens, J.A. (1985) A Fuzzy K-Nearest Neighbor Algorithm. *IEEE Transactions on Systems, Man and Cybernetics*, **SMC-15**(4), 580–585.

Kendall, M. and Stuart, A. (1977) *The Advanced Theory of Statistics*, Griffin and Co., London.

Mari L. (2001) The Duality of Sensing and Coding as a Fundamental Topic of Information Science, *Proceedings of the World Conference on Systemics, Cybernetics and Informatics (SCI 2001)*, Vol. XVI, Orlando, FL (pp. 177–180).

Mauris, G., Benoit, E. and Foulloy, L. (1996) The Aggregation of Complementary Information via Fuzzy Sensors. *International Journal of Measurement*, **17**(4), 235–249.

Mauris, G., Lasserre, V. and Foulloy, L. (2001) A Fuzzy Approach for the Expression of Uncertainty in Measurement. *Measurement*, **29**(3), 165–177.

Moore, R. (1966) *Interval Analysis*, Prentice Hall, Englewood Cliffs, NJ.

Ree and Jager (1993) Control-Data Representation at Knowledge Level, in *IEEE International Conference on Systems, Man and Cybernetics*, Vol. 4 Le Touquet (pp. 702–706).

Reznik, L. and Dabke, D. (1999) Evaluation of Uncertainty in Measurement; a Proposal for Application of Intelligent Methods, in *XV Imeko World Congress*, Osaka (pp. 93–100).

Sugeno, M. (1974) *Theory of Fuzzy Integrals and its Applications*, Doctoral thesis, Tokyo Institute of Technology.

Zadeh, L.A. (1965) Fuzzy Sets. *Information and Control*, **8**, 338–353.

Zadeh, L.A. (1971) Quantitative Fuzzy Semantics. *Information Sciences*, **3**, 159–176.

Zadeh, L.A. (1975) The Concept of a Linguistic Variable and its Application to Approximate Reasoning. *Information Sciences*, Part 1: **8**(3), 199–249; Part 2: **8**, 301–357; Part 3: **9**, 43–80.

13: Signals, Information and Knowledge, and Meaning

Qing Ping Yang
Brunel University, Uxbridge, UK

1 Nature of Cognitive Understanding and Relationship to Knowledge and Ideas 68
2 Definition and Application of Data 68
3 Definition and Application of Information 69
4 Definition and Application of Knowledge 71
5 Definition and Application of Wisdom 72
6 Definition and Application of a Signal 72
7 Human and Machine Signals 73
8 Role of Cognitive Psychology 74
References 74

1 NATURE OF COGNITIVE UNDERSTANDING AND RELATIONSHIP TO KNOWLEDGE AND IDEAS

The real world exists independently of our perception, and one is aware of the outside real world through seeing, hearing, and other senses. Human perception organizes sensory stimulation into percepts of various objects and phenomena. Sufficient experience of external objects and phenomena enables one to categorize and characterize them by means of defined properties or attributes. Basic knowledge is thus formed as ideas, concepts (including conceptual relations) about these attributes and the associated matters. Figure 1 shows the basic relation of human experience or cognition.

In fact, human cognition may be split into two processes, as shown in Figure 2. The first one is essentially a perception process, while the second one involves some memory and learning of the empirical relations.

Human cognition is an empirical process, subjective to possible a priori propositions and individual experience (Yang and Butler, 1997). Although the limitation of human senses often results in unsatisfactory knowledge, human perception, and cognition always plays a fundamental role in measurement, in particular, in the definition of a measurand and its associated measurement scale.

Modern sciences have aimed at systematically organizing objectively verifiable empirical experience and knowledge as scientific knowledge. Scientific knowledge is often quantifiable, precise, concise and thus easy to process and transform. Measurement plays a central role in the acquisition of scientific knowledge. Measuring tools, that is, instruments, are used to extend and improve the power of human senses. The machine perception performed by measuring instruments transforms the identified measurable properties or attributes (i.e. measurands) into some predefined numbers or symbols through symbolic representation as in Finkelstein (1975). The information carried by these symbols can be extracted through data processing, and scientific knowledge about the observed system can then be developed on the basis of this information. Similar to human cognition shown in Figure 1, this process may be illustrated in Figure 3.

2 DEFINITION AND APPLICATION OF DATA

Data is a plural form of datum and generally refers to 'known facts or things used as a basis for inference or reckoning' (Pearsell, 1999). In information systems including

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

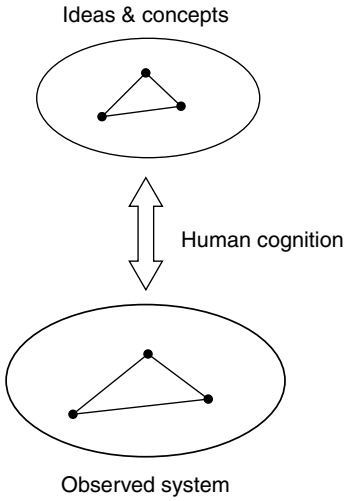


Figure 1. Human cognition.

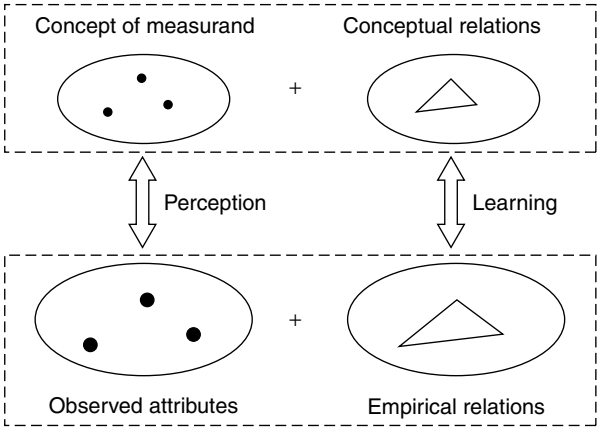


Figure 2. Perception and learning in human cognition.

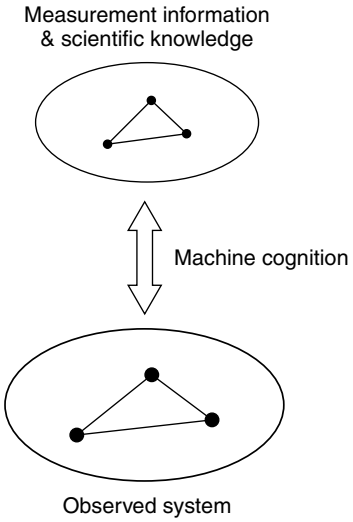


Figure 3. Machine cognition.

measurement, data may be defined as 'the quantities, characters, or symbols on which operations are performed by computers and other automatic equipment, and which may be stored or transmitted in the form of electrical signals, records on magnetic tape or punched cards, etc.', according to the Oxford English Dictionary (Simpson and Weiner, 2002). Some common examples include numbers, characters, images, and sounds.

Data are one kind of information carrier, and information is an integral part of data. Data are the mixture or fusion of carrier and information. Information or meaning can only be extracted and understood through some kind of data processing (e.g. brain, computer).

In measurement, data are usually recorded numbers or signals stored in various forms. Data have to be collected and processed before information and knowledge can be extracted. Typically, data is collected through a sampling process, both spatial and temporal, as discussed in Yang and Butler (1997, 1998). They have to be representative, that is, the information they carry should be essentially the same as or significantly close to that exhibited by the observed system.

It is often necessary to separate data collection or acquisition from data processing, although an integrated process would be more desirable and has indeed become increasingly common. The collected data that have only gone through minimum conditioning and processing are often called *raw data*.

Since computers can easily process digital data (both discrete-valued and discrete-time), raw data should ideally be in digital format. Continuous-time (typically also continuous-valued) signals become discrete-time and continuous-valued data (or sampled data) through sampling in time domain. If raw data are already discrete-time, but continuous-valued, they can be converted to digital data using an (A to D) analog-to-digital converter. The raw data can be simply continuous-time and continuous-valued signals, for example, recorded on magnetic tape, in which case both sampling and A to D conversion is required to generate digital data.

3 DEFINITION AND APPLICATION OF INFORMATION

Information may be defined as something or knowledge communicated concerning some particular fact, subject, or event. It can be separated from the person informed and can produce different responses in something, and is capable of being stored in, transferred by, and communicated to inanimate things (Simpson and Weiner 2002).

Information exists objectively and, unlike knowledge, it does not require a human subject, although it ultimately

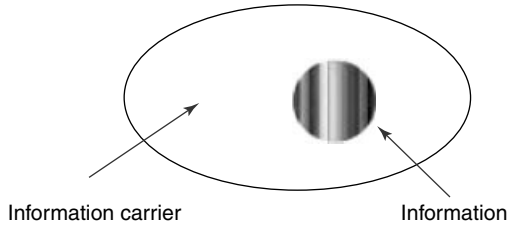


Figure 4. Information cell.

serves human beings. Information becomes knowledge once it is understood and possessed by the informed subject.

Information cannot exist on its own and it necessarily requires some kind of information carrier. Information is usually carried and conveyed by signals in various energy forms, for example, mechanical, electrical, optical, and so on. It can also be stored in various media, for example, printed text, electronic file, and even in the human brain. The relationship between information and information carrier may be generally presented as an information cell model, as shown in Figure 4.

The cell may manifest itself in various forms, for example, electronic signal, data, text, sound, image, brain, and even real-world objects. The cell body denotes the information carrier, while the nucleus actually represents the information carried and conveyed.

3.1 Information and measurement

On the basis of the information cell model, machine perception together with human perception may be represented as shown in Figure 5. Measuring instruments therefore essentially function as an information flow channel, or information machine as discussed in Draper, McKay and Lees (1952), Finkelstein (1982) and Sydenham, Hancock and Thorn (1989). Measurement is an information acquisition process, primarily concerned with the capture (or sampling), enhancement, extraction, and presentation of information. Ideally, the information should be preserved as much as possible in the measuring chain and could be easily separated or extracted from the carrier.

In measurement, information is carried and conveyed by measurement signals or recorded data. The information in a measurement is ultimately knowledge about the measured object. It is primarily concerned with the magnitudes and the relations of the measurand(s), represented on some predefined scales, as in the representation theory of measurement discussed in Pfanzagl (1968), Krantz *et al.* (1971), Narens (1985), Finkelstein (1982).

The information may be regarded as the meaning of the measurement data, and, as such, has to be extracted using various data transforming techniques. There are two distinct types of data transforming involved in a measurement process, as shown in Figure 6. The first transformation (AB) converts the measurand(s) to signals in various forms, which can then be conditioned and enhanced. The central function of this transformation is the generation of machine percepts or representation of measurand(s) in the form of numbers or symbols. The second transformation (BC) converts the symbols into meaningful information, although the latter also usually takes the form of numbers or symbols. The key process of the second transformation is calibration. The first transformation is data-driven, relating the measurand(s) to the measurement numbers or symbols, whereas the second one is information-driven, relating the symbol relations back to empirical relations.

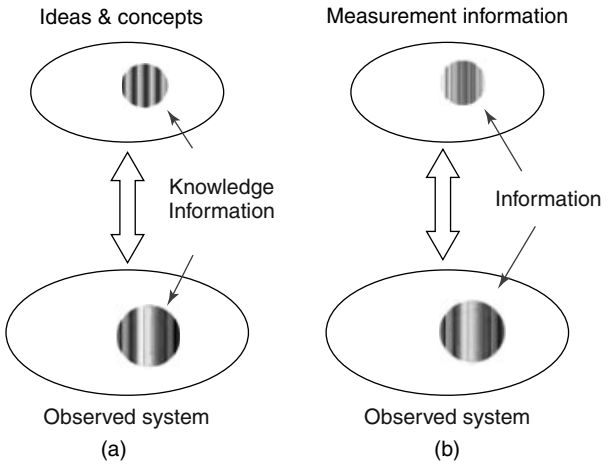


Figure 5. (a) Human machine and (b) machine perception.

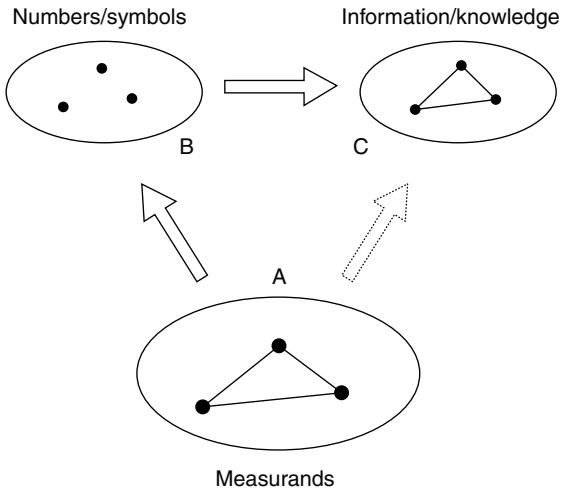


Figure 6. Extraction of measurement information.

4 DEFINITION AND APPLICATION OF KNOWLEDGE

Knowledge can be generally defined as the familiarity gained by experience of a fact, thing, or person. Knowledge is usually in reference to a subject or person, and information becomes knowledge as information is interpreted and possessed by a person (Sydenham, Hancock and Thorn, 1989). In fact, human knowledge could also be represented and stored as information.

From epistemology's point of view, knowledge can be obtained through direct and indirect experience. Ideas and concepts are formed from human perception and from the observer's direct experience of the real world. Human knowledge gained also comes from the experience of the manifestation of the information carried by the observed system.

Indirect experience can be any experience other than direct experience. For example, this may be the experience of somebody else, or one's internal experience and mental activities, that is, reasoning and deduction, which result in rational knowledge. Indirect knowledge should also include the knowledge acquired through machine perception. The role of instruments and measurement is not only extension of the range and capability of human senses but also the enhancement of indirect experience and knowledge.

Measurement starts with human direct experience, and as basic concepts of attributes or measurands are formed, some basic and fuzzy knowledge is developed. Measurement serves to quantify and improve this kind of knowledge, generating, and refining scientific knowledge, which is usually precise and easy to process and apply.

4.1 Direct and indirect knowledge in measurement

Of course, human knowledge comes from both direct and indirect experiences. Figure 7 shows the combined relations between human perception and machine perception, where the observed object or system is the key link of the two perception processes. Since machine perception and measurement may be regarded as a kind of indirect experience, Figure 7 also models the interactions between direct and indirect experiences. It is widely recognized that one's direct experience plays an important role generally in his or her cognition process, knowledge acquisition and, in particular, indirect experience.

4.2 Level of abstraction and knowledge pyramid

Both human and machine perceptions aim at extracting information and knowledge, but the two have different

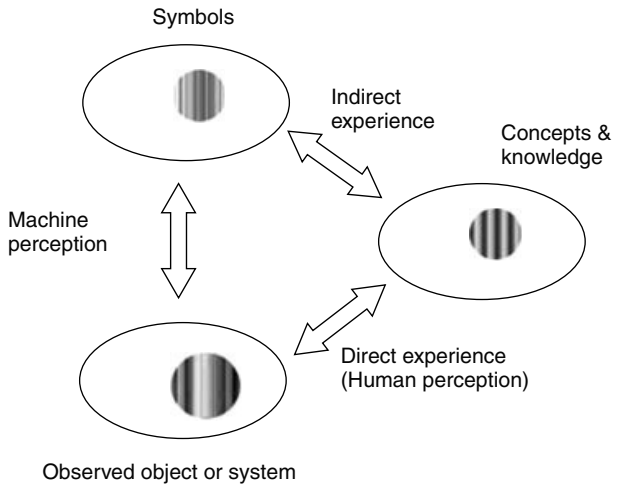


Figure 7. Direct and indirect knowledge.

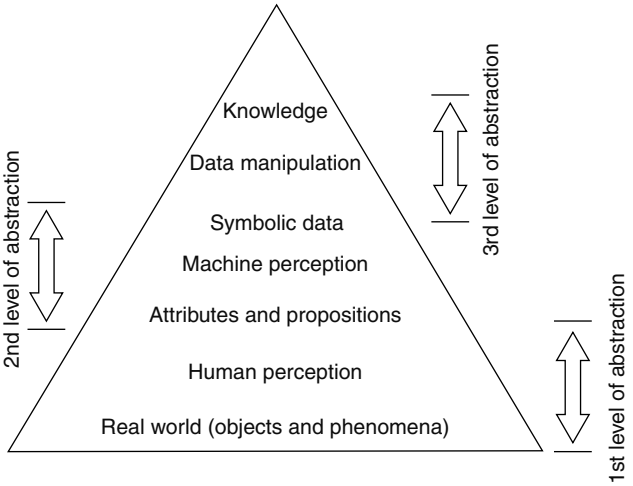


Figure 8. Knowledge pyramid.

levels of abstraction. In terms of the level of abstraction, the relationship between knowledge and human/machine perceptions can be presented as a knowledge pyramid, as in Yang and Butler (1997) and as shown in Figure 8.

The first level of abstraction is the concept of attributes and their relations formed after human perception through direct experience. The second level of abstraction transforms these attributes to predefined symbols and it represents the observed system in symbols. The third level of abstraction further transforms the symbols into information and knowledge about the observed system through general data manipulation.

4.3 Measurement and other branches of science

Measurement science requires the underpinning of some other branches of science, for example, mathematics and

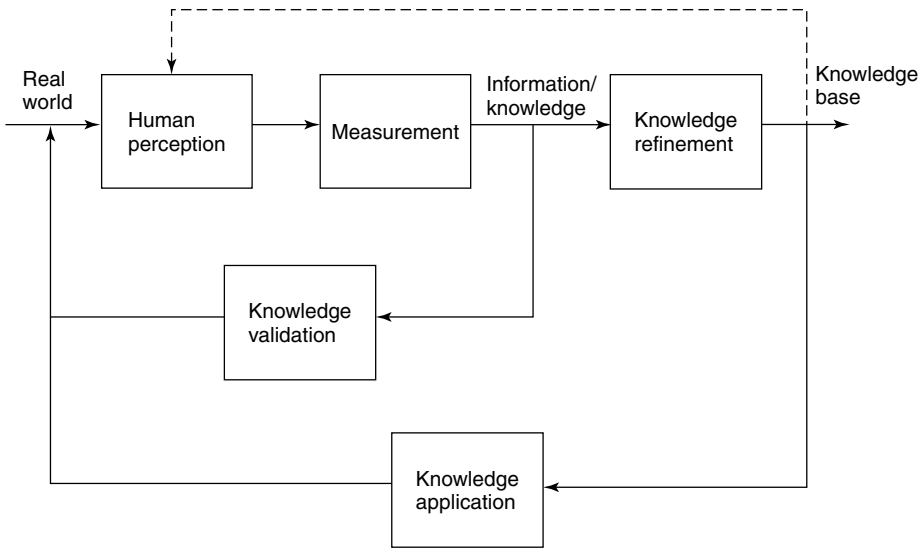


Figure 9. Measurement and other branches of science.

other fields of information science including data processing, communication, and control. Measurement also underpins many other branches of science, for example, physical, chemical, and biological science. Domain experts rely upon measurement and instrumentation to study their problems. But they work on a larger framework, as shown in Figure 9, where the knowledge obtained from measurement about the measured attributes help them understand the objects or phenomenon in the system concerned. This understanding leads to knowledge at a higher level, which may be further refined through validation and which constitutes the science of the subject domain concerned. The objective of measurement is knowledge acquisition, whereas those of the domain experts are primarily concerned with the direct practical experience, knowledge validation, and knowledge refinement (Yang and Butler, 1997).

5 DEFINITION AND APPLICATION OF WISDOM

Wisdom may be defined as special or advanced knowledge that is developed from wider experience and is often not easily or naturally associated or applied.

As discussed in the previous section, measurement extends our experience and knowledge. It often functions as part of an overall system that applies and makes use of the knowledge generated in measurement. It underpins many other branches of science. Measurement can thus be a source of intelligence or wisdom, especially in a complicated system or process in which advanced knowledge is required and/or should be applied. For instance, it can be argued that the development of our knowledge in physics generally involves some kind of measurement, including those very advanced ones representing the wisdom of mankind, for example, Einstein's relativity theory was based upon the measurement of the speed of light.

Having discussed various related concepts like data, information, knowledge, and wisdom, it is common wisdom to use simple examples to clarify the relations and differences of related concepts!

Suppose we have received the following parts of texts:

A: 'house, prices, UK, rise, fastest, annual, market, last, peaked, 25.3%, 2002, 1989'.
B: 'House prices in the UK rose by 25.3% over 2002, the fastest annual rise since the market last peaked in 1989.'
C: 'Property can be a very good investment in the UK.'
D: 'What goes up can come down, and usually quite sharply.'

Part A could best be described as data because it mainly contains individual words and their relations are not explicitly given. Part B could best be described as information with the relations between words and concepts clearly stated. Part C could best be described as knowledge derived from the interpretation of the information (Part B). Of course, additional knowledge can also be developed. Part D could best be described as wisdom.

6 DEFINITION AND APPLICATION OF A SIGNAL

Signals may be defined as 'a modulation of an electric current, electromagnetic wave, or the like by means of which

information is conveyed from one place to another', according to the Oxford English Dictionary (Simpson and Weiner, 2002). Signals are closely related to data, and in fact they become data once recorded. But by definition they have wider existence in natural and man-made systems, and thus play a more important role. Like data, they are information carriers. Information is carried by the magnitude or attributes of the variation of signals (Finkelstein, 1994).

A signal is essentially the change or modulation of an energy state, which can be detected by a human observer or instrument. Signals can be regarded as factors of energy or power, and energy as the signal carrier. Signals are generated from energy connections and interactions between the objects and matter in the system concerned. Various energy forms may be involved, for example, electrical, mechanical, thermal, optical, magnetic, and chemical. Each element of the measuring chain can be regarded as a transducer and the energy flow may change form in each transducer. Energy flow determines the signal flow in a measuring instrument.

6.1 Signal processing

Since the energy associated with a signal is influenced or modulated not only by the measurand but by other effects also, noises will always be present in a measuring chain together with signals. It is thus necessary to carry out signal conditioning and signal processing to increase the signal-to-noise ratio.

Signal processing is a technology concerned with the representation, transformation, and manipulation of signals and the information they contain, in such a way as to make them more useful in a given application. Signal processing has a long history. The last four decades have seen its extremely rapid growth, partly due to availability of the ever-increasing power of devices and computers, both hardware and software, and partly due to its prevalence in the advances of many branches of science and technology (Yang, 2001).

Signals may be represented and processed in various domains, for example, time domain, frequency domain, and time-frequency domain. The ultimate goal of signal processing is information enhancement and extraction, for example, noise reduction, classification, detection, estimation, prediction, feature extraction, and so on.

7 HUMAN AND MACHINE SIGNALS

Machine signals are predominately used nowadays, and they are generated by and also mainly intended for machines. But signals are closely related to the signs intended for humans and may be utilized as human signals. In fact, one definition refers to a signal as 'a sign or notice, perceptible by sight or hearing, given especially for the purpose of conveying warning, direction, or information' (Simpson and Weiner, 2002).

7.1 Human signals and machine signals in measurement

In instrumentation and measurement, machine signals, and human signals are closely related in several ways. Firstly, human sensation and perception based upon the direct detection or sense of human signals are an important and necessary part of measurement and knowledge acquisition. Although they tend to have limited range and power, human perception can directly lead to the understanding of measurand(s) and their empirical relations, which are the basis of measurement.

Secondly, although machine signals may be quite different from human signals, they are actually often based upon human signals. This is necessary because they should carry and convey the same information. The similarity may extend to the signal generation mechanisms and energy forms.

Thirdly, machine perception extends the range and power of human senses. Measuring instruments or machines are able to handle more signals with greater capability. They can generate quantified numerical results, which subsequently lead to refined scientific knowledge.

Finally, the output signals of a measuring instrument or machine perception often have to be presented to a human observer. Indicating or displaying instruments have been a very important part of a measurement system. As measurement is often intended for the human observer, the interpretation of the output signals becomes an integral part of measurement.

7.2 Signals and semiotics

Indeed, as measurement signals may be regarded as symbols or signs, their interpretation may be explained using a semiotic triadic model or semiotic triangle, in which the meaning of the sign is related to both the specific sign presented and the observer's empirical experience of the real-world object.

Similar explanations can be found in measurement. In fact, Figures 6 and 7 presented in the previous sections are essentially a triadic model and may be referred to as a measurement triangle, as shown in Figure 10. Further discussions on semiotic issues in measurement are given in **Article 114, Models of the Sensor Interface, Volume 2**,

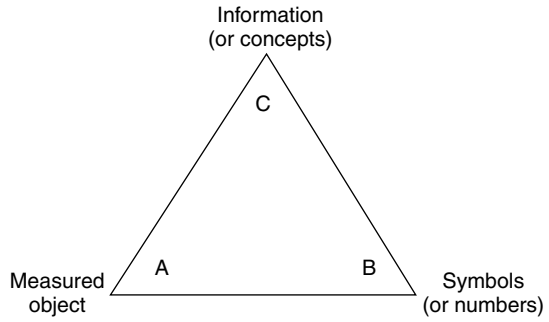


Figure 10. Measurement triangle.

but it is interesting to note here that measurement is primarily concerned with how measurement symbols (signals or data) are related to the observed object(s), whereas semiotics centers around how the symbols are interpreted. Both of them require the human observer's direct experience of the related system or process.

8 ROLE OF COGNITIVE PSYCHOLOGY

As an empirical science, cognitive psychology is concerned with mental processes and their effects on human behavior. It studies mental processes such as attention, sensation, perception, memory, language, reasoning, learning, problem solving, and decision making.

Human sensation and perception are the basis of measurement. As discussed in the previous sections, concepts, and definitions of measurands as well as their measurement scales are based upon experience and understanding of empirical relations developed from human sensation and perception.

Since human perception is subject to the limitation of human senses and direct experience, the concept of measurand and its measurement may go through continuous development and evolution. For example, the length standard has been redefined several times to reflect the latest understanding of the measurand.

The study of human sensation and perception could also provide useful examples, references, and knowledge for sensing and measurement. In some applications, sensors may even need to be integrated with human senses, for example, hearing aids.

At a higher level, the understanding of memory, language, reasoning, learning, and decision making can make a great contribution to information extraction, knowledge representation, manipulation, storing, and retrieval in measurement. It may also lead to suitable models for intelligent sensors and instrumentation.

REFERENCES

Draper, C.S., McKay, W. and Lees, S. (1952) *Instrument Engineering*, Vols. 1–3, McGraw-Hill, New York.

Finkelstein, L. (1975) Representation by Symbols as an Extension of the Concept of Measurement. *Kybernetes*, **4**, 215–223.

Finkelstein, L. (1982) Theory and Philosophy of Measurement, in *Handbook of Measurement Science*, *Volume 1 Theoretical Fundamentals* (ed P.H. Sydenham), John Wiley & Sons, Chichester (pp. 1–30).

Finkelstein, L. (1994) Measurement and Instrumentation Science – An Analytical Review. *Measurement*, **14**, 3–14.

Krantz, D.R., Luce, R.D., Suppes, P. and Tversky, A. (1971) *Foundations of Measurement*, Academic Press, New York.

Narens, L. (1985) *Abstract Measurement Theory*, MIT Press, Cambridge, MA.

Pearsell, J. (1999) *The Concise Oxford Dictionary*, 10th edn, Oxford University Press, Oxford.

Pfanzagl, J. (1968) *Theory of Measurement*, Physica-Verlag, Würzburg, Vienna.

Simpson, J. and Weiner, E. (2002) *The Oxford English Dictionary*, 2nd edn, Oxford University Press, Oxford. URL: http://dictionary.oed.com/.

Sydenham, P.H., Hancock, N.H. and Thorn, R. (1989) *Introduction to Measurement Science and Engineering*, John Wiley & Sons, Chichester.

Yang, Q. (2001) Introduction to the Special Issue on Signal Processing. *Measurement and Control*, **34**(8), 228.

Yang, Q. and Butler, C. (1997) On the Framework of Measurement Science, in *Proceedings XIV IMEKO World Congress*, Vol. 5, Tampere (pp. 101–106).

Yang, Q. and Butler, C. (1998) Object-Oriented Model of Measurement Systems. *IEEE Transactions on Instrumentation and Measurement*, **47**(1), 104–107.

14: Hierarchical Aspects of Measurement Systems

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 The Systems Hierarchy of Measurement Systems 75
2 Machines Hierarchy of Measurement Systems 75
3 Secondary Functions or Orders of Measurement Systems 77
4 Formal Data Measurement Components 77
5 Boundary Perspective of the Four Information Machines 78
Related Articles 78
References 78

1 THE SYSTEMS HIERARCHY OF MEASUREMENT SYSTEMS

As pointed out in **Article 5, Humans in the Real World, Volume 1**; **Article 2, Organization of Instrument Science, Volume 1**; and **Article 6, Substructure of Human–Machine Systems, Volume 1** (McGhee, Henderson and Sankowski, 1986 and McGhee *et al*., 1996), the wide view that Systems Engineering permits, highlights the placing of information and information systems within the structure of human–machine systems. The substructure of human–machine systems (**Article 6, Substructure of Human–Machine Systems, Volume 1**); (McGhee, Henderson and Sankowski, 1986; McGhee *et al*., 1996) may be represented by the hierarchical dendrograph or tree diagram given in the classification of Figure 1, where it can be seen that technical instruments, also called *instrumentation*, are clearly located as elements within the information subsystem of human–machine systems. Functional and structural characterization plays a primary role in this systems approach. Thus, the functions and boundary structure of information machines may be linked to the hierarchical structure of the complete system. Technical instrument systems, also called *technical instruments*, perform a number of different functions. Four primary functions, which are also shown in Figure 1, distinguish between instruments for measuring, calculating, communicating, and controlling.

2 MACHINES HIERARCHY OF MEASUREMENT SYSTEMS

Technical instruments are one of the two types of human-made machines. It is important to visualize the place occupied by these machines in a hierarchical classification of entities in the physical universe as illustrated in Figure 2. This will be recognized as similar to the hierarchy used in everyday situations by bioscientists and earth scientists who are interested in the ordering of the variety of living and nonliving things. The grouping of the physical and life sciences allows a clear division between these two important sciences. Two classes or groups of things, which may be called *superkingdoms*, make up the terrestrial *hyperkingdom*. The *superkingdom of living things* is made up of animals, plants, and other life-forms such as fungi and microbiological organisms. In the *superkingdom of nonliving things* it is possible to identify the separate groups or *kingdoms* of materials and machines. Materials that are important in all engineering activities may occur naturally as earth and minerals. There is also a diversity of human-made materials.

Only the *division* of the machine kingdom called the *division of information machines* is of immediate interest in this article. These machines may be grouped under

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

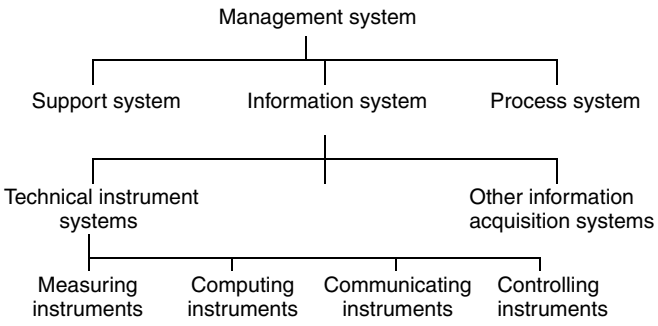


Figure 1. Instruments in a systems hierarchy.

the *subdivisions* of *deductive* or *inductive* types. Deductive types of machines perform mathematical calculations that inexorably lead to only one possible calculated result. On the other hand, inductive machines use rules for reasoning so that logical inference is used to form a conclusion, which is not generally numeric. Further distinctions lead to the four *classes* of information machines, which allow measuring, calculating, communicating, or controlling operations to be performed. A significant feature of these diagrams is the representation of the calculating function as providing an important primary support for the other three functions. This does not imply that calculation or processing cannot exist independently of the others. Indeed, the emphasis is meant to indicate the overall importance, which this data processing function possesses.

Further analysis indicates that lower levels of classification can be proposed in the same way as for biological systems. Thus, the lower level groups are called the *order*, *family*, *species*, and *subspecies* of the *machine kingdom*. The *order* and *family* groups will be considered below. Lower orders of this kingdom can be built up by considering specific kinds of sensors and other elements, that Solopchenko (1994) has called *formal data measurement components*, or FDMCs.

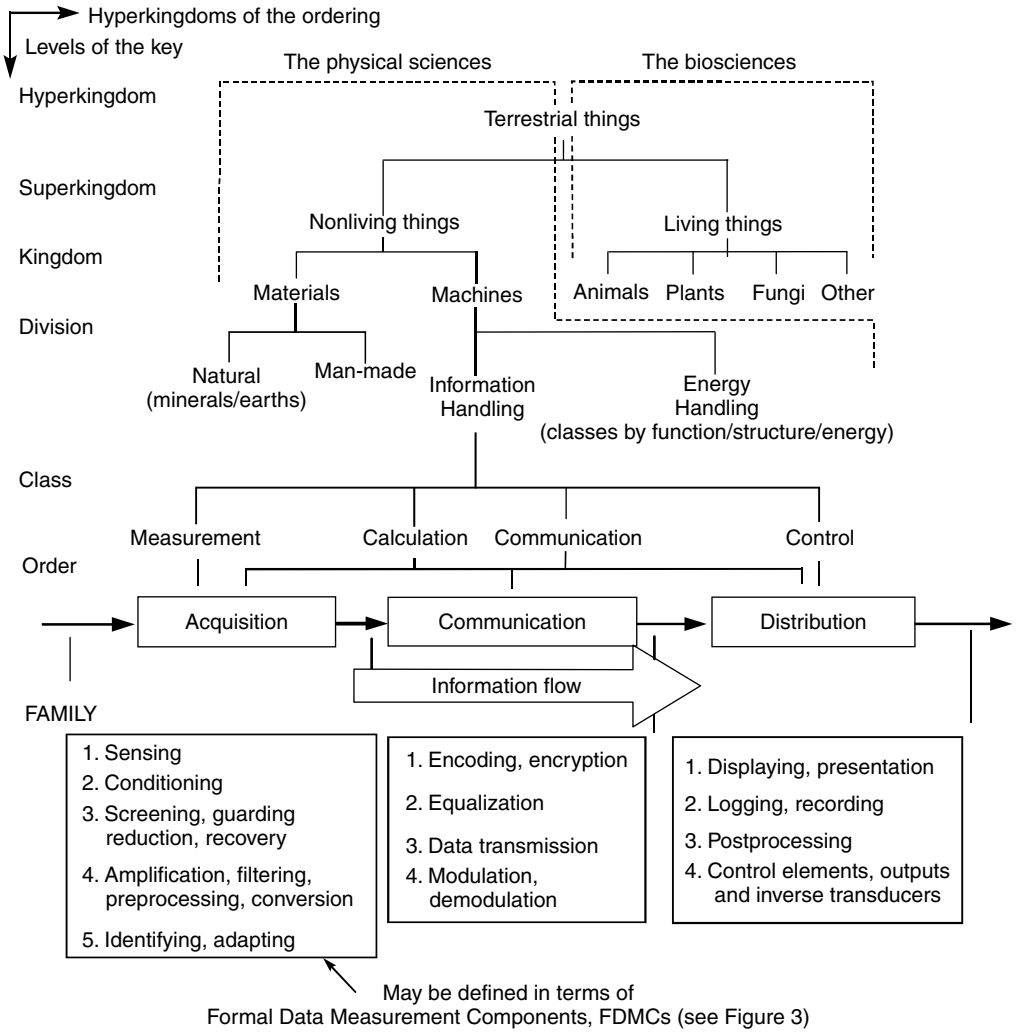


Figure 2. Instruments in a machines hierarchy. From the physical and biosciences to the ORDERS of technical instruments.

3 SECONDARY FUNCTIONS OR ORDERS OF MEASUREMENT SYSTEMS

Before any data can be processed, an ACQUISITION SUBSYSTEM acquires it. Within the context of measuring systems, further subdivision by function is possible as shown in Figure 2. Hence, a separation of the sensing and transduction functions, from those performed by any conditioning of the sensor output, is apparent. *Conditioning functions* may be galvanic isolation, screening, and guarding against electromagnetic penetration, circuit completion, and supply of energy to modulating types of sensors. Preamplification, which is also a subsidiary function appropriate to measurement, emphasizes the importance of electronic circuits and systems in modern instrumentation (Lang, 1987).

Information *Communication and Distribution* are important from the broader view of measurement in particular and information technology in general. Some subsidiary functions appropriate to communication systems are given in Figure 2. Aspects of distribution systems, such as data presentation and recording, and human–machine interfaces, are relevant for measuring systems. The dendrograph of Figure 2 also provides a clear illustration of the place occupied by materials science, which is important for sensors, in the ordering of the physical sciences.

4 FORMAL DATA MEASUREMENT COMPONENTS

Measurement systems are formed from the interconnection of FDMCs (McGhee *et al*., 2001; Solopchenko, 1994), which are their subsystems, elements, or building blocks. The term Data Measurement is used by Solopchenko (1994). A clear view of the place occupied by information machines, one of the two types of human-made machines in the physical universe, is given in Figure 2, which has been built upon the ideas of classification. The relationship between the various key levels of the division of information machines as far as the individual Formal Data Measurement Components, or FDMC, is shown in Figure 2. In the present context, they are referred to as *Data Measurement components*, or DMCs, to emphasize that analog and digital signals are used to carry information in the form of data from the sensor to the output of the instrument.

Extending Figure 2 allows the classification of FDMCs given in Figure 3. It can be seen from this figure that analog sensors are classified as Formal Analog DMCs, or FADMCs, while digital sensors are classified as Formal Analog-to-Digital DMCs, or FA-to-DDMCs. Each of the FDMCs shown in the tree diagram may be represented by a universal, coherent model (McGhee *et al*., 2001; Solopchenko, 1994), which allows the errors at each stage of a measurement chain to be taken into account. Since such an approach allows a single block diagram to be used to represent widely different DM components, the primary measurement aim of taking account of errors, their sources, and their quantification is achieved. The classification given in Figure 3 is supported by the definition of Data Measurement systems in Solopchenko (1994) and McGhee *et al*. (2001). It can be seen that measuring instruments perform a number of functions such as sensing, conditioning, screening, and guarding, amplification, and filtering as well as others. These functions may be described in terms of FDMCs as given in the tree diagram of Figure 3.

The physical world is dominantly analog in nature. Since interfacing with the real world is usually accomplished using analog means, it is clear that there is a group of FADMCs. This group of components is appropriately used

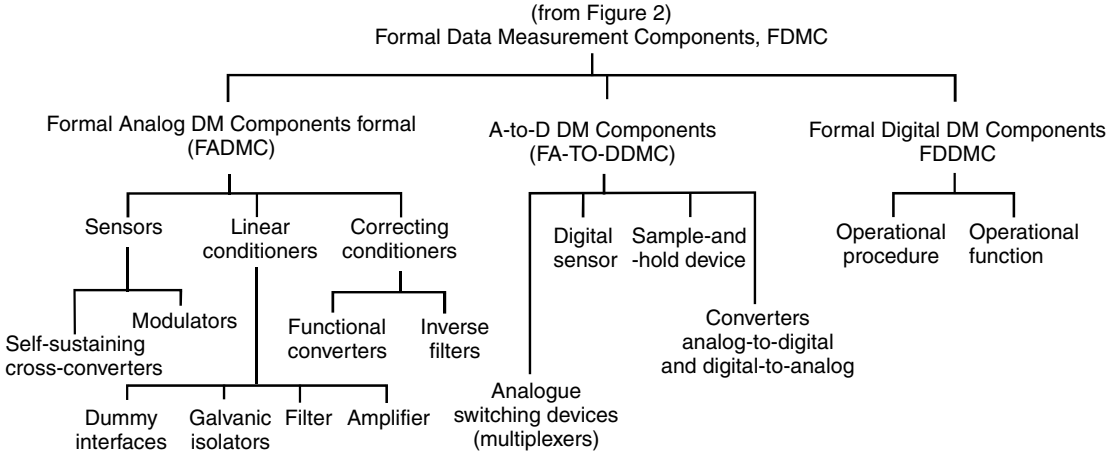


Figure 3. The classification of formal data measurement components, FDMCs.

when analog information handling operations are to be performed on the information bearing signals. In accordance with the predominance of digital instrumentation, most contemporary instrumentation uses a minimum of analog circuits. The group of formal components used in this case are named FA-to-DDMCs. Thus, all forms of analog-to-digital and digital-to-analog converters belong within this group. Other components are used for multiplexing purposes and for sample-and-hold operations.

It has been observed that software is the most important group of components in modern instrumentation systems. The two main groups of FDDMCs, may be either operational procedures or operational functions.

5 BOUNDARY PERSPECTIVE OF THE FOUR INFORMATION MACHINES

A combination of the tree diagrams of Figures 1 and 2 with the boundary view of human–machine systems leads to the boundary view of the four information machines given in Figure 4. This approach again illustrates the recursive aspect of the total, or holistic, view, which is a feature of the systems engineering method (McGhee *et al*., 1996; McGhee, Henderson and Sankowski, 1986). Only one output is shown in these diagrams. However, the totality of influencing impacts is evident.

A distinction is also seen between the control instrumentation system and the other three. Power flow is ultimately modulated or regulated by information feedback in control instrumentation systems, whereas power flow supports information flow in the other three. The information input or output at the system boundary may also be in any of the COMETMAN energy forms described in **Article 109, Structure and Energy in Sensor Systems, Volume 2**. In the case of the measurement system, typical variables to be measured in different energy forms could be mechanical displacement, fluid flow rate, or thermodynamic temperature. As a complete visualization is complex, a simpler consideration is appropriate. The information and energy tetrahedron, described for systems in **Article 6, Substructure of Human–Machine Systems, Volume 1** and more fully for sensors in **Article 109, Structure and Energy in Sensor Systems, Volume 2**, is an illustration of such a simplification for the resource, contamination, and energy or signal forms of the input. Again it should be emphasized that, when this figure represents a measuring system with electrical information carrying output signals, the importance of the disturbance inputs is immediately seen. Contamination of the electrical output signal may be due to any of the energy forms shown, although only the temperature and chemical form are considered. Thus, the holistic aspect of the systems engineering approach, when applied to contamination using the information and energy tetrahedron, allows the introduction of the concepts relating to error sources and error characterization. Solopchenko (1994) has emphasized that taking errors and their sources into account is a primary aim in the analysis and design of measurement systems.

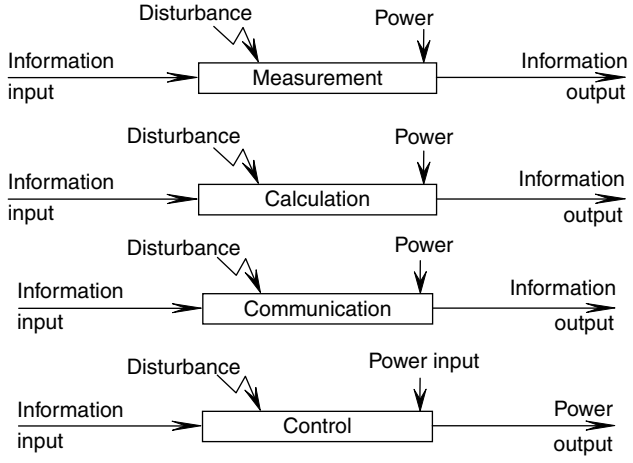


Figure 4. Boundary perspective of the four information machines.

RELATED ARTICLES

Article 2, Organization of Instrument Science, Volume 1; **Article 5, Humans in the Real World, Volume 1**; **Article 6, Substructure of Human–Machine Systems, Volume 1**.

REFERENCES

Lang, T.T. (1987) *Electronics of Measuring Systems*, John Wiley & Sons, Chichester.

McGhee, J., Henderson, I.A. and Sankowski, D. (1986) Functions and Structures in Measurement Systems: A Systems Engineering Context for Instrumentation. *Measurement*, **4**(3), 111–119.

McGhee, J., Henderson, I.A., Kulesza, W. and Korczynski, J. (2001) *Measurement Data Handling: Theoretical Techniques*, ISBN 83 7283 007 X, A.C.G.M. Lodart S.A., Lodz.

McGhee, J., Korczynski, M.J., Kulesza, W. and Henderson, I.A. (1996) *Scientific Metrology*, ISBN 83 904299 9 3, A.C.G.M. Lodart, Lodz.

Solopchenko, G.N. (1994) Formal Metrological Components of Measuring Systems. *Measurement*, **13**, 1–12.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

15: Typical Measurement Systems Architectures

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Generalized Architecture of a Measurement System 79
2 Microprocessor-based Instrumentation 79
3 Process Control Instrumentation 80
4 Petroleum and Gas Process Instrumentation 82
5 Supervisory Control and Data Acquisition System 82
6 Remote Transmitter Units 82
Related Articles 83
References 84

1 GENERALIZED ARCHITECTURE OF A MEASUREMENT SYSTEM

A measuring system has the general structure shown in Figure 1 (McGhee *et al*., 1996). There are three main divisions within the system. Input elements are primary sensing elements and transducers. This distinction is not trivial as there are many cases in which the sensing element is fashioned for one energy domain, while the transducer as its name implies converts the output of the sensing element into a signal in another energy form. The intermediate elements are usually electronic circuits and systems that make extensive use of *large scale integration* and possibly distributed *computing,* where *microprocessors* and *personal computers* are dispersed throughout the system. *Data transmission* and *telemetry* methods would be used if the measuring point is remote from the observation or display point. *Signal recovery/detection* may be necessary if the information is transmitted using a suitable form of modulation. Output elements could consist of some *postprocessing* before the output of the measuring system is used to provide a soft copy in a *display*, a hard copy in a *plotter*, or *printer*, or a stored record in a *data-logger*.

2 MICROPROCESSOR-BASED INSTRUMENTATION

The hierarchical subdivision of measurement systems reveals the existence of four classes of *Information Machines* (**Article 14, Hierarchical Aspects of Measurement Systems, Volume 1**; McGhee *et al*., 1996; McGhee, Henderson and Sankowski, 1986) that perform the primary functions of

1. measurement,
2. calculation,
3. communication,
4. control.

Essential support to the remaining three classes of information machines is provided by the calculating machine. They may also be used as central elements in microprocessor-based instrumentation (**Article 158, Microcontroller Systems, Volume 3**), as shown in Figure 2. Note the different structures at the input end of the system (Lang, 1987; McGhee *et al*., 1996, 2001; McGhee, Henderson and Sankowski, 1986). Among the many elements, it is important to note the **s**ample-and-hold, or S&H, units, the multiplexers and the data converters. The first p inputs are analog signal inputs, which are analog multiplexed before being **s**ampled and held, or S&H. This S&H operation is sequentially performed on each signal. By contrast, each of the analog inputs from $(p + 1)$ to q are synchronously sampled and held before conversion. A

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

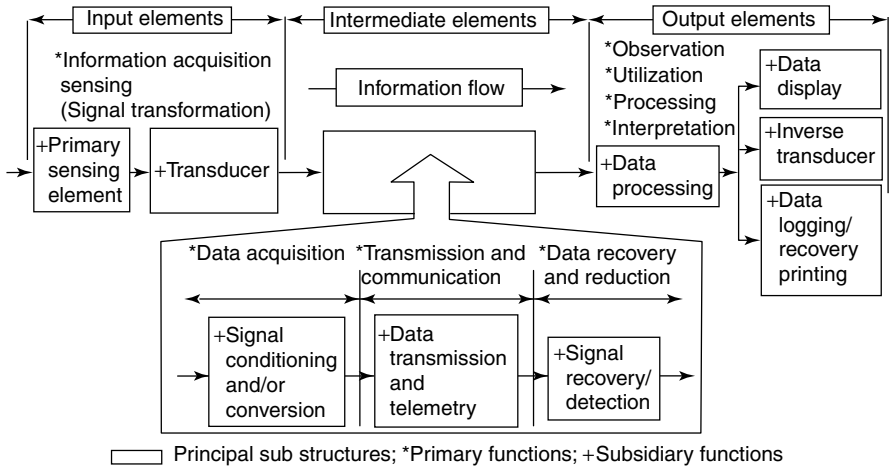


Figure 1. Generalized architecture for a measuring system. (Each system must be examined to allow consideration of particular features such as the absence of certain elements or the existence of more complex interconnection of structures).

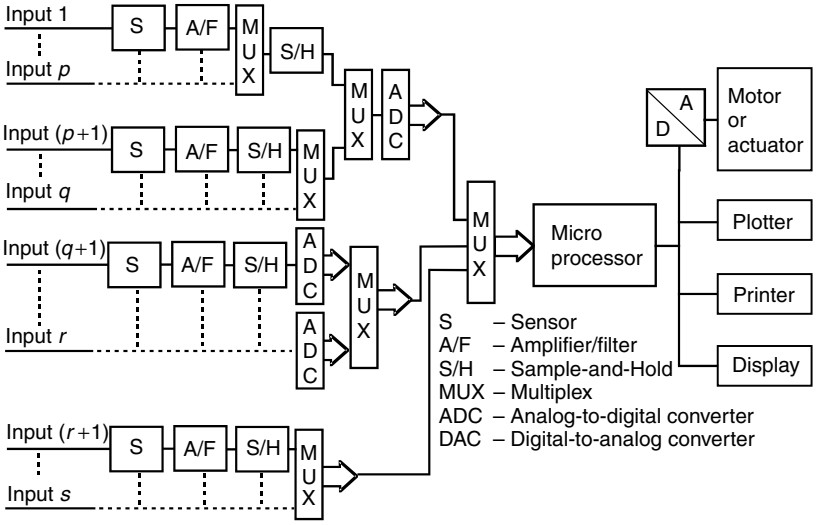


Figure 2. Architecture of microprocessor-based instrumentation.

possibility, which is shown for inputs $(q + 1)$ to r, is for individual channels to consist of a sensor, an amplifier/filter an S&H and a dedicated converter. All of the $(r + 1)$ to s inputs are assumed to be binary signals, which may be generated from switch contacts or other event detectors. In these lines, the amplifiers/filters may also include some kind of binary signal conditioner such as a Schmitt trigger.

Calculating machines, more commonly referred to as *computers, minicomputers* or *microprocessors*, which are considered in **Article 152, Fundamentals of the Stored Program Digital Computer, Volume 3**; **Article 153, Single Address Instruction Microcomputer, Volume 3**; **Article 154, Internal Operation of the Microprocessor, Volume 3**; **Article 155, External Operation of the Microprocessor, Volume 3**; **Article 156, Memory Management in the Microprocessor, Volume 3**; **Article 157, Data Acceleration in Computers, Volume 3**; and **Article 158, Microcontroller Systems, Volume 3**, process information using binary operations. These binary operations may belong to either the arithmetic group of operations or the Boolean group of operations. Arithmetic operations are addition, subtraction, multiplication, and division. Boolean operations consist of conjunction, or logical AND, disjunction, or logical NAND, and exclusive OR, or logical XOR. Microprocessors also allow the generation of diverse signal forms for use in system testing.

3 PROCESS CONTROL INSTRUMENTATION

A Process System is a complex combination of men, money, machines, and materials (**Article 5, Humans in**

the Real World, Volume 1, McGhee, Henderson and Sankowski, 1986) with the specific aim of producing some resource output having enhanced or surplus value. Instrumentation is required in process systems (Sydenham, 1983) to assist the humans involved in operating the process for assessing process behavior and so extending and refining their ability to monitor and control it.

In most cases, humans will be replaced by instruments for the following:

- *Measurement* – to provide more accurate, reliable data.
- *Calculation* – to process large quantities of data more quickly and efficiently.
- *Communication* – to transmit information for a wide range of uses, frequently between remotely separated transmitters and receivers.
- *Control* – to improve the quality of the process resource output with a minimal involvement of humans

A schematic of a typical process control measurement loop is shown in Figure 3. There are a number of identifiable stages of a process control system. The most commonly occurring structure is called a *loop*. In a loop, a sensor, also called a *gauge*, a primary transducer, or pickup, measures a physical variable whose value is to be controlled. As this signal is rarely in a form that is convenient for easy handling, it must be conditioned by using some appropriate technique. Such conditioning is most commonly performed using electronic circuitry consisting of amplifiers, filters, and data converters. All of this circuitry must be designed to be intrinsically safe if the instrument is to be used in a hazardous location.

Subsequently, this conditioned signal is transmitted or relayed to a *controller*. The controller compares the conditioned received signal with the process *set-point*. From this comparison, it generates a control signal, which is then amplified, if necessary, before being applied to an *actuator* and *final control element*. The actuator and final control element cause a manipulation that appropriately regulates the flow of energy in the form of say a fluid-flow rate, a pressure, a level, a temperature, or a chemical composition.

A loop will also consist of some means of monitoring its behavior by displaying the value of the set-point, the error, and any operational limits. Such facilities may also include a graphical mimic diagram of the process, as well as appropriate means for logging and recording of process data. Status, alarm, and warning indicators are also normally included.

Some severe problems, which must be dealt with using process instrumentation, are as follows:

- Generally, the instruments of whatever class must remain in operation for 24 h every day with attention only being provided during scheduled maintenance periods.
- The reliability of the instruments must be high to ensure that they perform their specified function. Failures cause lost production, which can be compounded if the failure endangers the safety of the plant or more seriously of the personnel or of the environment.
- Instrumentation may operate in a wide range of environments. They may operate in hazardous locations. They may also be subjected to extremely high or low temperatures, dry or humid. Under failure conditions, the instrumentation should always '*fail-to-safety*'. This means that hazardous processes should be rendered safe in a controlled manner, and in such a way as to present no dangerous occurrences or conditions, which could

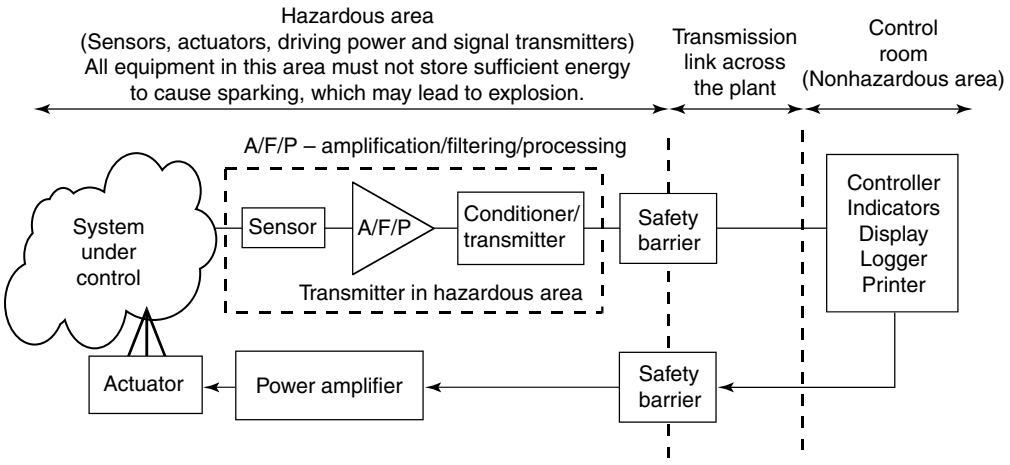


Figure 3. Architecture of a process control measurement loop.

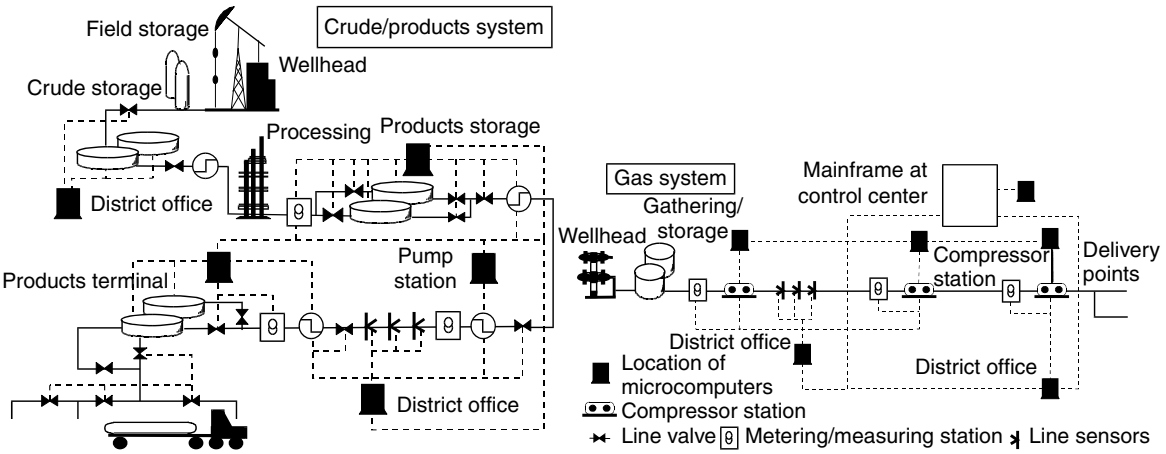


Figure 4. Schematic diagram of the instrumentation in a crude/petroleum products system and a gas process system.

lead to damage of equipment or, more seriously, to injury or death of personnel.

4 PETROLEUM AND GAS PROCESS INSTRUMENTATION

Computer-based instrumentation is now regarded as indispensable in these industries. Figure 4 shows the main parts of the instrumentation for both petroleum and gas process systems. Although petroleum and gas process systems handle different materials, they have the following common features:

- Both have wellheads, which are used to extract the raw resources.
- Secure storage facilities are required in both cases.
- Distribution is achieved through piping, which may be extensive.
- Security against dishonest fiscal accounting and custodial delivery is essential.

The figures emphasize the important role played by computers as well as the diversity of the systems and noteworthiness of their distributed nature. There are many remote measuring points. These may be in locations, such as at the wellhead storage, where large numbers of temperature, level, flow, and pressure variables need to be measured. It is necessary to integrate these measurements to provide effective control of such complex systems. Pipelines, which are an indispensable part of both petroleum and gas process systems, require supervision and control just as much as the process systems themselves. Indeed, the pipeline system could be considered a complex process in its own right.

5 SUPERVISORY CONTROL AND DATA ACQUISITION SYSTEM

To assist in the overall control of the process system and its distributing pipeline requires the incorporation of an important system known as a Supervisory Control And Data Acquisition, or SCADA, system. Figure 5 illustrates the block diagram of a typical SCADA system. The master station includes fully redundant computers that communicate over suitable links such as microwave or cable links to Remote Transmitter Units (RTUs).

The master station instrumentation also consists of what is called a Human–Machine Interface, or MMI. This interface consists of Visual Display Units, or VDUs, printers, and data loggers. Other graphic facilities will include a hard mimic hydraulic profile of the process and its associated pipeline. This MMI also includes a control panel for process/pipeline dispatchers. It is also usual to include redundant communication channels between the various pumping/compressor stages of the pipeline, and also between the various important elements of the process.

Master station computers process all-important data at regular intervals so that hourly or daily reports may be printed automatically. Data to be printed on these summary reports may be inventories of process/pipeline performance, volumes of various materials received and delivered. Volumes and leakage information may also be indicated. This data may be used for direct invoicing.

6 REMOTE TRANSMITTER UNITS

Remote Transmitter Units, which are microprocessor-based units for digitally multiplexing data before transmission to the master station, are connected in a SCADA system as

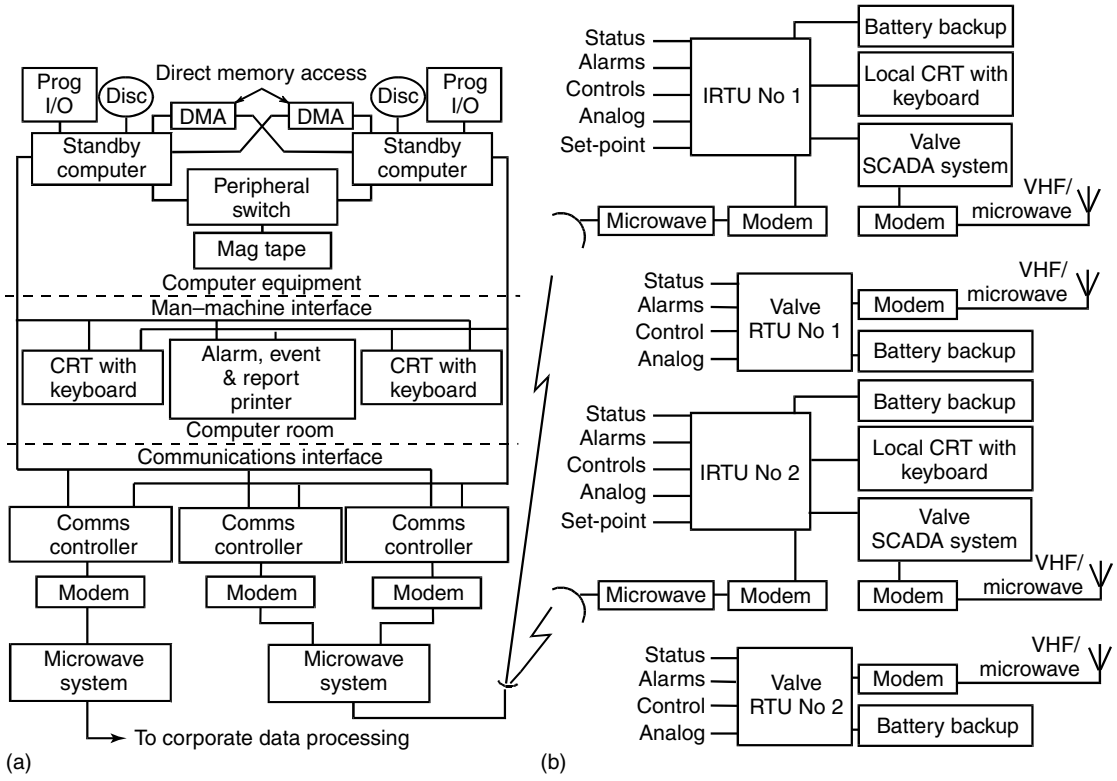


Figure 5. A Supervisory Control and Data Acquisition, or SCADA, system: (a) computer equipment, human–machine interface and communications interface, (b) field located Intelligent Remote Transmitter Units (IRTUs) in a SCADA system.

shown in Figure 5. It is also possible that these RTUs are independent, in so far as they implement local supervision, data acquisition, and control in their own right. RTUs may also have some *intelligence*, in which case they are referred to as IRTUs.

Typically, an RTU may acquire data from pumping or compressor stations on a pipeline or perhaps from a local process. Variables such as suction, pump case temperature, discharge and process temperatures and pressures are typical of the wide range, which are monitored. Flow rate is also important, especially if the discharge is an origin or destination location for custodial transfer of the product. As volume accumulation is also important, then tank levels and temperatures must be measured, so that temperature compensated tank volumes can be computed both locally and at the master station. Leaks may also be detected if flow rates are monitored at separate locations in the product flow system. Measurement of energy consumption in kilowatt-hours is also essential.

Status and alarm signals are required to ensure safe and reliable operation of the system. It is normal to indicate the open or closed status of valves, on or off status as well as high levels of vibration, shaft bearing temperature and so on, of pumps and motors, low suction pressures, power failures, relief valve open, leak detection, fire and smoke detection, high sump level, and many more. Of the many control actions, the most common are starting and stopping pumps and motors as well as opening and closing valves. Others are set-point control of many system variables such as valve throttling, turbine speed, vessel fluid level, pressure, temperature, flow rate, and discharge pressure regulation, not forgetting station reset or shutdown.

Technology convergence has meant that programmable controllers and/or IRTUs now almost exclusively perform many process operations involving logical sequence control. Continuous analog control uses direct digital control strategies, which provide fast control response while processing or in transit. Design changes are easily implemented using these programmable controllers, with downloading of program revision to the RTUs by the master station.

RELATED ARTICLES

Article 14, Hierarchical Aspects of Measurement Systems, Volume 1; **Article 16, Reduction of Influence Factors, Volume 1**; **Article 132, Data Acquisition Systems (DAS) in General, Volume 3**; **Article 158, Microcontroller Systems, Volume 3**; **Article 165, Introduction to**

Networked Instrumentation, Volume 3; **Part 5, Measuring System Design, Volume 2**.

REFERENCES

Lang, T.T. (1987) *Electronics of Measuring Systems*, John Wiley & Sons, Chichester.

McGhee, J., Henderson, I.A. and Sankowski, D. (1986) Functions and Structures in Measurement Systems: A Systems Engineering Context for Instrumentation. *Measurement*, **4**(3), 11–119.

McGhee, J., Korczynski, M.J., Kulesza, W. and Henderson, I.A. (1996) *Scientific Metrology*, ISBN 83 904299 9 3, A.C.G.M. Lodart, Lodz.

McGhee, J., Henderson, I.A., Kulesza, W. and Korczynski, J. (2001) *Measurement Data Handling: Theoretical Technique*, ISBN 83 7283 007 X, A.C.G.M. Lodart S.A., Lodz.

Sydenham, P.H. (1983) Chapter 16.3, in *Handbook of measurement science: Vol 2* (ed. P.H. Sydenham), John Wiley & Sons, Chichester.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

16: Reduction of Influence Factors

Paul P.L. Regtien
University of Twente, Enschede, The Netherlands

1 Influence Effects 87
2 Isolation of Influence Effects 88
3 Compensation for Influence Effects 89
4 Feedback Methods for Suppressing Influence Effects 90
5 Correction Methods to Reduce Influence Effects 91
References 92

1 INFLUENCE EFFECTS

Any measurement system has imperfections and any act of measurement is liable to errors. Measurement errors either originate from system deficiencies (for instance system noise, quantization, and drift), or are due to environmental influences such as thermal, electromagnetic, and mechanical interference. Manufacturers of measuring instruments try to minimize such intrinsic errors by a proper design of the instrument structure; the remaining imperfections should be given in the data sheets of the system. The user of a measurement system should minimize additional errors that could arise from improper use and faulty interfacing.

Besides influence effects from outside the system, instruments may suffer from many other imperfections. They influence the magnitude of the instrument transfer (*multiplicative error*) or give rise to unwanted output signals (*additive errors*). Both error types may occur as a deterministic or a stochastic error. Figure 1 shows a simplified model of a measurement system, with an indication of these error sources. In this figure, x_m is the measurand and y_o the output signal. Additive error signals are modeled as additional input signals: x_d and x_s represent deterministic and stochastic signals respectively. They model all kinds of interference from the environment as well as the equivalent input error signals due to system offset and noise. Additive error signals are always expressed as (equivalent) input signals, which facilitates comparison with the actual input signal.

The error inputs ε_d and ε_s symbolize deterministic and stochastic multiplicative errors; they mainly originate from environmental influences: temperature changes, electric, magnetic, or electromagnetic fields, mechanical vibrations, shocks, and nuclear radiation. For this simplified model, the output signal of a sensor can be written according to:

$$y_o = S(1 + \varepsilon_d + \varepsilon_s)(x_m + x_d + x_s) \tag{1}$$

where S is the nominal sensitivity (transfer) of the measurement system. This model is useful for evaluating system errors, influence factors, and the various error-reduction methods. When all influence factors are eliminated, the output of the system satisfies:

$$y_o = S \times x_m \tag{2}$$

This situation is hypothetical and will never occur in practice. Therefore, it is important to know what is the deviation from this ideal situation and to quantify it. It is equally important to know the type of error: deterministic or stochastic; additive or multiplicative, in order to effectively suppress them when necessary. Error suppression can be obtained by a sophisticated design, or by just some additional signal processing. Some of these methods will reduce mainly additive sensors, others minimize multiplicative errors. We will discuss some basic error reduction methods in the next sections.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

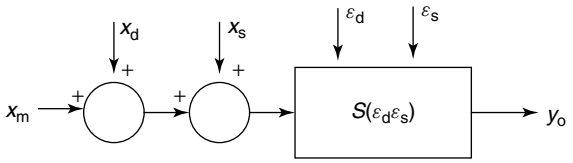


Figure 1. Simplified sensor model including error signals.

2 ISOLATION OF INFLUENCE EFFECTS

Error signals can be reduced or eliminated by isolating the measurement instrument from the interference source. The error signals may be kept outside the measurement system by several isolation techniques, depending on the type of signal. The isolation techniques are also referred to as *shielding*, *guarding*, or *(pre)filtering*, according to the implementation. We briefly review some causes and the associated isolation methods.

2.1 Capacitively induced errors

Capacitively induced errors arise by electric injection of spurious signals, through a capacitive connection between the error voltage source (for instance, wires of the mains) and the measurement system, notably the electrical parts at its input (Figure 2a).

Capacitive coupling (in the figure, the distributed capacitances are modeled with only lumped capacitors) results in an equivalent error signal:

$$V_{n,o} = \frac{C_s}{C_g} V_{n,i} \tag{3}$$

with the assumption $C_s < C_g$. Its effect can be reduced by a *grounded* shield around the sensor structure and the input leads (Figure 2b). The equivalent error signal is reduced to

$$V_{n,o} = \frac{C_{rest}}{C_{sh}} V_{n,i} \tag{4}$$

This approaches zero when the rest capacitance can be made very small (so an effective enclosure around the sensitive parts of the system is provided).

2.2 Magnetically induced errors

Magnetically induced signals arise by a time varying magnetic flux through the loop made up by the input circuit.

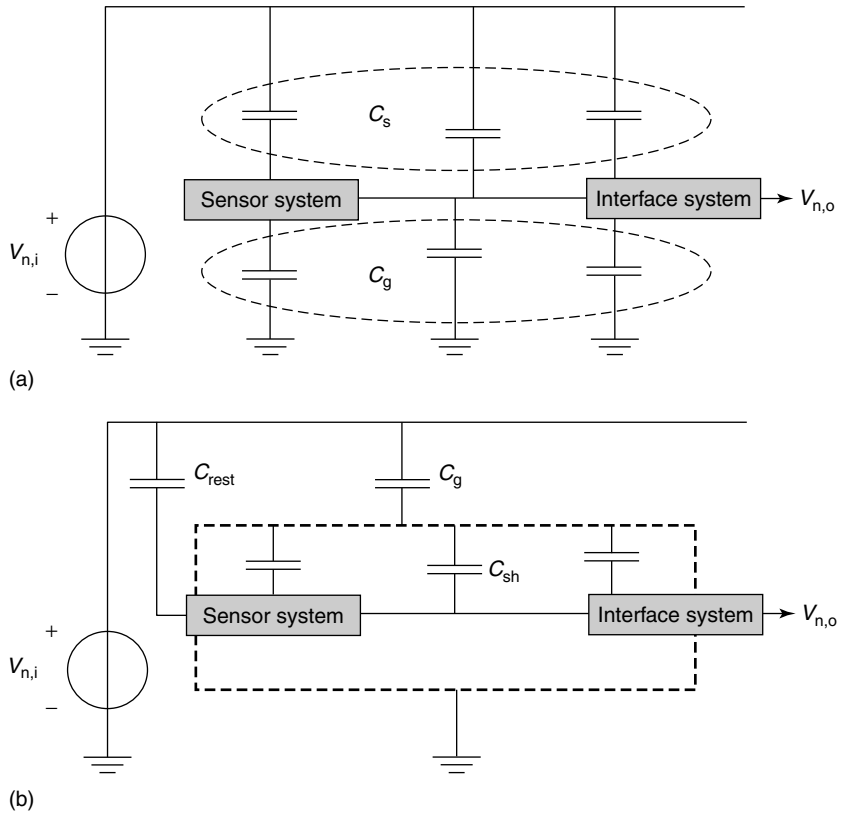


Figure 2. Capacitively induced errors: before and after shielding.

The induced error signal amounts to

$$V_{\mathrm{n,o}} = \frac{\mathrm{d}\Phi}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\oint_A B \times \mathrm{d}A\right) = A\frac{\mathrm{d}B}{\mathrm{d}t} \tag{5}$$

and is proportional to the surface area of the loop, the strength of the magnetic field, and the frequency of the field (for instance 50 or 60 Hz from the mains). The interference can be reduced by a shield made of a material with a high magnetic permeability (to reduce the induction B), and by minimizing the area of the loop A (for instance, by placing the input leads close to each other or by twisting).

2.3 Temperature induced errors

Changes in environmental temperature may introduce both additive errors (for instance, offset drift) and multiplicative errors (for instance, the temperature coefficient of the transfer). This interference can be reduced by thermal shielding (an encapsulation with high thermal resistance or a temperature controlled housing).

2.4 Optically induced errors

Unwanted optical input can be stopped by optical filters; many filter types are available on the market. Measurand and interferences should have a substantial difference in wavelength to make the filtering sufficiently effective. Optical interference originating from environmental light (daylight, lamplight) has a low frequency (DC, 50/60 Hz and harmonics). If the frequency of the optical measurand is in the same band, filtering is not a solution. In that case, modulation may help reduce interference. If the measurement system contains an auxiliary light source (as with many optical systems), that signal source can be modulated (using an LED or laser diode). If the source itself cannot be modulated, optical modulation can be performed by, for instance, a chopping wheel, as is applied in many pyroelectric measurement systems.

2.5 Mechanically induced errors

Mechanical disturbances (shocks, vibrations) are reduced by a proper mechanical construction with elastic mounting, performing suitable damping of the vibrations. The method requires knowledge about the most important resonance frequencies of the structure.

3 COMPENSATION FOR INFLUENCE EFFECTS

Compensation is a simple and effective method of minimizing additive errors due to interference signals. The basic idea is as illustrated in Figure 3, where the technique is applied on the sensor level.

In Figure 3(a), the output of the sensor is y_1, which contains unwanted signal components, for instance, due to interference x_i or offset. From this output, a compensation voltage y_c is subtracted. The condition for full compensation is $y_\mathrm{c} = S \times x_\mathrm{i}$, making $y_\mathrm{o} = S \times x_\mathrm{m}$, independent of x_i. For correct compensation, the interference signal as well as the sensor transfer should be known. One way of accomplishing compensation is by an adjustable compensation signal: at zero input, y_c is (manually) adjusted to a value at which the output is zero. A more elegant way of compensation is by applying a second sensor, as illustrated in Figure 3(b). The measurement signal, x_m, is supplied simultaneously to both sensors, which have equal but opposite sensitivity to the measurement input (for example, two strain gauges, one loaded on compressive stress, the other on tensile stress). A minus sign represents the opposite sensitivity of the sensors. The two output signals are subtracted by a proper electronic circuit. Because of the antisymmetric structure with respect to the measurement signal only, many interference signals appear as *common* output signals, and are thus eliminated by taking the difference of the two outputs.

The effectiveness of the method depends on the degree of symmetry of the double-sensor or differential sensor structure. From Figure 3 it follows:

$$y_1 = S_{m1}x_\mathrm{m} + S_{i1}x_\mathrm{i} \tag{6}$$

$$y_2 = -S_{m2}x_\mathrm{m} + S_{i2}x_\mathrm{i} \tag{7}$$

where S_{mk} and S_{ik} are the sensitivities for the measurement signal and the interference signal. So, the output of the total

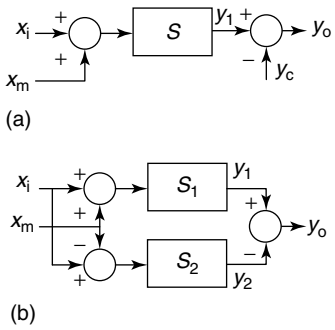


Figure 3. General layout of compensation; (a) compensation signal and (b) balanced configuration.

sensor system equals

$$y_o = (S_{m1} + S_{m2})x_m + (S_{i1} - S_{i2})x_i \tag{8}$$

The sensitivities S_{m1} and S_{m2} are about equal, so the output signal becomes

$$y_o = 2S_m x_m + \Delta S_i x_i = 2S_m \left(x_m + \frac{\Delta S_i}{2S_m} x_i \right) \tag{9}$$

Analogous to the definition of the rejection ratio for differential amplifiers, we can define a quality measure for the imbalance of the differential sensor:

$$H = \frac{2S_m}{\Delta S_i} \tag{10}$$

a parameter characterizing the system's ability to distinguish between measurand and interfering signals.

The method is illustrated with a two-active element Wheatstone bridge in Figure 4.

In this half-bridge mode, R_3 and R_4 are fixed resistances; R_1 and R_2 are resistive sensors: their resistance values change with a particular physical quantity as well as an interfering signal, according to

$$R_1 = R(1 + S_{m1}x_{m1} + S_{i1}x_{i1}) \tag{11}$$

$$R_2 = R(1 + S_{m2}x_{m2} + S_{i2}x_{i2}) \tag{12}$$

Here, S_{mk} is the sensitivity of sensor k ($k = 1, 2$) to the measurand (for instance, deformation), and S_{ik} is the sensitivity to the interference signal (for instance, temperature). Note that in this system, the sensor resistance is just R at a particular reference value of the measurand, and at zero interference. Assuming both sensor parts experience the same measurement signal, and have equal but opposite sensitivities to the measurand, the bridge output voltage approximately satisfies, the equation:

$$\frac{V_o}{V_i} = \frac{1}{2} S_m x_m + \frac{1}{4} \Delta(S_i x_i) \tag{13}$$

If both sensor parts have equal sensitivity to interference (by a symmetric sensor design), and both sensor parts experience the interference equally, the error term is zero, and the interference is completely eliminated. The equation is useful for making a quick assessment of the error due to asymmetry, relative to the measurement signal.

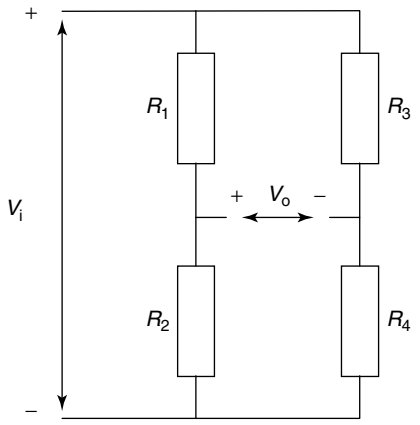


Figure 4. Measurement bridge with two active elements.

4 FEEDBACK METHODS FOR SUPPRESSING INFLUENCE EFFECTS

Feedback is an error reduction method originating from the early amplifiers with vacuum tubes. Their instable operation was a real problem until the systematic application of feedback, which reduces in particular multiplicative errors. Figure 5 shows the general idea.

The sensor has a nominal transfer S, but due to multiplicative interference, it has changed to $S(1 + \varepsilon_i)$. The feedback is accomplished by an actuator with an inverse transduction effect, and a transfer k. From classical control theory, the error reduction factor can easily be found. The transfer of the total system, S_f, is given by

$$S_f = \frac{S}{1 + S \times k} \tag{14}$$

A relative change dS in the sensor transfer S causes a relative change dS_f in S_f according to

$$\frac{dS_f}{S_f} = \frac{1}{1 + kS} \times \frac{dS}{S} \tag{15}$$

So, the relative error in the feed-forward part is reduced by a factor equal to the loop gain $1 + kS$ of the system. The penalty for this improvement is a reduction of the overall sensitivity with the same factor.

Feedback also reduces additive interference signals to a degree that depends on the point of injection in the system.

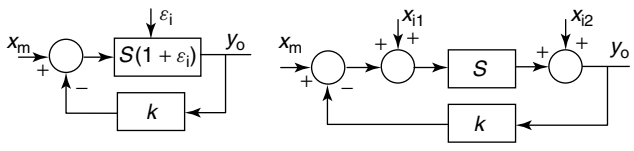


Figure 5. Interference signals in a feedback system.

Two cases are discussed (Figure 5). The output due to two interfering signals x_{i1} and x_{i2} equals:

$$y_o = \frac{S}{1+kS} \times x_{i1} + \frac{1}{1+kS} \times x_{i2} \tag{16}$$

Obviously, signals entering at the input of the system are reduced by feedback as much as the measurement signal (so the SNR is not better); interfering signals injected at the output of the sensing system are reduced by a factor S more than the measurement signal.

Feedback reduces errors in the forward signal path: the transfer is mainly determined by the feedback path. Prerequisites for an effective error reduction are

- high forward path transfer,
- stable feedback path transfer.

The application of this method to sensors requires a feedback element with a transfer that is the inverse of the sensor transfer. The imperfections of the sensor are reduced; however, the demands on the actuator are high. The method is illustrated with an example of a capacitive accelerometer system in which two error reduction methods are combined: compensation by a balanced sensor construction and feedback by an inverse transducer (Figure 6).

A displacement of the seismic mass m results in a capacitance difference ΔC; this value is converted to a voltage that is compared with a reference value (here this value is zero). The amplified voltage difference is supplied to an electromagnetic actuator that drives the mass back toward its neutral position. When properly designed, the system reaches a state of equilibrium where the applied inertial force is compensated by the electromagnetic force from the actuator. The current required to keep equilibrium is a measure of the applied force or acceleration.

A more detailed model of this system, for instance, for stability analysis, is depicted in Figure 7. All transduction steps are visualized in separate articles.

Obviously, the feedback is performed in the mechanical domain, by counteracting the inertial force F_i with the electromagnetic force F_a of the actuator. The transfer function of the whole system can easily be derived from this model,

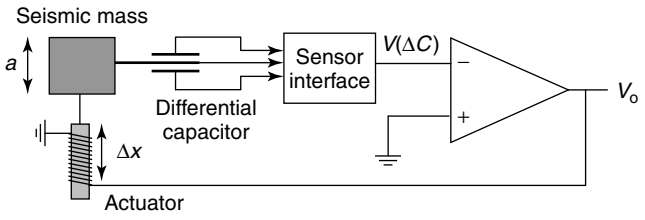


Figure 6. Illustration of a feedback system to reduce sensor errors: capacitive accelerometer.

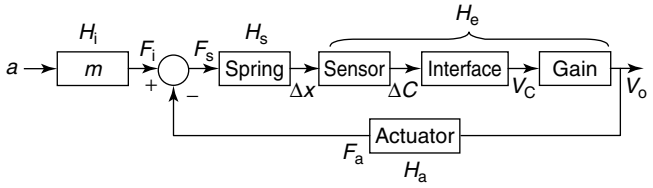


Figure 7. Model of the differential capacitive accelerometer with feedback.

and is given in the next equation:

$$V_o = \frac{H_s H_e}{1 + H_a H_s H_e} H_i a \tag{17}$$

In this equation, H_i is the transfer from applied acceleration a to inertial force F_i. H_s is the transfer of the mechanical spring: from force to displacement. The transfer of the electrical system: capacitive sensor, electronic interface, and amplifier together) is H_e. Finally, H_a is the transfer of the actuator. For $H_a H_s H_e \gg 1$ the transfer function of the total sensor system becomes:

$$V_o = \frac{H_i}{H_a} a \tag{18}$$

Indeed, this is independent of the spring stiffness, the sensor transfer, the interface, and the gain of the amplifier, according to the feedback principle. In equilibrium, the mass is at its center position. Hence, no particular demands have to be made on the spring, the sensor, and the interface circuit; the only requirement is a low zero error. Actually, the requirements with respect to the sensor quality have been transferred to those of the actuator. The system transfer only depends on the seismic mass m and the actuator transfer H_a.

There are various sensors on the market that are based on the feedback principle, which is discussed in more detail in **Article 113, Force-feedback Sensors, Volume 2**.

5 CORRECTION METHODS TO REDUCE INFLUENCE EFFECTS

An erroneous sensor signal can be corrected if knowledge about the causes of the errors or the value of the errors is available. Two different strategies can be distinguished:

- static correction
- dynamic correction.

In the first class of strategies, correction is performed while leaving the sensor as it is. Figure 8 shows the general configuration for two approaches. In the model-based

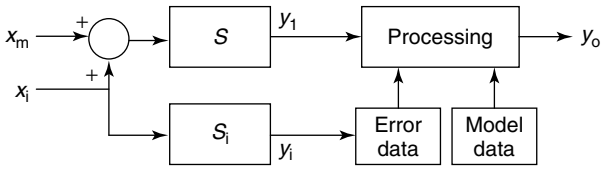


Figure 8. Model-based correction and measurement-based correction.

approach, the sensor signal is corrected on the basis of prior knowledge about the origin of the error, for instance, nonlinearity or a calibration curve, stored in a look-up table. If the errors are unknown (interference), the error signal could be measured separately by additional sensors. The output of these sensors is used to correct the original sensor signal. The method is straightforward, but requires additional sensors, at least one for each type of interference.

Dynamic correction involves a particular sensor design. Basically, the method involves a series of measurements of the same quantity in such a way as to eliminate the errors by postprocessing. This can be done in various ways, according to the type of quantity and error:

- multiple input signal measurements, eliminating scale, and offset errors;
- cyclic interchanging of components (dynamic matching);
- cyclic changing in the sensitivity direction of the sensor (flipping).

The first method is illustrated in Figure 9. Alternately, the input of the system is connected to the measurand (yielding an output y_{o1}), to 'ground' (giving output y_{o2}) and to a reference (resulting in the output y_{o3}). Multiplicative errors due to the interfering signal x_s are represented by the relative error ε_s.

The three output signals are:

$$y_{o1} = S(1 + \varepsilon_s)(x_m + x_{off}) \tag{19}$$

$$y_{o2} = S(1 + \varepsilon_s)(x_{off}) \tag{20}$$

$$y_{o3} = S(1 + \varepsilon_s)(x_{ref} + x_{off}) \tag{21}$$

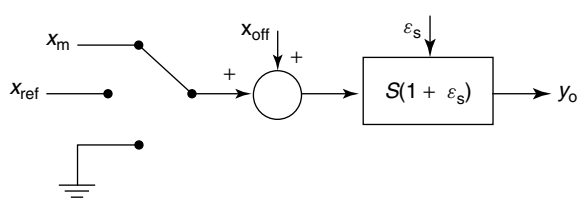


Figure 9. Three-point measurement to eliminate scale and offset errors.

From these equations, the measurand can be calculated:

$$x_m = \frac{y_{o1} - y_{o2}}{y_{o3} - y_{o2}} x_{ref} \tag{22}$$

Offset and scale errors are completely eliminated if the errors do not change during the sequence of the three measurements. The method requires a reference of the measurand type, as well as the possibility of completely isolating the input from the measurand to find the offset error contribution. For most electrical quantities, like voltage, capacitance, and resistance, this is quite an easy task. This is not the case for many other physical quantities: repetitively short circuiting the input of a magnetic field sensor or a force sensor, for instance, without removing the measurand itself requires complicated and therefore impractical shielding techniques.

Modulation, as discussed before, is a very effective way of eliminating offset and the effects of influence factors. The basic idea is to shift the (low frequency) information in the input signal to another, higher frequency band. After high-pass filtering, the offset and drift are effectively eliminated. However, modulation of the input signal is not always possible. An alternative way of reducing offset is flipping the directivity of the sensor, without changing the sign and value of the offset (Van Putten, Van Putten and Van Putten, 1994). Clearly, the method can only be applied to sensors that are sensitive to the direction of the input quantity (which should be a vector). Another technique, applicable to modulating sensors, is spinning. An example where this technique is applied is the spinning-current Hall sensor (Bellekom, 1998). Here, the auxiliary current is electronically rotated stepwise with respect to the geometry, yielding a number of outputs; each output contains an offset signal, originating from layout asymmetries of the chip. However, going around the average of the offset is zero, and by averaging the outputs over a full cycle, offset is almost completely eliminated.

REFERENCES

Bellekom, A.A. (1998) *Origin of Offset in Conventional and Spinning-Current Hall Plates*, PhD thesis, Delft University of Technology, Delft.

Van Putten, M.J.A.M., Van Putten, M.H.P.M. and Van Putten, A.F.P. (1994) Full Additive Drift Elimination in Vector Sensors Using the Alternating Direction method (ADM). *Sensors and Actuators A*, **44**, 13–17.

17: EMC and EMI

Kim R. Fowler

Technical Consultant, Cool Stream, Windsor Mill, MD, USA

1 Definitions and Examples of EMC and EMI 93
2 Basic Principles of Energy Coupling 93
3 Mechanisms of Noise Energy Coupling 94
4 Electrostatic Discharge (ESD) 95
5 General Guidelines for EMC and EMI 96
6 Standards for EMC 96
References 97

1 DEFINITIONS AND EXAMPLES OF EMC AND EMI

Electromagnetic interference (EMI) occurs when one electrical circuit interferes with another. EMI may exist between components within an electronic circuit, between electronic, and electrical subsystems, or between separate instruments. Electromagnetic compatibility (EMC) is the art and science of reducing EMI to acceptable levels.

General and specific accounts that review EMI and EMC, provide practical solutions to EMI and EMC, and provide key references are Fowler (1996), Johnson and Graham (1993), and Ott (1988).

Typically, the electromagnetic interference is called *noise*, which is any undesired electrical activity coupled from one circuit into another. EMI has three necessary components: a noise source, a susceptible circuit, and an energy coupling mechanism. EMC is concerned with reducing or eliminating the interference or noise to acceptable levels.

Examples of noise sources include

- power lines or mains
- motors
- high-voltage equipment (e.g. spark plugs, igniters)
- discharges and sparks (e.g. lightning, static electricity)
- high-current equipment (e.g. arc welders).

Examples of susceptibility within receiver circuits include

- crosstalk on inputs in digital logic causing unintentional changes in logic states
- crackle in the reception of radio
- static discharge that destroys components, such as keyboard circuits.

2 BASIC PRINCIPLES OF ENERGY COUPLING

EMI has three necessary components: a noise source, a susceptible circuit, and an energy coupling mechanism. Typically, neither the noise source nor the susceptible circuit may be eliminated; this leaves you to reduce or remove the energy coupling mechanism between the two.

2.1 Reciprocity

Anything that reduces emissions usually makes the circuit less susceptible to disruption. This means that the same techniques to reduce noise or remove a noise source will probably work to reduce susceptibility within receiver circuits. Reciprocity is a useful guideline but not an absolute rule.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.

© 2005 John Wiley & Sons, Ltd.

ISBN: 0-470-02143-8.

2.2 Impedance versus resistance

One component of understanding energy coupling and interference in EMI and EMC is to understand how electrical current flows. Current always follows the path of lowest impedance Z (*see* (1)), not necessarily the path of lowest resistance R. Remember that impedance includes reactive terms for inductance L and capacitance C. Consequently, electrical charge follows the path that represents the minimum of resistance and inductive reactance and maximum capacitive reactance; generally, this is true for signal frequencies above the audio range (>3 kHz).

$$Z = \sqrt{R^2 + \left[\varpi L - \left(\frac{1}{\varpi C}\right)\right]^2} \tag{1}$$

2.3 Pseudo-impedance

A very useful diagnostic tool for signal frequencies above the audio range (>3 kHz) is Pseudo-impedance as explained by Van Doren (1991). Pseudo-impedance is the diagnostic ratio of the rates of change in voltage to current.

$$\text{Diagnostic ratio} = \frac{\left(\frac{dv}{dt}\right)}{\left(\frac{di}{dt}\right)} \tag{2}$$

3 MECHANISMS OF NOISE ENERGY COUPLING

Noise may couple between circuits through four primary mechanisms: conductive, inductive, capacitive, and electromagnetic.

3.1 Conductive coupling

Conductive coupling, see Figure 1, requires a connection between the source and the receiver that completes a continuous circuit. Often these connections are inadvertent; when they are, they are called *sneak circuits*.

A ground loop is a particular form of conductive coupling, which has two primary characteristics. It has multiple ground connections providing multiple return paths that cause significant current flow in the grounding structure. It unbalances the desired circuit. (*See* Figure 5 in **Article 177, Grounding and Shielding, Volume 3**.)

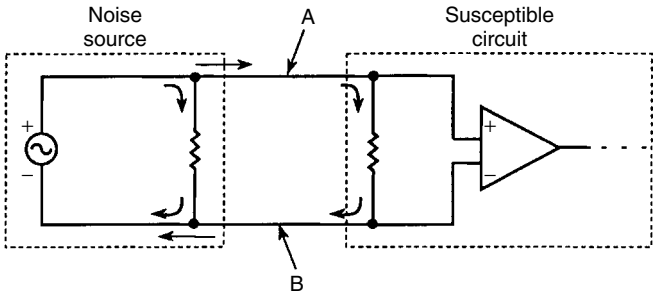


Figure 1. Conductive coupling. If either connection A or B is removed, the conductive noise is eliminated. (Reproduced from *Electronic Instrument Design: Architecting for the Life Cycle*, Fowler, K.R. Copyright (1996) © Oxford University Press.)

3.2 Inductive coupling

Inductive coupling requires a receptive loop circuit and a changing magnetic flux through the loop. A current bearing a loop circuit may be the source of the changing magnetic flux. The coupled or induced noise voltage is proportional to the time rate of change in current in the source and the loop area of the circuit(s). The Pseudo-impedance or diagnostic ratio in (3) gives a value much less than free-space impedance.

$$\text{Inductive coupling} = \frac{\left(\frac{dv}{dt}\right)}{\left(\frac{di}{dt}\right)} \ll 377\,\Omega \tag{3}$$

Straight wires create small loops that can magnetically couple. Twisted wire eliminates the effective loop area of the cables and magnetic coupling, thus eliminating inductive coupling (see Figure 2).

Inductive coupling meets the principle of current flow through the path of least impedance in designing circuit boards. The return current mirrors the topology of the signal current; that is, it flows in the return plane immediately adjacent to the signal line. This means that the return plane of a circuit board should not have any slots in it that cut across the signal path. A slot, Figure 3, increases the loop area, hence, it increases the inductive reactance and the generation of a magnetic field.

3.3 Capacitive coupling

Capacitive coupling requires both proximity between circuits and a changing voltage. The coupled or induced noise voltage is proportional to the time rate of change in the voltage in the source and inversely related to separation distance

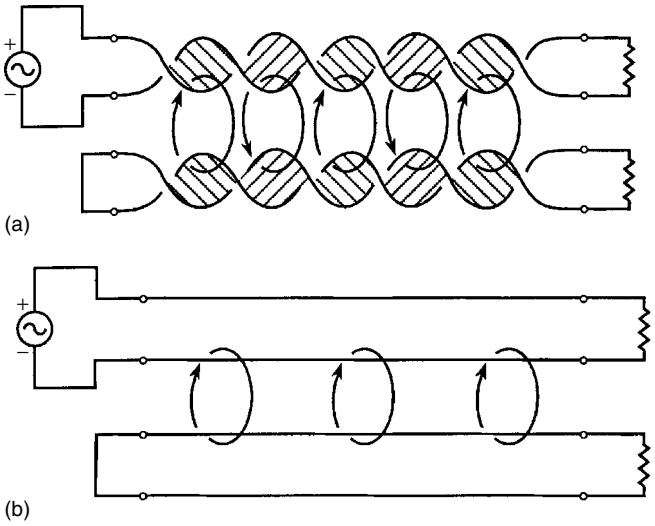


Figure 2. Effect of twisted wire. (a) Straight wires create small loops that can couple magnetically and (b) twisted wire eliminates the effective loop area of cables and magnetic coupling. (Reproduced from *Electronic Instrument Design: Architecting for the Life Cycle*, Fowler, K.R. Copyright (1996) © Oxford University Press.)

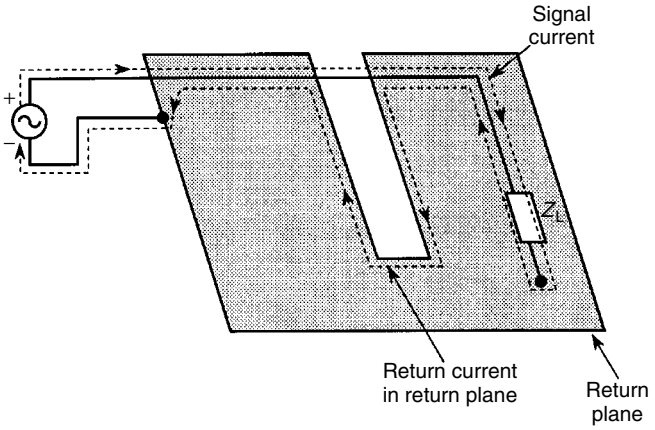


Figure 3. A slot in the return plane of a circuit board increases the current loop area and the self-inductance. (Reproduced from *Electronic Instrument Design: Architecting for the Life Cycle*, Fowler, K.R. Copyright (1996) © Oxford University Press.)

between the circuits. The Pseudo-impedance or diagnostic ratio in (4) gives a value much larger than free-space impedance.

$$\text{Capacitive coupling} = \frac{\left(\frac{dv}{dt}\right)}{\left(\frac{di}{dt}\right)} \gg 377\,\Omega \qquad (4)$$

Generally, a properly connected shield will reduce capacitive coupling (see Figure 4).

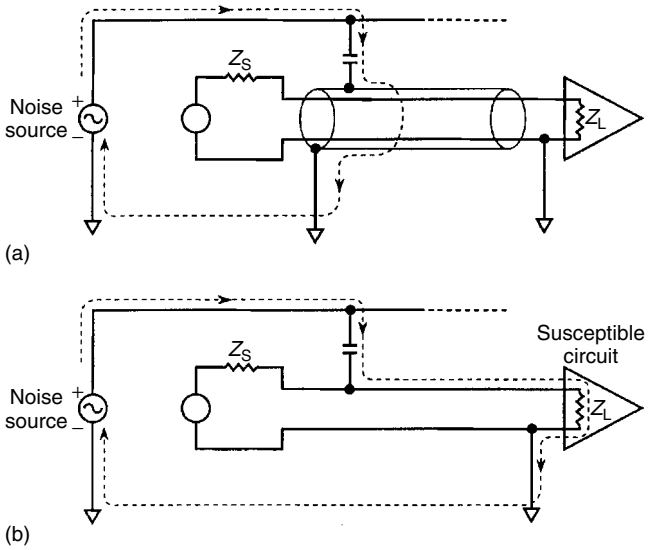


Figure 4. Effect of shielding. (a) Without the shield, stray currents can disrupt susceptible circuits and (b) a properly connected shield can divert capacitively coupled current from susceptible circuits. (Reproduced from *Electronic Instrument Design: Architecting for the Life Cycle*, Fowler, K.R. Copyright (1996) © Oxford University Press.)

3.4 Electromagnetic coupling

Radiated electromagnetic energy requires an antenna in both the noise and susceptible circuits. The antenna must be an appreciable portion of the wavelength. Electromagnetic coupling usually occurs for frequencies greater than 20 MHz and for signal conductors that are longer than a small fraction of the fundamental wavelength ($>\lambda/20$). The Pseudo-impedance for electromagnetic coupling in (5) is around the value for free-space impedance or 377 Ω; practically, any value between 100 and 500 Ω is a reasonable value for electromagnetic coupling.

$$\text{Electromagnetic coupling} = \frac{\left(\frac{dv}{dt}\right)}{\left(\frac{di}{dt}\right)} \cong 377\,\Omega \qquad (5)$$

EMI through electromagnetic coupling always begins as conductive (i.e. current in wires), becomes radiative, and ends as conductive, that is, fields interact with circuitry. Cables can be primary sources and receivers for frequency of operation above 200 MHz.

4 ELECTROSTATIC DISCHARGE (ESD)

Electrostatic discharge (ESD) is a transfer of electrical charge at very high voltage, ranging from hundreds to thousands of volts (Table 1), and very low current. Generally,

Table 1. Some typical values for ESD.

Static generation	Static voltage (V)	
	10% relative humidity	65% relative humidity
Walking across carpet	35 000	1500
Walking over vinyl floor	12 000	250
Common plastic bag picked up from bench	20 000	1200
Work chair padded with polyurethane foam	18 000	1500

Source: Reproduced from *Electronic Instrument Design: Architecting for the Life Cycle*, Fowler, K.R. Copyright (1996) © Oxford University Press.

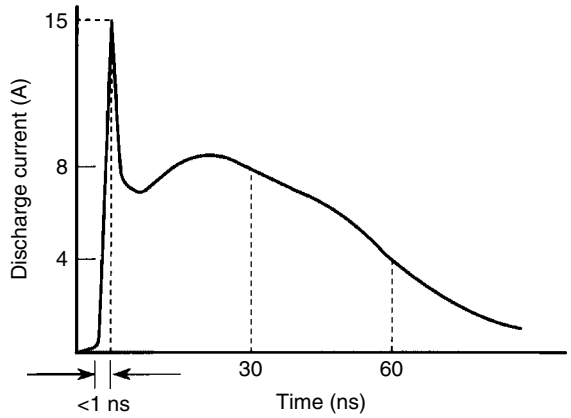


Figure 5. ESD discharge waveform at 4 KV; this is one of several different waveforms for testing circuits against ESD. (Reproduced from *Electronic Instrument Design: Architecting for the Life Cycle*, Fowler, K.R. Copyright (1996) © Oxford University Press.)

ESD has very fast rise times (<1 ns) and short duration (~90 ns), as shown in Figure 5. ESD can easily damage sensitive electronic circuits. Any instrument containing integrated circuits is susceptible to ESD if not protected.

ESD transfers charge in three stages: pickup, storage, and discharge. Usually mechanical rubbing between dry, insulated materials picks up and transfers the charge from a source to storage. Sources include plastic work surfaces, carpeted floors, wool, cotton or synthetic garments, vinyl or fiberglass chairs, common plastic bags, trays and bubble wrap, spray cleaners, and electrostatic copiers. Humans often can be the storage means for the electrical charge until its discharge occurs.

Control methods for ESD eliminate activities and materials that create high static charge and provide transient protection to sensitive circuits. Eliminating high static charge includes appropriate grounding, using protective handling and materials, and maintaining humidity. A checklist in **Article 177, Grounding and Shielding, Volume 3** lists some of the activities that will reduce problems with ESD. Figure 13 in **Article 177, Grounding and Shielding, Volume 3** illustrates one way to protective sensitive circuits by shunting ESD to chassis ground.

5 GENERAL GUIDELINES FOR EMC AND EMI

Characterize your instrument or circuit for EMI and EMC in the following ways:

- Establish: grounding options, source and load impedances, and frequency bandwidth needed.
- Determine the predominant coupling mechanisms.
- Diagram the circuit paths and reduce loops, such as ground loops and inductive loops.

Use good design techniques to reduce EMI and improve EMC:

- reduce frequency bandwidth
- balance currents
- route signals for self-shielding: use return planes, short traces, and decoupling capacitors
- add metal enclosure shielding only when necessary.

Remember reciprocity, anything that reduces emissions usually makes the circuit less susceptible to disruption, as well. If you decrease the inductive loop of a circuit to reduce emission, then you will also reduce the mutual coupling that could receive noise.

Susceptibility often traces to incorrect return paths or long signal lines that are not properly shielded. Remember that current follows the path of lowest impedance. Understand mechanisms of noise energy transmission:

- conductive coupling
- inductive coupling
- capacitive coupling
- electromagnetic coupling.

When you do encounter EMI, begin the search for the coupling mechanisms using the diagnostic clues in Table 2.

6 STANDARDS FOR EMC

Many different regulations and standards exist around the world for designing instruments to meet EMC requirements. These standards provide levels that tests must pass for instruments to avoid interference. Commercial regulations cover the frequency band from 9 kHz to 1 GHz.

Typically, these standards specify conducted and radiated environments for interference and susceptibility. Clearly, conductive coupling falls into the conducted environment.

Table 2. Diagnostic clues for noise-coupling mechanisms.

Coupling mechanism	Diagnostic clues
Conductive	1. Complete conductive path (such as multiple ground connections) 2. Frequency < 10 MHz
Inductive	1. $(dv/dt)/(di/dt) \ll 377\,\Omega$ 2. Ringing on pulse edges 3. Low source impedance and high load impedance 4. Laying conductive foil over cable decreases noise level (acts as an eddy current shield) 5. Large circuit loops
Capacitive	1. $dv/dt)/(di/dt) \gg 377\,\Omega$ 2. Rounding of pulse edges 3. High source impedance and high load impedance 4. Laying conductive foil over cable increases noise level (acts as a coupling capacitor)
Electromagnetic	1. Distance between noise source and susceptible circuit $> \lambda$ 2. Frequency > 20 MHz

Source: Reproduced from *Electronic Instrument Design: Architecting for the Life Cycle*, Fowler, K.R. Copyright (1996) © Oxford University Press.

Table 3. Diagnostic clues for noise-coupling mechanisms.

Type	Standard	Country
Commercial	CNELEC	Europe
	IEC 871, 801, 555	Europe
	FCC	United States
	VCCI	Japan
	VDE	Germany
Military	GAM-EG-13	France
	VG NORM	Germany
	DEF STAN 59-41	United Kingdom
	MIL STD-461E	United States

Source: Reproduced from *Electronic Instrument Design: Architecting for the Life Cycle*, Fowler, K.R. Copyright (1996) © Oxford University Press.

Generally, the three remaining types of coupling, inductive coupling, capacitive coupling, and electromagnetic coupling, falls under the domain of the radiated environment. Table 3 lists examples of some of these standards.

REFERENCES

Fowler, K.R. (1996) *Electronic Instrument Design: Architecting For The Life Cycle*, Oxford University Press, New York.

Johnson, H.W. and Graham, M. (1993) *High-Speed Digital Design: A Handbook of Black Magic*, PTR Prentice Hall, Englewood Cliffs, NJ.

Ott, H.W. (1988) *Noise Reduction Techniques in Electronic Systems*, 2nd edn, John Wiley & Sons, New York.

Van Doren, T. (1991) *Grounding and Shielding Electronic Systems*, Notes from video seminar on the NTU satellite network.

18: Nature and Scope of Closed-loop Systems

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 History of Closed-loop Control Systems 101
2 Illustration of an Exemplar Closed-loop Control System 102
3 Setting up the Mathematical Model for an Open-loop System 103
4 Mathematical Model of the Closed-loop System 103
5 Effects of Friction in Real Feedback Systems 104
6 Nature of Closed-loop Systems used in Measurement Systems 105
Related Articles 106
References 106

1 HISTORY OF CLOSED-LOOP CONTROL SYSTEMS

This account of linear closed-loop control principles and theory is now a mature topic, wherein the control theory has greatly advanced to allow for sophistication of multivariable, stochastic, nonlinear, and digital applications.

The design of measuring systems does at times demand knowledge of such advanced ideas but the starting point, which often suffices, is the classical theory and its foundation. This set of articles is revised and edited from material first published in brief by Atkinson in Sydenham (1982). A detailed explanation is available in Atkinson (1972).

Modern closed-loop control systems can be traced back to James Watt, who invented the first rotating steam engine (patented in 1781). Initially, the speed of these engines was controlled manually using a throttle valve on the steam inlet.

In 1788, Boulton, Watt's coprincipal, in a company manufacturing the new steam engines, had visited the Albion Mill and described to Watt a form of centrifugal governor that was being used to regulate the grinding speed of the millstones driven by wind. Watt adapted this mechanical governor to measure and automatically control the speed of his steam engines (Atkinson, 1972).

Since those days, closed-loop control and measurement have been inexorably linked; indeed to control a physical quantity accurately and rapidly in the presence of changing demand, changing internal parameters, and changing load, measurement of the control quantity coupled with feedback control is essential. Such measurement can, of course, only be accomplished by means of instruments.

Another development down this path is that feedback itself is often used extensively within measuring instruments to improve the accuracy of the measurement.

In the nineteenth century, vast numbers of regulated steam engines were in operation in factories throughout Britain. These closed-loop systems were by no means perfect in that the precision of speed control was often rather poorly achieved with large and unpredictable fluctuations being the norm.

Engineers of that time had little real insight into the dynamics of systems. They attempted to improve the performance by making smaller, lighter, and better lubricated governors. Today, we can easily see from our understanding of the theory of closed-loop systems that this was not the way to go.

Much to their consternation these expected improvements normally led to an unforeseen associated difficulty in that the steam-engine speed then tended to oscillate or

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

hunt continuously about the demanded value. This particular form of *instability*, as it is now called, has plagued designers of feedback control systems of all kinds ever since. It is the main limitation on performance. Handled from an understanding of the physics involved, it can usually be limited to allow good dynamic performance – *see* **Article 19, Dynamic Behavior of Closed-loop Systems, Volume 1** and **Article 62, Second-order System Dynamics, Volume 1**.

The practical importance of this difficulty stimulated no less a person than Maxwell (1868) to investigate the problem in great depth. He brought mathematical insight to bear by relating the existence of instability to the presence of positive real parts in the complex roots of the characteristic equation of the system. His work did not, however, greatly influence matters at that time.

The First World War (1914–1918) caused military engineers to realize that to win wars it is necessary to position heavy masses (e.g. ships and guns) precisely and quickly. Classic work performed by Minorsky (1922) in the United States was on automatic ship steering and it was Hazen (1934) who defined a servomechanism for the first time.

The concepts of automatic control, as they developed, are covered in Bennett (1979). Sydenham (1979) gives an overview of the historical development of feedback control in its use in early instruments and the chart recorders, wherein it was first recognized intuitively that velocity feedback could aid loop response times.

2 ILLUSTRATION OF AN EXEMPLAR CLOSED-LOOP CONTROL SYSTEM

The operation of a mechanical position control provides a good example for illustrating the need for feedback in control. The following principles apply to all energy domains but the electromechanical one is more intuitively obvious than in others.

Suppose there is a need to control the angular position of a heavy rotating mass, such as a gun turret or a machine tool turntable. The power resource of mechanical or electrical power assistance, rather than total reliance on muscle power, allows the obvious advantage of rapid control.

To simplify the problem, suppose an ideal frictionless electric motor is available. To achieve maximum acceleration of the rotating load, the motor is coupled to the load through an ideal frictionless step-down gearbox. Friction is always present in real systems but for the moment it is ignored – see Section 5.

It is assumed that the motor produces a torque at the load, which is directly proportional to its supply voltage. In order to control the supply voltage to the motor, an ideal power amplifier is connected, which receives at its input a control voltage, v_i, that is directly proportional to an angular positional signal, θ_i.

This signal is applied manually through a light handwheel connected to a position-to-voltage transducer (such as a rotary potentiometer).

The notional arrangement is illustrated in Figure 1. This system will produce rapid acceleration of the rotating mass in response to small and effortless motion applied manually to the handwheel. When the handwheel is at a nominal zero position, the mass will cease accelerating; a change in the handwheel position in one direction will produce acceleration in one direction, and a change in position in the opposite direction will produce acceleration in the other direction.

This, the open-loop system will, however, be found to be quite unsatisfactory as a controller. To show why, the next section discusses its mathematical model.

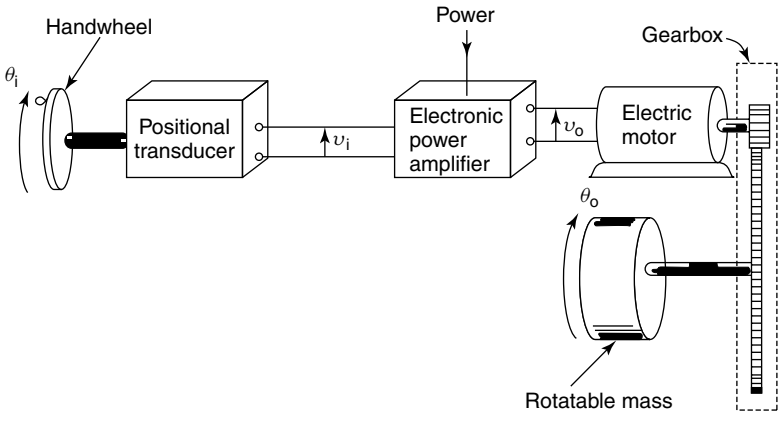


Figure 1. Position control system without feedback.

3 SETTING UP THE MATHEMATICAL MODEL FOR AN OPEN-LOOP SYSTEM

A simple mathematical analysis for the open-loop system is as follows.

Let the effective moment of inertia of the moving parts, referred to the position of the rotating mass, be J. Also, let

$$\upsilon_i = k_t \theta_i$$
$$\upsilon_a = k_a \upsilon_i \tag{1}$$

and the effective torque acting on the mass T be given by

$$T = k_m \upsilon_a \tag{2}$$

where k_t, k_a, and k_m are constants.

Thus,

$$T = k_m k_a k_t \theta_i = K \theta_i \tag{3}$$

where K is a composite system constant.

Expressed in calculus notation, according to Newton's second law of motion it is given by

$$J \frac{d^2 \theta_o}{dt^2} = T = K \theta_i \tag{4}$$

The response to a small change in θ_i from zero to a fixed positive angle is shown in Figure 2. This can be seen to be a first-order system – *see* **Article 61, First-order System Dynamics, Volume 1**.

In order to change the position of the mass from one fixed position to another in the shortest time, it is necessary to first accelerate the mass with all the torque available but then quickly retard it with a reversed torque such that at the instant the required position has been reached, both the velocity and acceleration are simultaneously reduced to zero. With this simple approach, the manual control problem is extraordinarily difficult to apply to advantage.

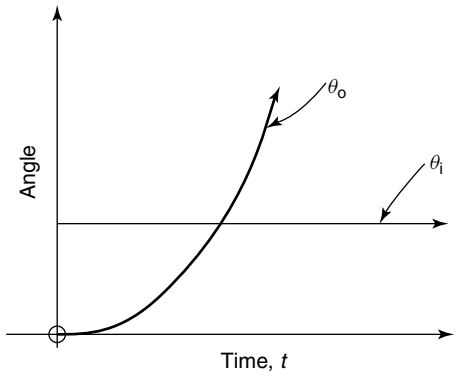


Figure 2. Response of position control system without feedback.

Worse still, in many circumstances (e.g. control of the position of a gun turret) there will be load disturbances present, for example, wind gusts which will cause the mass to deviate from the desired position in a random manner.

The result is that to get workable control of the position using open-loop control, only small torques can be applied and the result is a sluggish motion that is prone to system disturbances and sloppy following action.

4 MATHEMATICAL MODEL OF THE CLOSED-LOOP SYSTEM

This position control problem is provided with a much better solution similar to the one in which Watt solved his speed control problem, that is, by adding feedback of a measured value of the controlled variable to the power amplifier so that the system power can be applied with as much power as is needed and in the right direction at the appropriate time.

The mathematical model for such a feedback or closed-loop system is now developed using Figure 3 as the feedback system arrangement.

The controlled output variable θ_o can be measured by means of a transducer identical to that monitoring the input command θ_i, so that a signal υ_o (given by $\upsilon_o = k_t \upsilon_o$) is available for comparison with υ_i. The notion is now that if θ_i is made equal to the required position of the rotating mass, the amplifier can be fed by a difference signal $(\upsilon_i - \upsilon_o)$, which is given by

$$\upsilon_i - \upsilon_o = k_t \theta_i - k_t \theta_o = k_t \varepsilon \tag{5}$$

where ε is defined as the positional error between the required position θ_i (termed the *command* or *input*) and the actual output position θ_o.

This error signal is now amplified as before and applied to the motor. Thus, a driving torque will always be present as long as θ_o is different from θ_i. When they are of the same value there will be no driving torque and the mass will hopefully stop moving at the point where it needs to be. An analysis of this system is very revealing.

The effective torque, T, acting on the mass is no longer $K\theta_i$, but is now equal to $K\varepsilon$. Thus, again applying Newton's second law, gives

$$J \frac{d^2 \theta_o}{dt^2} = K \varepsilon \tag{6}$$

But $\varepsilon = \theta_i - \theta_o$, hence

$$J \frac{d^2 \theta_o}{dt^2} = K (\theta_i - \theta_o) \tag{7}$$

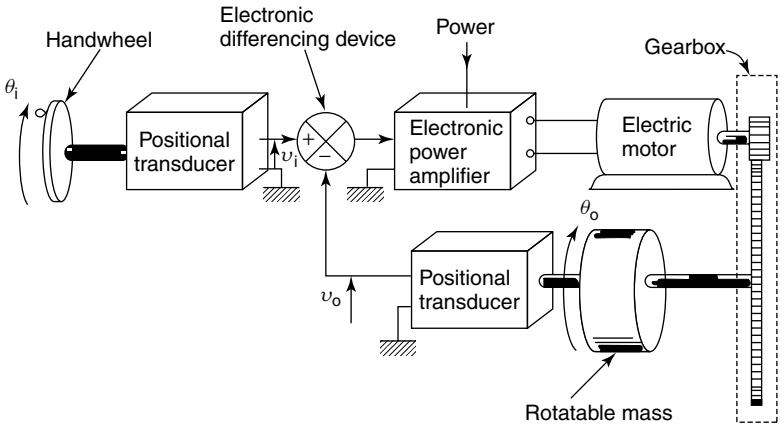


Figure 3. Position control system with feedback.

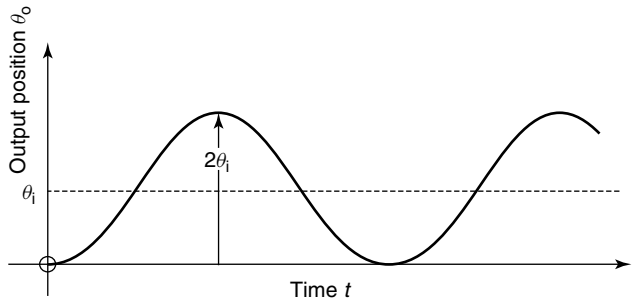


Figure 4. Response of position control system with feedback.

and so

$$J\frac{d^2\theta_o}{dt^2} + K\theta_o = K\theta_i \tag{8}$$

This will be recognized as the equation for simple harmonic motion – *see* **Article 62, Second-order System Dynamics, Volume 1**.

Hence, a sudden displacement in the handwheel position will lead to continuous oscillations as illustrated in Figure 4. Mathematicians define this system situation as *critically stable* because its oscillations will neither increase in magnitude nor decay; control engineers tend to regard such a system as unstable for this is not the desired state of control due to the continuous oscillation. This system is exhibiting exactly the same type of behavior as Watt's regulated steam engines did and it is entirely unsatisfactory for practical position control systems when compared with what can be achieved.

5 EFFECTS OF FRICTION IN REAL FEEDBACK SYSTEMS

To show how a more stable feedback system can be made, let us now look at the apparently nonideal practical system that has some friction present in the bearings, gearing contacts, and seals.

Friction almost always acts against motion regardless of direction and this will cause the oscillation to decay eventually. There are, however, various forms of friction including:

- Stiction (the torque necessary to just cause motion to begin)
- Coulomb friction (a constant torque independent of velocity) and
- Viscous friction (a torque which is directly proportional to velocity).

Stiction and Coulomb friction both cause undesirable side effects (stiction producing stick–slip motion when the system is commanded to follow a constant velocity input and Coulomb friction producing a constant offset in response to a constant input). It is thus essential to minimize stiction and Coulomb friction by good mechanical design before any form of electronic compensation is applied.

As will now be shown, it is useful to ensure that the viscous component dominates or that a similar kind of dominating effect is reproduced by other means.

The effect of viscous friction on the form of differential equation of motion for the system is to add an extra term proportional to output angular velocity, thus

$$J\frac{d^2\theta_o}{dt^2} + F\frac{d\theta_o}{dt} + K\theta_o = K\theta_i \tag{9}$$

where F is the viscous frictional torque per unit of angular velocity.

Practically, the required viscous damping can be achieved by either attaching a physical viscous damper to the rotating mass or more readily done by feeding back an extra signal that is derived from another transducer – this time

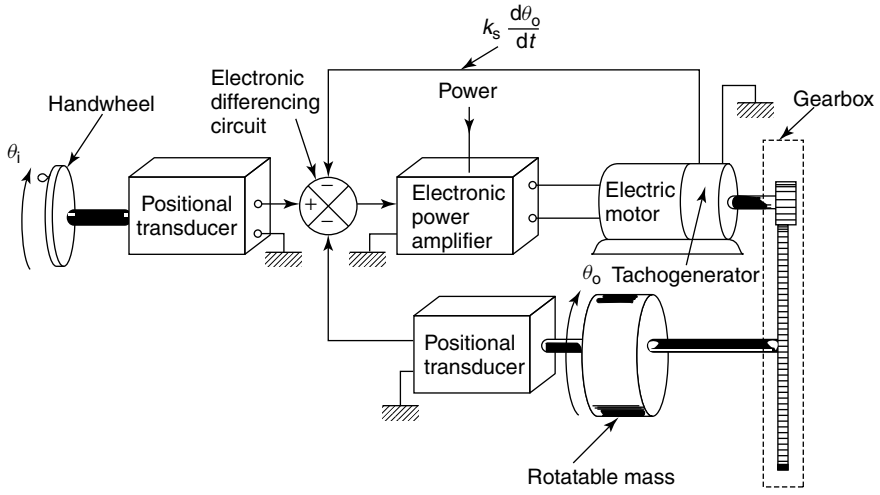


Figure 5. Position control system stabilized with velocity feedback.

a tachogenerator, which is directly proportional to angular velocity.

The second of these arrangements is shown in Figure 5. Electric motor-tachogenerators built as a combined unit are commercially available for this purpose, or a precision tachogenerator can be added to the gear train, or other output shaft.

The electronic differencing circuit will now produce a signal

$$k_t\varepsilon - k_s\frac{d\theta_o}{dt} \tag{10}$$

where k_s is the tachogenerator constant.

Hence, the drive torque T produced by the motor is given by

$$T = k_m k_a\left(k_t\varepsilon - k_s\frac{d\theta_o}{dt}\right) \tag{11}$$

Again assuming that all forms of unwanted friction are negligible, and using Newton's second law

$$\begin{aligned} J\frac{d^2\theta_o}{dt^2} &= k_m k_a\left(k_t\varepsilon - k_s\frac{d\theta_o}{dt}\right) \\ &= K\varepsilon - k_m k_a k_s\frac{d\theta_o}{dt} \end{aligned} \tag{12}$$

Hence,

$$J\frac{d^2\theta_o}{dt^2} + k_m k_a k_s\frac{d\theta_o}{dt} + K\theta_o = K\theta_i \tag{13}$$

The term $k_m k_a k_s$ can be regarded as the equivalent viscous frictional constant, F.

This can be recognized as a second-order harmonic system, but this time with damping – *see* **Article 62, Second-order System Dynamics, Volume 1**.

6 NATURE OF CLOSED-LOOP SYSTEMS USED IN MEASUREMENT SYSTEMS

The analysis of the response of this system to various inputs – *see* **Article 60, Zero-order System Dynamics, Volume 1**; **Article 61, First-order System Dynamics, Volume 1**; and **Article 62, Second-order System Dynamics, Volume 1** – is important for two dominant reasons; firstly, the arrangement forms the basis of most chart recording instruments, printers, plotters, cruise control in automobiles, and numerous other industrial applications involving measuring systems used in practice; and secondly, it represents the embodiment of the second-order system which is used as an important reference in the design of higher-order systems.

Feedback systems are more likely to be associated with the larger power systems but they also find extensive application in fine mechanisms. For example, the miniature laser reading head in a CD-ROM head has three positional servos acting to follow the disk. It is also used in feedback type sensors – *see* **Article 113, Force-feedback Sensors, Volume 2** and Jones (1979).

While the implementation of servos is usually done with digital signals today, the underlying thinking is as given above. The difference in digital systems is that the analog linear theory given above has to be transposed into the discrete mathematical domain due to the sampling carried out by the digitizer.

When designing instrument feedback arrangements, it is wise to make them linear systems if at all possible, for then

their behavior can be well modeled and predicted. If not, then modeling is considerably harder, but still feasible.

Texts on classical linear control theory matured by the 1960's; today, the topic does not attract much attention by publishers in new titles. However, most titles on process control engineering will cover the topic. A time-honored source is Levine (1996). Some titles appearing under control theory are best left for those who need and can handle the complex mathematics other than simple linear systems.

Instrument designers will usually make use of modeling tools to set up a system – *see* **Article 105, Modeling with LabVIEW™, Volume 2**.

RELATED ARTICLES

Article 37, Relationship Between Systems in the Time and Frequency Domain, Volume 1; **Article 38, Stability Issues, Volume 1**; **Article 59, Introduction to the Dynamic Regime of Measurement Systems, Volume 1**; **Article 60, Zero-order System Dynamics, Volume 1**; **Article 61, First-order System Dynamics, Volume 1**; **Article 62, Second-order System Dynamics, Volume 1**.

REFERENCES

Atkinson, P. (1972) *Feedback Control Theory for Engineers*, 1st edn Heinemann, London.

Bennett, S. (1979) *A History of Control Engineering: 1800–1930*, Peter Peregrinus, London.

Hazen, H.L. (1934) Theory of Servo-mechanisms. *Journal of the Franklin Institute*, **218**(3), 279–331.

Jones, B.E. (1979) Feedback Instruments. *Journal of Physics E-Scientific Instruments*, **12**, 145–158.

Levine, W.S. (1996) *The Control Handbook*, CRC Press, Boca Raton, FL.

Maxwell, J.C. (1868) On Governors. *Proceedings of the Royal Society of London*, **16**, 270–283.

Minorsky, N. (1922) Directional Stability of Automatically Steered Bodies. *Journal of the American Society of Naval Engineers*, **42**, 280–309.

Sydenham, P.H. (1979) *Measuring Instruments: Tools of Knowledge and Control*, Peter Peregrinus, Stevenage.

Sydenham, P.H. (1982) *Handbook of Measurement Science*, Wiley, Chichester.

This article is based in part on a chapter by P. Atkinson originally published in Handbook of Measurement Science Vol 1, P.H. Sydenham, Wiley, 1982.

19: Dynamic Behavior of Closed-loop Systems

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Use of the Laplace Transform in Control Systems Investigations 107
2 Analysis of Simple Position Controller with Damping and Velocity Feedback 108
3 Concept of Damping Ratio and Undamped Natural Angular Frequency 109
4 Frequency Response of Closed Loop Systems 110
5 Second-order Correlations 111
Related Articles 111
References 111

1 USE OF THE LAPLACE TRANSFORM IN CONTROL SYSTEMS INVESTIGATIONS

This article continues the discussion on instrument feedback systems and is based on original material in Atkinson (1982). As shown in **Article 18, Nature and Scope of Closed-loop Systems, Volume 1**, the differential equation of a simple linear feedback control system can be derived by the application of physical laws (Newton's second law of motion in the example given).

However, in order to determine the behavior of the system in response to certain inputs, that is, as it responds to a change in input and has not settled down to the steady state, we need to have available a method of analysis for such systems when they are disturbed with a range of inputs; the method of Laplace transforms now used is also covered in **Article 36, Systems in the Frequency Domain, Volume 1** and **Article 143, DFT and FFTs, Volume 3**.

The Laplace transform of a signal $\theta_i(t)$ is formally defined as $\Theta_i(s)$, in which s (p is also often used) is the complex variable $\sigma + j\omega$ and

$$\Theta_i(s) = \int_{0-}^{\infty} \theta_i(t) e^{-st}\, dt \tag{1}$$

and

$$\theta_i(t) = \frac{1}{2\pi j} \int_{c-j\omega}^{c+j\omega} \Theta_i(s) e^{st}\, ds \tag{2}$$

Here, c is chosen to be larger than the real parts of all the singularities of $\Theta_i(s)$.

Fortunately, there is rarely any need to evaluate these integrals in practice because they have been tabulated in transform pairs to aid the rapid solution of differential equations – see Table 1 for the commonly seen transforms. Modeling tools will usually have them embedded in their data libraries.

In the presence of zero initial conditions, we may transform derivatives by the rule

$$\mathcal{L}\frac{d^n f(t)}{dt^n} = s^n F(s) \tag{3}$$

and integrals by the rule

$$\mathcal{L}\int_0^t f(t)dt = \frac{F(s)}{s} \tag{4}$$

in which $\mathcal{L}$ represents the operation of taking Laplace transforms and $F(s)$ is the Laplace transform of $f(t)$.

In situations in which the initial conditions are nonzero, then

$$\mathcal{L}\frac{d^n}{dt^n}[f(t)] = s^n F(s) - s^{n-1} f(0-) - s^{n-2} f^1(0-) \cdots - f^{n-1}(0-) \tag{5}$$

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

Table 1. Some functions and their Laplace transforms.

$f(t)$	$F(s) = \mathcal{L}[f(t)]$
Unit impulse function $\delta(t)$	1
Unit step function $u(t)$	$\frac{1}{s}$
Ramp function $u(t)t$	$\frac{1}{s^2}$
Exponential delay $u(t)\mathrm{e}^{-\alpha t}$	$\frac{1}{s+\alpha}$
Exponential rise $u(t)\ (1-\mathrm{e}^{-\alpha t})$	$\frac{\alpha}{s(s+\alpha)}$
$u(t)t\ \mathrm{e}^{-\alpha t}$	$\frac{1}{(s+\alpha)^2}$
$u(t)\sin(\omega_{\mathrm{rt}}t)$	$\frac{\omega_{\mathrm{rt}}}{s^2+\omega_{\mathrm{rt}}^2}$
$u(t)\mathrm{e}^{-\alpha t}\ \sin(\omega_{\mathrm{rt}}t)$	$\frac{\omega_{\mathrm{rt}}}{(s+\alpha)^2+\omega_{\mathrm{rt}}^2}$
$u(t)\mathrm{e}^{-\alpha t}\ \cos(\omega_{\mathrm{rt}}t)$	$\frac{s+\alpha}{(s+\alpha)+\omega_{\mathrm{rt}}^2}$
$f(t-\tau)$	$\mathrm{e}^{-s\tau}F(s)$

and

$$\mathcal{L}\int_0^t f(t)\mathrm{d}t = \frac{1}{s}F(s) + \frac{f^{-1}(0-)}{s} \tag{6}$$

where $f(0-),\ f^1(0-), \ldots, f^{n-1}(0-)$ are the values of the function, and its $n-1$ derivatives and $f^{-1}(0-)$ is the value of the time integral of $f(t)$ just prior to the application of the signal at $t = 0$.

It should be noted that the limit $t = 0-$ is used here, whereas in rigorous mathematical texts, in which the derivatives of discontinuous functions are not legitimate functions, the lower limit is quoted as $t = 0+$. However, in practical engineering analysis in which the unit impulse function ($\delta(t)$) is used, a more consistent methodology results by using a lower limit $t = 0-$.

Notice that when control engineers use the single-sided Laplace transform, all the driving signals are considered to operate after $t = 0$; they are defined as zero prior to this instant and this may be conveniently represented as multiplying all the time functions by the unit step $u(t)$.

2 ANALYSIS OF SIMPLE POSITION CONTROLLER WITH DAMPING AND VELOCITY FEEDBACK

Laplace transforms that have now been introduced allow analysis of a typical closed loop controller having damping and velocity feedback.

Returning to the simple position system described in **Article 18, Nature and Scope of Closed-loop Systems, Volume 1**, it is necessary to define the form of the input signal and the all-important initial conditions.

Assuming zero initial conditions, both sides of the differential equation can be transformed, thus

$$Js^2\Theta_{\mathrm{o}}(s) + Fs\Theta_{\mathrm{o}}(s) + K\Theta_{\mathrm{o}}(s) = K\Theta_{\mathrm{i}}(s) \tag{7}$$

Hence,

$$\Theta_{\mathrm{o}}(s) = \frac{K\Theta_{\mathrm{i}}(s)}{Js^2 + Fs + K} = \frac{(K/J)\Theta_{\mathrm{i}}(s)}{s^2 + (F/J)s + K/J} \tag{8}$$

Consider the response of this characteristic equation to various input signals.

2.1 Response to the unit impulse as an input demand

If $\theta_{\mathrm{i}}(t) = \delta(t)$, then $\Theta_{\mathrm{i}}(S) = 1$, and so

$$\Theta_{\mathrm{o}}(s) = \frac{K/J}{s^2 + (F/J)s + K/J} \tag{9}$$

If we take the case in which $F < 2\sqrt{(JK)}$ and complete the square in the denominator; then

$$\Theta_{\mathrm{o}}(s) = \frac{K/J}{(s+\alpha)^2 + \omega_{\mathrm{rt}}^2} \tag{10}$$

where

$$\alpha = \frac{F}{2J} \text{ and } \omega_{\mathrm{rt}} = \left(\frac{K}{J} - \frac{F^2}{4J^2}\right)^{1/2}$$

This does not quite agree with the form of any of the tabulated Laplace transforms, but with slight manipulation it becomes

$$\Theta_{\mathrm{o}}(s) = \frac{K/J}{\omega_{\mathrm{rt}}}\frac{\omega_{\mathrm{rt}}}{(s+\alpha)^2 + \omega_{\mathrm{rt}}^2} \tag{11}$$

which may be inverse transformed to yield

$$\theta_{\mathrm{o}}(t) = u(t)\frac{K/J}{\omega_{\mathrm{rt}}}\mathrm{e}^{-\alpha t}\sin(\omega_{\mathrm{rt}}t) \tag{12}$$

This is the unit impulse response of the system.

2.2 Response to the unit step as an input demand

If $\theta_{\mathrm{i}}(t) = u(t)$, then $\Theta_{\mathrm{i}}(s) = 1/s$, thus

$$\Theta_{\mathrm{o}}(s) = \frac{K/J}{s[s^2 + (F/J)s + K/J]} \tag{13}$$

This must be broken into partial fractions (done using special mathematical techniques not covered here) to find matching transforms, thus

$$\Theta_o(s) = \frac{1}{s} - \frac{\alpha}{\omega_{rt}} \frac{\omega_{rt}}{(s+\alpha)^2 + \omega_{rt}^2} - \frac{s+\alpha}{(s+\alpha)^2 + \omega_{rt}^2} \quad (14)$$

Each term has a recognizable inverse Laplace transform leading to the solution in the time domain of

$$\begin{aligned}\theta_o(t) &= u(t) - \frac{\alpha}{\omega_{rt}} u(t)\, e^{-\alpha t} \sin(\omega_{rt} t) - u(t)\, e^{-\alpha t} \cos(\omega_{rt} t) \\ &= u(t)\{1 - e^{-\alpha t} \sqrt{[1 + (\alpha/\omega_{rt})^2]} \\ &\quad \times \sin[\omega_{rt} t + \tan^{-1}(\omega_{rt}/\alpha)]\} \end{aligned} \quad (15)$$

This is the unit step response.

2.3 Response to the unit ramp as an input demand

If $\theta_i(t) = u(t)$, then $\Theta_i(s) = 1/s^2$, and so

$$\Theta_o(s) = \frac{K/J}{s^2[s^2 + (F/J)s + K/J]} \quad (16)$$

This must be broken into partial fractions to yield the response

$$\theta_o(t) = u(t) \left(t - \frac{F}{K} + \frac{K e^{-\alpha t}}{F \omega_{rt}} \sin(\omega_{rt} t + \phi) \right) \quad (17)$$

where

$$\phi = \tan^{-1} \left(\frac{2\alpha\omega_{rt}}{2\alpha^2 - K/J} \right)$$

This is the unit ramp response.
These responses are plotted in Figure 1.

The quantities M_{pt}, T_p, ε_{ss}, and ε_p form some of the basic means of performance specification.

A similar set of responses are analyzed in **Article 62, Second-order System Dynamics, Volume 1**, where systems in general are covered. The different terminologies, differing parameter, and process of analyses given will perhaps be found useful depending upon the discipline to which they are being applied.

3 CONCEPT OF DAMPING RATIO AND UNDAMPED NATURAL ANGULAR FREQUENCY

The value of the damping term F relative to the terms J and K governs the dynamic behavior of the system. There are four possibilities:

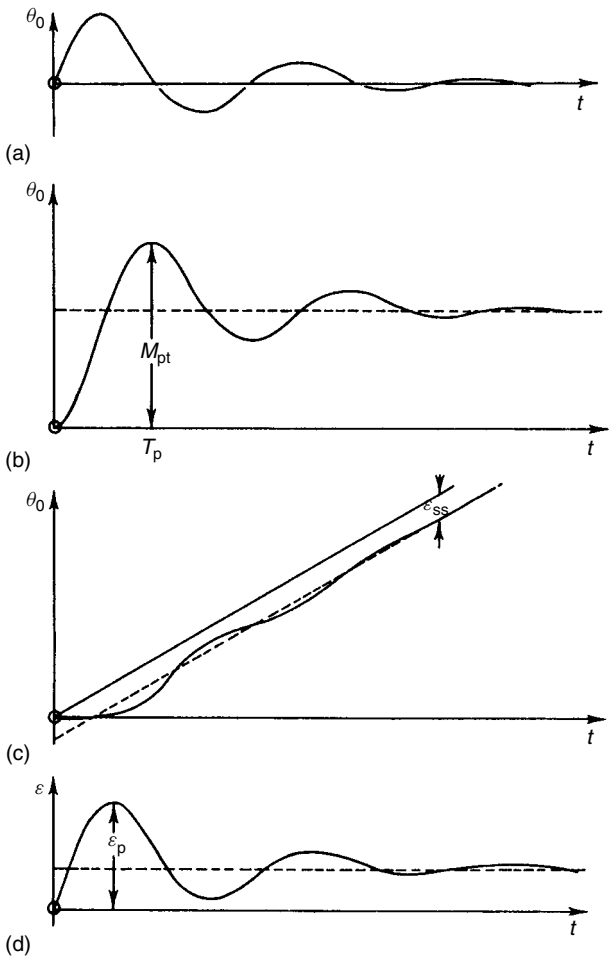


Figure 1. Time domain responses of a stabilized position control system: (a) impulse, (b) step, (c) ramp output, (d) ramp error.

1. $F = 0$, in which case the system will oscillate continuously with sinusoidal oscillations of angular frequency $\sqrt{(K/J)}$. This quantity is termed the *natural undamped angular frequency*. This response is termed *critically stable*.
2. $F < 2\sqrt{(JK)}$, in which case the response contains an exponentially damped sinusoidal mode and will exhibit overshoot in response to a step input. This is called the *underdamped response*.
3. $F = 2\sqrt{(JK)}$, for which the response is *critically damped*, that is, it does not quite overshoot in response to a step input.
4. $F > 2\sqrt{(JK)}$, for which the response is a double exponential rise in response to a step input. This is called the *overdamped response*.

The simple example given in article **Article 18, Nature and Scope of Closed-loop Systems, Volume 1** concentrates on condition 2 because this is in practice the most important case as it allows the system to settle within

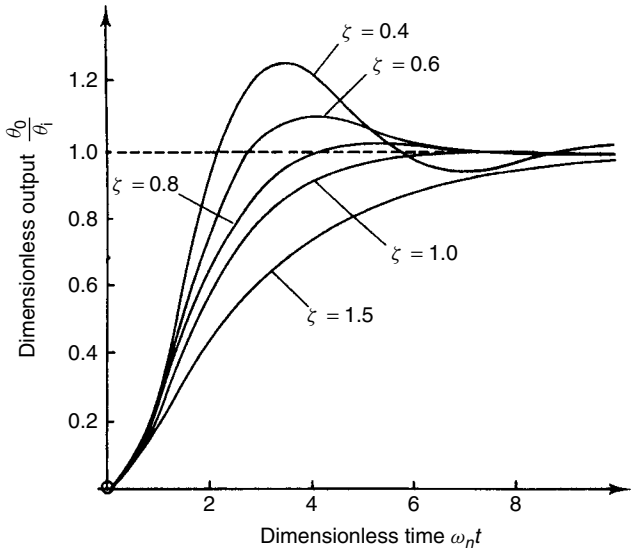


Figure 2. Dimensionless unit step response.

a given tolerance band around the desired value faster than any other.

It is convenient to normalize the effect of the damping term F by relating it to the value F_c required to achieve critical damping where $F_c = 2\sqrt{(JK)}$.

The damping ratio ζ is defined as

$$\zeta = \frac{F}{F_c} = \frac{F}{2\sqrt{(JK)}} \tag{18}$$

so that

$\zeta < 1$ is an underdamped system
$\zeta = 1$ is a critically damped system
$\zeta > 1$ is an overdamped system

The differential equation may be rewritten in terms of ω_n and ζ thus

$$\frac{d^2\theta_o}{dt^2} + 2\zeta\omega_n\frac{d\theta_o}{dt} + \omega_n^2\theta_o = \omega_n^2\theta_i \tag{19}$$

The response to a unit step input for various values of ζ is illustrated in Figure 2.

4 FREQUENCY RESPONSE OF CLOSED LOOP SYSTEMS

The steady state behavior of the system in response to a sinusoidal input is of considerable practical importance. When we refer to the frequency response of a system, we mean the variation of the phase and magnitude of the steady state output of the system, as the frequency of the input sinusoid is varied over the range of interest.

The Laplace transform of the output of the system is related to the Laplace transform of the input by the equation

$$s^2\Theta_o(s) + 2\zeta\omega_n s\Theta_o(s) + \omega_n^2\Theta_o(s) = \omega_n^2\Theta_i(s) \tag{20}$$

For steady state sinusoids, it is permissible to substitute $j\omega = s$, where ω is the angular frequency of the input, and produce the following operational relationship between $\Theta_i(j\omega)$ and $\Theta_o(j\omega)$:

$$\begin{aligned}\frac{\Theta_o(j\omega)}{\Theta_i(j\omega)} &= \frac{\omega_n^2}{(j\omega)^2 + 2\zeta\omega_n(j\omega) + \omega_n^2} \\ &= \frac{\omega_n^2}{\omega_n^2 - \omega^2 + j2\zeta\omega_n\omega} \\ &= \frac{1}{1 - (\omega/\omega_n)^2 + j2\zeta\omega/\omega_n}\end{aligned} \tag{21}$$

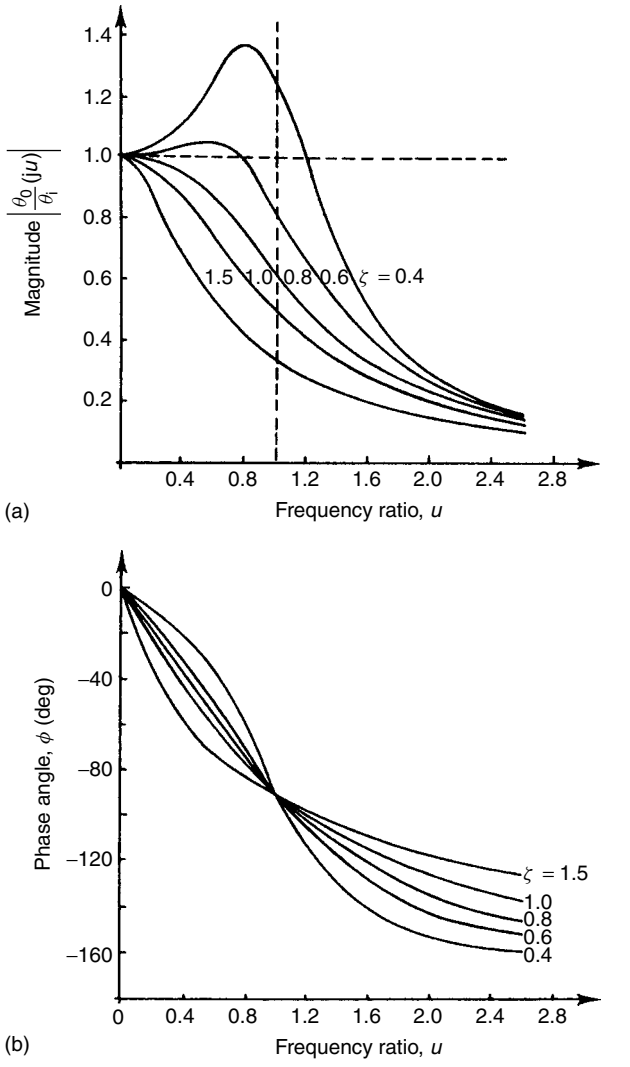


Figure 3. Frequency response characteristics: (a) magnitude, (b) phase.

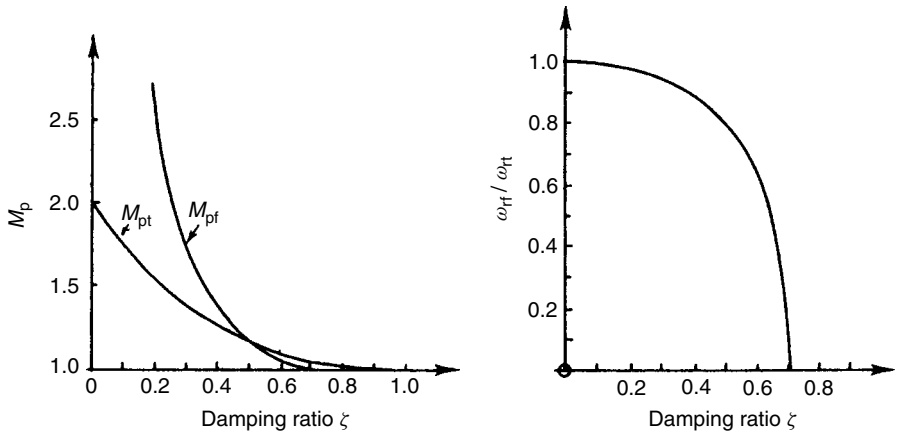


Figure 4. Time and frequency domain characteristics of the second-order system.

It is simpler to work in terms of a nondimensional frequency ratio $u = \omega/\omega_n$ for which

$$\frac{\Theta_o(ju)}{\Theta_i(ju)} = \frac{1}{(1-u^2) + j2\zeta u} \tag{22}$$

From this expression, we can determine the modulus, M, its peak value, M_{pf} (if any), and the phase, ϕ, which is given by

$$M = \frac{1}{\sqrt{[(1-u)^2 + (2\zeta u)^2]}} \tag{23}$$

$$M_{pf} = \frac{1}{2\zeta\sqrt{(1-\zeta^2)}} \tag{24}$$

and

$$\phi = -\tan^{-1}\left(\frac{2\zeta u}{1-u^2}\right) \tag{25}$$

The magnitude and phase characteristics are illustrated in Figure 3. The angular frequency at which the frequency response has its peak value is designated by the symbol ω_{rf}. It may be shown that

$$\omega_{rf} = \omega_n\sqrt{(1-2\zeta^2)} \tag{26}$$

5 SECOND-ORDER CORRELATIONS

The time domain (step response) and frequency response of the second-order system are connected through correlating equations; these may be combined as graphs, which are useful in the approximate design of higher-order systems (Atkinson, 1968). These graphs are given in Figure 4.

Levine (1996) provides a general background to control systems analysis.

RELATED ARTICLES

Article 18, Nature and Scope of Closed-loop Systems, Volume 1; **Article 20, Closed-loop Sampled Data Systems, Volume 1**; **Article 21, Nonlinear Closed-loop Systems, Volume 1**; **Article 27, Signals in the Frequency Domain, Volume 1**; **Article 28, Signals in the Time Domain, Volume 1**; **Article 29, Relationship Between Signals in the Time and Frequency Domain, Volume 1**.

REFERENCES

Atkinson, P. (1968) *Feedback Control Theory for Engineers*, 1st edn, Heinemann, London.

Atkinson, P. (1982) Closed loop systems, in *Handbook of Measurement Systems*, Wiley, Chichester.

Levine, W.S. (1996) *The Control Handbook*, CRC Press, Boca Raton, FL.

This article is based in part on a chapter by P. Atkinson originally published in Handbook of Measurement Science Vol 1, P.H. Sydenham, Wiley, 1982.

20: Closed-loop Sampled Data Systems

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Introduction to Closed-loop Sampled Data Systems 112
2 Use of Sample and Hold Circuits 112
3 Frequency Response Methods Applied to Sampled Closed-loop Systems 113
4 Time Domain Analysis of Sampled Data Closed-loop Systems 114
5 Stability of Closed-loop Sampled Data Systems Using the Z-transform 116
6 Compensation of Sampled Data Closed-loop Systems 116
Related Articles 117
References 117

1 INTRODUCTION TO CLOSED-LOOP SAMPLED DATA SYSTEMS

This article continues the discussion of instrument feedback systems and is based on original material in Atkinson (1982). Today many practical control systems use discrete digital signals. Their operation is an extension of analog principles usually by addition of sampling of the signal stream. The prior discussion of closed-loop systems in **Article 18, Nature and Scope of Closed-loop Systems, Volume 1** and **Article 19, Dynamic Behavior of Closed-loop Systems, Volume 1** have assumed operation using analog continuous signals by an A/D converter – *see* **Article 138, Sampling Theory Relevant to DAS, Volume 3** and **Article 139, Analog-to-Digital (A/D) Converters, Volume 3**.

To understand the complications that can arise and how the digital controller system can be analyzed in depth, it is necessary to understand the sampling process. It is not usually expected that a measuring systems designer be familiar with carrying out this level of analysis but its principles do need to be generally appreciated.

Sampling is defined in general as the generation of a sequence of impulses at the sampling instants, with the area or magnitude of each impulse being equal to the original signal value at that time (Figure 1).

The notation $x^*(t)$ represents the ideal impulse sampled version of the original signal $x(t)$ entering the sampler.

Feedback systems can be effectively represented by the block diagram shown in Figure 2, in which the continuous elements have a transfer function $H(s)$. The precise analysis of this type of system can be very involved; however, in many practical situations, the sampling interval T_s is short compared with the time for a transient oscillation of the whole system. Under these circumstances, it is possible to replace the impulse sampler theoretically by a pure gain of value $1/T_s$ and analyze the circuit by normal continuous theory or simulation.

2 USE OF SAMPLE AND HOLD CIRCUITS

In the arrangement illustrated in Figure 2, the impulse sampler is placed in the error channel of the feedback system. This is a fairly normal situation in control systems; in a properly designed system, the main elements then act as a low-pass filter, which smoothes the sampled error so that the output $\theta_o(t)$ follows the input $\theta_i(t)$ over the required profile of inputs.

Even in a system that has been designed correctly and that has a reasonably high sampling rate, the response will not be exactly the same as that of the equivalent linear

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

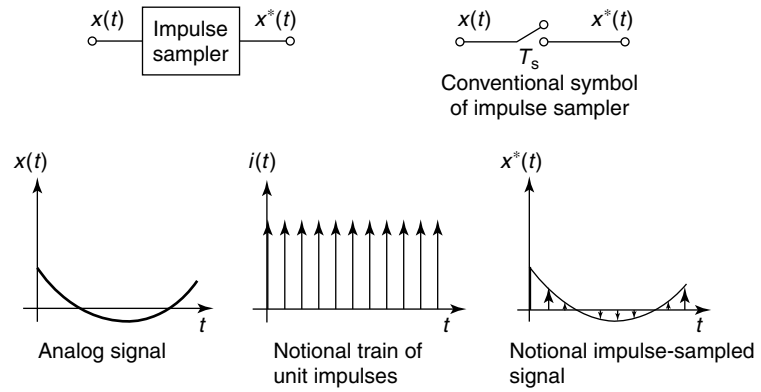


Figure 1. Action of a theoretical impulse sampler.

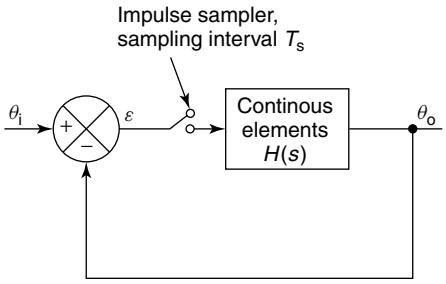


Figure 2. Closed-loop sampled data system.

system. It is often most economical to use low sampling rates, in which case the intersample ripple on the output may be very severe. In order to make the behavior of the sampled data system more like that of the continuous system and particularly to reduce intersampling ripple on the output, various forms of filter are used between the sampler and the continuous elements.

The simplest type of filter is the zero-order hold or clamp. The actions of an impulse sampler with a zero-order hold are shown in Figure 3.

The output is held at the previous sampled value until the next sampling instant. The transfer function of the zero-order hold takes the form:

$$\frac{1-\exp(-sT_s)}{s}$$

The total response of a sample and hold circuit approximates on average (ignoring harmonics) to a pure time delay of $T_s/2$. This approximation is valid down to quite low sampling rates and allows design to be performed by classical feedback control methods in the frequency domain. The approximation can also be used for approximate transient analysis even at quite low sampling rates.

Further improvements in signal smoothing can be achieved by using predictive hold circuits, which use the past two values to predict (or estimate) the slope of the curve to the next sampling instant.

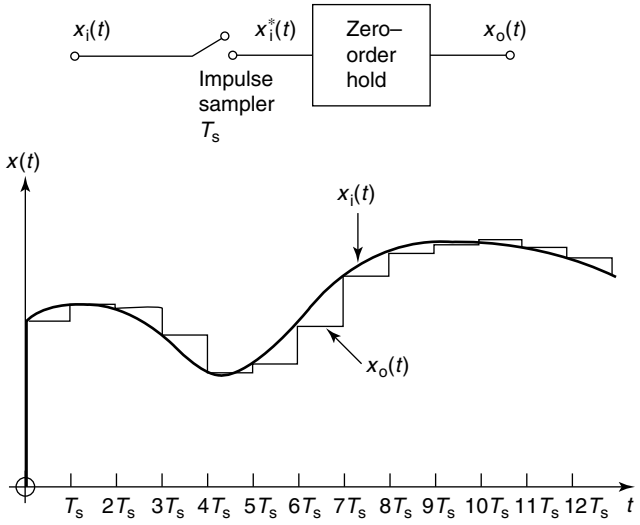


Figure 3. Action of the zero-order hold.

3 FREQUENCY RESPONSE METHODS APPLIED TO SAMPLED CLOSED-LOOP SYSTEMS

Although for many purposes frequency domain analysis may be employed for sampled data systems using the approximate continuous equivalents described in Sections 1 and 2 above, at very low sampling rates, or in cases where more accurate analysis is required, it is necessary to use a more accurate technique. Linvill (1951) developed a formula for computing the frequency response of a sampled data system by making a vector addition of all the harmonics produced by sampling.

Given a continuous signal of Laplace transform $E(s)$, then the Laplace transform $E^*(s)$ of the sampled signal is given by

$$E^*(s) = \frac{1}{T_s} \sum_{n=-\infty}^{+\infty} E(s + jn\omega_s) \tag{1}$$

where n is an integer and ω_s is the angular sampling frequency. The frequency response is then written in the usual way by substituting $s = j\omega$, that is,

$$E^*(j\omega) = \frac{1}{T_s} \sum_{n=-\infty}^{+\infty} E(j\omega + jn\omega_s) \tag{2}$$

Frequency response loci (i.e. Nyquist diagrams) can be constructed using this series to investigate stability of the loop.

4 TIME DOMAIN ANALYSIS OF SAMPLED DATA CLOSED-LOOP SYSTEMS

The method of z-transforms provides a basis for time-domain analysis and stability analysis of a closed-loop sampled data system. Its main limitation is that it provides information about signal amplitudes at the sampling instants only. It, therefore, provides no information regarding the intersample ripple.

An impulse-sampled signal $\varepsilon^*(t)$ has a Laplace transform $\varepsilon^*(s)$, which contains s in the form $\exp(sT_s)$; the z-transform is obtained by making the substitution $z = \exp(sT_s)$. The z-transform can be represented as a series

$$\varepsilon(z) = \sum_{n=0}^{\infty} \varepsilon(nT_s) z^{-n} \tag{3}$$

where n is an integer.

We can interpret z^{-1} as a delay operator of T_s seconds, z^{-2} as a delay operator of $2T_s$ seconds, and so on. The summation will take a general form

$$\varepsilon(z) = k_0 + k_1 z^{-1} + k_2 z^{-2} + k_3 z^{-3} + \cdots + k_n z^{-n} \tag{4}$$

Each term in the series contains the amplitude k_n at the sampling instant nT_s in the form $k_n z^{-n}$ (See Figure 4). The z-transforms for various functions of time are given in Table 1 for reference.

Although the z-transform can be inverted in a number of ways, it is generally best to expand the expression into a power series in powers of z^{-1} by algebraic long division. The coefficient of z^{-n} corresponds to the value of the time function at the nth sampling instant.

Classic control theory texts, such as Truxall (1955), show how to calculate the response of a closed-loop sampled data control system containing an impulse sampler in the error channel (Figure 5).

The z-transfer (or pulse transfer function) of the system is given by

$$\frac{\Theta_o(z)}{\Theta_i(z)} = \frac{G(z)}{1 + HG(z)} \tag{5}$$

It should be noted that $HG(z) \neq H(z)G(z)$.

The application of the closed-loop z-transform is illustrated by an example (Figure 6) in which the input is a unit

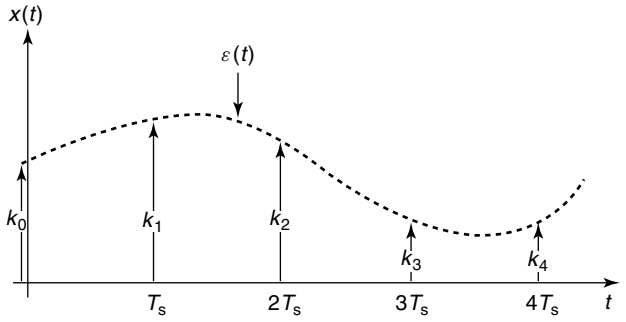


Figure 4. Illustrating z-transform series representation of a sampled signal.

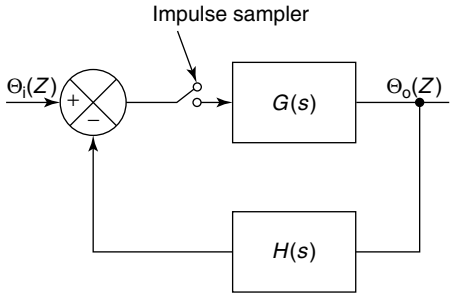


Figure 5. Sampled-data control system.

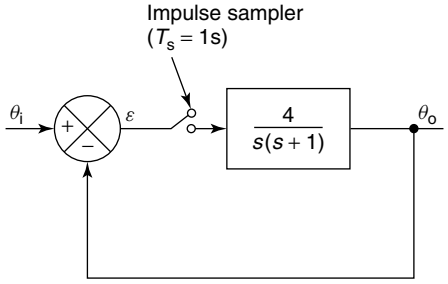


Figure 6. Sampled-data system for worked example.

Table 1. Laplace and z transforms of commonly met time functions.

Time function $f(t)$	Laplace transform	Z-transform
Unit step $u(t)$	$\frac{1}{s}$	$\frac{z}{z-1}$
Unit ramp $u(t)t$	$\frac{1}{s^2}$	$\frac{T_S z}{(z-1)^2}$
Acceleration function $u(t)t^2/2$	$\frac{1}{s^3}$	$\frac{T_s^2 z(z+1)}{2(z-1)^3}$
$u(t)\frac{t^n}{n!}$	$\frac{1}{s^{n+1}}$	$\lim_{\alpha\to 0}\frac{(-1)^n}{n!}\frac{\partial^n}{\partial\alpha^n}\left(\frac{z}{z-\exp(-\alpha T_s)}\right)$
$u(t)\mathrm{e}^{-\alpha t}$	$\frac{1}{s+\alpha}$	$\frac{z}{z-\exp(-\alpha T_s)}$
$u(t)t\ \mathrm{e}^{-\alpha t}$	$\frac{1}{(s+a)^2}$	$\frac{T_s z\exp(-\alpha T_s)}{[z-\exp(-\alpha T_s)]^2}$
$u(t)\ (1-\mathrm{e}^{-\alpha t})$	$\frac{\alpha}{s(s+\alpha)}$	$\frac{[1-\exp(-\alpha T_s)]z}{(z-1)[z-\exp(-\alpha T_s)]}$
$u(t)\sin(\omega_{rt}t)$	$\frac{\omega_{rt}}{s^2+\omega_{rt}^2}$	$\frac{z\sin(\omega_{rt}T_s)}{z^2-2z\cos(\omega_{rt}T_s)+1}$
$u(t)\mathrm{e}^{-\alpha t}\ \sin(\omega_{rt}t)$	$\frac{\omega_{rt}}{(s+\alpha)^2+\omega_{rt}^2}$	$\frac{z\exp(\alpha T_s)\sin(\omega_{rt}T_s)}{z^2\exp(2\alpha T_s)-2z\exp(\alpha T_s)\cos(\omega_{rt}T_s)+1}$
$u(t)\cos(\omega_{rt}t)$	$\frac{s}{s^2+\omega_{rt}^2}$	$\frac{z[z-\cos(\omega_{rt}T_s)]}{z^2-2z\cos(\omega_{rt}T_s)+1}$
$u(t)\mathrm{e}^{-\alpha t}\ \cos(\omega_{rt}t)$	$\frac{s+\alpha}{(s+\alpha)^2+\omega_{rt}^2}$	$\frac{z^2-z\exp(-\alpha T_s)\cos(\omega_{rt}T_s)}{z^2-2z\exp(-\alpha T_s)\cos(\omega_{rt}T_s)+\exp(-2\alpha T_s)}$

step function and it is required to calculate the output.

$$H(s)=\frac{4}{s(s+1)}=\frac{4}{s}-\frac{4}{s+1} \tag{6}$$

From the table of z-transforms (Table 1)

$$H(z)=\frac{4z}{z-1}-\frac{4z}{z-\exp(-\alpha T_s)} \tag{7}$$

where $\alpha T_s=1$. Thus,

$$H(z)=\frac{2.53z}{z^2-1.37z+0.368} \tag{8}$$

and so

$$\frac{\Theta_o(z)}{\Theta_i(z)}=\frac{H(z)}{1+H(z)}=\frac{2.53z}{z^2+1.16z+0.368} \tag{9}$$

Now $\theta_i(t)=u(t)$. Hence, again using the table of transforms:

$$\begin{aligned}\Theta_i(z)&=\frac{z}{z-1}\\ \Theta_o(z)&=\frac{2.53z}{z^2+1.16z+0.368}\ \frac{z}{z-1}\\ &=\frac{2.53z^2}{z^3+0.16z^2-0.793z-0.368}\end{aligned} \tag{10}$$

By algebraic long division this expression yields:

$$\begin{aligned}\Theta_o(z)=2.53z^{-1}&-0.406z^{-2}+2.03z^{-3}+0.339z^{-4}\\ &+1.37z^{-5}+0.827z^{-6}+1.06z^{-7}+1.00z^{-8}+\cdots\end{aligned} \tag{11}$$

As can be seen from this response, although the system is highly undamped, its output eventually converges towards the input.

5 STABILITY OF CLOSED-LOOP SAMPLED DATA SYSTEMS USING THE Z-TRANSFORM

For a sampled data control system, the z-transfer function is given by

$$\frac{\Theta_o(z)}{\Theta_i(z)} = \frac{G(z)}{1 + HG(z)} \tag{12}$$

The stability of the system depends on the positions of the zeros of $[1 + HG(z)]$ in the s-plane. However, the transformation of $\exp sT_s = z$ has been made and the positions of the zeros can be mapped in the z-plane. The mapping, $z = \exp sT_s$, maps the imaginary axis of the s-plane into the unit circle about the origin of the z-plane and the left half of the s-plane into the interior of the unit circle.

A sampled data control system will only be absolutely stable if all the zeros of $1 + HG(z)$ lie inside the unit circle.

Direct application of this criterion can be tedious; use of the bilinear transform $z = (1 + w)/(1 - w)$ maps the inside of the circle in the z-plane onto the left-hand side of the w-plane. It is then possible to use the Routh–Hurwitz or the Nyquist stability criterion directly.

6 COMPENSATION OF SAMPLED DATA CLOSED-LOOP SYSTEMS

Sampled data (i.e. digital) systems offer a wider range of possibilities for compensation than do continuous signal systems. Series and parallel or series/parallel compensation using continuous transfer functions can be designed in a similar way using a Nyquist or inverse Nyquist diagram based on the Linvill approximation (Section 3 above).

Alternative strategies involve series digital compensators alone (Figure 7a) or in combination with a parallel continuous compensator (Figure 7b).

Digital compensators can be arranged to produce a phase-lead for which $H_c(z) = (z - a)/(z + b)$, with the zero lying to the right of the pole in the z-plane, or a phase lag for which $H_c(z) = (z + a)/(z - b)$, with the zero lying to the left of the pole.

Many other forms of a digital compensator could be realized by construction from hard-wired logical circuitry or, more usually today, by means of software implemented on a digital microprocessor almost always associated with a closed-loop controller; software implementation is clearly very advantageous.

The arrangement illustrating control and digital compensation of a continuous process is shown in Figure 8; the error signal is generated by the computer and is

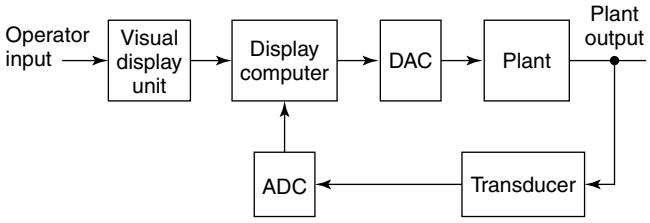


Figure 8. Control and compensation using a digital computer.

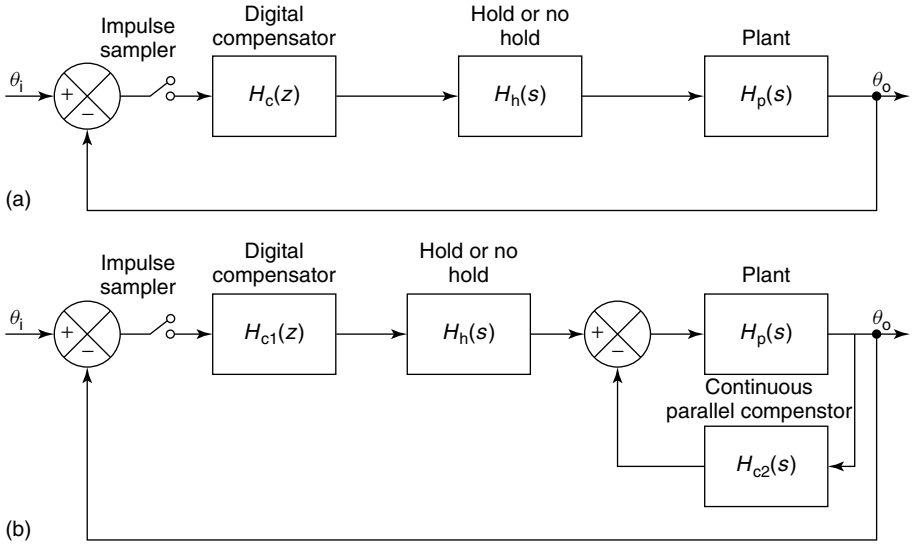


Figure 7. Forms of digital compensation: (a) series digital compensation; (b) series digital with parallel continuous compensation.

then operated on by a suitable algorithm before it is outputted to the plant via a digital-to-analog converter (DAC).

This article has only introduced the foundational approach to understanding the operation and tuning of digital closed loops. More detail can be found in Franklin *et al*. (1997), Nekoogar and Moriarty (1998), Ogata (2001), Stubberud, Williams and DiStefano (1994), and Wittenmark (1990). Application is now assisted using the current, widely available circuit system simulation tools, Bishop (1997).

RELATED ARTICLES

Article 18, Nature and Scope of Closed-loop Systems, Volume 1; **Article 19, Dynamic Behavior of Closed-loop Systems, Volume 1**; **Article 37, Relationship Between Systems in the Time and Frequency Domain, Volume 1**; **Article 38, Stability Issues, Volume 1**.

REFERENCES

Atkinson, P. (1982) Closed-loop systems, in *Handbook of Measurement Science*, Vol. 1 (ed. P.H. Sydenham), Wiley, Chichester (pp. 591–640).

Bishop, R.H. (1997) *Modern Control Systems Analysis and Design Using Matlab and Simulink*, Addison-Wesley.

Franklin, G.F., Powell, D.J., Workman, M.L. and Powell, D. (1997) *Digital Control of Dynamic Systems*, Addison-Wesley.

Linvill, W.K. (1951) Sampled-data Control Systems Studies through Comparison of Sampling with Amplitude Modulation. *Transactions AIEE*, **70**, Part II, 1779–88.

Nekoogar, F. and Moriarty, G. (1998) *Digital Control Using Digital Signal Processing*, Prentice Hall.

Ogata, K. (2001) *Modern Control Engineering*, Prentice Hall.

Stubberud, A.A., Williams, A.J. and DiStefano, J.J. (1994) *Schaum's Outline of Feedback and Control Systems*, McGraw-Hill Trade.

Truxall, J.G. (1955) *Introductory System Engineering*, McGraw-Hill, Tokyo.

Wittenmark, B. (1990) *Computer Controlled Systems*, Prentice Hall.

This article is based in part on a chapter by P. Atkinson originally published in Handbook of Measurement Science Vol 1, P.H. Sydenham, Wiley, 1982.

21: Nonlinear Closed-loop Systems

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Introduction to Nonlinear Systems 118
2 Methods for Studying Nonlinear Systems 119
3 Phase-plane Analysis 120
4 Describing Function Method for Investigating Nonlinear Closed-loop Systems 120
5 How to Approach the Design of Closed Loops in Measurement Systems 122
Related Articles 123
References 123

1 INTRODUCTION TO NONLINEAR SYSTEMS

This article continues the discussion of instrument feedback systems and is based on original material in Atkinson (1982).

Although certain nonlinearities (Coulomb friction and stiction) were mentioned in connection with the positional servomechanism in **Article 18, Nature and Scope of Closed-loop Systems, Volume 1**, the theory covered there and in **Article 19, Dynamic Behavior of Closed-loop Systems, Volume 1** and **Article 20, Closed-loop Sampled Data Systems, Volume 1** describes the class of closed loop that are called *linear systems*.

Linear systems are those which obey the principle of superposition. Although linear theory is an indispensable design aid, it must be realized that all practical systems do inevitably contain nonlinearities to some degree.

For example, an amplifier can exhibit saturation if overdriven; when the error signal in a control system becomes large the control signal driving the equipment under control will be limited in amplitude. The result of this will normally be that the response time of the system will be longer than predicted by means of linear theory, which is based on the assumption that saturation is not present.

There are many types of nonlinearity. They were first distinguished in Gelb and Vander Velde (1968) as explicit or implicit nonlinearities.

Explicit implies that the output of the nonlinearity is explicitly determined in terms of the required input variables, whereas implicit implies a more complicated relationship between input and output through, for example, an algebraic or differential equation.

With explicit nonlinearities, we have to distinguish between static and dynamic forms, in which dynamic implies that the output is related not only through the input but also through the derivatives of the input. Among the explicit, static nonlinearities, we must again divide between single-value (memoryless) nonlinearities such as saturation and dead-space, and multiple-value (memory) nonlinearities such as hysteresis.

The analysis and design of nonlinear systems is vastly more complex than that of linear systems, every nonlinear system being a miniature universe. Fortunately, simulation tools are available for such work that make design easier than by long hand formal methods.

Although some nonlinearities are intentionally inserted by the designer to obtain improved system performance, for the most part, they are a nuisance causing undesirable side effects when the error signal is large (e.g. the effect of saturation) or when it is small (such as, backlash in gears that may cause tick to occur). System errors of this kind are discussed in **Article 57, Static Considerations of General Instrumentation, Volume 1**.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

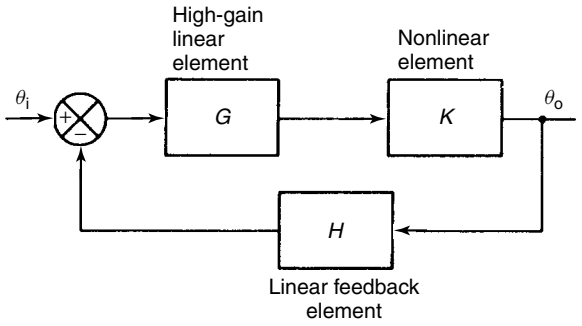


Figure 1. Nonlinear element in high-gain feedback loop.

In some circumstances (as with hysteresis), it may be possible to effectively eliminate the effects of the nonlinearity by means of a high-gain stable local feedback loop. This procedure is strongly recommended where it is possible. The procedure is based on the idea that the nonlinear element has a variable static gain, K. It is driven from a high-gain linear amplifier of gain, G, and feedback is applied through an element of gain, H, as shown in Figure 1.

Analysis shows that the transfer function in such cases is given by

$$\frac{\theta_o}{\theta_i} = \frac{GK}{1 + HGK} \tag{1}$$

The method relies on making $HGK > 1$ for all possible values of K, so that

$$\frac{\theta_o}{\theta_i} > \frac{1}{H} \tag{2}$$

The effects of the nonlinearity are thus made quite negligible so long as none of the elements saturate, this point often being overlooked.

This arrangement is only of use so long as the loop is relatively stable, so careful design of the loop is essential. This is probably best performed in the frequency domain using the describing function method given in Section 4.

Stiction can cause stick–slip motion in feedback mechanisms following constant velocity (ramp) inputs. This problem can again be overcome by means of a high-gain local feedback loop. In this case, velocity feedback is used around a motor–amplifier series combination; the gain of the amplifier is made very high so that the arrangement behaves like a perfect integrator of transfer function $1/k_s s$ where, k_s, is the velocity feedback constant. The arrangement is again identical to that shown in Figure 1; K represents the nonlinear motor characteristic $H = k_s s$, and G is the high gain of the amplifier. There is usually very little problem in designing this loop to have adequate relative stability.

2 METHODS FOR STUDYING NONLINEAR SYSTEMS

There are no completely satisfactory methods of studying nonlinear systems. Computer simulations are of great importance; to perform a simulation it is, however, necessary to choose a model to represent the important characteristics of a system.

A computer tool study of any nonlinear system is an essential final step before the real system is built.

Trial and error design based entirely on computer simulation is not advisable as it may not give a complete insight into the system behavior and the strategies which are available.

In instrumentation systems design, it is often possible to avoid nonlinearities by choice of components and their working regimes. For efficient lowest cost design it is, however, often better to make use of nonlinear effects meaning that more investigation into the behavior will be needed.

The first step in designing a nonlinear system is usually to attempt to linearize the essential nonlinearities for small signals about a working point.

For example, if we have a nonlinear device with an input signal $x(t)$ and an output $y(t)$ such that $y(t) = kx^2(t)$ then $\partial y/\partial x = 2kx$

If we are working at some particular working point x_p, then for small departures from the working point, the effective gain of the device is $2kx_p$.

For working points at values of x greater than x_p, the gain will be greater than $2kx_p$ and for working points at values of x less than x_p, the gain will be less.

Linear systems design application, combined with a sensitivity analysis and a computer simulation should again prove adequate.

Phase-plane analysis is a useful tool in the analysis of nonlinear second-order systems – *see* **Article 62, Second-order System Dynamics, Volume 1** – subjected to step (and ramp) inputs.

The technique involves the determination of the response in terms of its derivative of output (or error) plotted as a function of its output (or error). This plot is called a *phase-plane diagram*. In formulating the relevant equations, time is removed explicitly. For many practical nonlinearities, where the nonlinearities can be represented by linear segmented characteristics, the phase plane can be divided into various regions each of which corresponds to motion on a particular linear segment of the nonlinearity (see Section 3).

For higher-order systems, on occasions where it is necessary only to determine whether the designed system will remain absolutely stable for the entire envelope of input

signals, the methods of Lyapunov or Popov will provide exact answers without solving the differential equations.

The method of describing functions based on the concept of quasi-linearization for a given class of input signals provides a sound basis for the analysis and design of nonlinear feedback-control systems (see Section 4).

3 PHASE-PLANE ANALYSIS

Phase-plane analysis considers any second-order differential equation of form:

$$\ddot{x} + A\dot{x} + Bx = F \tag{3}$$

where A, B, and F are not necessarily constant. A phase portrait consists of a number of phase trajectories in the $\dot{x}$-versus x-plane.

If we define $y = \dot{x}$, then

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt} = \frac{F - Bx - Ay}{y} \tag{4}$$

This equation represents the slope of the phase trajectory in terms of functions of x and y. In general B, A, and F may be functions of x and y. To find the phase trajectory itself, this equation must be integrated; sometimes this can be performed analytically, but usually it is better done numerically using a digital computer or, the original long hand way, graphically using the method of isoclines.

By way of example, consider the simple viscously damped second-order servomechanism described in **Article 18, Nature and Scope of Closed-loop Systems, Volume 1** but now with the additional Coulomb friction.

Consider the response of the system with a step input and let C be the magnitude of the Coulomb frictional torque, which always opposes motion.

The instantaneous accelerating torque in $K\varepsilon$ and the retarding torque is $F\dot{\theta}_o + C\sin\dot{\theta}_o$, where $\sin\dot{\theta}_o$ is positive for $\dot{\theta}_o > 0$ and negative for $\dot{\theta}_o < 0$. Thus, applying Newton's second law, we have

$$J\ddot{\theta}_o = K\varepsilon - F\dot{\theta}_o - C\sin\dot{\theta}_o \tag{5}$$

Now $\varepsilon = \theta_i - \theta_o$, thus $\theta_o = \theta_i - \varepsilon$; also for a step input $d\theta_i/dt = 0$ and $d^2\theta_i/dt^2 = 0$ under steady state conditions.

It is possible to translate the above equation into the error form

$$J\ddot{\varepsilon} + F\dot{\varepsilon} + K\varepsilon + C\sin\dot{\varepsilon} = 0 \tag{6}$$

From this equation we can deduce

$$\dot{\varepsilon} = -\frac{K\varepsilon}{NJ + F} - \frac{C\sin\dot{\varepsilon}}{NJ + F} \tag{7}$$

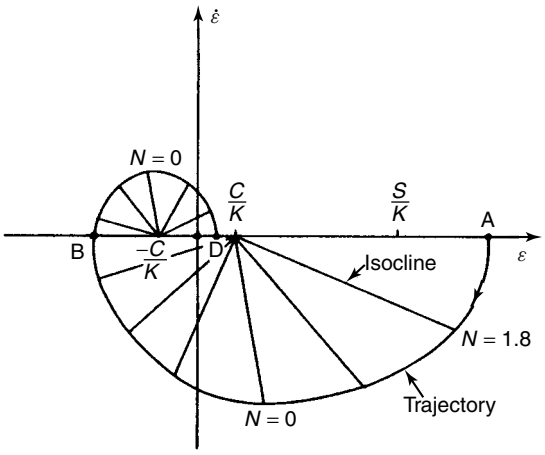


Figure 2. Phase trajectory for simple positional servomechanism with Coulomb friction.

in which N is the slope of the phase trajectory where it crosses the isocline.

The first term defines the family of isoclines for the linear system, whereas the second term introduces the effect of the nonlinearity.

The focal point is changed from $\varepsilon = +C/K$ to $\varepsilon = -C/K$ as $\dot{\varepsilon}$ changes from a negative value to a positive value.

If a trajectory begins at $\varepsilon = A$ (where A is equivalent to the value of the input step), it transverses through the phase plane as shown in Figure 2. The slope of the phase trajectory is given by the value of N as each isocline is crossed. The determination of the passage of the trajectory through the isoclines to point B when $\dot{\varepsilon}$ becomes negative, thence to point D when it again becomes positive is a simple matter. Motion ceases at D because the generated torque is less than C.

The phase-plane technique can be used to analyze the behavior of the second-order system for a variety of commonly encountered nonlinearities.

The main disadvantage of the phase-plane method is that it cannot be extended to higher-order systems in a satisfactory manner.

4 DESCRIBING FUNCTION METHOD FOR INVESTIGATING NONLINEAR CLOSED-LOOP SYSTEMS

In its most simple form, the describing function method is an extension of ordinary transfer function analysis to take into account the effect of single nonlinearities in systems excited by sinusoidal input.

It is particularly useful as a method for predicting the amplitude and frequency of limit cycles, but it can also be

used to assess relative stability by conventional frequency response methods.

In essence, the describing function of a nonlinear element is the gain of the element in terms of the ratio of its fundamental component output to a sinusoidal input of given magnitude and frequency.

Whereas the gain of a linear element is only ever frequency dependent the gain of a nonlinear element is always amplitude dependent and, additionally, can also be frequency dependent.

Here we will confine our study to amplitude dependent describing functions. In order to illustrate this point, consider a very common form of nonlinearity, namely, *saturation*.

The output is linearly related to the input for small positive or negative excursions of the input, but the output reaches a limiting value for large excursions (see Figure 3). The response of this nonlinearity to a sinusoid will be sinusoid for small signals, a clipped sinusoid for medium size signals, and tending to a square wave of magnitude KE_s for very large signals.

The gain of the element, based on the ratio of fundamental output to input thus is constant at a value K for small inputs; beginning to decrease as the input goes beyond E_s and eventually trailing off towards zero as the fundamental output tends towards its limiting value of $4KE_s/\pi$ as the input tends towards infinity (see Figure 4).

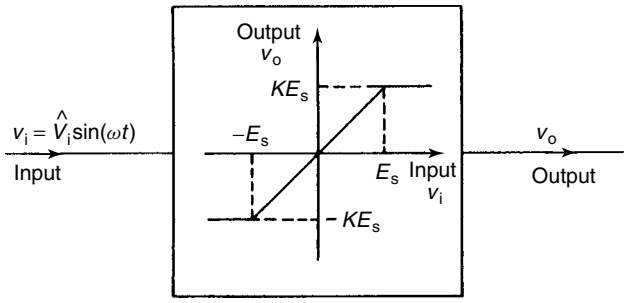


Figure 3. Illustrating a saturation nonlinearity.

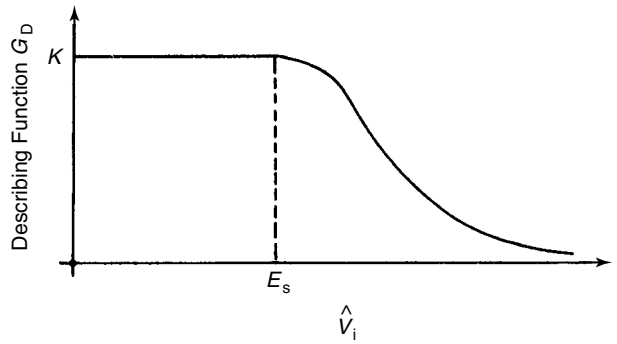


Figure 4. Describing function G_D for a saturation nonlinearity.

For memoryless nonlinearities the describing function is purely real, while memory-type nonlinearities introduce a phase shift, causing the describing function to be complex.

In general, if the fundamental output is expressed as a complex operator

$$V_o(j\omega) = \hat{V}_o \cos\theta + j\hat{V}_o \sin\theta \tag{8}$$

where $\hat{V}_o$ is the peak fundamental output and θ is the phase shift, then the describing function G_D is given by

$$G_D = \frac{\hat{V}_o \cos\theta}{\hat{V}_i} + j\frac{\hat{V}_o \sin\theta}{\hat{V}_i} \tag{9}$$

where $\hat{V}_i$ is the peak input.

To derive a describing function from this, we must use the Fourier series representation for the output

$$\begin{aligned} v_o(t) = {} & \tfrac{1}{2}A_o + A_1\cos(\omega t) + A_2\cos(2\omega t) \\ & + A_3\cos(3\,\omega t) + \cdots + B_1\sin(\omega t) + B_2\sin(2\,\omega t) \\ & + B_3\sin(3\,\omega t) + \cdots \end{aligned} \tag{10}$$

where

$$A_N = \frac{2}{\pi}\int_0^{\pi} v_o(t)\cos(N\omega t)\,\mathrm{d}(\omega t)$$

and

$$B_N = \frac{2}{\pi}\int_0^{\pi} v_o(t)\sin(N\omega t)\,\mathrm{d}(\omega t)$$

The definition of the describing function requires $A_o = 0$ and A_N and B_N for $N > 1$ to be negligible.

The fact that most control systems contain low-pass filtering elements, which filter out the harmonics to a substantial degree usually justifies these assumptions.

The describing function is then given by

$$\frac{B_1}{\hat{V}_i} + j\frac{A_1}{\hat{V}_i} \quad \text{in coordinate form}$$

or

$$\frac{\sqrt{B_1^2 + A_1^2}}{\hat{V}_i} \Big/ \tan^{-1}\frac{A_1}{B_1} \quad \text{in polar form}$$

Evaluation of the Fourier integrals may be quite simple, as in the case of the saturation nonlinearity but, sometimes, extremely difficult.

Fortunately, the describing functions of some commonly encountered nonlinearities have long been tabulated in the past, Thaler and Pastel (1962) and Gelb and Vander Velde (1968).

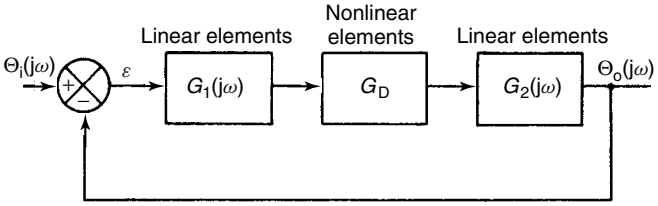


Figure 5. Nonlinear feedback system schematic.

The describing function for the saturation element is given by

$$G_D = \frac{2K}{\pi}[(\sin^{-1} R) + R\sqrt{(1 - R^2)}]\angle 0^\circ \qquad (11)$$

The Nyquist stability criterion – *see* **Article 38, Stability Issues, Volume 1** – may be involved for the determination of the stability of systems containing a nonlinear element as shown in Figure 5.

The closed-loop transfer function of this system is given by

$$\frac{\Theta_o}{\Theta_i} = \frac{G_1(j\omega)G_D G_2(j\omega)}{1 + G_1(j\omega)G_D G_2(j\omega)} \qquad (12)$$

The Nyquist criterion is based on the characteristic equation

$$1 + G_1(j\omega)G_D G_2(j\omega) = 0 \qquad (13)$$

To avoid the need to plot numerous Nyquist diagrams for every value of G_D, we may reform the characteristic equation as

$$G_1(j\omega)G_2(j\omega) = -\frac{1}{G_D} \qquad (14)$$

The encirclement of the locus $-1/G_D$ by the function $G_1(j\omega)G_2(j\omega)$ plotted as a polar diagram now indicates absolute instability.

Relative stability can be assessed by treating the locus $-1/G_D$ as the equivalent of the critical point $(-1, j0)$ used in linear systems design.

The amplitude and frequency of a limit cycle can readily be assessed by the point at which the frequency locus

$$G_1(j\omega)G_2(j\omega)$$

intersects the locus of $-1/G_D$. This is illustrated in Figure 6, which shows the Nyquist diagram of an unstable system that must limit cycle at angular frequency ω_c, where

$$\angle G_1(j\omega_c)G_2(j\omega_c) = -180^\circ \qquad (15)$$

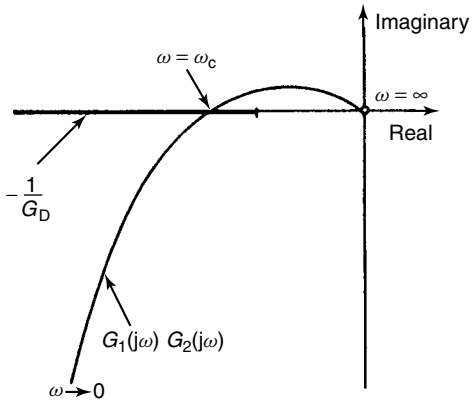


Figure 6. Nyquist diagram for a typical nonlinear system showing $-1/G_D$ (unstable situation).

from which ω_c may be calculated (generally by iterative trial and error).

The amplitude of the limit cycle can now be determined from

$$|G_1(j\omega)||G_D||G_2(j\omega)| = 1 \qquad (16)$$

which is the Nyquist amplitude condition for continuous oscillation.

The value of G_D can now be determined and the value of $\hat{V}_i$, the signal input to the nonlinear element, can be estimated by trial and error.

The error signal entering the first linear element can hence be computed from $\hat{V}_i/|G_1(j\omega_c)|$.

In the absence of any input to the system itself, the error and the output are identical in magnitude so that $\theta_o = \hat{V}_i/|G_1(j\omega_c)|$. This is the magnitude of the resultant limit cycle.

In general, it is necessary to determine the conditions graphically rather than analytically or by iterative trial and error.

Atherton (1975) has shown how the method of describing functions can be used to design compensating elements for nonlinear systems and be extended to treat multiple nonlinearities, to handle transient oscillations, dual sinusoidal inputs, and random inputs. Systems modeling tools often include means for investigating nonlinear systems.

5 HOW TO APPROACH THE DESIGN OF CLOSED LOOPS IN MEASUREMENT SYSTEMS

Many lengthy textbooks have been written on closed-loop systems so that it is hardly surprising that the information presented here merely touches the surface of a vast subject.

Material described here is based on the transfer function approach because this certainly offers the designer the most comprehensive set of techniques.

It should be understood, however, that many authors prefer to integrate the classical transfer function approach with the more recent state-variable approach.

In the state-variable approach the system model is described in terms of n first-order differential equations, each equation being a separate description of the behavior of a particular state and its connection with the other states and the driving inputs.

This form of description allows the equations for the system states to be condensed into the form of a single vector, $\mathbf{x}(t)$, and related to the driving input vector, $\mathbf{u}(t)$, by the equation

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \tag{17}$$

where $\mathbf{A}$ and $\mathbf{B}$ are coefficient matrices.

The system outputs can then usually be related to the internal states by the vector equation

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) \tag{18}$$

where, $\mathbf{C}$, is another coefficient matrix and, $\mathbf{y}(t)$, is the output vector.

The solution to the first of these matrix equations can be determined by various powerful computerized matrix methods and hence the response $\mathbf{y}(t)$ for a given $\mathbf{u}(t)$ can be determined quickly and accurately.

Apart from their application in time-domain analysis, state-variable techniques form the basis of the Lyapunov stability analysis and of optimal control theory. It can also be used in other design techniques, such as pole assignment, and pole assignment and decoupling methods, using state vector feedback.

In order to obtain greater sensitivity and better accuracy of measurement, it is likely that more and more instruments will incorporate feedback. The designers of instruments of all kinds will, therefore, have to become more familiar with the analysis and design of feedback systems.

In essence, the design of measuring systems loops is best kept manageable by avoiding nonlinear effects in any elements under the designer's control.

Those nonlinear effects that cannot be eliminated or lessened at source should be reduced by the use of local feedback solutions.

Manufacturing them in hardware form, even with a control prototyping kit, is not recommended until they have been investigated thoroughly with an appropriate modeling tool, of which many are available including Matlab® and Simulink – see Bishop (1997).

RELATED ARTICLES

Article 18, Nature and Scope of Closed-loop Systems, Volume 1; **Article 19, Dynamic Behavior of Closed-loop Systems, Volume 1**; **Article 26, Signal Classification, Volume 1**; **Article 88, Error Sources in Fine Mechanics, Volume 2**.

REFERENCES

Atherton, D.P. (1975) *Non-linear Control Engineering*, Van Nostrand Reinhold, London.

Atkinson, P. (1982) Closed Loop Systems, in *Handbook of Measurement Science*, Wiley, Chichester, 591–640.

Bishop, R.H. (1997) *Modern Control Systems Analysis and Design Using Matlab and Simulink*, Addison-Wesley.

Gelb, A. and Vander Velde, W.E. (1968) *Multiple-input Describing Functions and Non-linear Systems*, McGraw-Hill, New York.

Thaler, G.J. and Pastel, M.P. (1962) *Analysis and Design of Nonlinear Control Systems*, McGraw-Hill, New York.

This article is based in part on a chapter by P. Atkinson originally published in Handbook of Measurement Science Vol 1, P.H. Sydenham, Wiley, 1982.

22: Characteristics and Theory of Knowledge

Luca P. Mari
Università Cattaneo, Castellanza (VA), Italy

1 The Problem of Knowledge 127
2 The Status of Realism 127
3 Semiotics of Knowledge 129
4 Pragmatic Classification of Models 129
5 The Evaluation of Quality of Knowledge 130
6 Data and Inference in Knowledge 131
7 Inexactness of Knowledge and Measurement 132
8 (Non)foundations of Knowledge 132
Related Articles 133
Further Reading 133

1 THE PROBLEM OF KNOWLEDGE

Human beings know but do not definitely know what knowledge is: traditions, prejudices, expectations, and projections are more or less always part of knowledge and make it a combination of objectivity and subjectivity. Rationality allows some critical control on knowledge, and the recognition of the limitations to which human knowledge is subject is precisely a result of rationality.

The interest in theorizing about knowledge arises from the observation that different persons have different beliefs, and ultimately that beliefs and facts are distinct: 'theory of knowledge is a product of doubt', as Bertrand Russell wrote. In the history of both western and eastern culture, such a doubt has stimulated an impressive amount of research, ideas, and philosophical systems, and nevertheless very different positions have been maintained and still remain on the nature of knowledge and its object (it is reasonable to hypothesize that, more than from the plethora of such positions, the complexity of the topic derives from its inherent reflexivity, due to the fact that the object of the analysis coincides with the tool by means of which the analysis is performed: to know knowledge, only knowledge can be employed). A basic dichotomy can be identified, whose elements play the role of competing attractors for an ideal continuum of positions: *objectivism* assumes that a world external to the subject exists independently of him and has predefined properties, existing as such before they are acquired by the perceptive-cognitive system of the subject, whose aim is to reconstruct them; on the other hand, *solipsism* asserts that the cognitive system of the subject projects his own world out of him, and the reality of such a world is just an image of the laws internal to the system.

The position currently supported by the majority of scientists and engineers can be plausibly characterized as a kind of 'pragmatic realism', close to but not coincident with objectivism, according to which the conjoint efforts of science and technology are aimed at reaching, and actually guarantee, better and better, that is, more and more objective knowledge of the world whose properties are therefore progressively discovered.

Measurement plays a crucial role in supporting this realism.

2 THE STATUS OF REALISM

In acquiring and processing information from the world, human beings constantly *produce models* (and sometimes theories: we will not emphasize here the distinction between models and theories, grounded on formal logic) of the world they observe, thus generating knowledge on it (see Figure 1).

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.



Figure 1. Human beings produce models of the world they observe.

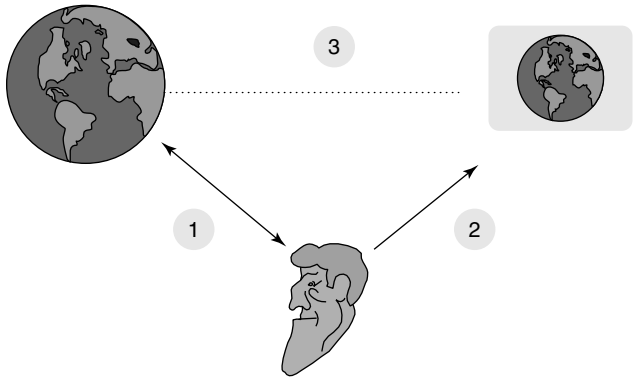


Figure 2. The relations among subjects, world, and models.

Such knowledge results from the relations among the three interacting entities, the subject, the world, and the model, so that the relation between the world and its model is not direct, but always mediated by the subject who produced the models themselves as shown in Figure 2.

Whenever it remains individual, knowledge is just tacit and usually implicit and as such it reduces to personal experience that can be communicated only by person-to-person means, as imitation.

Critical is therefore the *objectivity*, that is, the independence from the subject, of the relation between the world and its models. Realism assumes two operative reasons for justifying the possibility of some objective knowledge:

- *Intersubjectivity*: were knowledge only subjective, mutual understanding would be an exception more than a rule (Figure 3).

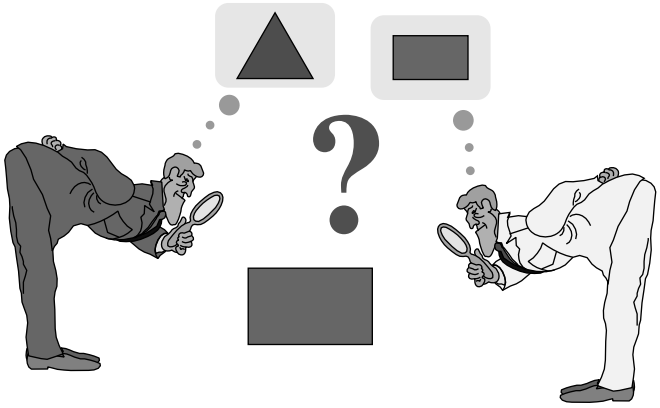


Figure 3. A justification for realism: mutual understanding.

- *Pragmatics*: were knowledge only subjective, our ability to effectively operate on the world would be an exception more than a rule (Figure 4).

Realism can be then interpreted as a weak form of objectivism: the world exists independently of us and both intersubjective and pragmatic experiences lead us to assume that it cannot be too different from our models of it.

On the other hand, to generate knowledge that can be shared, subjective models must be expressed in some socially understandable and usable form, such as statements in a natural language or mathematical laws. This points out a further basic issue on knowledge: 'how can it be



Figure 4. A justification for realism: effectiveness.

that mathematics, a product of human thought independent of experience, is so admirably adapted to the objects of reality'? in the words of Albert Einstein.

Philosophers and scientists have formulated different opinions in this regard, more or less explicitly in reference to a basic dichotomy: either 'scientific laws faithfully describe how the world is' or 'scientific laws are just synthetic means to express information about events in an aggregate way'. The former position implies a *metaphysical* hypothesis on the nature of the world, classically stated as 'numbers are in the world' (Kepler) or by assuming that 'the great book of nature' cannot be understood 'but by learning its language and knowing the characters in which it is written: it is written in mathematical terms' (Galileo); in contrast, the latter position suggests the *economic* nature of science: since 'in Nature the law of refraction does not exist at all, but only different cases of refraction', by means of such a law 'we do not have to keep in mind the countless phenomena of refraction in the various compositions of matter and under the various incidence angles, but only the rule that we call 'law of refraction', a much easier thing' (Mach).

Measurement has been often adopted to justify the former position.

3 SEMIOTICS OF KNOWLEDGE

Knowledge can be *about* the physical world but it is not *part of* it. Given the realistic assumption of the independence of the physical world from the subject, both subjective and objective knowledge can be interpreted in an evolutionary context as the results of mankind to adapt to his (firstly only physical and then also social) environment. In this regard, Karl Popper has suggestively proposed to identify 'some stages of the cosmic evolution' as organized in three 'worlds', as follows:

World 1

0. hydrogen and helium
1. heavier elements; liquids and crystals
2. living organisms

World 2

3. sensitivity (animal conscience)
4. conscience of self and death

World 3

5. human language; theories of self and death
6. products of art, technology, and science.

In this framework, knowledge (whose object can belong to either Worlds, and finally could even become knowledge itself...) is a rather advanced entity, appearing initially within World 2, in the form of subjective experiences, and then fully evolving in the context of World 3. The transition from World 2 to World 3 corresponds to the social ability to communicate, and therefore to share, experience: that is why the availability of a (textual or nontextual) language is considered the first step within World 3. Furthermore, the usage of a language gives knowledge a syntax and makes it a semiotic entity (*see also* **Article 23, Principles of Semiotics as Related to Measurement, Volume 1**).

Given the complexity of the concept of knowledge and its fuzzy characterization, rather than trying a definition of it we suggest that the (possible) presence and the relative importance of the semiotic components, syntax, semantics, and pragmatics, can be adopted as a criterion to distinguish among the different entities that are commonly deemed to be (related to) knowledge. In particular,

- the exclusive availability of *pragmatic* information ('to know how to do'), such as the competence shown by many craftsmen, appears to be a limited kind of knowledge, if knowledge at all;
- the exclusive availability of *syntactical* information and the ability of purely symbolic (i.e. only syntactical) processing, as performed by most automatic devices, appears to be a limited kind of knowledge, if knowledge at all.

The designation of 'knowledge-based' for the systems operating on the basis of an explicit semantics is a further argument to support the hypothesis that meanings are critical for the emergence of 'proper' knowledge, and therefore that socially communicable knowledge ('World 3 knowledge') is an entity spanning all the semiotic components.

4 PRAGMATIC CLASSIFICATION OF MODELS

If the pragmatic component is taken into account, different purposes for knowledge can be recognized: models can be adopted for description, explanation, prevision, prescription.

It is usual that the first stages of the development of a new field of knowledge are devoted to the production of models aimed at the *description* of the system under analysis. Typical outcomes of this work are the identification of properties relevant to describe the system and their evaluation to classify the system itself into more or less rough categories.

To overcome the conventionality of taxonomies and whenever the available knowledge allows it, some relations among properties are identified, so that each property is

embedded in a network of dependencies. In such cases, the relational information that is (explicitly or implicitly) conveyed by properties can be referred to in order to obtain an *explanation* of the system state/behavior: the value of the property x_1 is v_1 because x_1 is connected to the properties $x_2, \ldots, x_n$ by the relation R, and the properties $x_2, \ldots, x_n$ have values $v_2, \ldots, v_n$ respectively, and $R(x_1, \ldots, x_n)$.

Sometimes, models can be further enhanced to include relations connecting properties with an explicit functional time dependence, $\forall i = 1, \ldots, n, x_i = x_i(t)$, for example, in the form (known as *canonic representation*, or *local state transition* in System Theory):

$$\frac{\mathrm{d}x_i(t)}{\mathrm{d}t} = f_i(x_1(t), \ldots, x_n(t)) \tag{1}$$

for time-continuous models, and

$$x_i(t + \Delta t) = x_i(t) + f_i(x_1(t), \ldots, x_n(t))\Delta t \tag{2}$$

for time-discrete models. Models can then be used also for *prevision*, in particular if the integral/time-global versions of the canonic representations is taken into account:

$$x_i(t) = x_i(t_0) + \int_{t_0}^{t} f_i(x_1(\tau), \ldots, x_n(\tau))\,\mathrm{d}\tau \tag{3}$$

$$x_i(t_n) = x_i(t_0) + \sum_{j=0}^{n-1} f_i(x_1(t_j), \ldots, x_n(t_j))\Delta t \tag{4}$$

allowing to compute the system state $\langle x_1(t), \ldots, x_n(t)\rangle$ at a generic (future or past) time t from a reference, initial state $\langle x_1(t_0), \ldots, x_n(t_0)\rangle$ and by means of the state transitions $\langle f_1, \ldots, f_n\rangle$.

Finally, if an external intervention is possible on the system, its spontaneous dynamics can be controlled to let the system evolve toward a required target. In this case, models are then aimed at *prescription*: given a generalized version of the local state transition function including in its domain both the current state and the user input, models specify how to provide such an input, and therefore become decision-making tools (see Figure 5).

According to the traditional paradigm of science and its relations with technology, by repeatedly following this

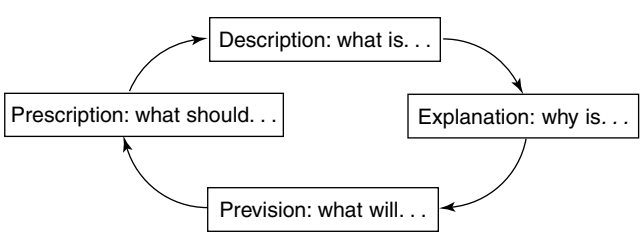


Figure 5. The knowledge loop among the four kinds of models.

knowledge loop, the quality of knowledge itself and the effectiveness of system control can always be enhanced.

On the other hand, in many situations prescriptions are required even when predictive, explanatory, and sometimes even socially agreed descriptive models are not available (let us quote the crucial examples of medicine and business administration). In these cases, experiences and expectations (i.e. World 2 knowledge) still play a critical role.

5 THE EVALUATION OF QUALITY OF KNOWLEDGE

Given the combination of subjectivity and objectivity so usually present in knowledge, it is not amazing that the evaluation of the quality of knowledge represents a basic issue in the process of acquisition of candidate items for their integration in an existing body of knowledge.

The quality of a model can only be evaluated in reference to the goals for which the model itself has been produced: the general criterion for this quality evaluation is therefore the *adequacy to goals*. Truth, traditionally thought of as 'correspondence to facts', is regarded as a specific case of adequacy, applicable whenever correspondence to facts is indeed considered an important issue (note how this position radically differs from the pragmatist definition of truth, according to which 'a sentence may be taken as a law of behavior in any environment containing certain characteristics; it will be "true" if the behavior leads to results that are satisfactory to the person concerned, and otherwise it will be "false"' (Russell)).

While adequacy is hardly the object of general treatment, the possibility of evaluating the truth of a model has been widely debated and is surely one of the most critical topics of the Philosophy of Science. Following Karl Popper, it can be suggested that the controversy is specifically related to two basic Problems:

1. how to compare (the statements of) competing models?
2. how to evaluate (the truth of) a model?

in reference to which three standpoints can be identified:

- *Verificationism*: typical of classical science and brought to its extreme consequences by the Neo-Positivistic school: Problem 2 admits a solution (and therefore the truth of a model can be determined), from which a solution to Problem 1 is derived: the reference to truth is the foundation allowing the advancement of science.
- *Falsificationism*: also called 'critical rationalism', as advocated by Popper himself: Problem 1 admits a solution (in the presence of competing models, the one that is chosen is the one that is not falsified and has the

greater empirical content), but a solution to Problem 2 cannot be derived from it: by means of conjectures and confutations, truth is approximated; the preference of a model over a competing one can be rationally motivated, but a model cannot be justified in itself.

- *Epistemic relativism*, also called 'irrationalism', supported by philosophers such as Thomas Kuhn and, in its extreme consequences, Paul Feyerabend: Problem 2 does not admit a solution ('the only principle that does not inhibit progress is anything goes. For example, we may use hypotheses that contradict well-confirmed theories and/or well-established experimental results. We may advance science by proceeding counter-inductively' (Feyerabend)), and therefore also Problem 1 cannot be solved: no criterion/method that is absolutely valid holds in scientific research.

6 DATA AND INFERENCE IN KNOWLEDGE

We get an insight into knowledge by considering its operational side of being a *faculty to solve problems*, and, in particular, to modify the state of systems according to given goals. As human beings, we constitutively have the ability to operate state transitions on the systems with which we interact by means of a 'World 2 strategy': we acquire data on the current state through our sensorial apparatus; by means of the brain, we perform inference on such data, and finally we use the data resulting from this process to drive our motor apparatus whose activity actually carries out the required state transition (these three steps correspond to the tripartite structure of the neural system: *sensorium*, brain, and *motorium*; note that more than 99% of the about 10^{10} neurons of human beings are part of the brain). In many cases, this strategy is manifestly both more efficient and more effective than a blind 'try-and-error' approach, although far more complex than it.

The same conditions, the availability of data and the ability to deal with them by means of inference, are also characteristic of the 'World 3 strategy' to problem solving, as shown in Figure 6.

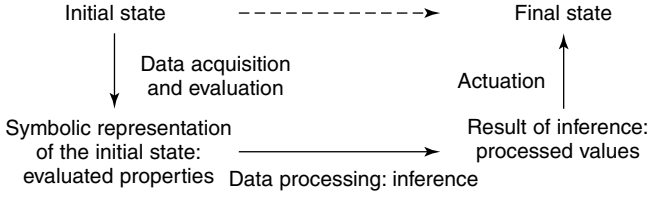


Figure 6. The 'World 3 strategy' to problem solving.

This scheme highlights the complementary role of data and inference in knowledge (*see also* **Article 10, Extensions of the Representational Theory of Measurement, Volume 1**).

Data, that is, evaluated properties, are aimed at being a *faithful representative* of the observed state, as obtained by either subjective or intersubjective and objective procedures, and can be expressed by means of either an informal or a formal language. The fundamental operation to empirically get formal data by means of an intersubjective and objective procedure is measurement: according to the representational point of view to measurement theory (*see also* **Article 8, Formal Theory of Measurement, Volume 1**), such a faithfulness is formalized by requiring that the mapping from empirical states to symbols be a homomorphism for the scale type in which states are measured; the existence of monomorphisms for the measurement scale type (i.e. admissible scale transformations) manifests the residual presence of conventionality in the selection of symbols. It is in this regard that one could wonder about *the truth* of symbols and the related sentences.

Inference is an operation aimed at obtaining new data 'conclusions' from the processing of the given inputs 'premises'. To understand the structure of inferential processes, the fundamental distinction between *singular* and *universal assertions* (sometimes called *facts* and *laws* respectively) must be taken into account. In set-theoretical terms, $a \in P$ (the element a belongs to the set P; the property P holds the element a) is singular, whereas $P \subset Q$ (P is a subset of Q; for all elements x, if the property P holds for x then also the property Q holds for it) is universal (it should be clear therefore that data obtained by means of measurement are singular).

Two kinds of inference are then traditionally considered, that in their simplest forms are as follows:

- From the singular $a \in P$ and the universal $P \subset Q$ by *deduction*, the singular $a \in Q$ is obtained; deduction is a truth-preserving inference that, strictly speaking, *does not lead to new knowledge*.
- From a collection of singular $a \in P$ and $a \in Q$ by *induction*, the universal $P \subset Q$ can be obtained; induction is a hypothetical inference that *does not lead to conclusive knowledge*.

The problem of foundation of empirical knowledge is traditionally ascribed to this circularity: deduction leads to true conclusions, but only if the truth of its (both singular and universal) premises can be assumed; induction is the only means to obtain new universal knowledge, but the truth of such knowledge cannot be definitely assumed.

7 INEXACTNESS OF KNOWLEDGE AND MEASUREMENT

We have already noted the relevance of language for World 3 knowledge: truth is a property of sentences (actually: of declarative ones), and 'science, though it seeks traits of reality independent of language, can neither get on without language nor aspire to linguistic neutrality. To some degree, nevertheless, the scientist can enhance objectivity and diminish the interference of language, by the very choice of language' (Quine). That is why formalization (i.e. the expression of knowledge in a form such that inferential processes can be entirely performed on the basis of the syntactical component of data) is often regarded as a critical requirement for scientific knowledge. On the other hand, formalized languages can be (and in many cases actually are) far too precise for expressing empirical knowledge: 'there are certain human activities which apparently have perfect sharpness. The realm of mathematics and of logic is such a realm, par excellence. Here we have yes-no sharpness. But (...) this yes-no sharpness is found only in the realm of things *we say*, as distinguished from the realm of things *we do*. (...) Nothing that happens in the laboratory corresponds to the statement that a given point is either on a given line or it is not' (Bridgman).

Hence, the same empirical knowledge can be expressed in sentences by balancing two basic components: *certainty* (a term for some aspects more general than truth) and *precision* (also called *specificity* or, at the opposite, vagueness). Therefore, 'all knowledge is more or less uncertain and more or less vague. These are, in a sense, opposing characters: vague knowledge has more likelihood of truth than precise knowledge, but is less useful. One of the aims of science is to increase precision without diminishing certainty' (Russell).

The fact that the length of the diagonal of a physical 1 m side square *cannot* be $\sqrt{2}$ m is an important consequence of metrological thinking: the information conveyed by real numbers (and the related concepts of continuity/differentiability) is too specific to be applicable, as is, to physical systems. By progressively enhancing the resolution of the measuring systems, and therefore by increasing the specificity of the measurement results, their uncertainty consequently grows, until the object of measurement itself becomes uncertain (in the previous example, at the atomic scale the concept of 'physical square' is meaningless), and an 'intrinsic uncertainty' (also called 'model uncertainty') is reached.

This reflects a basic feature of the relation that by means of knowledge it is established between World 1 (to which the object of knowledge belongs) and World 3 (to which the sentence that expresses knowledge belongs): if symbols are not generally so specific to univocally denote (properties of) things ($2 + 2 = 4$ holds for both apples and aircraft carriers), at the same time, things are too complex to be fully described by means of symbols.

8 (NON)FOUNDATIONS OF KNOWLEDGE

Philosophy of Knowledge (and Philosophy of Science, in particular) has always quested for a *foundation* of knowledge, that is, the elements on which the 'building of knowledge' can be firmly erected. In the course of history, such a foundation has been found in natural elements (for example, Thales of Miletus affirmed that the principle that causes all the things is water, while Heraclitus of Ephesus found it in fire), in physical or conceptual structures (atoms according to Democritus of Abdera, numbers in the conception of Pythagoric school), in metaphysical principles (such as the hypothesis that Nature is simple), in methodological assumptions (in particular, the postulation that any empirical knowledge cannot be derived but from sense data). The usage of the metaphor of foundations is not conceptually neutral: the architectural image of 'foundations' reveals the hypothesis that scientific research can make knowledge incrementally grow from its bases, where measurement has been traditionally recognized as the operation that is able to produce the objective data playing the role of such bases.

In the last decades, this confidence in the progressive development of scientific knowledge has been questioned by concentric objections, all emphasizing that definitive foundations are beyond the reach of the means human beings adopt to know. Complementary to the abovementioned philosophical positions of epistemic relativism (according to which raw sense data do not exist because data are always theory laden), an important area of scientific research is currently devoted to the systems that exhibit relevant *structural complexity*, a characteristic that makes such systems irreducible from the classical paradigm of reduction to simplicity through the hypotheses of linearity, principle of superposition of effects, and so on. Knowledge is recognized to be an always evolving process, where 'there is never an absolute beginning. We can never get back to the point where we can say, 'Here is the very beginning of logical structures' (Piaget). More than the actual availability of data, knowledge is recognized to be a potentiality (what is 'stored' in our brain is how to compute multiplications, not the results of operations such as 1234×5678), information is always under reconfiguration.

The role assigned to measurement is paradigmatic of the shift toward what could be called *reticular* (and therefore without foundations) knowledge. Indeed, according to the current standpoints of philosophy of measurement:

- Since measurement results depend on standards through a traceability chain, standards themselves could be thought of as 'realizations of true values,' than playing the role of actual foundations for measurement; on the other hand, standards must be indeed 'realized' by primary laboratories, who maintain their quality by means of interlaboratory comparisons: therefore, this claimed 'path towards foundations' cannot but include a component of conventionality.
- Any measurement result depends for its evaluation on the previous measurement of a set of influence quantities, in their turn being new measurands so that in principle such a dependence should be recursively applied, with the consequence that a 'well-founded measurement' would be impossible to be completed; the usual operative choice to assume that the quantities influencing the initial measurand are not influenced by other quantities, and therefore, that they can be 'directly measured', highlights the conceptual approximation inherent in any measurement.
- While the adequacy of empirical models is controlled by means of measurement, the quality of measurement results depends on the quality of mathematical models used for designing measuring systems.

This complexity makes knowledge the most versatile tool available to human beings and a fascinating object for knowledge itself.

RELATED ARTICLES

Article 13, Signals, Information and Knowledge, and Meaning, Volume 1; **Article 23, Principles of Semiotics as Related to Measurement, Volume 1**; **Article 24, Principles of Epistemology as Related to Measurement, Volume 1**; **Article 39, Characteristics of Data, Information, Knowledge, and Wisdom, Volume 1**; **Article 63, Outline of Systems Thinking, Volume 2**.

FURTHER READING

The scientific literature devoted to this topic is so wide that any selection of items is partial. What follows is a (short and subjectively selected) list of general references to the issues covered here.

Hacking, I. (1983) *Representing and Intervening*, Cambridge University Press, Cambridge.

Morin, E. (1986) *La connaissance de la connaissance*, Vol. III of *La méthode*, Seuil, Paris.

Piaget, J. (1970) *Genetic Epistemology*, Columbia University Press, New York.

Popper, K.R. (1994) *Knowledge and the Body-Mind Problem – In Defence of Interaction*, Routledge, London.

23: Principles of Semiotics as Related to Measurement

Luca P. Mari
Università Cattaneo, Castellanza (VA), Italy

1 Signs as Entities that 'Stand for' Something 134
2 Coding and Decoding 135
3 Conventionality of Signs 136
4 The Opposition Analog/Digital in a Semiotic Perspective 136
5 The Systemic Nature of Signs 137
6 Syntax, Semantics, and Pragmatics 137
7 Semiotics and Communication 138
8 Applying Semiotic Principles to Measurement Systems 139
Related Articles 139
Further Reading 139

1 SIGNS AS ENTITIES THAT 'STAND FOR' SOMETHING

Semiotics is commonly defined as the *doctrine of signs*, a sign being 'something which stands to somebody for something in some respect or capacity' according to Charles S. Peirce, one of the seminal thinkers about semiotics itself. The emphasis is here on the relation of 'standing for', which, in its simplest form, can be modeled as follows.

When a purpose is assigned to, or recognized proper of, things, they can be evaluated in their ability to satisfy it. Any given purpose induces a relation of *functional substitutability* S on the set of considered things T such that $\forall x, y \in T$, $S(x, y)$ if and only if x is a substitute of y with respect to the purpose, that is, x is as able as y to satisfy the purpose itself.

Instead of investigating here the general properties of the relation S (but at least the observation should be made that in many cases functional substitutability is not a matter of a yes–no alternative, and therefore that S could be usually formalized as a fuzzy relation: *see also* **Article 12, Fuzzy Approaches for Measurement, Volume 1**), let us devote our attention to the specific relation of *identification*. The functional substitution implied in the identification is such that x identifies a given y if and only if x operates as the selector of y in a set of candidate things $y_1, y_2, \ldots$ the exhibition of x being considered functionally equivalent to the selection of y and the nonselection of any other $y_i \neq y$ in the candidate set. In such a relation $S_{id}(x, y)$ let us call x and y the *identifier* and the *identified entity* respectively: x *stands for* y. For example, the sound c-h-a-i-r (a physical thing as a space-time event produced by the utterance of a speaker), could be the identifier chosen to select a chair instead of any other nonchair object (note that no restrictions have been imposed on the set of the entities that are the object of identification: y can be a physical object but also an informational entity).

It is a common observation that different $x_1, x_2, \ldots$ can be adopted as identifiers for the same entity y, $S_{id}(x_1, y)$, $S_{id}(x_2, y), \ldots$ (a chair can be identified by different sounds, possibly pronounced by different persons in different languages in different times, but also by writings, drawings, gestures, ...). In this case, the $x_1, x_2, \ldots$ are functionally substitutable with each other in their role of identifiers for y, and therefore a derived relation S_{id-y} holds among them. Formally (we will continue to forget the fact that also S_{id-y} could be fuzzy) $S_{id-y}(x_1, x_2)$ if and only if $S_{id}(x_1, y) \wedge S_{id}(x_2, y)$.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

A fundamental step is taken when the *class* $\hat{x}$ of all the identifiers x for which the relation S_{id-y} holds is abstractly considered as the identifying entity for y, thus recognizing that signs, although instantiated in physical things, are *information entities* (*see also* **Article 22, Characteristics and Theory of Knowledge, Volume 1**).

2 CODING AND DECODING

Signs generally result from the relation between two elements: identifiers and identified entities. Such a relation is operatively realized and performed in two phases:

- for a given entity y to be identified, an identifier is obtained by means of an operation of *coding*: an information entity $\hat{x}$ is at first associated with y, and then an identifier x is selected such that $x \in \hat{x}$; for example, see Figure 1,
- for a given identifier x, the identified entity is obtained by means of an operation of *decoding*: an information entity $\hat{x}$ is at first identified, by means of a pattern recognition, as the class to which x belongs, and then an entity y is selected as associated with $\hat{x}$; for example, see Figure 2, (*see also* **Article 33, Coding Theory and its Application to Measurement, Volume 1**).

Figures 1 and 2 are instances of a more general 'semiotic triangle', in which the 'standing for' relation is represented as in Figure 3: where the dotted line expresses that the relation is indirect and a 'mediator' is usually required to connect identifiers and identified entities.

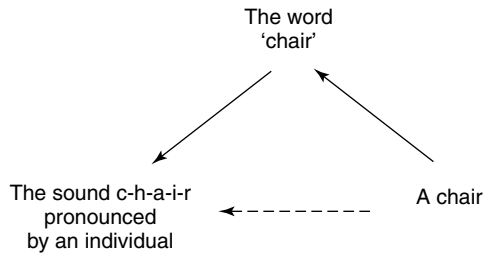


Figure 1. An exemplification of the structure of a coding operation.

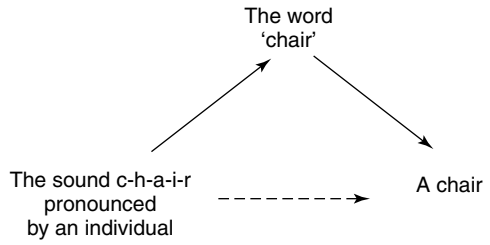


Figure 2. An exemplification of the structure of a decoding operation.

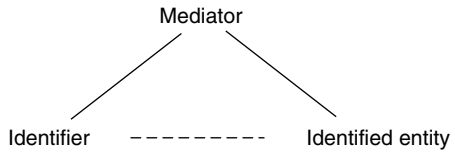


Figure 3. The 'semiotic triangle'.

In the history of Semiotics, diagrams of this kind have been widely adopted to present and generalize the relation that we have introduced as between identifiers and identified entities. For example, de Saussure defined it in terms of 'signifiers' and 'signifieds' (and called 'signification' the relation itself), while Hjelmslev used the terms 'expression' and 'content' respectively.

These diversities witness the different interpretations and emphasis put on the elements of the relation. For example, the mediator has been thought of as either the *sense* of the identifier (chairs are identified by means of the term 'chair' because of the meaning associated with such a term, that is, the set of features that are shared by everything to which the term applies, the so-called *intension* of the term), or the *set of the entities*, the entities to what the identifier stands for (the so-called *extension* of the term), or the *subject* with the competence to maintain the relation, or the *social context* of the individuals who agreed to associate the identifier with the identified entity.

In the case in which the entity to be identified belongs to the physical world (let us mention again that it could be a purely informational entity, as in the case when one is talking about words or numbers), a common, although surely not necessary, situation is such that the relation between the physical thing chosen as identifier and the identified entity is mediated by *two* informational entities, for example, (see Figure 4) so that each arrow in Figure 5 represents a partial realization of the 'standing for' relation.

Semiotics has been in particular working on the *informational component* of the 'standing for' relation, thereby often minimizing, or even neglecting, the analysis of the relation between the symbols and the physical things adopted as support for them. From now on we will accept

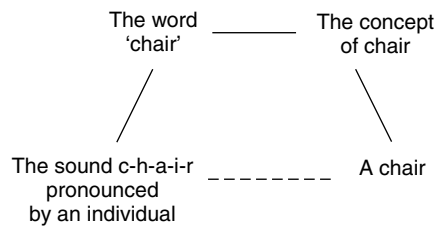


Figure 4. The structure of the semiotic relations.

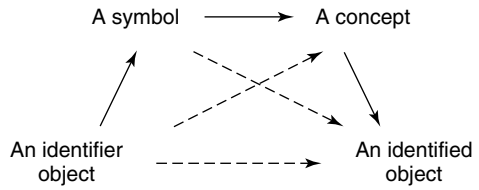


Figure 5. The 'standing for' relations.

this general standpoint, and follow the terminology proposed by Ogden and Richards who describe the 'standing for' relation in terms of *symbols* that stand for *referents* (*see also* **Article 8, Formal Theory of Measurement, Volume 1**).

3 CONVENTIONALITY OF SIGNS

The 'standing for' relation is a complex one: the same symbol could stand for different referents, and different symbols could stand for the same referent (e.g. in the case of the linguistic phenomena of *polysemy* and *synonymy* respectively). This suggests that such a relation is not inherent to the entities involved in it: an entity becomes a symbol only as the result of a choice.

The issue of the (degree of) arbitrariness of signs has been thoroughly inquired by many philosophers, who noticed its fundamental implications in terms of autonomy of symbols in relation to referents and therefore, generally speaking, of language in relation to reality. For example, in Plato's Cratylus, the problem of 'right names' for things is discussed, and it is concluded that 'whatever name you give to a thing is its right name; and if you give up that name and change it for another, the later name is no less correct than the earlier, just as we change the name of our servants; for I think no name belongs to a particular thing by nature'.

Following Peirce, signs are usually distinguished in three broad categories, characterized by their increasing degree in conventionality:

- *Indexes*, such as 'natural signs' (smoke standing for a yet unseen fire) and 'signals' (a phone ringing standing for a waiting call), for which the symbol is causally connected to the referent, so that every subject informed on the connection is able to infer the existence of the relation.
- *Icons*, such as images or onomatopoeic words, for which the relation symbol-referent is based on some mutual resemblance or imitation of the related entities; in this case, the relation can be intensively learned (e.g. the higher the sound the angrier the speaker) and is easily, while often implicitly, shared among cultural communities.
- *Symbols* (in specific sense), such as those constituting many elements of textual languages, for which the relation symbol-referent is purely conventional (obtained sometimes by an explicit ruling convention, sometimes by usage), so that it must be learned according to an extensive strategy, that is, by explicitly listing the symbol-referent pairs.

4 THE OPPOSITION ANALOG/DIGITAL IN A SEMIOTIC PERSPECTIVE

The latter two categories, icons and symbols, can be meaningfully expressed in terms of the *opposition between analog and digital*, as traced back to the concept of *structure* modeled and formalized in Measurement Theory (*see also* **Article 8, Formal Theory of Measurement, Volume 1**). The opposition A/D concerns the strategy adopted for *coding and decoding the metainformation* that complements the information that symbols convey on referents. Indeed, together with the information enabling the selection of referents, in many cases, some structural information must be maintained in coding and recognized in decoding. For example, if the referent is a grade in $\{A, \ldots, E\}$ the observation of a physical support coding the symbol 'B' should bring both the information 'is B' (and 'is not A, and not C, ...') and the (ordinal) metainformation 'is less than A but more than C, ...'. Therefore:

- *analog* is the strategy by which the metainformation is coded in the support configuration, so that both coding and decoding correspond to the application of a homomorphic (i.e. structure preserving) mapping;
- *digital* is the strategy by which the metainformation is maintained in the coding rule, so that the physical support is only required to be able to assume at least two distinguishable configurations (those usually symbolized as '0' and '1'), as specified in Shannon's Theory of Information.

This characterization accounts for the nature of *opposition* of the strategies of analog and digital coding:

- The definition of the code rule can be intensive in analog cases, whereas it must be extensive in digital cases: while analog codes can be analytically defined, the lack of structure forces digital codes to be defined by explicitly and completely listing the occurrences symbol-referent.
- The set of information entities to code can be incompletely preidentified in analog cases, whereas it must be

preidentified in digital cases: for example, to increase the cardinality of the set of the possible symbols to code on a physical support, an analog code can be adopted as is, whereas a digital code must be redefined.

On the other hand;

- analog coding can be adopted only if some metainformation is available, whereas digital coding is always available for finite sets of symbols;
- supports adopted in coding must be able to maintain the metainformation in physically distinguishable configurations in analog cases, whereas it can be very simple since only two distinguishable configurations are in principle required in digital cases.

It should be noted that mixed (partly analog, partly digital) coding rules are common, as in the case of the usual numerical notation, in which the single digits are digitally coded while the positional rule is analog.

5 THE SYSTEMIC NATURE OF SIGNS

The 'standing for' relations are seldom defined as single and independent coding rules. Rather, their conventionality is considerably restricted by the effects derived by their *systematic definition* (as an example, consider the possibility of reconstructing the meaning of a linguistic term, that is, 'decoding' it, by means of its etymological analysis).

Natural languages, such as English or Italian, are far more complex than artificial languages, such as the formalism of mathematical logic or computer programming languages, also because they include a huge amount of exceptions, that is, irregularities, in their coding and decoding rules. Nevertheless, the fact that some systematic effects progressively emerge from historical usage instead of explicit decision, as indeed in the case of natural languages, does not reduce their relevance but only the uniformity of the system of rules.

This systemic component was called *language* (*langue*, in French) by Saussure, who contrasted it with *speech* (*parole*, in French), regarded as the individual act of selection and actualization of symbols that stand for intended referents by means of some coding rules of the language. Any specific film would be therefore the 'speech' of the 'language' of cinema, an example highlighting that *langue*/*parole* is actually the dichotomy code/instance or schema/usage, as Hjelmslev termed it. 'Each of these two terms achieves its full definition only in the dialectical process which unites one to the other: there is no language without speech, and no speech outside language: it is in this exchange that the real linguistic praxis is situated' (Barthes).

Dialectical is also the process by which the decoding of composite structures of symbols, for example, sentences, is performed: paradigmatically, whenever coding rules are *context-sensitive* (a typical characteristic of natural languages), not only can the meaning of a sentence be derived from the meaning of its constituting parts, but the meaning of such parts could also depend on their role in the sentence, so that it can be determined only after some hypothesis of the meaning of the whole sentence itself. This generally makes the recognition of the 'standing for' relations a complex, recursive process, and again this explains why the constructs of artificial languages are defined as context-free whenever recognized as adequate (as an example of the role of context in rule evaluation, consider two possible definitions of the disjunction operator OR: if $v(x)$ is the truth value of the sentence x, in classical logic, such an operator is context-free, since $v(x \vee y) = f(v(x), v(y)) = \max(v(x), v(y))$; on the other hand, in the case of probabilistic logic $v(x \vee y) = v(x) + v(y) - v(x \wedge y)$ and therefore the operator is context-sensitive, because $v(x \vee y) \neq f(v(x), v(y))$, with the term $v(x \wedge y)$ playing the role of context).

6 SYNTAX, SEMANTICS, AND PRAGMATICS

A fundamental classification to isolate the different contributions to the complexity of the 'standing for' relation was proposed by Charles W. Morris, who suggested three basic subdisciplines as the constituting components of Semiotics: *syntactics* (also, and more commonly, called *syntax*), *semantics*, and *pragmatics*. Despite their large reciprocal autonomy, such disciplines can be presented in terms of progressive extension of scope:

- *Syntactic* is the information dealt with as *data*, taking into account the collection of available signs and its structure; a basic issue of syntax is *parsing*, that is, the check that a sentence is well formed according to a given set of syntactical rules.
- *Semantic* is the information dealt with as data provided *with meaning*, taking into account (also) the entities to what the signs stand for; a basic issue of semantics is *truth evaluation*, that is, the check of the correspondence between the content of a sentence and the actual state of the reality (it should be noted that the evaluation of the truth of a sentence does not always require the interpretation, that is, the 'semantization' of the sentence itself; for example, tautologies in propositional logic (e.g. $A \vee \neg A$) are true for any interpretation of A; this establishes a distinction between *linguistic* and *empirical truth*).

- *Pragmatic* is the information dealt with as data provided with meaning *and value*, taking into account (also) the relation of signs to interpreters and therefore issues related to behaviors, subjective interests, utilities, . . .; a basic issue of pragmatics is *relevance assessment*, that is, the check that a sentence is actually useful for its deemed receivers.

The distinction among these disciplines and their goals can be exemplified by means of that particular system of signs that is mathematics:

- The formula '$2)\sqrt{(+ \leq)x}$' is not well formed, so that a fortiori neither its meaning and truth nor its utility for a given subject can be evaluated.
- The formula '$2 + 3 = 4$' is well formed, its meaning can be evaluated, and it is actually false in the usual interpretation of its constituting signs.
- The formula '$1 = 1$' is well formed and it is true, but plausibly useless for most subjects.

Given the centrality of the 'standing for' relation, semantics can be considered the core component of semiotics. Nevertheless, the threshold between syntax and semantics is not always well defined, and often actually a matter of 'point of view'. For example, in the case of Morse code, the physical support is an electric current, whose patterns are interpreted as sequences of 'dots' and 'dashes', whose patterns are in their turn interpreted as sequences of alphanumeric characters; the sequence 'dot-dash' is then a semantic entity with respect to an electric signal but it is a syntactical entity with respect to its deemed interpretation, that is, the character 'a'.

7 SEMIOTICS AND COMMUNICATION

An important area of application of Semiotics is *communication*, that is, the transfer of messages conveying some sort of information from a sender to a receiver through a channel, (see Figure 6). The standard model for a basic communication system has been defined by Shannon: the message generated by the sender is coded into a pattern of signs, here called a *signal*, that is transmitted by the channel and finally decoded again into a message for making it accessible to the receiver.

As formalized by Shannon, the communication problem – how to maximize the probability that the received message is the same as the one generated by the sender even in the presence of a noisy channel – specifically relates to syntax. The general semiotic nature of the problem has been shown by Roman Jakobson, who suggested that each of the six components of a communication system shown in Figure 7: is associated with a specific *function*, more or less present and emphasized in each communication act. Indeed, if the prominence is usually given to the referent (i.e. the 'content' of the message, corresponding to the so-called *referential*, or *cognitive*, *function*), in many messages, the role of the other functions must also be taken into account to fully understand the communication act itself. According to Jakobson, a communication can differently stress:

- the position of the sender (*emotive function*) on the communication subject (e.g. in terms of rage or irony);
- the orientation toward the receiver (*conative function*), as typically in the case of imperative sentences, whose aim is indeed to convey commands to receivers;
- the role of the channel (*phatic function*), whenever a support to the management of the communication itself is required, for example, to check whether the channel is still operative between the sender and the receiver (e.g. 'are you still there'?);
- the formal structure of the message itself (*poetic function*), for example, when homophonies, rhymes, . . . are adopted;
- the requirements on the code (*metalinguistic function*), typically whenever the sender and the receiver want to check whether they are using the same coding rules (e.g. 'what do you mean'?).

The semantic component is critical in communications with a prevailing referential function, and the problem of the truth evaluation of their contents can be generally posed. The other functions are instead oriented to the pragmatics of the communication: messages such as 'fantastic', or 'excuse me', or 'repeat please', or . . . are conveyed to obtain some communicational aim more than to state a meaning.

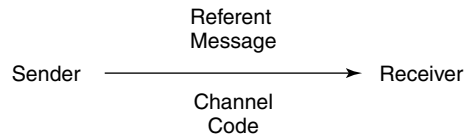


Figure 7. The components of a communication system.

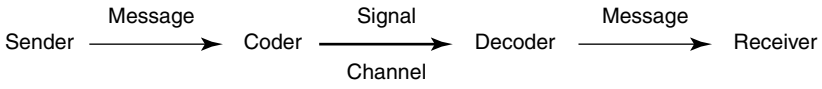


Figure 6. The structure of a communication system.

8 APPLYING SEMIOTIC PRINCIPLES TO MEASUREMENT SYSTEMS

As an operation aimed at expressing in symbols the information empirically obtained on a system about a quantity, measurement can be meaningfully analyzed in its semiotic character. In comparison with other forms of judgment, there are two general grounds of peculiarity for measurement:

- the mediator between the referent (i.e. the measurand) and the symbol (i.e. the measurement result) is an empirical entity, external to both the measured thing and the measurer subject: the measurement system;
- the symbols adopted as identifiers for the measurands are chosen in a formal language, whose composition and structure are explicitly known.

While for a general semiotic system only the syntactical component can be usually formalized (the attempt was made by logics and philosophers of science such as Bar-Hillel, Carnap, and Hintikka to quantify the semantic information conveyed by propositions as their 'quantity of content': with such a broad connotation, the problem remained largely unsolved and was substantially left aside), these characteristics of measurement allow to consider it in some more specific terms:

- *From the syntactical point of view*: measurement can be thought of as an operation of selection of a symbol from a set, the actual granularity of such a set (as usually formalized in terms of either number of significant digits or expanded uncertainty) depending on the resolution of the sensing device; the usual Shannon's concept of quantity of information can be adopted in this case, such that the quantity of information conveyed by a measurement result increases as its uncertainty decreases.
- *From the semantic point of view*: measurands are always evaluated relative to a reference, that is, explicitly reported in measurement results in terms of a measurement scale (and specifically measurement unit whenever applicable) and that expresses the actual meaning for the (usually) numerical symbols by which the measurand is quantified. Each measurement scale is characterized by a type, the most common scale types (nominal, ordinal, interval, ratio, absolute) being linearly ordered according to the algebraic structure they imply on the symbol set. The degree of semantic information conveyed by a measurement result depends thus on the degree of richness in algebraic structure of its scale type (formally, the semantic information increases as the class of admissible transformations for the scale type becomes more specific).
- *From the pragmatic point of view*: because of the existence of functional relations connecting them (the typical case of physical laws), measurands are embedded in a network of pragmatic information allowing new measurement results to be obtained by computation, that is, by derived measurement. While a completely disconnected measurand can be defined in a totally arbitrary way, and therefore its evaluation is pragmatically useless, the more the measurand is connected (e.g. the greater the number of functional relations in which it is present), the higher is the degree of pragmatic information conveyed by its values.

Even this summary presentation shows how the semiotic perspective can be useful to understand some fundamental characteristics of measurement (uncertainty, scale types, and derived measurement) in a general conceptual framework.

RELATED ARTICLES

Article 22, Characteristics and Theory of Knowledge, Volume 1; **Article 24, Principles of Epistemology as Related to Measurement, Volume 1**; **Article 25, Introduction to Signals in Physical Systems, Volume 1**; **Article 63, Outline of Systems Thinking, Volume 2**.

FURTHER READING

Chandler, D. (2001) *Semiotics: The Basics*, Routledge, London.

Devlin, K. (1991) *Logic and Information*, Cambridge University Press, Cambridge.

Eco, U. (1976) *A Theory of Semiotics*, Indiana University Press, Bloomington, IL.

Kyburg, H.E. (1984) *Theory and Measurement*, Cambridge University Press, Cambridge.

Mari, L. (1999) Notes Towards a Qualitative Analysis of Information in Measurement Results. *Measurement*, **25**(3), 183–192.

24: Principles of Epistemology as Related to Measurement

Timothy Lindsay John Ferris
University of South Australia, Mawson Lakes, South Australia, Australia

1 Definition of Epistemology 140
2 Explanation of Epistemology 140
3 Relevance of Epistemology to Measurement 141
4 Applying Epistemological Principles to Measurement Systems 142
5 Examples of the Use of Epistemology in Measurement Systems 142
References 143

1 DEFINITION OF EPISTEMOLOGY

Epistemology is the branch of philosophy that is concerned with the study of how knowledge is possible. The basic problem of epistemology is that, in common language, people make claims to know certain things, and some basis for determination of whether the claims are reasonable or appropriate is necessary in order to ensure that any action taken consequent upon the belief that one knows the matter is appropriately founded.

The analysis of knowledge requires distinction between that which is known and the knowledge of it. That which is known is either a tangible thing or phenomenon that is observable in nature, or an idea that belongs as part of an intellectual construction. The knowledge of the matter is distinct from the matter in that it is a representation of that which is known without actually being that which is known. In the case of the matter being an idea, there is still a distinction between the idea and knowledge of the idea in that the idea is an existential part of the person and experience of the person who has the idea, whereas knowledge of the idea has the self-consciousness of enabling the knower to be aware of what the idea is, and of the fact that they know the idea. Thus, knowledge is always a representation of what is known and is never what is known in and of itself.

The relation of things and their representation in knowledge leads us to the central problem of epistemology, expressed by the question: 'How do I know that reality is what I conceive and assert it to be in cognitive experience?' (Russell, 1898). Epistemology aims at identifying the relation between matters in reality and the subject's consciousness of these matters (Perry, 1931), and so it explores the problem of how judgments joining various pieces of information may be made. Epistemology, in relation to measurement, is particularly concerned with the issue of how empirical information can be used to form knowledge, that is, how it is possible to make synthetic, as distinct from analytic, judgments (Rescher, 1980). Kant distinguished analytic judgments, judgments relying only on pure reason and *a priori* information with synthetic judgments, which rely on *a posteriori*, empirical information.

Epistemology develops theories of truth, evidence (Hall, 1943), meaning, and verification (Carnap, 1936) in order to clarify each of these concepts and to make clear when it is reasonable to trust information that one has received, and when not to trust information.

2 EXPLANATION OF EPISTEMOLOGY

The classical account of knowledge states that S knows that p if and only if

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

1. p is true;
2. S believes that p; and
3. S is justified in believing that p (Gettier, 1963).

The classical account of knowledge has three fundamental elements: truth, belief, and justification, and the only serious challenge to this understanding has been that which is raised in Gettier (1963). The classical account of knowledge is always satisfied in the case of matters known analytically. The classical account was developed to determine when matters known synthetically could be asserted as known.

A theory of truth develops an understanding of what it means to assert 'p is true'. The assertion that p is true must be explained so that it is clear. In brief, 'true' means that something is as it is represented to be in the statement that is described as 'true'. The assertion that p is true is thus an assertion of correspondence between what is the case of the matter and how the matter is represented in the information that is the subject of the assertion of truth.

Knowledge 'that p' requires that the subject, S, believes that the matter corresponds to the representation embodied in p. 'Believe' is a verb with several distinct uses, including 'believe that p', which pertains to the mental assent that p is true. Mental assent does not require commitment to the truth of that which is believed, but it does require that the believer intellectually accepts that the assertion believed is in fact true and that there is no confusion or error introduced by any factor such as may introduce any distortion of perception of the observed.

The requirement for justification in the classical account of knowledge demands a theory of evidence as the basis for determining those signs that are something other than the signs themselves and that are necessary to determine that which is signified as true. Evidence may be directly related to the matter and explicit evidence of something, or it may only indicate what is claimed, or it may be circumstantial, resulting in the matter being concluded as probable.

The physical sciences assume that there is a physical world, which is fundamentally and substantially as it appears to be. On some occasions, what is seen is not real, but the cause of this difference is distorting circumstances that result in seeing something different than what is present. The physical sciences assume that such distorting observational circumstances can be explained through the physical phenomena of the world, and that such illusions can be overcome.

3 RELEVANCE OF EPISTEMOLOGY TO MEASUREMENT

There are many definitions of measurement in the literature:

1. 'Measurement is the process of assigning numbers to represent qualities: the object of measurement is to enable the powerful weapon of mathematical analysis to be applied to the subject matter of science' (Campbell, 1920).
2. 'Measuring in a narrow sense (measurement) is the experimental comparison of a measured quantity with a metrological standard. Measuring in a wide sense (classification) is the experimental comparison of a measured object with a particular standard (pattern)' (Hofmann, 1983).
3. Measurement is 'the quantitative determination of a physical magnitude by comparison with a fixed magnitude adopted as the standard, or by means of a calibrated instrument' (Sydenham, 1976).
4. 'Measurement is the assignment of numbers or other symbols by an objective, empirical process to attributes of objects or events of the real world in such a way as to describe them' (Finkelstein, 1990).

Each of these definitions of measurement emphasizes that measurement is a process of observation of the real world that results in representation of the observed in the structured framework of the scale used for the investigation. The intention of measurement is to convert observation of the world into a form amenable to the application of the analytical processes of mathematics. The intention of conversion from synthetic knowledge to information that can be operated upon in an analytic manner is a major leap that demands the analysis of epistemology. The reason is that the results obtained from the analytic processes obtain a status, in ordinary thought, linked to their derivation by an analytic process, rather than the uncertainty associated with immediate recognition of the empirical nature of the observation process.

Epistemology emphasizes the limitations that are placed on claims to knowledge by the method and circumstances of the gathering of the information concerning the matter. Thus, epistemology encourages skepticism concerning whether what appears to be the case is actually the case. In the measurement situation, this subject matter of epistemology concerns the development of understanding of the possible sources and kinds of erroneous measures that may be obtained. This interest is at the level of verification of the data in the measurement system.

Epistemology is also of relevance in the theory of measurement in developing awareness that the ultimate purpose of measurement is to obtain understanding of the things that are observed. This awareness prompts the understanding that measurement concerns the development of knowledge about the observed and forms the basis for wise action based, in part, on knowledge of the observed, and on other knowledge.

4 APPLYING EPISTEMOLOGICAL PRINCIPLES TO MEASUREMENT SYSTEMS

Measurement systems apply instruments to systems that are considered interesting. This situation produces an interface between the observed and the instrument, which is an information boundary across which the measurement information must flow. The development of measurement processes has respected the skepticism concerning the quality of information as a basis for knowledge presented in epistemology. Traditional concerns in the development of measurement process theory have included the issues of accuracy and calibration. The ideas of both accuracy and calibration address different issues associated with the perspective on measurement provided through epistemology.

Accuracy: The definitions of measurement of Sydenham (1976) and Hofmann (1983), Section 3, above, involve comparison of the observed manifestation with some kind of scale or standard. If a measurement were entirely accurate, there would be an exact correlation between the real manifestation and the measure of it obtained. The discussion of accuracy is premised on the idea that there is a 'correct value' of the manifestation in the observed, and that that correct value is present to be obtained, should one have sufficiently good means to measure it. It is not possible to prove that there is such a thing as one true value of things.

Inaccuracy: The assumption that there is a true value arises from a variety of effects including influence effects, noise effects, instrument imperfections such as hysteresis or wear, error in implementation of the physical effect of the sensor system, and loading of the observed by the sensor system. Influence effects result from the sensitivity of the sensor to effects other than the target effect. Noise relates to the presence of random events in the observed and sensor system affecting the output. Instrument imperfections result in the instrument systematically producing erroneous results, either in a simple relationship to true value or in a relationship dependent on the history of the instrument. Error in the implementation of the system results in all output being misleading because it is a measure of something other than what was intended. Loading of the observed by the sensor is a particular class of poor design, in which the combined system of the sensor and the observed results in a different manifestation being present than would be present if the instrument were removed.

Calibration is the simplest approach to the accuracy problem. A set of measurements of known values of the manifestation is performed with the instrument, and results are recorded, and used to provide a conversion between instrument output and true value. Calibration can remove inaccuracy related to the instrument characteristics, but cannot remove inaccuracy arising from noise or influence factors or loading effects different than those at the time of calibration.

Awareness of epistemology has led to investigation of the issue of accuracy, and the development of the concept of calibration, to address the epistemological problems of illusion and hallucination. Nicholas (1994) has a discussion of several specific examples of these effects.

5 EXAMPLES OF THE USE OF EPISTEMOLOGY IN MEASUREMENT SYSTEMS

Section 4 discussed the application of epistemology to measurement systems at the data level. In this section, the application of epistemology to measurement systems at the levels of knowledge and wisdom is discussed.

Sacerdoti, Paggi and Sacerdoti (1991) describe measurement processes as 'the means by which man dialogues with the external world in order to enrich his knowledge. The processes pass through dialectic counterreacted phases, that is, 'historical phase', during which the 'receiver' learns to recognize forms and situations; organization of the experimental situations on the basis of results; representation of the results'. This description of measurement expresses the idea that measurement concerns the gaining of knowledge about the observed.

Sacerdoti, Paggi and Sacerdoti (1991) provides a frame leading to the Object Oriented Model of Measurement (OOMM) of Ferris (1997). The OOMM provides a structure for the analysis of measurement involving a class of entity, which is characterized as having relations into which it enters, attributes that it manifests, an attribute observation process, and the belief that repeatable and verifiable measurements are possible. The model of the class then controls the kind of measurements that are applied to the observed and the interpretation of the results, leading to the creation of knowledge. The major problem is that nothing in the structure enables one to determine that the observed is not a member of the class to which it is assumed to belong.

An example of the application of epistemology to measurement at the level of knowledge and wisdom is found in the testing and evaluation of complex products and systems. The design and development of product requires measurement of the prototypes of the product by which it is determined whether the product can perform its function.

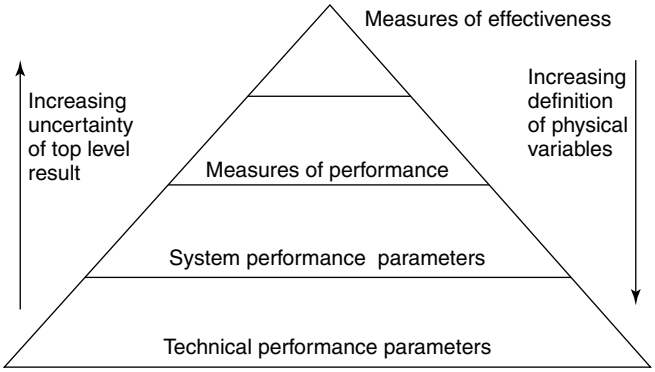


Figure 1. The pyramid of measures involved in the measurement of a product or system intended to perform some function. (Adapted from Crouch, V. and Sydenham, P. (1994) Relationship between T&E and M&I, *3rd Australasian Instrumentation and Measurement Conference*, 26–29 April, Adelaide, Institution of Engineers Australia, Barton (pp. 1–6).)

The test and evaluation measurement scenario is expressed in Figure 1.

In Figure 1, the technical performance measures refer to the measures of physical quantities of fundamental kinds of manifestation such as length, mass, temperature, and time, expressed in terms of standard units. The system performance parameters describe system performance in terms of derived units, which combine basis units, and describe aspects of the performance of the whole systems, such as velocity, or power consumption. Measures of performance concern the performance of the system in performing its design tasks, and include measures such as reliability, and maintainability. Measures of effectiveness concern whether the system satisfies the need for which it was conceived and developed.

The pyramid of Figure 1 shows a progression from the lowest level of data, through information, knowledge to wisdom, at each higher level in the pyramid.

REFERENCES

Campbell, N.R. (1920) *Physics The Elements*, Cambridge University, Cambridge.

Carnap, R. (1936) Testability and Meaning. *Philosophy of Science*, **3**(4), 419–471.

Crouch, V. and Sydenham, P. (1994) Relationship between T&E and M&I, *3rd Australasian Instrumentation and Measurement Conference*, 26–29 April, Adelaide, Institution of Engineers Australia, Barton (pp. 1–6).

Ferris, T.L.J. (1997) The Concept of Leap in Measurement Interpretation. *Measurement Journal of the International Measurement Confederation*, **21**(4), 137–146.

Finkelstein, L. (1990) Measurement and Instrumentation as a Discipline in the Framework of Information Technology, *Knowledge Based Measurement – Application, Research and Education*, IMEKO Secretariat, Karlsruhe (pp. 257–263).

Gettier, E.L. (1963) Is Justified True Belief Knowledge? *Analysis*, **23**, 121–123.

Hall, E.W. (1943) Perception as Fact and as Knowledge. *The Philosophical Review*, **52**(5), 468–489.

Hofmann, D. (1983) Automatic Testing with Intelligent Sensor Systems Measuring or Classification? *Technological and Methodological Advances in Measurement Acta IMEKO 1982*, North Holland, Amsterdam (pp. 545–551).

Nicholas, J.V. (1994) Measurement Illusions, *3rd Australasian Instrumentation and Measurement Conference*, 26–29 April, Adelaide, Institution of Engineers Australia, Barton (pp. 75–80).

Perry, C.M. (1931) Epistemology Re-Examined. *The Philosophical Review*, **40**(5), 444–458.

Rescher, N. (1980) *Induction*, Basil Blackwell, Oxford.

Russell, J.E. (1898) Epistemology and Mental States. *The Philosophical Review*, **7**(4), 394–396.

Sacerdoti, G., Paggi, R. and Sacerdoti, D. (1991) Measurements and Representation, *IMEKO XII Measurement and Progress*, Vol 1, IMEKO Secretariat, Beijing (pp. I-159–I-160).

Sydenham, P.H. (1976) Introductory Address, *Proceedings of the Regional Colloquium on the Nature and Scope of Measurement Science*, IMEKO Secretariat, Budapest (pp. 17–29).

25: Introduction to Signals in Physical Systems

Eugen Georg Woschni
Technical University Chemnitz, Chemnitz, Germany

1 General Remarks on Signals 147
2 Task of Measurement and Mathematical Description 147
3 Two Fundamental Problems: Signal and System Identification 148
Further Reading 148

1 GENERAL REMARKS ON SIGNALS

Within measuring systems, communications take place between the various parts. These communications take place on signal carriers; today, these are predominantly electrical, but other energy formats are sometimes used. The theory given here can be applied to them also.

In addition to the useful signal **x**, interfering signals **z** occur (Figure 1). Both types of signals are carriers of information: wanted information in the case of the useful signal, and unwanted in the case of the interfering signal. Both signals can be treated the same way.

In a signal $\mathbf{x}(a_r)$, for the purpose of information transmission as found in measuring systems, it must be possible to change these information parameters. Information parameters are those parameters of the signal upon which the behavior of the information to be transmitted is mapped. Thus:

$$\text{Transmitted information I} = \text{Fnc. signal } \mathbf{x}(a_r) \quad (1)$$

2 TASK OF MEASUREMENT AND MATHEMATICAL DESCRIPTION

As represented in Figure 1, the task of measurement technology is to establish the characteristics of actually occurring input signals, $\mathbf{x}_r$, from known (measured) output signals $\mathbf{y}_r$. Signals having several components, $\mathbf{x}_r$, $\mathbf{y}_r$, for example, can be summarized to form signal vectors **x, y, z**, in the same way as the disturbances $\mathbf{z}_r$. The output quantities are a function of the input quantities, and the relationship between them is given by the behavior of the measuring system.

The aim is that the behavior should be of a certain type, which is represented by an ideal mathematical operation $\mathbf{O}_{\text{id}}$:

$$\mathbf{y}_{\text{id}} = \mathbf{O}_{\text{id}}\{\mathbf{x}\} \quad (2a)$$

In the individual case, this operation may be a constant with only one dimension, to give an example, but it may also be a differentiation or integration, as for instance, in the case of measuring devices used for averaging.

The practical measuring system with the connection $\mathbf{O}_{\text{real}}$ between its output and input quantities also includes real, falsified output quantities that depend on the disturbances **z**

$$\mathbf{y}_{\text{real}} = \mathbf{O}_{\text{real}}\{\mathbf{x}; \mathbf{y}\} \quad (2b)$$

Hence, an error $\boldsymbol{\varepsilon}$ occurs given by

$$\boldsymbol{\varepsilon} = \mathbf{y}_{\text{id}} - \mathbf{y}_{\text{real}} \quad (2c)$$

These relationships will be dealt with in more detail in later articles (*see* **Article 19, Dynamic Behavior of Closed-loop Systems, Volume 1**; **Article 27, Signals in the Frequency Domain, Volume 1**; **Article 28, Signals in the**

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

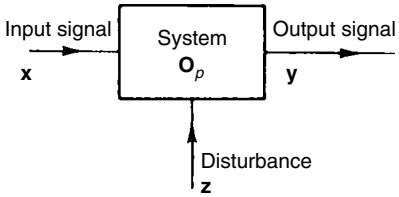


Figure 1. Task of measurement.

Time Domain, Volume 1; and **Article 29, Relationship Between Signals in the Time and Frequency Domain, Volume 1**) of this topic, in particular, those with time-varying signals. The discussion has the following four objectives:

1. description of the signals by their characteristic values and functions;
2. description of measuring systems by means of characteristic values and functions;
3. description of the errors and deduction of quality criteria;
4. means for optimizing the system, that is, for minimizing errors.

Objective 1. is discussed in **Article 27, Signals in the Frequency Domain, Volume 1**; **Article 28, Signals in the Time Domain, Volume 1**; **Article 29, Relationship Between Signals in the Time and Frequency Domain, Volume 1**; **Article 30, Statistical Signal Representations, Volume 1**; **Article 31, Discrete Signal Theory, Volume 1**; **Article 32, Geometrical Signal Representations, Volume 1**, the other objectives, which are based upon the first, are dealt with in **Article 31, Discrete Signal Theory, Volume 1**; **Article 32, Geometrical Signal Representations, Volume 1**; **Article 35, Systems in the Time Domain, Volume 1**; **Article 36, Systems in the Frequency Domain, Volume 1**; **Article 37, Relationship Between Systems in the Time and Frequency Domain, Volume 1** and **Article 38, Stability Issues, Volume 1**. Reference is made in each case to descriptions in the time and frequency domains.

3 TWO FUNDAMENTAL PROBLEMS: SIGNAL AND SYSTEM IDENTIFICATION

Figure 2 shows the objectives that have been outlined above and that are tailored especially for the following two fundamental problems of measurement technology.

Signal identification. In Figure 2(a), the aim is to determine the signal parameters of the input signals in terms of known (measured) output signals **y**, where, if possible, disturbances should have no influence.

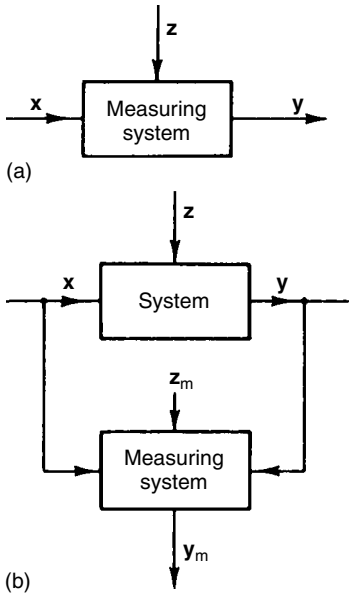


Figure 2. (a) Identification of signals and (b) identification of systems.

System identification. Figure 2(b), on the other hand, depicts measurement of the parameters of a system. In that case, input test signals **x** and associated output signals **y** are both fed to the measuring system. The parameters that are characteristic of the system have to be determined from the output quantities of this measuring system $\mathbf{y}_m$.

FURTHER READING

Many texts have been published on this topic of linear systems and circuit theory. Searches for relevant books with 'signal' as a search word tend to be in the minority now, being largely overtaken by those on digital signals and systems – which are not the same. Digital systems are, in fact, partially based on the application of this linear systems theory. A selection on linear signals and systems is as follows:

Buck, J.R., Daniel, M.M. and Singer, A.C. (2001) *Computer Explorations in Signals and Systems Using MATLAB*, 2nd edn, Prentice Hall, NJ.

Hsu, H.P. and Hse, H.P. (1995) *Schaum's Outline of Signals and Systems*, McGraw–Hill Trade.

Karu, Z.Z. (1995) *Signals and Systems Made Ridiculously Simple*, ZiZi Press, Huntsville, AL.

Oppenheim, A.V., Willsky, A.S. and Hamid, S. (1996) *Signals and Systems*, Prentice Hall, NJ.

Woschni, E.-G. (1988) *Informations Technik*, Hüthig, Heidelberg.

Woschni, E.-G. (1994) Signal Theory in Measurement and Instrumentation, *Concise Encyclopedia of Measurement & Instrumentation*, Pergamon Press, Oxford.

Zadeh, L.A. and Desoer, C.A. (1963) *Linear System Theory*, McGraw–Hill, New York.

26: Signal Classification

Eugen Georg Woschni
Technical University Chemnitz, Chemnitz, Germany

1 Signals with and without Quantization Over Time 149
2 Table of Signal Classifications 149
3 Signal-flow Graphs 149
References 151
Further Reading 151

1 SIGNALS WITH AND WITHOUT QUANTIZATION OVER TIME

A distinction is made between *analog* signals, which are signals having no quantization of the information parameter, and *discrete* signals in which, because of quantization, the information can assume a finite number of values. An important special case of discrete signals is that involving binary signals in which the information parameter can take only two discrete values: 0 or 1.

In time-dependent signals – most signals are time dependent, or are converted to become time-dependent signals by scanning, as is the case in television – the information parameter can either change at any time (continuous signals), or changes are possible at given cycle times only, due to time quantization.

If the entire behavior of the signal, including future behavior, is known, the signal is said to be a *determined* signal. The transmission or measurement of this signal, naturally, does not produce any gain in information. This type of signal plays a major role as test signals (e.g. impulse function, step function). Contrary to this, a signal to be measured has little a priori information. Signals with unknown characteristics of flow are called *nondetermined* signals. If they are described by a probability distribution, they may also be called *stochastic* signals.

2 TABLE OF SIGNAL CLASSIFICATIONS

Table 1 gives a survey of signal designations as described in detail in Section 1 (Woschni, 1988; Woschni and Krauss, 1976).

3 SIGNAL-FLOW GRAPHS

To trace and represent the flow of the signal, signal-flow graphs are used. Figure 1 gives a survey of the graphical

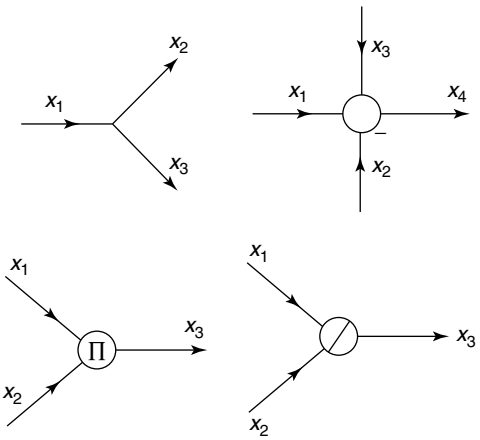
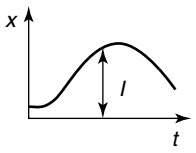
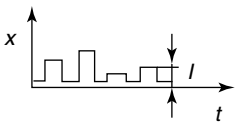
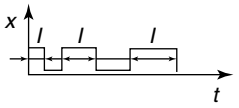
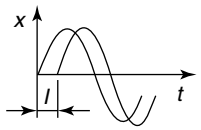
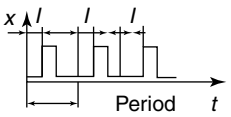
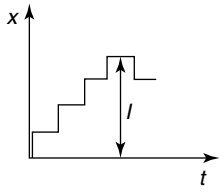
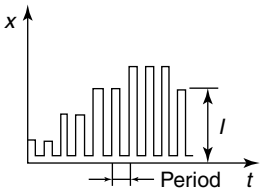
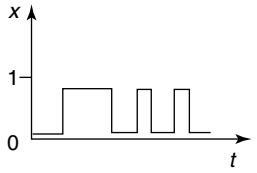
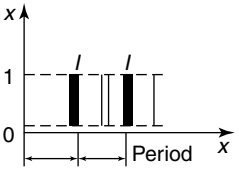


Figure 1. (a) Branching: $x_2 = x_3 = x_1$; (b) adding or subtraction: $x_4 = x_1 + x_3 - x_2$ (c) multiplication: $x_3 = x_1 \times x_2$; (d) division: $x_3 = x_1/x_2$.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

Table 1. Classification of signals.

Characteristic			Without quantization of time (continuous)	With quantization of time (discontinuous)
Without quantization of the information parameter	Analog		Continuous analog signal  Information parameter *I*: signal amplitude	Discontinuous Analog signal  Information parameter *I*: amplitude of the pulse sequence
				 Information parameter *I*: width of the pulse sequence
			 Information parameter *I*: phase shift or frequency	 Information parameter *I*: phase shift of the pulse sequence
With quantization of the information parameters	Discrete	Multipoint signal	Continuous multipoint signal  Information parameter *I*: signal amplitude can have *n* discrete quantities	Discontinuous multipoint signal  Information parameter *I*: signal amplitude can have *n* discrete quantities
		Binary	Continuous binary signal  Information parameter *I*: can have only quantities 0 and 1	Discontinuous binary signal  Information parameter *I*: is given as the code word of the 0, 1 signals of one period

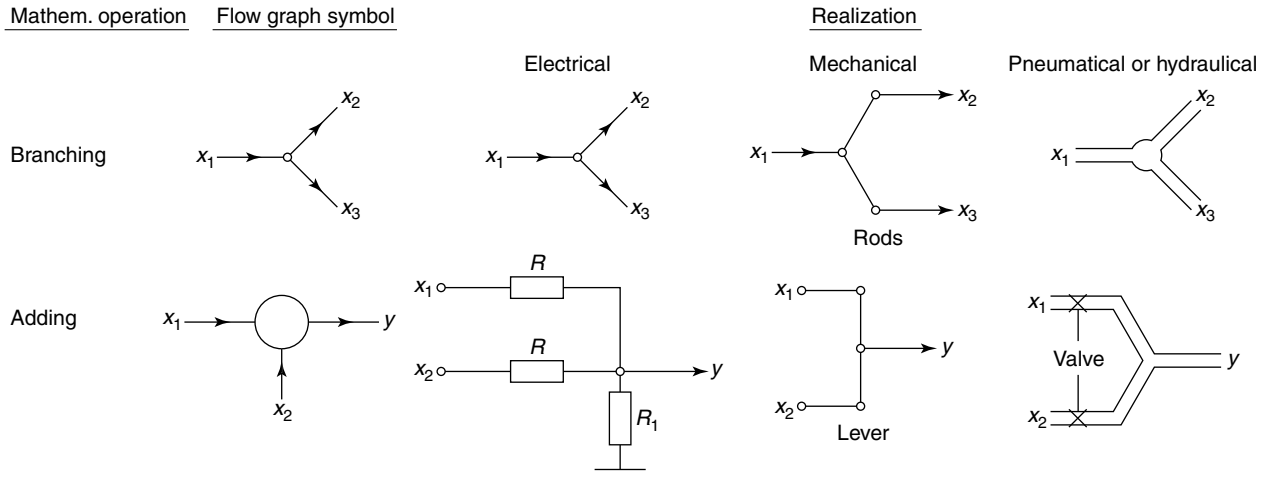


Figure 2. Examples of technical realization of branching and adding.

representation of the branching, addition, subtraction, multiplication, or division of signals. The arrows point to the direction of signal flow.

Figure 2 shows some examples of technical realization. Signals may be represented in the time and frequency (spectral) domains. For the formation of statistical characteristic values, time or statistical means are used. Furthermore, signal representations are applied by making use of geometrical relationships.

More information on this material is given in subsequent articles.

REFERENCES

Books on the topic need to cover linear and digital signals and systems. Digital titles are added to the linear titles given in **Article 25, Introduction to Signals in Physical Systems, Volume 1**.

Woschni, E.-G. (1988) *Informations Technik*, Huthig, Heidelberg.

Woschni, E.-G. and Krauss, M. (1976) *Informations Technik, Arbeitsbuch*, Verlag Technik, Berlin.

FURTHER READING

Buck, J.R., Daniel, M.M. and Singer, A.C. (2001) *Computer Explorations in Signals and Systems Using MATLAB*, 2nd edn, Prentice Hall, NJ.

Grover, D. and Deller, J. (1998) *Digital Signal Processing and the Microcontroller*, Prentice Hall PTR, NJ.

Hayes, M.H. (1998) *Schaum's Outline of Digital Signal Processing*, McGraw-Hill Trade, New York.

Hsu, H.P. and Hse, H.P. (1995) *Schaum's Outline of Signals and Systems*, McGraw-Hill Trade.

Karu, Z.Z. (1995) *Signals and Systems Made Ridiculously Simple*, ZiZi Press, Huntsville, AL.

Levine, W.S. (1996) *The Control Handbook*, CRC Press, NJ.

Lyons, R.G. (1996) *Understanding Digital Signal Processing*, Prentice Hall PTR, NJ.

Oppenheim, A.V., Willsky, A.S. and Hamid, S. (1996) *Signals and Systems*, Prentice Hall, NJ.

Williamson, D. (1999) *Discrete-Time Signal Processing: An Algebraic Approach*, Springer-Verlag, Berlin.

Zadeh, L.A. and Desoer, C.A. (1963) *Linear System Theory*, McGraw-Hill, New York.

Ziemer, R.E., Tranter, W.H. and Fannin, D.R. (1998) *Signals and Systems: Continuous and Discrete*, Prentice Hall, NJ.

27: Signals in the Frequency Domain

Eugen Georg Woschni
Technical University Chemnitz, Chemnitz, Germany

1 Simplest Case – Harmonic Oscillation 152
2 Periodic Signals – Fourier Spectrum 152
3 Periodic Signals – Walsh Spectrum 154
4 Nonperiodic Signals – Fourier Transform 154
5 Spectral Power Density 155
6 Practical Investigations to Study Signals 156
7 Example of Rectangular Pulse Signal 157
References 157
Further Reading 157

1 SIMPLEST CASE – HARMONIC OSCILLATION

To study the nature of the frequency domain of a signal, the simplest case may be used as a foundation for the explanation. A very important, determined fundamental signal, which is also of major importance as a test signal, is harmonic oscillation

$$x = \widehat{X}\sin(\omega t + \varphi) \tag{1a}$$

where $\widehat{X}$ is the amplitude, $\omega = 2\pi f$ is the angular frequency, $T = 1/f$ is the oscillation period and φ is the phase angle (often zero). Representation in the complex plane, as shown in Figure 1(a), yields according to the so-called 'symbolic method,' complex and oriented indicators x and, consequently, the relationship

$$x = \widehat{X}\exp[\,\mathrm{j}(\omega t + \varphi)] = \widehat{X}\cos(\omega t + \varphi) + \mathrm{j}\widehat{X}\sin(\omega t + \varphi) = \widehat{X}\mathrm{e}^{\mathrm{j}\omega t} \tag{1b}$$

which is equivalent to (1a) and which has the oriented complex amplitude

$$\widehat{X} = \widehat{X}\mathrm{e}^{\mathrm{j}\varphi} \tag{1c}$$

According to Figure 1, the directed quantity $\widehat{X}$ can be split into real and imaginary parts:

$$\widehat{X} = A + \mathrm{j}B \quad |\widehat{X}| = \widehat{X} = \sqrt{(A^2 + B^2)} \quad \varphi = \arctan\left(\frac{B}{A}\right) \tag{1d}$$

The advantage of the symbolic method consists, above all, of the possibility of having a simple and easily understandable addition of several partial oscillations having the same frequency (Woschni, 1988). Periodic signals are particularly useful as test signals, because the same signal characteristic is repeated after each cycle period T and thus can be observed on an oscilloscope synchronized with T.

2 PERIODIC SIGNALS – FOURIER SPECTRUM

According to Fourier, it is possible to represent this type of signal having a cyclic time behavior $x(t)$ by a series of sinusoidal and cosinusoidal oscillations with frequencies that are multiples of the fundamental frequency ω_0 given by

$$\omega_0 = 2\pi f_0 = \frac{2\pi}{T} \tag{2a}$$

This is shown in Figure 2, which depicts two sinusoids (of many) that form the cyclic rectangular oscillation (square

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

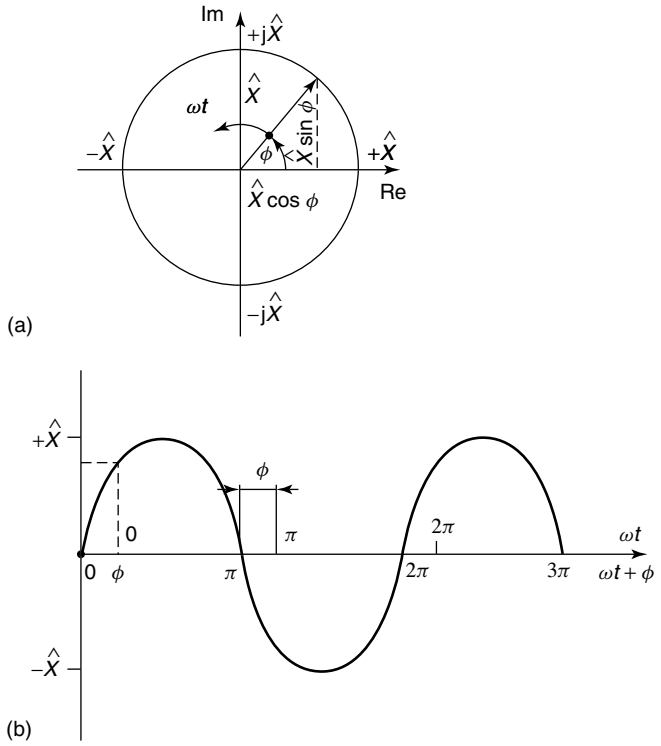


Figure 1. Harmonic oscillation: (a) indicator representation and (b) time representation.

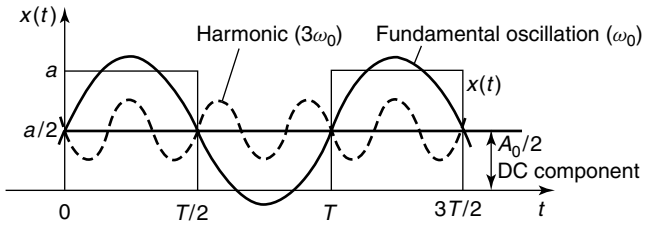


Figure 2. Representation of the rectangular oscillation by sinusoids.

wave signal, see also Figure 7)

$$x(t) = \frac{1}{2}A_0 + \sum_{n=1}^{\infty}[A_n \cos(n\omega_0\, t) + B_n \sin(n\omega_0\, t)] \quad (2b)$$

with the amplitude spectrum

$$C_n = \sqrt{(A_n^2 + B_n^2)} \quad (2c)$$

The Fourier coefficients can be calculated from the relationships

$$A_n = \frac{2}{T}\int_{-T/2}^{+T/2} x(t)\cos(n\omega_0\, t)\,\mathrm{d}t \quad (3a)$$

$$B_n = \frac{2}{T}\int_{-T/2}^{+T/2} x(t)\sin(n\omega_0\, t)\,\mathrm{d}t \quad (3b)$$

From (3a, b), one can see that there are only cosine terms for the even time functions $x(t) = x(-t)$, and only sine terms for the odd functions $x(t) = -x(-t)$.

Transformation with the aid of Euler's theorem, leads to the complex Fourier series

$$x(t) = \frac{1}{T}\sum_{n=-\infty}^{+\infty} \widehat{X}(\mathrm{j}n\omega_0)\mathrm{e}^{\mathrm{j}n\omega_0 t} \quad (4a)$$

with the complex coefficient

$$\widehat{X}(\mathrm{j}n\omega_0) = \int_{-T/2}^{+T/2} x(t)\mathrm{e}^{-\mathrm{j}n\omega_0 t}\,\mathrm{d}t \quad (4b)$$

and the amplitude spectrum $|C_{nk}|$ given by

$$|C_{nk}| = \left|\frac{1}{T}\widehat{X}(\mathrm{j}n\omega_0)\right| \quad (4c)$$

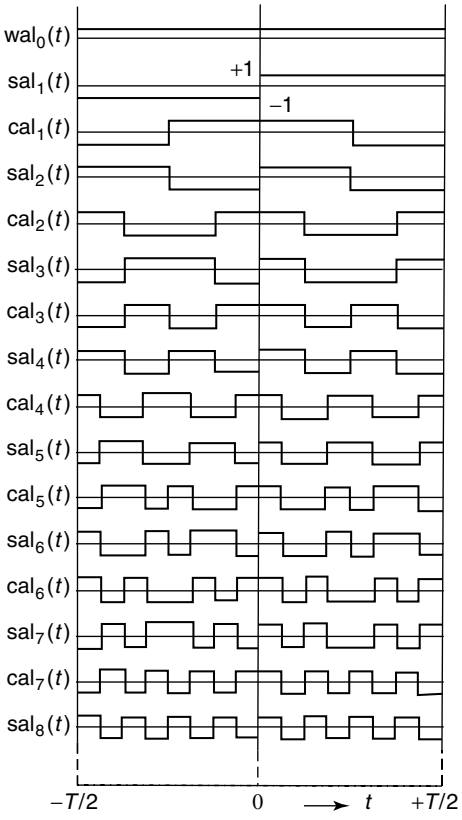


Figure 3. Walsh functions up to eighth order.

3 PERIODIC SIGNALS – WALSH SPECTRUM

In addition to the system of orthogonal functions based upon sine and cosine functions, other orthogonal systems have been introduced, in particular the Walsh functions (Harmuth, 1970), which are shown up to eighth order in Figure 3.

From (2) and (3) respectively, the corresponding relationships are developed:

$$x(t) = W_0 + \sum_{n=1}^{\infty} [W_{c_n} \mathrm{cal}_n(t) + W_{s_n} \mathrm{sal}_n(t)] \qquad (5a)$$

with the Walsh coefficients W_{c_n} and W_{s_n}:

$$W_{c_n} = \frac{1}{T} \int_{-T/2}^{+T/2} x(t)\ \mathrm{cal}_n(t)\, \mathrm{d}t \qquad (5b)$$

$$W_{s_n} = \frac{1}{T} \int_{-T/2}^{+T/2} x(t)\ \mathrm{sal}_n(t)\, \mathrm{d}t \qquad (5c)$$

As one can see, the Walsh spectra, which are also called *sequency spectra*, are superior to the Fourier spectra, in that the multiplication with sine and cosine functions respectively, is obviated by a simple reversal of signs. Therefore, these spectra can be determined more easily by experiment than the Fourier spectra (Figure 4).

4 NONPERIODIC SIGNALS – FOURIER TRANSFORM

Nonperiodic functions are of great importance both as determined signals, that is, test signals (step function, impulse function) and as nondetermined signals (unknown signals that are to be measured).

A discrete Fourier spectrum exists for periodic signals, whereas a continuous spectrum follows for nonperiodic signals. It is obtained from the Fourier series – see (4) – by passing to the limit:

$$T \longrightarrow \infty \quad \omega_0 = \frac{2\pi}{T} \longrightarrow \mathrm{d}\omega$$

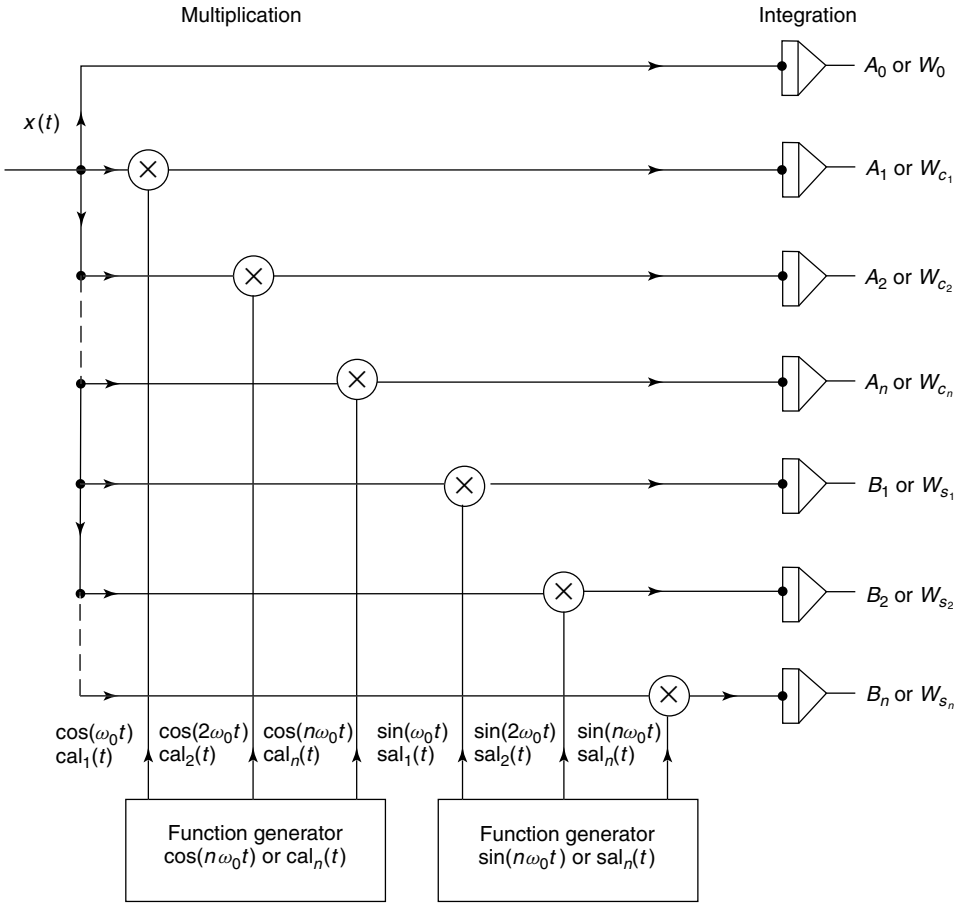


Figure 4. Experimental determination of the Fourier and Walsh coefficients.

$$\frac{1}{T} \longrightarrow \frac{\mathrm{d}\omega}{2\pi} \quad n\omega_0 \longrightarrow \omega \tag{6a}$$

In this case, it yields

$$\widehat{X}(\mathrm{j}\omega) = \int_{-\infty}^{+\infty} x(t)\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t = F\{x(t)\} \tag{6b}$$

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \widehat{X}(\mathrm{j}\omega)\mathrm{e}^{\mathrm{j}\omega t}\,\mathrm{d}\omega = F^{-1}\{\widehat{X}(\mathrm{j}\omega)\} \tag{6c}$$

Here the integrals are to be understood as the Cauchy principal value

$$\lim_{c\to\infty}\int_{-c}^{+c}\cdots$$

F and F^{-1} are the Fourier and inverse-Fourier transforms respectively. Physically, $\widehat{X}(\mathrm{j}\omega)$ represents the complex amplitude related to $\mathrm{d}\omega$, and is therefore also called *spectral amplitude density* having the dimension of amplitude per frequency interval, that is, V s and V Hz^{-1} respectively. Depending on whether the frequency scale [Hz] or the angular frequency scale [s^{-1}] is chosen, the values will differ by a factor 2π.

An identical calculation can be made for Walsh functions. This leads to the sequential amplitude density (Harmuth, 1970).

The Fourier transform has the following important properties and theorems. The transform is linear, that is,

$$\sum_{\nu} a_\nu x_\nu(t)\ \mathrm{O}^- \sum_{\nu} a_\nu \widehat{X}_\nu(\mathrm{j}\omega) \tag{7a}$$

(where O^- is the sign for 'assignment'). For a change in the timescale, the relationship is

$$x(at)\ \mathrm{O}^- \frac{1}{|a|}\widehat{X}\left(\frac{\mathrm{j}\omega}{a}\right) \tag{7b}$$

Particular importance should also be attributed to the displacement theorems, namely, the time displacement

$$x(t-t_0)\ \mathrm{O}^-\widehat{X}(\mathrm{j}\omega)\exp(-\mathrm{j}\omega t_0) \tag{7c}$$

and the frequency shift

$$x(t)\exp(\mathrm{j}\omega t_0)\ \mathrm{O}^-\widehat{X}[\mathrm{j}(\omega-\omega_0)] \tag{7d}$$

For differentiation, one obtains

$$\frac{\mathrm{d}^n x(t)}{\mathrm{d}t^n}\ \mathrm{O}^-(\mathrm{j}\omega)^n\widehat{X}(\mathrm{j}\omega) \tag{7e}$$

and for convolution

$$\int_{-\infty}^{+\infty} x_1(\tau)x_2(t-\tau)\,\mathrm{d}\tau\ \mathrm{O}^-\widehat{X}_1(\mathrm{j}\omega)\widehat{X}_2(\mathrm{j}\omega) \tag{7f}$$

5 SPECTRAL POWER DENSITY

For nondeterministic signals, for instance noise, let us assume in the following text that they are stationary, that is, that their time averages are not time-dependent quantities. To identify such signals $x(t)$, the spectral power density $S_{xx}(\omega)$ is used. It is defined as the part of the power ΔP, which falls into a differentially small frequency range $\Delta\omega$, that is,

$$S_{xx}(\omega) = \lim_{\Delta\omega\to 0}\frac{\Delta P}{\Delta\omega} = \frac{\mathrm{d}P}{\mathrm{d}\omega} \tag{8a}$$

In contrast to the spectral amplitude density, which cannot be determined in the case of random signals, the spectral power density is a real-valued function of the frequency ω. It does not contain any phase information. The latter is lost in the calculation of the average value, which is necessary for the formation of the power. This can also be seen from the relationship existing on the basis of Parseval's equation, by averaging over a time domain T (Zadeh and Desoer, 1963; Woschni, 1988):

$$S_{xx}(\omega) = \frac{1}{2\pi}\lim_{T\to\infty}\frac{|\widehat{X}(\mathrm{j}\omega)|^2}{2T} \tag{8b}$$

Consequently, the spectral power density is real and always positive, and it is an even function for which $S_{xx}(\omega) = S_{xx}(-\omega)$. Since the phase angle is missing, $S_{xx}(\omega)$ does not contain the full information about $x(t)$; a reverse calculation is not possible. The power P of the entire signal existing in the whole frequency domain, can be calculated on the basis of Parseval's equation, for the energy W

$$W = \int_{-\infty}^{+\infty} x^2(t)\,\mathrm{d}t = \frac{1}{2\pi}\int_{-\infty}^{+\infty}|\widehat{X}(\mathrm{j}\omega)|^2\,\mathrm{d}\omega$$

(Zadeh and Desoer, 1963; Woschni, 1988):

$$\begin{aligned} P = \overline{x^2(t)} &= \lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{+T} x^2(t)\,\mathrm{d}t \\ &= \int_{-\infty}^{+\infty}\lim_{T\to\infty}\frac{1}{2\pi}\frac{|\widehat{X}(\mathrm{j}\omega)|^2}{2T}\,\mathrm{d}\omega \\ &= \int_{-\infty}^{+\infty} S_{xx}\,\mathrm{d}\omega \end{aligned} \tag{8c}$$

From (8a, b) it can be concluded that a random phenomenon contains a periodic component of frequency ω_0

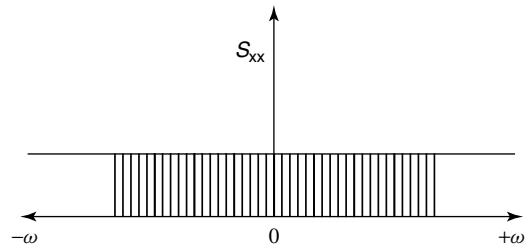


Figure 5. Spectral power density of a white-noise signal.

with amplitude $\widehat{X}$. Dirac delta functions will develop in $S_{xx}(\omega)$ at the frequencies $\pm\omega_0$ (Woschni and Krauss, 1976):

$$S_{xx}(\omega)|_{\omega_0} = \tfrac{1}{2}\widehat{X}^2\delta(|\omega| - \omega_0) \tag{8d}$$

Furthermore, from (8c) it follows that the power density $S_{xx}(\omega)$ must decrease rapidly from a certain critical frequency and must vanish at higher frequencies because of the requirement of boundedness of the power P. Depending on the critical frequency ω_c, a distinction is made between narrowband and wideband signals.

A typical wideband (broadband) signal is the white-noise signal with a constant power density, as shown in Figure 5.

There is a relation between time spectral power density and the auto correlation function, *see* **Article 29, Relationship Between Signals in the Time and Frequency Domain, Volume 1**.

6 PRACTICAL INVESTIGATIONS TO STUDY SIGNALS

The foundation for investigations in practice is the possibility of determining the characteristic functions and values by experiment. For the registration of spectral amplitude density and power density respectively, one applies the same principles. In the case of filtering, several filters of bandwidth $\Delta\omega$ are used, which are staggered in the frequency, and whose outputs will be connected one after the other to a display unit for the voltage that is proportional to $|\widehat{X}(\mathrm{j}\omega)|$. According to Figure 6(a), it is likewise possible to synchronize the switch with a sweep voltage that deflects the beam of an oscilloscope in the x-direction, proportionally to ω; consequently, the spectrum $|\widehat{X}(\mathrm{j}\omega)| = f(\omega)$ can be recorded.

Figure 6(b) shows another method that is characterized by the fact that only one filter of bandwidth $\Delta\omega$ and of center frequency ω_{IF} (which is also called *intermediate frequency*) is required. Tuning is carried out by mixing with the continuously tunable auxiliary frequency ω_{A}, where the following frequencies will be allowed to pass:

$$\omega = \omega_{\mathrm{IF}} - \omega_{\mathrm{A}}; \quad \omega^* = \omega_{\mathrm{IF}} + \omega_{\mathrm{A}} \tag{9a, b}$$

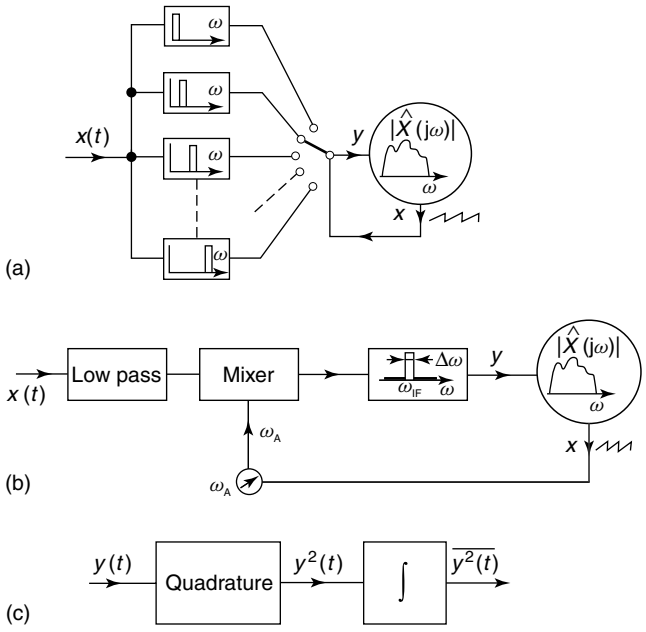


Figure 6. Spectral analyzer: (a) switched-filter principle of operation; (b) variable center frequency of single filter method; and (c) formation of the power density.

The low-pass filter at the input eliminates the image frequency $\omega*$, so that one can cover the entire frequency domain required by tuning ω_{A}. Synchronizing the sweep signal with the x-axis sweep will generate the amplitude spectrum on the oscilloscope display. It is important to ensure that the filter has enough time to respond to the starting surge. For the transient time t_{tr} of the filter, according to Shannon's sampling theorem, *see* **Article 31, Discrete Signal Theory, Volume 1**:

$$t_{\mathrm{tr}} = \frac{1}{2\Delta f} = \frac{\pi}{\Delta\omega} \tag{9c}$$

The sampling of the spectrum must, therefore, be carried out relatively slowly (low sweep frequency of the oscilloscope), for which reason long-persistence cathode ray oscilloscope methods are required. During the scanning run, the spectrum must be practically stable; it must be a steady signal in the statistical sense.

In addition to filtering and variable frequency methods, increasing techniques are used today, which make use of computers to implement the Fourier transform directly according to (6b). This will be discussed in detail in **Article 29, Relationship Between Signals in the Time and Frequency Domain, Volume 1**, where specially adapted techniques for the fast Fourier transform will be dealt with. The computing advantages that arise when the amplitude sequential spectrum defined by the Walsh functions is used instead of the amplitude frequency

spectrum, have already been pointed out in connection with Figure 4. While the techniques described in Figure 6(a, b) supply the amplitude response of the spectrum $|\widehat{X}(\mathrm{j}\omega)|$ only, it is possible to obtain additional phase information from the relationship $\varphi = \arg[\widehat{X}(\mathrm{j}\omega)]$ using computing methods.

To display the power spectrum, the formulation $\Delta P/\Delta\omega = \overline{\Delta x^2(t)}/\Delta\omega$ is implemented, as outlined in Figure 6(c), between the filter output and the input of the display unit and the oscilloscope.

7 EXAMPLE OF RECTANGULAR PULSE SIGNAL

For a periodic sequence of rectangular pulses having the pulse width ΔT or the repetition ratio $\tau = \Delta T/T$ (see Figure 7a), after substitution into (3a, b) and (2c) or (4b, c) and after an elementary calculation, one obtains pure cosinusoidal oscillations having the amplitudes

$$C_n = A_n = 2\tau \frac{\sin(n\pi\tau)}{n\pi\tau} = 2\tau \sin(n\pi\tau) \qquad (10a)$$

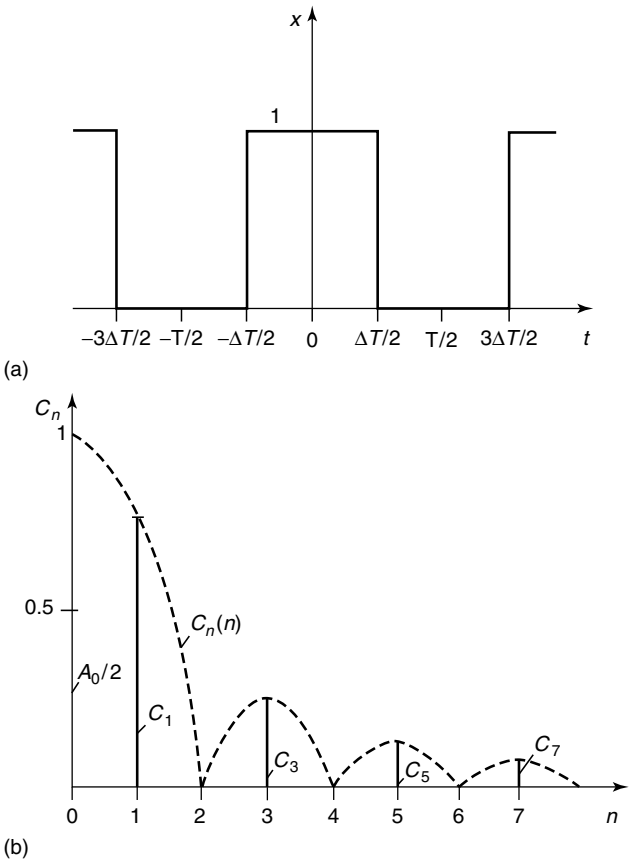


Figure 7. (a) Sequency of pulses and (b) spectrum for $\tau = 0.5$.

Figure 7 shows the spectrum as well as the envelope curve for $\tau = 0.5$. This signal was already treated in Section 2 of this article.

The relationship with Figure 2 becomes immediately evident: the smaller the pulse width, that is, $t_{\mathrm{tr}} = \Delta T$, the more complicated the problem is in general. The pulse height is still correctly indicated in the region up to the first zero of the envelope curve. For $\tau \to 0$, a constant spectrum results because the first zero shifts towards $\omega = \infty$. This case is significant as a test signal. For $\tau \to 0$, the unit impulse becomes the Dirac function $\delta(t)$ with the normalization

$$\int_{-\infty}^{+\infty} \delta(t)\,\mathrm{d}t = \int_{-0}^{+0} \delta(t)\,\mathrm{d}t = 1$$

According to (4b), this has a spectral amplitude density $\widehat{X}(\mathrm{j}\omega) = 1$, with only cosine oscillations of constant amplitude occurring.

Some more signals that are used as test signals as well as the approximate representation of measurement signals are treated in **Article 28, Signals in the Time Domain, Volume 1**.

REFERENCES

Harmuth, H.F. (1970) *Transmission of Information by Orthogonal Functions*, Springer, Berlin.

Woschni, E.-G. (1988) *Informations Technik*, Huthig, Heidelberg.

Woschni, E.-G. and Krauss, M. (1976) *Informations Technik, Arbeitsbuch*, Verlag Technik, Berlin.

Zadeh, L.A. and Desoer, C.A. (1963) *Linear System Theory*, McGraw-Hill, New York.

FURTHER READING

Buck, J.R., Daniel, M.M. and Singer, A.C. (2001) *Computer Explorations in Signals and Systems Using MATLAB*, 2nd edn, Prentice Hall, NJ.

Hsu, H.P. and Hse, H.P. (1995) *Schaum's Outline of Signals and Systems*, McGraw-Hill Trade.

Karu, Z.Z. (1995) *Signals and Systems Made Ridiculously Simple*, ZiZi Press, Huntsville, AL.

Oppenheim, A.V., Willsky, A.S. and Hamid, S. (1996) *Signals and Systems*, Prentice Hall, NJ.

Tolimieri, R., An, M. and Lu, C. (1997) *Algorithms for Discrete Fourier Transform and Convolution*, 2nd edn, Springer, New York.

Ziemer, R.E., Tranter, W.H. and Fannin, D.R. (1998) *Signals and Systems: Continuous and Discrete*, Prentice Hall, NJ.

28: Signals in the Time Domain

Eugen Georg Woschni

Technical University Chemnitz, Chemnitz, Germany

1 Mean Values 158
2 Correlation Function 158
3 Relationships to Spectral Power Density 159
4 Practical Investigations in the Signal Time Domain 160
5 Examples of Mean Value and Correlation Functions 161
References 162
Further Reading 162

1 MEAN VALUES

A time function $x(t)$ can be characterized by time averages of the nth order, which are also called moments of nth order:

$$\overline{x^n(t)} = \frac{1}{2T}\int_{-T}^{+T} x^n(t)\,\mathrm{d}t \tag{1a}$$

or, for nonperiodic signals:

$$\overline{x^n(t)} = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{+T} x^n(t)\,\mathrm{d}t \tag{1b}$$

For $n = 1$, one obtains the (arithmetic) linear average, which, from the physical point of view, can be interpreted to be the zero-frequency component (DC value) of the signal or, according to **Article 27, Signals in the Frequency Domain, Volume 1**, the Fourier coefficient $A_0/2$.

Of particular importance is the average value for $n = 2$, which, as a mean square value according to (8c) of **Article 27, Signals in the Frequency Domain, Volume 1**, represents a measure of the power. The root of the mean square value is the effective value X_{eff} given by

$$X_{\mathrm{eff}} = \sqrt{\overline{[x^2(t)]}} \tag{2a}$$

The following relationship exists between the component $\overline{x(t)}$, the AC component $\underset{\sim}{x}(t)$, and the mean square value X_{eff}:

$$X_{\mathrm{eff}} = \overline{x^2(t)} = [\overline{x(t)}]^2 + \overline{\underset{\sim}{x}^2(t)} \tag{2b}$$

For a harmonic oscillation, one obtains $X_{\mathrm{eff}} = \widehat{X}/\sqrt{2}$ for the effective value, see Section 5. If the Ergodic-theorem is fulfilled, normally in practice the effective value X_{eff} is equal to the standard deviation δ and the mean square value X_{eff}^2 to the variance δ^2, *see* **Article 30, Statistical Signal Representations, Volume 1**.

2 CORRELATION FUNCTION

The correlation function $\psi(\tau)$ represents a generalized mean square value, where a function is multiplied by the function displaced by time τ, and then the mean value is formed. If this function is the same function x, we call it an autocorrelation function $\psi_{xx}(\tau)$,

$$\psi_{xx}(\tau) = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{+T} x(t)x(t+\tau)\,\mathrm{d}t = \overline{x(t)x(t+\tau)} \tag{3a}$$

as Figure 1 shows.

It is suitable for making statistical statements about the internal relationships between function sections, as is now shown in a survey of its typical properties:

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.

© 2005 John Wiley & Sons, Ltd.

ISBN: 0-470-02143-8.

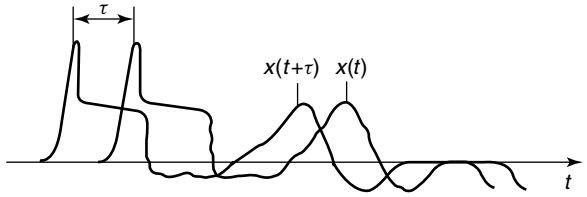


Figure 1. Definition of the autocorrelation function $\psi_{xx}(\delta)$.

1. In averaging, the phase information is lost, as is the case for the spectral power density (**Article 27, Signals in the Frequency Domain, Volume 1**). Therefore, there are also direct relationships between the autocorrelation function $\psi_{xx}(\tau)$ and the spectral power density $S_{xx}(\omega)$, which is discussed in detail in **Article 27, Signals in the Frequency Domain, Volume 1**. However, periodic components in the signal $x(t)$ will be maintained without giving consideration to the phase position, because the following expression applies to the autocorrelation function of the harmonic oscillation, independent of the phase position:

$$\psi_{xx}(\tau) = \tfrac{1}{2}\widehat{X}^2 \cos(\omega\tau) = X_{\text{eff}}^2 \cos(\omega\tau) \tag{3b}$$

2. The value for $\tau = 0$ represents, according to (3a), the mean square value and is the maximum value of the autocorrelation function

$$\psi_{xx}(0) = \overline{x^2(t)} = X_{\text{eff}}^2 \tag{3c}$$

The other threshold value for $\tau \to \infty$ is the square of the linear mean value:

$$\lim_{\tau\to\infty} \psi_{xx}(\tau) = [\overline{x(t)}]^2 \tag{3d}$$

3. Since it is of no significance whether the function $x(t)$ in (3a) is displaced toward positive or negative times, the autocorrelation function is an even function:

$$\psi_{xx}(\tau) = \psi_{xx}(-\tau) = \overline{x(t)x(t+\tau)} = \overline{x(t)x(t-\tau)} \tag{3e}$$

If two different signals $x(t)$, $y(t)$ are being compared one with the other, the measure used for the statistical relationship between them is the cross-correlation function $\psi_{xy}(\tau)$ according to the definition:

$$\psi_{xy}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} x(t)y(t+\tau)\,\mathrm{d}t = \overline{x(t)y(t+\tau)} \tag{4a}$$

In measurement technology, the cross-correlation function plays a major role in solving system identification tasks during normal operation by means of the disturbances. The solution of this measurement problem is the foundation for adaptive systems (Davies, 1970).

In contrast to the autocorrelation function, the cross-correlation function has the following features:

1. It has no even functions, but the following relationship holds:

$$\psi_{xy}(\tau) = \psi_{yx}(-\tau) \tag{4b}$$

2. It contains relative phase information concerning the two events $x(t)$, $y(t)$. In particular, the cross-correlation function of two harmonic signals with the same frequency disappears if the phase shift is $\pm\pi/2$, as can be seen following substitution into (4a). Likewise, the cross-correlation of two harmonic oscillations is zero if the frequencies are unequal.
3. The limiting cases are

$$\psi_{xy}(0) = \psi_{yx}(0) = \overline{x(t)y(t)} \tag{4c}$$

$$\lim_{\tau\to\infty} \psi_{xy}(\tau) = \overline{x(t)} \times \overline{y(t)} \tag{4d}$$

Experimental registration of the correlation function is discussed in Section 4 of this article.

3 RELATIONSHIPS TO SPECTRAL POWER DENSITY

The autocorrelation function, like the spectral power density, contains no phase information; this is lost in both cases because of the averaging operation. There is a relationship between both functions, as is the case between the time behavior of the signal $x(t)$ and the corresponding spectral amplitude density $X(\mathrm{j}\omega)$, via the Fourier transform; this relationship is known as the Wiener–Chinchine theorem (Woschni, 1988; Davies, 1970):

$$S_{xx}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \psi_{xx}(\tau) \mathrm{e}^{-\mathrm{j}\omega\tau}\,\mathrm{d}\tau = \frac{1}{2\pi} F\{\psi_{xx}(\tau)\} \tag{5a}$$

$$\psi_{xx}(\tau) = \int_{-\infty}^{+\infty} S_{xx}(\omega) \mathrm{e}^{\mathrm{j}\omega\tau}\,\mathrm{d}\omega = 2\pi F^{-1}\{S_{xx}(\omega)\} \tag{5b}$$

Since the autocorrelation function is an even function (see Section 2 of this article), cosinusoidal oscillations only occur. Consequently, (5a, b) can be rewritten

$$S_{xx}(\omega) = \frac{1}{\pi} \int_0^{\infty} \psi_{xx}(\tau) \cos(\omega\tau)\,\mathrm{d}\tau \tag{5c}$$

$$\psi_{xx}(\tau) = 2 \int_0^{\infty} S_{xx} \cos(\omega\tau)\,\mathrm{d}\omega \tag{5d}$$

If the autocorrelation function in (5a, b) is substituted by the cross-correlation function $\psi_{xy}(\tau)$, then the corresponding relationships with the cross-power density will be obtained:

$$S_{xy}(\mathrm{j}\omega) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \psi_{xy}(\tau)\mathrm{e}^{-\mathrm{j}\omega\tau}\,\mathrm{d}\tau = \frac{1}{2\pi}F\{\psi_{xy}(\tau)\} \quad (6a)$$

$$\psi_{xy}(\tau) = \int_{-\infty}^{+\infty} S_{xy}(\mathrm{j}\omega)\mathrm{e}^{\mathrm{j}\omega\tau}\,\mathrm{d}\omega = 2\pi F^{-1}\{S_{xy}(\mathrm{j}\omega)\} \quad (6b)$$

A survey of these relations is given in **Article 29, Relationship Between Signals in the Time and Frequency Domain, Volume 1**.

4 PRACTICAL INVESTIGATIONS IN THE SIGNAL TIME DOMAIN

The characteristic functions and values that have been introduced can be found by experiment.

For the measurement of the linear mean value, $\overline{x(t)}$, either moving-coil instruments or transistor voltmeters having a series-connected integration link is used; compare Figure 2(a), with Figure 2(b) showing a simple analog circuit for finding the average value according to the equation

$$u_\mathrm{o} = \frac{1}{C}\int i_C\,\mathrm{d}t \approx \frac{1}{RC}\int u_\mathrm{i}\,\mathrm{d}t \quad (7)$$

using the assumption $R \gg 1/\omega C$.

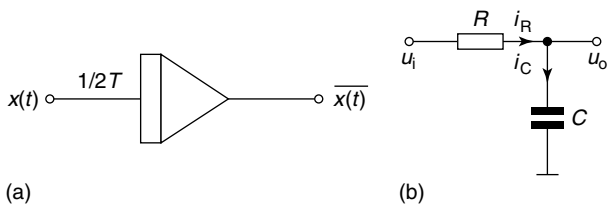


Figure 2. Measurement of the linear mean value: (a) basic circuit and (b) simple realization.

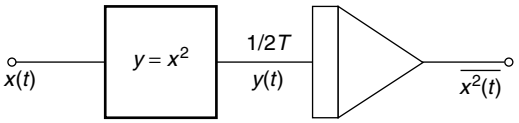


Figure 3. Measurement of the mean square value.

The mean square value can be found by means of the principle shown in Figure 3. To obtain the square of a value, either electronic circuits with a square-law characteristic, for example, diodes or transistors (as in the case of transistor voltmeters), or a measuring device having a square-law characteristic, such as a soft-iron or hot-wire instrument (heating $\approx P \approx I^2R$) is used.

Figure 4 shows the classical basic system for determining the autocorrelation function or the cross-correlation function. Delay is due to a delay section, multiplication, and averaging yield, according to (3a), the autocorrelation function (switch in position A) or, according to (4a), the cross-correlation function (switch in position B). Since the integration time T cannot be chosen to be infinite, only the short-term correlation function

$$\psi_\mathrm{k}(\tau) = \frac{1}{2T}\int_{-T}^{+T} \cdots\,\mathrm{d}t \quad (8)$$

will be measured in practice; under certain circumstances it reflects the actual behavior of $\psi(\tau)$ with sufficient accuracy.

The just treated classical analog methods have been increasingly displaced by digital methods using the principles of digital signal processing. There the signal function $x(t)$ is sampled leading to sampled values $x(rT_\mathrm{s})$ with the sampling time T_s. These values are processed due to the algorithms given by (1a), (3a), or (4a), leading to the mean values or the correlation functions.

To fulfil the sampling theorem conditions, high sampling rates are necessary to avoid aliasing errors, *see* **Article 31, Discrete Signal Theory, Volume 1**. On the other hand, online processing that is often demanded in signal processing, means that the necessary processing time has to be

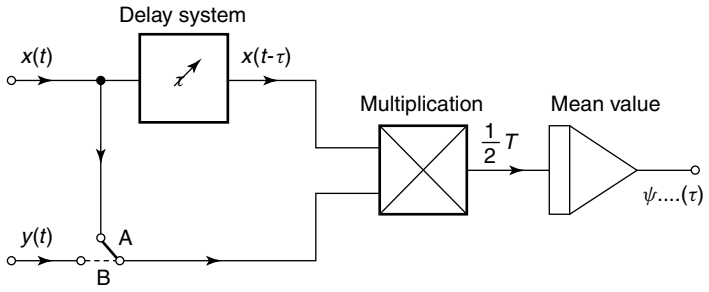


Figure 4. Experimental determination of the correlation function: (a) autocorrelation function $\psi_{xx}(\tau)$ and (b) cross-correlation function $\psi_{xy}(\tau)$.

smaller than the sampling time. Therefore, in general, the numbers of samples n are chosen due to $n = 2^s$ so that the time for the division through n can be small, because instead of the time-consuming division now only a shifting is necessary.

Methods based on the application of Fast Fourier Transform (FFT) are used to calculate the correlation functions from power density functions and vice versa, *see* **Article 27, Signals in the Frequency Domain, Volume 1**; **Article 29, Relationship Between Signals in the Time and Frequency Domain, Volume 1**. *See also* **Article 143, DFT and FFTs, Volume 3**.

5 EXAMPLES OF MEAN VALUE AND CORRELATION FUNCTIONS

As a first example, we find the mean values of a harmonic oscillation

$$\overline{x(t)} = \frac{1}{2T}\int_{-T}^{+T} \widehat{X}\sin\omega t\,\mathrm{d}t = 0 \tag{9a}$$

$$\overline{x^2(t)} = \frac{1}{2T}\int_{-T}^{+T} \widehat{X}\sin^2\omega t\,\mathrm{d}t = \frac{\widehat{X}^2}{2} \tag{9b}$$

This means the effective value is

$$X_{\text{eff}} = \widehat{X} = \frac{\widehat{X}}{\sqrt{2}} \tag{9c}$$

The next example is the rectifier unit as used in measurement instruments. The signal is shown in Figure 5 for half-wave (thin) or full-wave (thick) rectification.

The results are

1. for half-wave rectification

$$\overline{x(t)} = \frac{a}{T}\int_0^{\frac{T}{2}} \sin\omega_0 t\,\mathrm{d}t = \frac{a}{\pi} \tag{10a}$$

$$\overline{x^2(t)} = \frac{a}{T}\int_0^{\frac{T}{2}} \sin^2\omega_0 t\,\mathrm{d}t = \frac{a^2}{4} \tag{10b}$$

$$X_{\text{eff}} = \sqrt{\overline{x^2(t)}} = \frac{a}{2} \tag{10c}$$

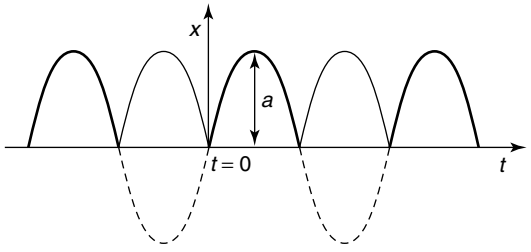


Figure 5. Half-wave (thin) and full-wave (thick) rectification.

2. for full-wave rectification

$$\overline{x(t)} = \frac{2a}{T}\int_0^{\frac{T}{2}} \sin\omega_0 t\,\mathrm{d}t = \frac{2a}{\pi} \tag{11a}$$

$$\overline{x^2(t)} = \frac{2a}{T}\int_0^{\frac{T}{2}} \sin^2\omega_0 t\,\mathrm{d}t = \frac{a^2}{2} \tag{11b}$$

$$X_{\text{eff}} = \sqrt{\overline{x^2(t)}} = \frac{a}{2} \tag{11c}$$

Let us deal now with an example of correlation functions for a rectangular pulse according to Figure 6(a). The

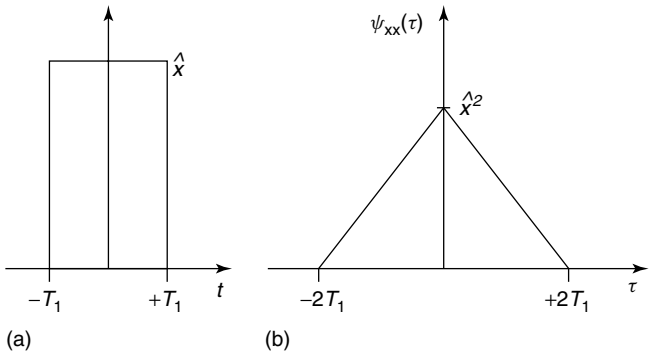


Figure 6. (a) Rectangular pulse and (b) corresponding autocorrelation function.

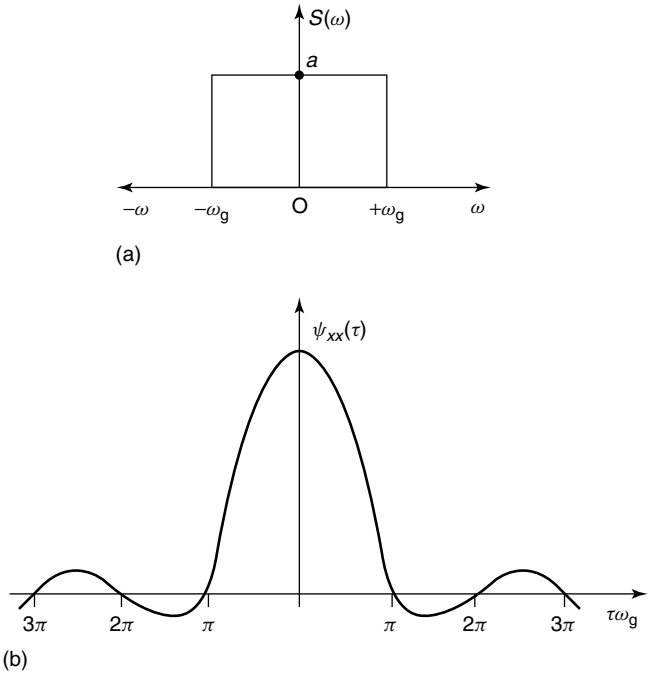


Figure 7. (a) Narrowband noise and (b) corresponding autocorrelation function.

autocorrelation function is given in (12).

$$\psi_{xx}(\tau) = \frac{1}{2T_1} \int_{-T_1+\tau}^{+T_1} \widehat{X} \times \widehat{X}\, \mathrm{d}t = \widehat{X}^2 \left(1 - \frac{|\tau|}{2T_1}\right) \quad (12)$$

The result shown in Figure 6(b) reveals that if the pulse width $2T_1$ decreases in the limiting case $T_1 \to 0$, that is, for the unit impulse $\delta(t)$, the autocorrelation function is also a delta function. This also follows immediately from the calculation of the autocorrelation function of white noise having a constant power spectrum $S_{xx}(\omega) = \text{constant}$, as is the case for unit impulse. For this calculation, (5a, b) are used.

The relationship between the width of the power spectrum and the corresponding autocorrelation function can also be seen if the autocorrelation function for narrowband noise is calculated in accordance with Figure 7(a):

$$\psi_{xx}(\tau) = a \int_{-\omega_g}^{+\omega_g} \mathrm{e}^{-\mathrm{j}\omega t}\, \mathrm{d}\omega = 2a\omega_g \frac{\sin(\omega_g \tau)}{\omega_g \tau} \quad (13)$$

As shown in Figure 7(b), the autocorrelation function becomes smaller as the noise bandwidth increases, degenerating into a delta function for white noise.

This shows clearly that statistical relationships cease to exist if two noise signals are slightly displaced relative to each other.

REFERENCES

Davies, W.D.T. (1970) *System Identification for Self-Adaptive Control*, Wiley, Chichester.

Woschni, E.-G. (1988) *Informations Technik*, Huthig, Heidelberg.

FURTHER READING

Buck, J.R., Daniel, M.M. and Singer, A.C (2001) *Computer Explorations in Signals and Systems Using MATLAB*, 2nd edn, Prentice Hall, NJ.

Grover, D. and Deller, J. (1998) *Digital Signal Processing and the Microcontroller*, Prentice Hall PTR, NJ.

Hayes, M.H. (1998) *Schaum's Outline of Digital Signal Processing*, McGraw-Hill Trade, New York.

Hsu, H.P. and Hse, H.P. (1995) *Schaum's Outline of Signals and Systems*, McGraw-Hill Trade.

Karu, Z.Z (1995) *Signals and Systems Made Ridiculously Simple*, ZiZi Press, Huntsville, AL.

Oppenheim, A.V., Willsky, A.S. and Hamid, S. (1996) *Signals and Systems*, Prentice Hall, NJ.

Williamson, D. (1999) *Discrete-Time Signal Processing: An Algebraic Approach*, Springer-Verlag, Berlin.

Ziemer, R.E., Tranter, W.H. and Fannin, D.R (1998) *Signals and Systems: Continuous and Discrete*, Prentice Hall, NJ.

29: Relationship Between Signals in the Time and Frequency Domain

Eugen Georg Woschni
Technical University Chemnitz, Chemnitz, Germany

1 Nature of Relationships and Transforms 163
2 Fourier Transform as the Means to Move Between the Domains 163
3 Theorems of Fourier Transform 164
4 Table of Time–Frequency Relationships 164
5 Example of Time–Frequency Transform 166
6 Fast Fourier Transform Concept and its Application 167
References 167
Further Reading 167

1 NATURE OF RELATIONSHIPS AND TRANSFORMS

To show the main ideas of time–frequency transform, the square wave time signal, given in Figure 1, may be considered.

Figure 1(a) shows the fundamental harmonic and the harmonic of first order with three times the basic frequency. The addition of these two frequencies as a first approximation already yields a similar function (Figure 1b). If one takes into consideration all harmonics, a reasonably exact square wave signal appears, which shows that the addition of frequencies leads to the appropriate time function. This means that the amplitude spectrum of sinusoidal functions, as given in Figure 2, is the equivalent representation in the frequency domain of Figure 1 in the time domain.

2 FOURIER TRANSFORM AS THE MEANS TO MOVE BETWEEN THE DOMAINS

As given in detail in **Article 27, Signals in the Frequency Domain, Volume 1**, the mathematical relation between the time and frequency domain and vice versa is given by the Fourier transform

$$\underline{\widehat{X}}(\mathrm{j}\omega) = \int_{t=-\infty}^{+\infty} x(t)\mathrm{e}^{\mathrm{j}\omega t}\,\mathrm{d}t = F\{x(t)\} \tag{1a}$$

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \underline{\widehat{X}}(\mathrm{j}\omega)\mathrm{e}^{\mathrm{j}\omega t}\,\mathrm{d}\omega = F^{-1}\{\widehat{X}(\mathrm{j}\omega)\} \tag{1b}$$

With regard to mathematics, this is a transformation between the original range – physically the time domain – and the image range – physically the frequency domain. This transformation is bidirectional, which means that it is possible to move into the frequency domain if the time domain is given and vice versa, as the example in Section 5 of this article demonstrates.

The fact that the convergence of this transformation is limited is known so that, especially in system theory, another transformation with no restriction due to the convergence is used, the Laplace transformation –

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

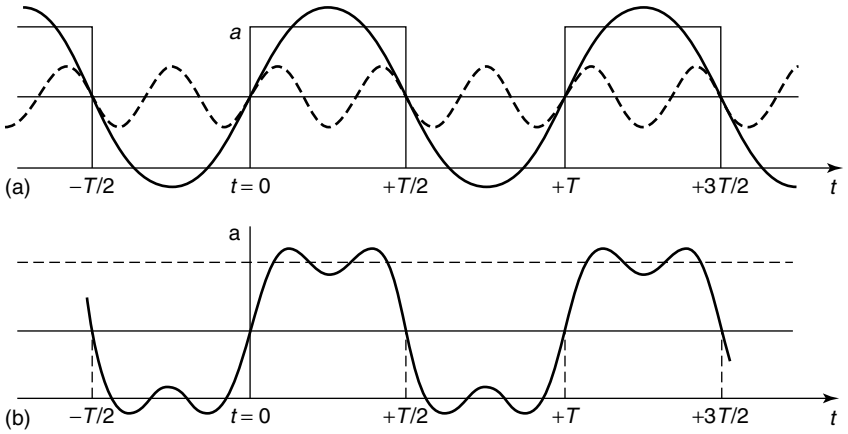


Figure 1. Square wave time signal (a) time signal with harmonics and (b) addition of the fundamental and first harmonic.

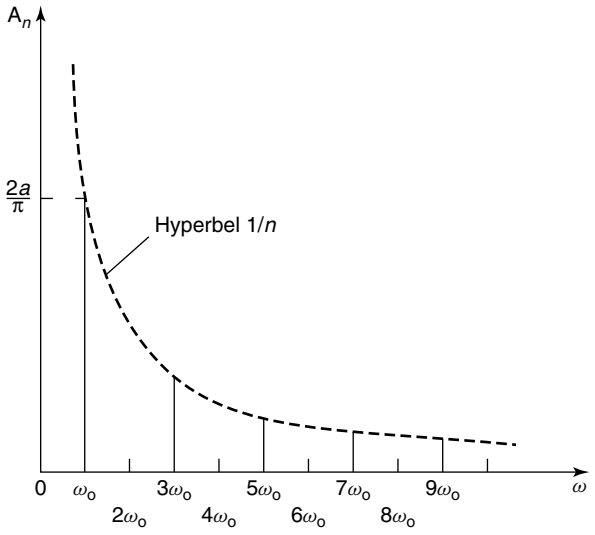


Figure 2. Amplitude spectrum of the time function of Figure 1(a).

see **Article 37, Relationship Between Systems in the Time and Frequency Domain, Volume 1**.

3 THEOREMS OF FOURIER TRANSFORM

A survey of the theorems of the Fourier transform is given in Table 1 and these correspond to the theorems of Laplace transformation as treated in **Article 37, Relationship Between Systems in the Time and Frequency Domain, Volume 1**.

Of great importance are the differentiation and the integration theorems. Corresponding to the differentiation of a sinusoidal function $\mathrm{d}/\mathrm{d}t\, e^{j\omega t} = j\omega e^{j\omega t}$, the multiplication with $j\omega$ in the frequency domain means a differentiation in the time domain and appropriate the division with $j\omega$ an integration. In **Article 36, Systems in the Frequency Domain, Volume 1** and **Article 37, Relationship Between Systems in the Time and Frequency Domain, Volume 1**, these results are used to describe the behavior of systems. The shifting theorem is the key to solve problems of discrete signal theory and system theory as sampled data systems as well as prediction or delay elements.

4 TABLE OF TIME–FREQUENCY RELATIONSHIPS

A summary is now given of the relationships established so far between the various signal representations in both the time domain and the frequency domain. Consideration is also given to the possibilities of conversion indicated above. Table 2 contains a survey of such relationships.

While the time function $x(t)$ and the spectral amplitude density $\widehat{X}(j\omega)$ contain the full information concerning the signal, this is not the case for the functions resulting from averaging (autocorrelation function and power density). Phase information is lost due to averaging. For this reason, conversions are possible only in the direction indicated by the arrow.

Conversions are possible, via the Fourier transform, between functions in the time domain and corresponding functions in the frequency domain. Therefore, it does not matter, as far as the significance of the statement is concerned, which of these two functions is measured; rather it is a question of convenience. For instance, it is useful to carry out measurements in optical communications predominantly in the time domain. In vibration measuring technology, however, it is preferable to carry out the measurements in the frequency domain.

The importance of the conversions has already been discussed. Mention should again be made of the fact that $x(t)$ and $X(j\omega)$ are related in the same way as are $\psi_{xx}(\tau)$

Table 1. Theorems of Fourier Transform.

(1) Addition theorem

$$\mathbf{F}\{f_1(t) + f_2(t)\} = \mathbf{F}\{f_1(t)\} + \mathbf{F}\{f_2(t)\}$$

(2) Multiplication theorem

$$\mathbf{F}\{af(t)\} = a\mathbf{F}\{f(t)\}$$

(3) Shifting theorem

$$\text{for } a > 0 \quad \mathbf{F}\{f(t-a)\} = \mathrm{e}^{-\mathrm{j}\omega a} F(\mathrm{j}\omega)$$

$$\text{or } \mathbf{F}\{f(t+a)\} = \mathrm{e}^{\mathrm{j}\omega a}\left(\mathbf{F}(\mathrm{j}\omega) - \int_0^a \mathrm{e}^{-\mathrm{j}\omega a} f(t)\,\mathrm{d}t\right)$$

(4) Likeness theorem

$$\mathbf{F}\{f(at)\} = \frac{1}{a}\mathbf{F}\left(\frac{p}{a}\right) \text{ if } a > 0$$

(5) Attenuation theorem

$$\mathbf{F}\{\mathrm{e}^{-at} f(t)\} = F(\mathrm{j}\omega + a)$$

(6) Limit theorem

$$\lim_{t\to\infty} f(t) = \lim_{\mathrm{j}\omega\to 0} \mathrm{j}\omega F(\mathrm{j}\omega) \lim_{t\to 0} f(t) = \lim_{\mathrm{j}\omega\to\infty} \mathrm{j}\omega F(\mathrm{j}\omega)$$

(7) Integration theorem

$$\mathbf{F}\left\{\int_0^t f(\tau)\,\mathrm{d}\tau\right\} = \frac{1}{\mathrm{j}\omega}\{f(t)\}$$

(8) Differentiation theorem

$$\mathbf{F}\{f^{(n)}(t)\} = \mathrm{j}\omega^n \mathbf{F}\{f(t)\} - \mathrm{j}\omega^{n-1} f(+0) - \cdots - f^{(n-1)}(+0)$$

if the limits

$$\lim_{t\to 0} f(t) = f(+0);\ \lim_{t\to 0} \dot{f}(t) = \dot{f}(+0); \ldots;$$

$$\lim_{t\to 0} f^{(n-1)}(t) = f^{(n-1)}(+0), \text{exist}$$

Note: Woschni, 1988.

Table 2. Relationships between time- and frequency-domain signals.

Time domain			Frequency domain	
	Function of time		Amplitude density	
Real function of a real variable	$x(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \widehat{X}(\mathrm{j}\omega)\mathrm{e}^{\mathrm{j}\omega t}\,\mathrm{d}\omega$	$\Longleftarrow F \Longrightarrow$	$\widehat{X}(\mathrm{j}\omega) = \int_{-\infty}^{+\infty} x(t)\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t$	Complex function of a real variable
Mean value operation; phase information is lost	$\psi_{xx}(\tau) = \lim_{\tau\to\infty}\frac{1}{2T}\int_{-T}^{+T} x(t)x(t+\tau)\,\mathrm{d}t$	$\downarrow$	$S_{xx}(\omega) = \frac{1}{2\pi}\lim_{T\to\infty}\frac{\lvert\widehat{X}(\mathrm{j}\omega)\rvert^2}{2T}$	Only unilateral conversion possible
	Autocorrelation function			
Real function of a real variable	$\psi_{xx}(\tau) = \int_{-\infty}^{+\infty} S_{xx}(\omega)\mathrm{e}^{\mathrm{j}\omega\tau}\,\mathrm{d}\omega$	$\Longleftarrow F \Longrightarrow$	$S_{xx}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \psi_{xx}(\tau)\mathrm{e}^{-\mathrm{j}\omega\tau}\,\mathrm{d}t$	Real function of a real variable

Note: Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.

and $S_{xx}(\omega)$. For instance, a constant amplitude density has the delta function, which is a time function, as an autocorrelation function, just as the constant power density in white noise has, *see also* **Article 28, Signals in the Time Domain, Volume 1**.

5 EXAMPLE OF TIME–FREQUENCY TRANSFORM

As an example with great relevance to practical application, let us deal with two signals used as test signals: The unit step $\omega(t)$ and the unit pulse or Delta function $\delta(t)$ as shown in Figure 3(a, b).

The spectral amplitude density of the unit pulse function is 1

$$\underline{\widehat{X}}(\mathrm{j}\omega)_{\vartheta} = F\{\delta(t)\} = \int_{-\infty}^{+\infty} \delta(t)\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t = 1 \qquad (2)$$

This means that only cosine functions with constant spectral amplitude appear, as also follows from the physical interpretation; only, in this case at $t = 0$ all spectral amplitude will be strengthened leading at least to the Dirac pulse (Figure 4a). The amplitude density of the unit step function is

$$\underline{\widehat{X}}(\mathrm{j}\omega)_{\omega} = F\{\omega(t)\} = \int_{0}^{+\infty} 1.\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t = 1 \qquad (3a)$$

This integral does not converge, so we have to solve the following integral

$$\lim_{\Delta\to 0}\int_0^{\infty} \mathrm{e}^{-\Delta t}\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t = \lim_{\Delta\to 0}\left[\frac{1}{-(\Delta+\mathrm{j}\omega)}\mathrm{e}^{-(\Delta+\mathrm{j}\omega)t}\right]_0^{\infty} = \frac{1}{\mathrm{j}\omega} \qquad (3b)$$

and we get a hyperbolic spectrum as shown in Figure 4(b), containing only sine functions: the step function is an uneven function so that only a sum of sine functions could approximate the course of the time function.

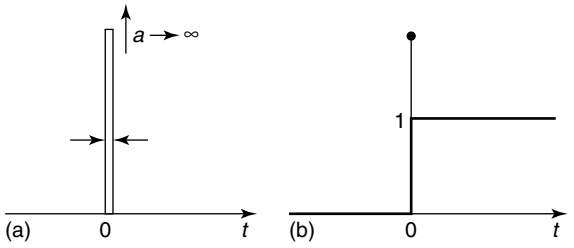


Figure 3. (a) Unit pulse function and (b) unit step function.

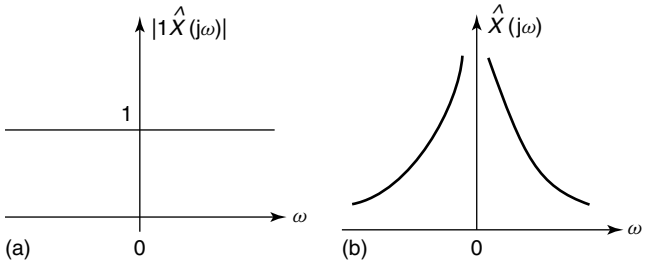


Figure 4. (a) Spectrum of a unit pulse function and (b) spectrum of a unit step function.

Another typical example is treated in **Article 28, Signals in the Time Domain, Volume 1**, Section 5.

6 FAST FOURIER TRANSFORM CONCEPT AND ITS APPLICATION

To accomplish the Fourier transform today, digital computers are used. The basis for this is the Fast Fourier Transform (FFT). Using a transform specially tailored to fit the way in which the digital computer operates, Discrete Fourier Transform (DFT) programs have been established, which save computing time, and where at least 1024 graph support points are quite usual. The calculation supplies the Fourier coefficients. As the number of support points is limited, the short-term correlation function is determined, which however, is practically identical with the correlation function, provided the correlation time is $\tau_k < T_s$.

To show the main ideas of FFT, we start from the equation of the Fourier transformation (1a), writing this equation for a sampled input signal with the sampling time T_s.

$$X^*(t) = \sum_{k=-\infty}^{+\infty} x(kT_s)\underline{\widehat{X}}^*(\mathrm{j}\omega) = T_s \sum_{k=-\infty}^{+\infty} x(kT_s)\mathrm{e}^{-\mathrm{j}\omega kT_s} \tag{4a}$$

Because it is not possible to use an infinite number of sampling values, a time boundary is realized, which means that instead of $k_{max} = \infty$ the maximum value of k is limited to N and instead of (4a) we get

$$\underline{\widehat{X}}^*(\mathrm{j}n\omega_0) = T_s \sum_{k=-\frac{N}{2}}^{+\frac{N}{2}} x(kT_s)\mathrm{e}^{-\mathrm{j}kn\omega_0 T_s} \tag{4b}$$

or with the fundamental frequency $\omega_0 = 2\pi/NT$

$$\underline{\widehat{X}}^*(\mathrm{j}n\omega_0) = \underline{\widehat{X}}^*\left(\mathrm{j}\frac{2\pi n}{NT_s}\right) = T_s \sum_{k=-\frac{N}{2}}^{+\frac{N}{2}} x(kT_s)\mathrm{e}^{-\frac{\mathrm{j}2\pi nk}{N}} \tag{4c}$$

From (4c) N complex Fourier coefficients arise. To get these coefficients due to (4c), N^2 multiplications are necessary. If $N = 2^S$ – that means N is a power of 2 – it is possible by means of division in subsumes and using an appropriate algorithm – the so-called Cooley–Tuckey Algorithm (Cooley and Tuckey, 1965) – to need only $N/2$ lb($N/2$) multiplications. For the often used value of $N =$ 1024, the number of multiplications thus can be reduced from 1048576 to 4608, which means that the calculation time is reduced to 1/227 of the original value!

Before microelectronics and microprocessors were developed, the application of Fourier transform was limited to special cases because it was extremely expensive to realize the transformation using analog methods. The classical application example in the beginning of the 1960s has been the testing of gears at the Volkswagen company using a spectral analyzer implemented in the analog technique. Today, the field of application is very broad. Signal analyzer, test equipment, and software tools now routinely make use of the FFT algorithms. *See also* **Article 143, DFT and FFTs, Volume 3**.

REFERENCES

Cooley, J.W. and Tuckey, J.W. (1965) An Algorithm for the Machine Calculation of Complex Fourier Series. *Mathematical Computation*, **19**, 279–301.

Woschni, E.-G. (1988) *Informations Technik*, Huthig, Heidelberg.

FURTHER READING

Buck, J.R., Daniel, M.M. and Singer, A.C. (2001) *Computer Explorations in Signals and Systems Using MATLAB*, 2nd edn, Prentice Hall, NJ.

Grover, D. and Deller, J. (1998) *Digital Signal Processing and the Microcontroller*, Prentice Hall PTR, NJ.

Hayes, M.H. (1998) *Schaum's Outline of Digital Signal Processing*, McGraw-Hill Trade, New York.

Tolimieri, R., An, M. and Lu, C. (1997) *Algorithms for Discrete Fourier Transform and Convolution*, 2nd edn, Springer, New York.

Williamson, D. (1999) *Discrete-Time Signal Processing: An Algebraic Approach*, Springer-Verlag, Berlin.

Woschni, E.-G. (1994) Signal Theory in Measurement and Instrumentation, *Concise Encyclopedia of Measurement & Instrumentation*, Pergamon Press, Oxford.

30: Statistical Signal Representations

Eugen Georg Woschni
Technical University Chemnitz, Chemnitz, Germany

1 Probability Distribution 168
2 Probability Density and Relation to Probability Distribution 168
3 Multidimensional Probability Functions 169
4 Gaussian Distribution 170
5 Relationship to Mean Values; Ergodic Theorem 170
6 Practical Investigations 171
7 Example of Use in Metrology 172
References 172

1 PROBABILITY DISTRIBUTION

In the following text, it is assumed that the probability functions are not dependent on time. In this situation, the process is a stationary stochastic one. In most cases – for instance, with noise – this proposition is met. If the stochastic parameters are changing relatively slowly, the process can be approximated by a stationary one for a determined time (Mendenhall and Sincich, 1992).

To describe randomly fluctuating events $\xi(t)$, use is made of characteristic functions, which are based on the theory of probabilities.

The probability distribution $W(x)$, which is also called the first-order distribution function, indicates the probability p that the signal $\xi(t)$ remains smaller than a barrier x,

$$W(x) = p[\xi(t) < x] \tag{1a}$$

as shown in Figure 1 with

$$W(x) = \frac{\sum \Delta t_r}{T} \tag{1b}$$

The limiting values of the probability distribution

$$\lim_{x \to -\infty} W(x) = 0 \tag{1c}$$

$$\lim_{x \to +\infty} W(x) = 1 \tag{1d}$$

are also clearly understandable, because they mean the impossibility of a value smaller than $-\infty$ as well as the certainty of the occurrence of any signal value $\xi(t) < \infty$. For continuous functions $\xi(t)$, the probability distribution $W(x)$ is a monotonically increasing function, see also Figure 3.

2 PROBABILITY DENSITY AND RELATION TO PROBABILITY DISTRIBUTION

Now we are interested in the probability $p(x)$ of the function $\xi(t)$ to be within a small region Δx near the value x

$$x \leq \xi(+) < x + \Delta x \tag{2a}$$

As Figure 2 shows, this probability is

$$p(x) = \frac{\sum \Delta t_r^*}{T} \tag{2b}$$

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

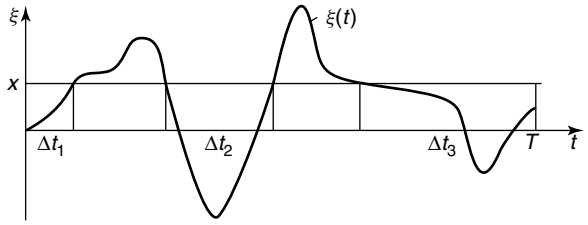


Figure 1. The definition of probability distribution.

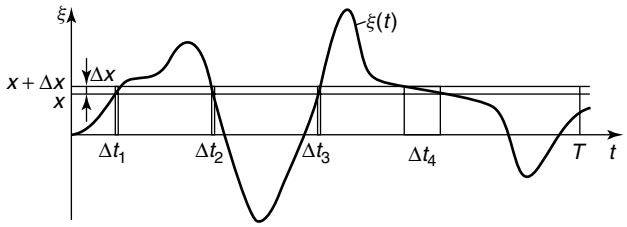


Figure 2. The definition of probability density.

This function is also a relation of Δx. Therefore, one relates to Δx and gains the probability density $w(x)$

$$w(x) = \frac{1}{\Delta x} p[x \leq \xi(t) < x + \Delta x]; \Delta x \to \mathrm{d}x \qquad (3a)$$

When comparing the Figures 3(a) and 3(b) it can be seen that the following relations exist

$$\int_{-\infty}^{x} w(u)\,\mathrm{d}u = W(x) \qquad (3b)$$

and

$$\frac{\mathrm{d}W(x)}{\mathrm{d}x} = w(x) \qquad (3c)$$

Taking into consideration the limit (1d), the normalization is

$$\int_{-\infty}^{+\infty} w(x)\,\mathrm{d}x = 1 \qquad (3d)$$

As shown in Figure 3, the probability that $\xi(t)$ lies within the interval x_2 to x_1 is calculated by

$$p[x_1 \leq \xi < x_2] = W(x_2) - W(x_1) = \int_{x_1}^{x} w(x)\,\mathrm{d}x \qquad (3e)$$

3 MULTIDIMENSIONAL PROBABILITY FUNCTIONS

If the stochastic process consists of not only one, but n fluctuating events $\xi(t)_1 \ldots \xi(t)_n$, then the same considerations

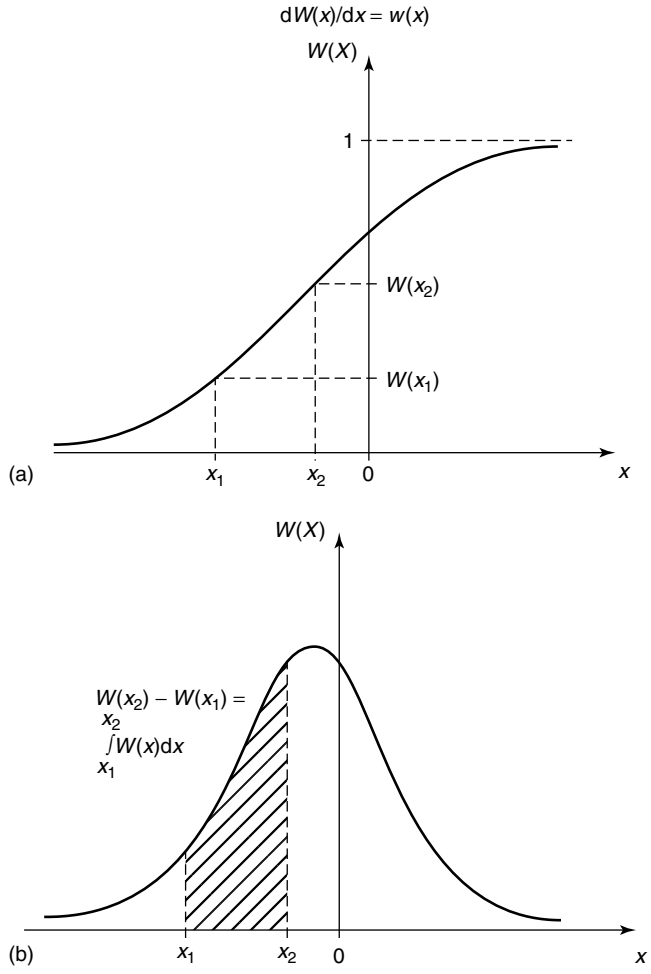


Figure 3. (a) Probability distribution and (b) corresponding probability density.

lead to n probability distribution functions or probability density functions with the limits $x_1 \ldots x_n$

$$W_1(x) \ldots W_n(x) \qquad (4)$$

and

$$w_1(x) \ldots w_n(x) \qquad (5)$$

To describe the relations between these functions, the compound probability distribution densities $w(x_1, x_2, \ldots, x_n)$ are introduced

$$W(x_1, x_2, \ldots, x_n) = p[\xi_1(t) < x_1, \xi_2(t) < x_2, \ldots, \xi_n(t) < x_n] \qquad (6a)$$

$$w(x_1, x_2, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \partial x_2, \ldots, \partial x_n} W(x_1, x_2, \ldots, x_n) \qquad (6b)$$

Furthermore, conditional probability distributions $W(x_1/x_2)$ and conditional probability distribution densities $w(x_1/x_2)$ are defined. They indicate the probability that the value x_1 occurs on condition that the value x_2 already exists. The following relationships hold for the compound probability density:

$$w(x, y) = w(x|y) \times w(y) = w(y|x)w(x) \qquad (6c)$$

4 GAUSSIAN DISTRIBUTION

Of utmost importance in practice is the Gaussian distribution density (Buja, 1991; Jaffe, 2000)

$$w(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(x-a)^2}{2\sigma^2}\right) \qquad (7a)$$

where $a = \overline{x(t)}$ is the linear mean value and σ the standard deviation, related to the square mean value, $\overline{x^2(t)}$, by

$$\sigma = \sqrt{[\overline{x^2(t)} - a^2]} \qquad (7b)$$

Figure 4 shows the Gaussian distribution density for $a = 0$.

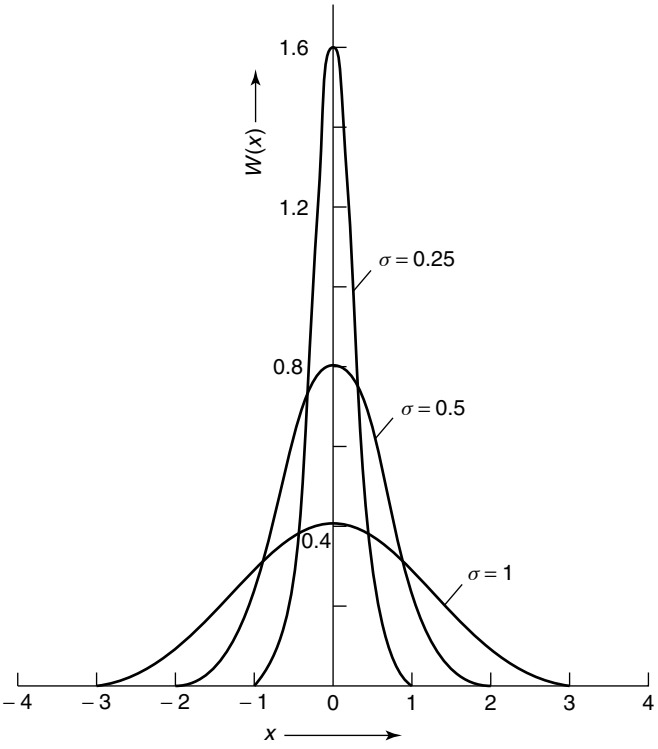


Figure 4. Gaussian distribution density.

Table 1. Gaussian probability integral $\Phi(x)$ evaluated for a range of x (Jahnke and Emde, 1966).

x	0	2	4	6	8
0.0	0.0000	0.0226	0.0451	0.0676	0.0901
0.1	0.1125	0.1348	0.1569	0.1790	0.2009
0.2	0.2227	0.2443	0.2657	0.2869	0.3079
0.3	0.3286	0.3491	0.3694	0.3893	0.4090
0.5	0.5205	0.5379	0.5549	0.5716	0.5879
0	0.0000	0.2227	0.4284	0.6039	0.7421
1	0.8427	0.9103	0.9523	0.9763	0.9891
2	0.9953	0.9981	0.9993	0.9998	0.9999

Source: Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.

The Gaussian distribution describes all stochastic processes represented by a very great number of single processes running independent of each other in the same overall event. Typical examples are the arrival of rain droplets or, in electronics, the thermal noise.

Very useful in the application of Gaussian distribution is the Gauss error integral in the form

$$\Phi(x) = \frac{2}{\pi} \int_0^x \mathrm{e}^{-\mu^2}\, \mathrm{d}\mu \qquad (7c)$$

with the values given in Table 1.

Some examples of importance in measurement are treated in Section 7.

5 RELATIONSHIP TO MEAN VALUES; ERGODIC THEOREM

The expectation value E of a function $f(x)$ is defined as follows:

$$E\{f(x)\} = \int_{-\infty}^{+\infty} f(x)w(x)\,\mathrm{d}x \qquad (8a)$$

For $f(x) = x^n$, the moment M_n of nth order is obtained. Of particular importance is the moment of first order, denoted as statistical linear mean value $\tilde{x}$:

$$\tilde{x} = M_1 = E\{x\} = \int_{-\infty}^{+\infty} x\,w(x)\,\mathrm{d}x \qquad (8b)$$

and the statistical square mean value phase space average:

$$\tilde{x}^2 = M_2 = E\{x^2\} = \int_{-\infty}^{+\infty} x^2\,w(x)\,\mathrm{d}x \qquad (8c)$$

Accordingly, the expectation value for the simultaneous occurrence of $x_1(t)$ and $x_2(t+\tau)$ is obtained:

$$\begin{aligned} &x_1(t)x_2(t+\tau) \\ &= E\{x_1(t), x_2(t+\tau)\} \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x_1(t)x_2(t+\tau)w[x_1(t), x(t+\tau)]\mathrm{d}x_1\mathrm{d}x_2 \end{aligned} \qquad (8d)$$

If this is an ergodic event, that is, if the ergodic theorem is satisfied, the statistical mean values $\tilde{x}^n$ and time mean values $\overline{x^n(t)}$ are equal

$$E\{x^n\} = \tilde{x}^n = \int_{-\infty}^{+\infty} x^n w(x)\,\mathrm{d}x = \overline{x^n(t)} = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{+T} x^n(t)\,\mathrm{d}t \qquad (9a)$$

With (8d), a definition can be obtained of the correlation function $\psi_{xy}(\tau)$, which is based on the generalized mean value:

$$\psi_{xy}(\tau) = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{+T} x(t)y(t+\tau)\,\mathrm{d}t = \overline{x(t)y(t+\tau)} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x(t)y(t+\tau)w[x(t),\, y(t+\tau)]\,\mathrm{d}x\,\mathrm{d}y \qquad (9b)$$

For the particularly important Gaussian distribution density according to (7a), one calculates

$$M_1 = \tilde{x} = \overline{x(t)} = \int_{-\infty}^{+\infty} \frac{x}{\sqrt{2\pi}\sigma}\exp\left(\frac{-(x-a)^2}{2\sigma^2}\right)\mathrm{d}x = a \qquad (9c)$$

$$M_2 = \tilde{x}^2 = \overline{x^2(t)} = \int_{-\infty}^{+\infty} \frac{x^2}{\sqrt{2\pi}\sigma}\exp\left(\frac{-(x-a)^2}{2\sigma^2}\right)\mathrm{d}x = a^2 + \sigma^2 \qquad (9d)$$

6 PRACTICAL INVESTIGATIONS

In order to display the probability distribution and density, electronic majority decision elements having an adjustable threshold value x are used, as outlined in Figure 5. Explanation in terms of analog processing assist appreciation of the process used.

This can be done in analog systems using either a triggering circuit or a voltage divider having a biased diode. With this, a normalization is to be carried out such that the corresponding conditions, (1d) and (3d), are observed. The arrangement has been for long also used for displaying these parameters on an oscilloscope screen, provided the voltage of the sweep generator for the x-deflection of the oscilloscope is used to control the threshold value x. The sweep frequency must be slow enough to ensure that adequate averaging occurs.

By coupling several installations in accordance with Figure 5(b), compound probability distributions $W(x, y)$ can also be recorded. For this purpose, the trigger outputs of one arrangement for each event x, y will be connected with an AND element and further processed as shown in Figure 5(b) (Woschni, 1988). Today, displays use digital processing to perform the task following digitization of the signal in an analog to digital (A/D) converter.

In practice it is possible to describe many events, at least approximately, by the Gaussian distribution of (7a). The probability for the fluctuation process to lie within the range $-x \le \xi(t) < +x$, or $a - x \le \xi(t) < a + x$, where a is a constant, is

$$p[-x \le \xi(t) < +x] = \frac{1}{\sqrt{2\pi}\sigma}\int_{-x}^{+x}\exp\left(\frac{-\xi^2}{2\sigma^2}\right)\mathrm{d}\xi \qquad (10a)$$

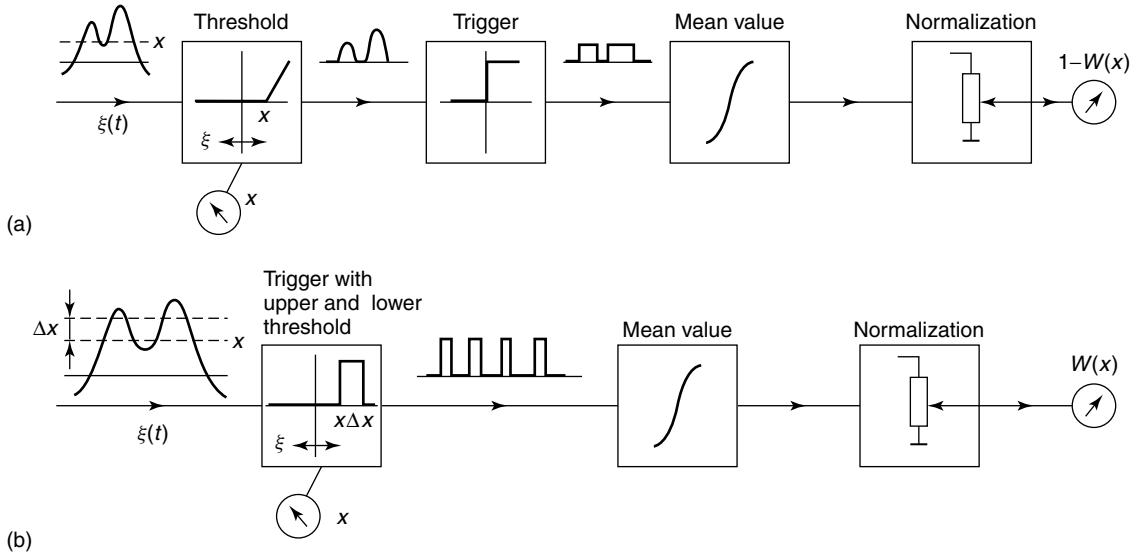


Figure 5. (a) Registration of the probability distribution and (b) probability density.

Table 2. Values of the example.

ϕ/rad	0.31	0.32	0.30	0.29	0.29	0.30	0.28	0.31	0.31	0.29
t/s^{-1}	2.39	2.40	2.41	2.43	2.40	2.38	2.39	2.41	2.37	2.42

To evaluate this, the probability integral is used, which is tabulated in the following form (Jahnke and Emde, 1966):

$$\phi(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-u^2)\, \mathrm{d}u \qquad (10b)$$

Extracts are shown in Table 1, see Section 4.

Some examples with importance in measurement will be treated in Section 7.

7 EXAMPLE OF USE IN METROLOGY

Where a number of measured values x_r are given, we are first interested in obtaining the standard deviation σ of the distribution as given by Hart, Lotze and Woschni (1997)

As an example, consider the rotation of the arm of an industrial robot being tested for preciseness of placement to the same position. The measured repeated values of the angle ϕ/rad and the time t s^{-1} are as shown in Table 2.

Using (10a) the standard deviation of the angle is $\sigma = 1.247 \times 10^{-2}$ rad and of the time $\sigma = 1.826 \times 10^{-2}$ s.

The mean value of the angular velocity is $\overline{\omega} = \overline{\varphi}/\overline{t} = 0.3/2.40 = 0.1250\,\mathrm{rad\ s^{-1}}$.

As an example, it may have been found by measurement that the length of workpieces having an average value of $a = 10\,\mathrm{cm}$ satisfies a Gaussian distribution and shows a standard deviation using (10a) of $\sigma = 3\,\mathrm{mm}$. What matters then might be the number of workpieces that lie within an admissible tolerance range of $10 \pm 4\,\mathrm{mm}$. Evaluation according to (10) shows that with the values of Table 1, some 82% of the pieces are within the tolerance range, and that the remaining pieces lie outside this range.

REFERENCES

Buja, A. (ed) (1991) *Computing and Graphics in Statistics*, Springer, New York, Berlin.

Hart, H., Lotze, W. and Woschni, E.-G. (1997) *Meßgenauigkeit*, Oldnbourg-Verlag, Munchen.

Jaffe, R.C. (2000) *Random Signals for Engineers using MATLAB and MATHCARD*, AIP Press, New York.

Jahnke, E. and Emde, F. (1966) *Tables of Higher Functions*, Teubner, Stuttgart.

Mendenhall, W. and Sincich, T. (1992) *Statistics for Engineering and the Sciences*, Dellen Publ. The Free Press, San Francisco, CA.

Woschni, E.-G. (1988) *Informationstechnik*, Huthig, Heidelberg.

31: Discrete Signal Theory

Eugen Georg Woschni

Technical University Chemnitz, Chemnitz, Germany

1 Discrete Time Sequences	173
2 Sampling Theorem	173
3 Aliasing Errors	175
4 Leakage Distortion	176
5 Frequency Domain Sampling	177
6 Discrete Fourier, Laplace, and Z-Transform	177
7 Spectral Analysis of Signal Samples	178
8 Examples of Signal Processing	179
References	180
Further Reading	181

1 DISCRETE TIME SEQUENCES

The principle of gaining a discrete time sequence $x^*(t)$ by sampling the original analog signal $x(t)$ is shown in Figure 1. There, the signal $x(t)$ is multiplied with a switching-function $s(t)$ leading to the sampled signal sequence $x^*(t)$. This signal is then converted into a digital form using the methods of coding as treated in **Article 33, Coding Theory and its Application to Measurement, Volume 1**.

The basic mathematical formulation is

$$x^*(t) = x(t)s(t) \tag{1a}$$

Thus, by means of pulse-amplitude-modulation, the original analog signal is transformed into a digital signal. The switching function may be written as

$$s(t) = \sum_{r=0}^{\infty} \delta(t - rT_s) \tag{1b}$$

with the sampling time T_s.

The time sequence that results is given by

$$x^*(t) = x(t) \sum_{r=0}^{\infty} \delta(t - rT_s) = \sum_{r=0}^{\infty} x(rT_s) \tag{1c}$$

The coding of the time sequences $x(rT_s)$ into a digital signal is treated in **Article 33, Coding Theory and its Application to Measurement, Volume 1**. The inverse process, to gain $x(t)$ again if $x^*(t)$ is known is shown in Figure 2. This smoothing function can be realized in electrical systems by means of a capacitor connected to a voltage holding circuit.

A key design issue is to decide how many samples are necessary to retain all of the information of the signal $x(t)$. This consideration leads to the sampling theorem, treated in the next section.

Another important issue is what happens if the sampling theorem is violated. In this case, aliasing errors appear, as treated in the relevant section of this article. Here also the signal processing following the sampling stage has to be taken into consideration as shown in this section.

2 SAMPLING THEOREM

There exist exact mathematic foundations for the sampling theorem (Johnson, 1989; Oppenheim and Schafer, 1989). Here, however, another approach using physical explanations is used because this is very suitable for dealing with the problem of aliasing errors.

If the spectral amplitude density $\underline{\hat{X}}(\mathrm{j}\omega)$ given in **Article 27, Signals in the Frequency Domain, Volume 1** is limited to the critical frequency ω_c as shown in Figure 3 we get, applying the Fourier transform, the corresponding

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.

© 2005 John Wiley & Sons, Ltd.

ISBN: 0-470-02143-8.

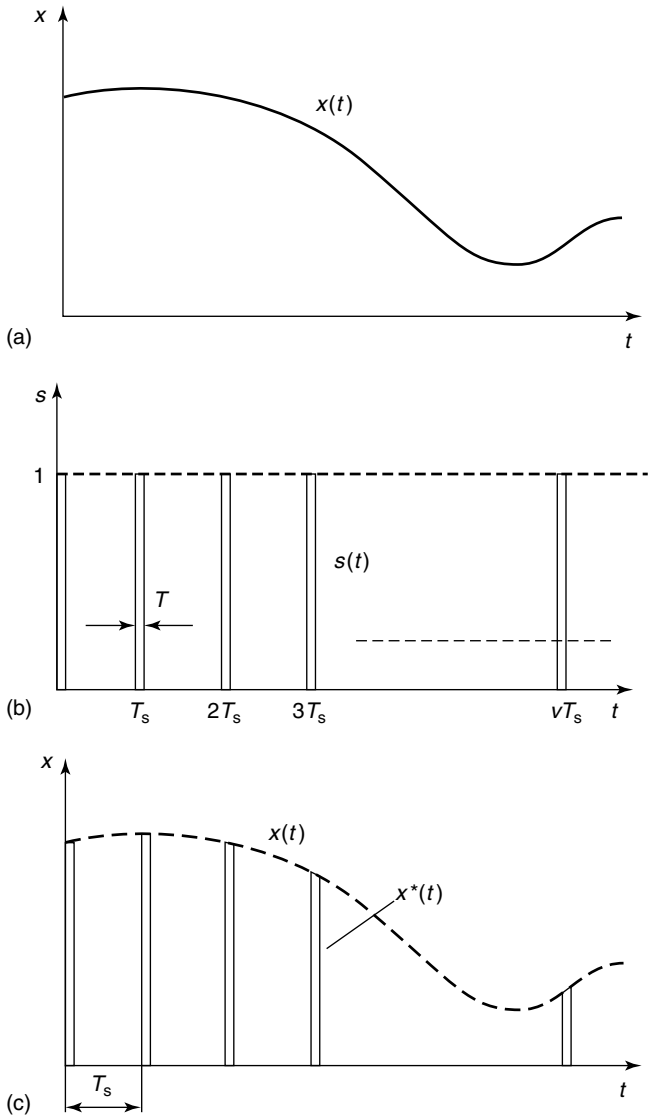


Figure 1. Principle of gaining the discrete signal sequence by means of sampling.

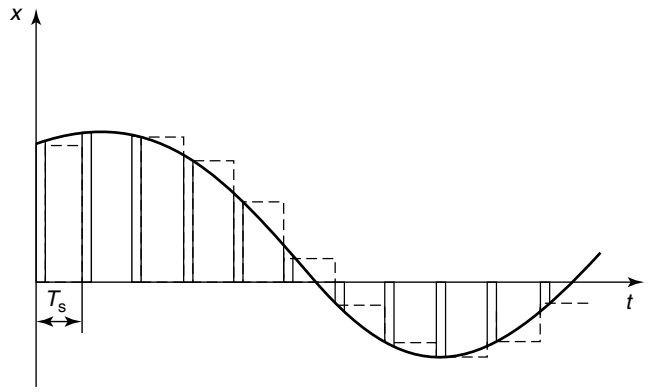


Figure 2. Smoothing with holding circuit.

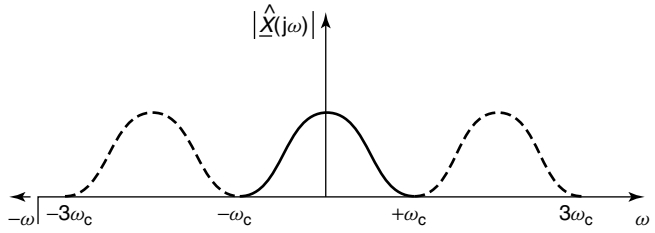


Figure 3. Band-limited spectral amplitude density.

time function $x(t)$ (Woschni, 1988).

$$x(t) = \frac{1}{2\pi} \int_{-\omega_c}^{+\omega_c} \hat{\underline{X}}(j\omega) e^{j\omega t} \, d\omega \tag{2a}$$

The spectral amplitude density function may be periodically continued as marked with dashed lines in Figure 3. Expansion of the series of this periodical frequency function yields

$$\underline{\hat{X}}_{\text{period}}(j\omega) = \frac{1}{2\,\omega_c} \sum_{r=-\infty}^{+\infty} x\left(\frac{r\pi}{\omega_c}\right) e^{-\frac{jr\pi\omega}{\omega_c}} \tag{2b}$$

In this series, the only values are

$$x\left(\frac{r\pi}{\omega_c}\right) = x(rT_s) \tag{2c}$$

Applying the sampling time (from the sampling theorem)

$$\frac{r\pi}{\omega_c} = rT_s \tag{2d}$$

where $T_s = \pi/\omega_c = 1/2f_c$

In reality, the amplitude density is not a band-limited entity as required by the sampling theorem. It is, therefore, necessary to carry out low-pass filtering before sampling called *antialiasing filtering*. Figure 4 shows that this leads to a cutoff error calculated due to system theory, where $S_{xx}(\omega)$ is the spectral power density of the signal (Woschni, 1988)

$$\overline{\varepsilon^2} = 2 \int_{\omega_c}^{\infty} S_{xx}(\omega) \, d\omega \tag{2e}$$

This error will be treated in detail together with the aliasing error in the next section.

Another approach also demonstrates the main points and leads to the aliasing errors if the sampling theorem is violated. We are starting with the switching function $s(t)$ due

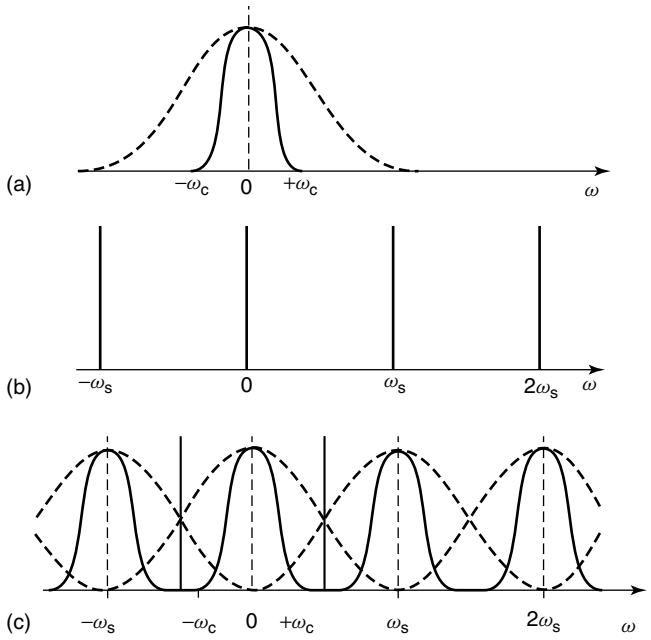


Figure 4. Explanation of sampling in the frequency domain.

to (1b). This periodic time function means a spectral function as shown in Figure 4(b).

$$s(t) = \frac{\tau}{T_s} \sum_{r=-\infty}^{+\infty} e^{jr\omega_s t} \tag{2f}$$

The multiplication with $x(t)$ due to (1a) represents an amplitude modulation as treated in **Article 34, Modulation Theory, Volume 1**. Figure 4(a) shows the spectral amplitude density of the signal $x(t)$, and the amplitude modulation has the spectral amplitude density as given in Figure 4(c). While in the case of a band-limited signal (thick line in Figure 4c), the spectra are separated from each other, in the case marked with dashed lines the spectra overlap. Here, the sampling theorem is violated and aliasing errors arise.

3 ALIASING ERRORS

Different error definitions are used in measurement. As shown in **Article 32, Geometrical Signal Representations, Volume 1** and especially in **Article 50, Errors in Signal Systems, Volume 1** the mean-square error introduced by Gauss has the advantage that it leads to closed mathematical solutions because of the validity of Euclidean geometry.

If the sampling theorem is fulfilled, the cutoff error due to the low-pass filtering is given by (2e). As shown in

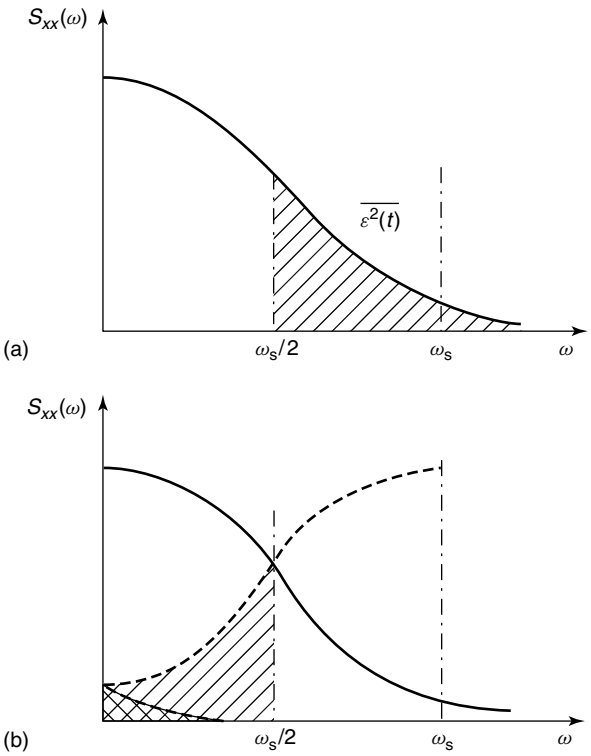


Figure 5. Calculation of the sampling errors in the frequency domain.

Figure 5(a), this error means the power of the cutoff part of the spectral power density of the signal $S_{xx}(\omega)$.

Figure 5 clearly shows the overlapping of the lower side band falling into the frequency range $0 \ldots \omega_s/2$ and leading to aliasing error.

$$\overline{\varepsilon^2} = 2 \int_{\frac{\omega_s}{2}}^{\infty} S_{xx}(\omega)\, d\omega \tag{3a}$$

Because the frequencies of the lower side band are mirrored, one can suppose the two error components – the cutoff and the aliasing errors – are not correlated to each other unless they originate from the same signal source. This means that the total sampling error is given by the sum of both components.

In measurement, these errors originate if it is not possible to realize a low-pass filtering before sampling. This case is typically for sensors with direct digital output, as, for instance, in the case treated in **Article 33, Coding Theory and its Application to Measurement, Volume 1**, because there a low-pass filtering is not feasible. These errors are becoming of decreasing importance because the development of microelectronics leads to ever higher pulse frequencies and then oversampling is possible (Woschni, 2001).

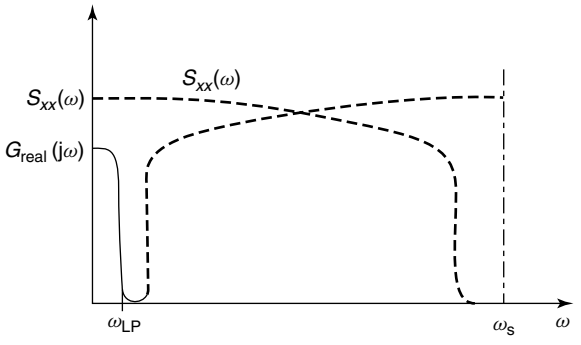


Figure 6. Mean-value operation after sampling.

Another example that is important in measurement science is the influence of a signal processing after sampling. A typical case arises in surface measurement. Here, instead of the whole information of the surface profile, only the average roughness is of interest. As demonstrated in Figure 6, low-pass filtering ω_{LP} after sampling is necessary to gain the mean value. This means only half of the value of the sampling frequency due to the sampling theorem is to be used without the generation of errors (Woschni, 1996a,b).

4 LEAKAGE DISTORTION

Another problem of analog-to-digital conversion is that of windowing. The signal function $x(t)$ has to be time limited, for example, by means of a time window instead of the infinite function. A time-limited function is used called the *truncation window*. This truncation in time leads to a variation of the spectrum, called *leakage distortion* (Thrane, 1979).

As shown in Figure 3, in the time domain a periodic time function appears leading to cutoff errors known as *leakage errors*. The main statements are demonstrated in Figure 7.

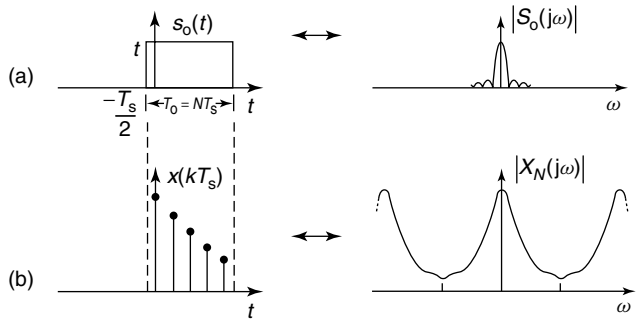


Figure 7. (a) Truncation window and its transform; (b) truncated sequence and its transform.

The truncated sampled signal function may be written as (Figure 7b)

$$x(kT_s) \quad k = 0, 1, 2, \ldots, N-1 \tag{4a}$$

This is the product of the infinite sequence $x_s(kT_s)$ and a rectangular time domain truncation window $s_o(t)$ as shown in Figure 7(a).

$$x_s(kT_s) = x(kT_s)s_o(t) \tag{4b}$$

With the switching function, as treated in Section 1 of this article, (1a)

$$x_s(kT_s) = x(t)s_o(t)s(t) \tag{4c}$$

The Fourier transform of this yields

$$X_N(j\omega) = X(j\omega)^* S_o(j\omega)^* S(j\omega) \tag{4d}$$

With the modulus of the Fourier transform of a rectangular time window being as treated in detail in Section 7

$$|S_o(j\omega)| = \mathrm{si}(\pi\alpha) = \frac{(\sin\pi\alpha)}{\pi\alpha} \tag{4e}$$

The results obtained mean that in Figure 7(a) the time window leads to the spectrum given by (4e) and the convolution with the spectrum of $x_s(kT_s)$ results in the spectrum demonstrated in Figure 7(b) where a rippling distortion can be seen.

Figure 8 shows the case of discrete components within the signal spectrum $X_s(j\omega)$. Convolution with the function $S_o(j\omega)$ would be replaced by a si-function due to (4e), which is explained later.

In general, the amount of leakage distortion depends on the length of the truncation window in comparison to the sampling interval as the following example shows.

A typical example is now given (Hart, Lotze and Woschni, 1997). A sinusoidal signal may be sampled with a normalized time truncation window Tf_c. Using the methods given above the relative mean-square error F is calculated as a function of k – the sampling time T_s due to the sampling theorem related to a shorter real sampling time T_s' –

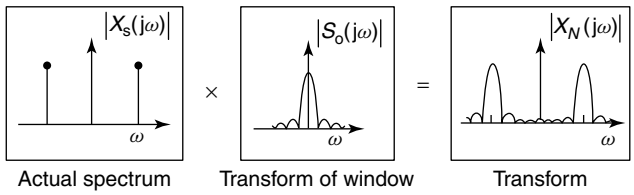


Figure 8. Leakage distortion of discrete components.

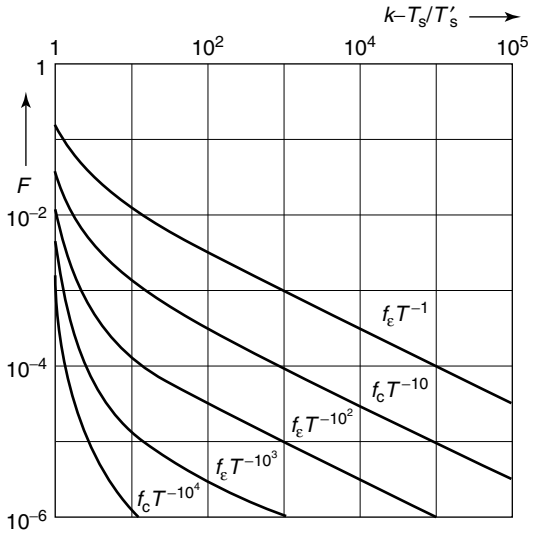


Figure 9. Errors as a function of the relative sampling time k with the relative truncation time as parameter.

Figure 9. It is evident that because of the leakage distortion shorter sampling times are necessary to avoid errors.

5 FREQUENCY DOMAIN SAMPLING

As treated in the last section, a truncated time sequence originates an infinite spectrum $X(\mathrm{j}\omega)$. Figure 7(b) demonstrates that a periodic spectrum appears. For computer application, this frequency function has to be sampled and limited to a finite number of samples. This procedure, therefore, is the same as the one realized in the last section; only time and frequency are now exchanged. Thus, here also the same problems of aliasing errors and leakage distortion arise.

Owing to (4a–d), $X(k\mathrm{j}\omega_\mathrm{s})$ may be the sampled frequency function gained from the original function $X(\mathrm{j}\omega)$ by means of a sampling function $S_\mathrm{f}(\mathrm{j}\omega)$ in the same way as in the time domain in Section 4

$$X(k\mathrm{j}\omega_\mathrm{s}) = X(\mathrm{j}\omega)S_\mathrm{f}(\mathrm{j}\omega) \tag{5a}$$

This infinite frequency function has to be frequency truncated to N numbers of samples by means of a frequency domain truncation window $S_\mathrm{o}(f)$ leading to the relation equivalent to (4c)

$$X_N(k\mathrm{j}\omega_\mathrm{s}) = X(\mathrm{j}\omega)S_\mathrm{o}(\mathrm{j}\omega)S_\mathrm{f}(\mathrm{j}\omega) \tag{5b}$$

This equation may be transformed to the time domain where the multiplication will be a convolution and $x(t)$ will be a

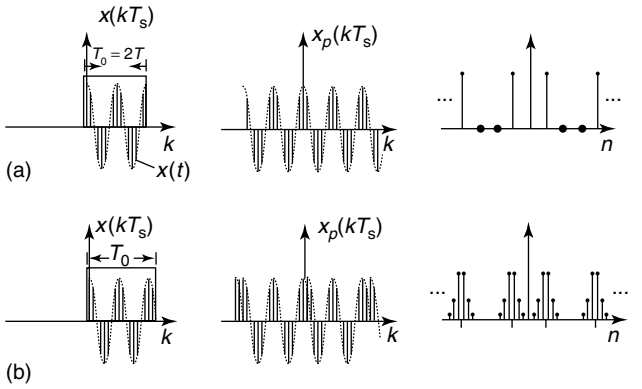


Figure 10. (a) Width of truncation window equal to two periods of $x(t)$; (b) width not equal to two periods.

periodic repetition of the sample function $x_p(kT_\mathrm{s})$

$$x_p(kT_\mathrm{s}) = x(t)^* s_\mathrm{o}(t)^* s_\mathrm{f}(t) \tag{5c}$$

So we get, in principle, the same relation between sampled spectrum and time function as shown in Figure 7(b), only now with a periodically repeated time function.

There exists an important case, found especially in measurement, when $x(t)$ is a periodic signal. Here, two different situations may be distinguished, as demonstrated in Figure 10. If the width of the truncation window contains an integral number of periods of the signal, then no error arises as Figure 10(a) shows. Otherwise errors appear (Figure 10b).

Other applications of frequency domain sampling concern the design of digital filters (Johnson, 1989).

6 DISCRETE FOURIER, LAPLACE, AND Z-TRANSFORM

As treated in **Article 27, Signals in the Frequency Domain, Volume 1**, between the time and the frequency domain there exists a relationship known as the *Fourier transform*. In the case of analog continuous signals, use of this transform yields (from (6b, c) of **Article 27, Signals in the Frequency Domain, Volume 1**)

$$\hat{X}(\mathrm{j}\omega) = \int_{-\infty}^{+\infty} x(t)\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t = F\{x(t)\} \tag{6a}$$

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{X}(\mathrm{j}\omega)\mathrm{e}^{\mathrm{j}\omega t}\,\mathrm{d}\omega = F^{-1}\{\hat{X}(\mathrm{j}\omega)\} \tag{6b}$$

In technical systems, it is reasonable to assume that the signals start at $t = 0$. Then, under the limit (6a) is zero

(so-called unilateral Fourier or Laplace transform)

$$\hat{X}(\mathrm{j}\omega) = \int_0^\infty x(t)\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t \tag{6c}$$

As shown in Section 1 of this article, instead of the time function $x(t)$ by means of sampling with T_s, the discrete time sequence

$$x^*(t) = x(t)\Sigma\delta(t - rT_s) = \Sigma x(rT_s) \tag{6d}$$

originates leading to the infinite sequence due to (6c)

$$\hat{X}^*(\mathrm{j}\omega) = \sum_{r=0}^{\infty} x(rT_s)\mathrm{e}^{-\mathrm{j}\omega rT_s} \tag{6e}$$

In Section 4, it is demonstrated that this function has to be time truncated. This means a limitation to N samples applies, that is, the function $x(t)$ is replaced by a periodic one. Instead of (6b), it follows

$$X(r) = \frac{1}{N}\sum_{k=0}^{N-1} X^*(k)\mathrm{e}^{\mathrm{j}(2\pi/N)kr} \tag{6f}$$

The Fourier coefficients are

$$X^*(k) = \sum_{k=0}^{N-1} X(r)\mathrm{e}^{\mathrm{j}(2\pi/N)kr} \tag{6g}$$

With the definition

$$W_N = \mathrm{e}^{-\mathrm{j}(2\pi/N)} \tag{6h}$$

The discrete Fourier transformation is (Oppenheim and Schafer, 1989)

$$X^*(k) = \sum_{r=0}^{N-1} X(r)W_N^{kr} \tag{6i}$$

$$X(r) = \frac{1}{N}\sum_{k=0}^{N-1} X^*(k)W_N^{-kr} \tag{6j}$$

Using this definition, the algorithms of Cooley and Tuckey (1965) now well known as Fast Fourier Transform (FFT) are introduced as treated in Section 6 of **Article 29, Relationship Between Signals in the Time and Frequency Domain, Volume 1**.

Using the same procedure, the Laplace transform may be applied to discrete signals leading to the relations – *see* **Article 37, Relationship Between Systems in the Time and Frequency Domain, Volume 1**.

$$X^*(s) = \sum_{r=0}^{\infty} X(rT_s)\mathrm{e}^{-rsT_s} \tag{6k}$$

$$X(rT_s) = \sum_{r=0}^{\infty} X^*(s)\mathrm{e}^{rsT_s} \tag{6l}$$

With the substitution

$$Z = \mathrm{e}^{sT_s} \tag{6m}$$

the Z-transformation results that is especially used in the theory of digital signal processing and filters – see Section 8 of this article.

7 SPECTRAL ANALYSIS OF SIGNAL SAMPLES

A periodic rectangular signal may be considered as shown in Figure 11. It is built by means of rectangular truncation windows in the time domain.

Using the method of Fourier spectrum as treated in Section 2 of **Article 27, Signals in the Frequency Domain, Volume 1** one obtains only cos functions with the coefficients

$$A_n = \frac{4}{T}\int_0^{\frac{\tau}{2}} a\cos n\omega_0 t\,\mathrm{d}t = \frac{4a}{Tn\omega_0}\sin\left(n\omega_0\frac{\tau}{2}\right) = \frac{2a\tau}{T}\mathrm{si}\left(n\omega_0\frac{\tau}{2}\right) \tag{7a}$$

or the complex Fourier coefficients

$$\hat{X}(\mathrm{j}n\omega_0) = a\int_{-\tau/2}^{+\tau/2} \mathrm{e}^{-\mathrm{j}n\omega_0 t}\,\mathrm{d}t = \frac{2a}{n\omega_0}\sin n\omega_0\frac{\tau}{2} = \frac{a\tau}{T}\mathrm{si}\left(n\omega_0\frac{\tau}{2}\right) \tag{7b}$$

The spectrum consists of harmonics with an envelope well known as the si(x) function that is demonstrated in

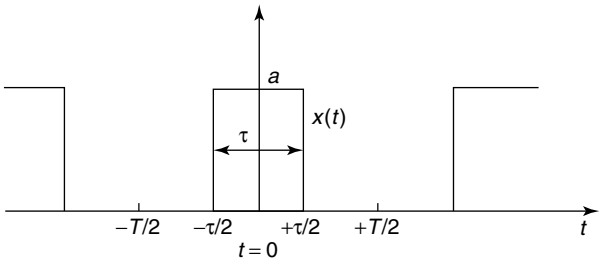


Figure 11. Periodic rectangular signal.

Figure 12. The smaller the pulse duty factor τ/T is the more harmonics that fall into the region up to the first zero point.

Because of the existence of the side maximums, the required spectrum is relatively wide. On the other hand, a truncation of the spectrum leads to a time function of $\mathrm{si}(x)$ form instead of the rectangular one of the original signal. To avoid these disadvantages, other time truncation windows with better performance, for example, smaller spectra are used (Johnson, 1989) or (Oppenheim and Schafer, 1989).

For instance, the bell-shaped window leading to the bell-shaped pulse is as shown in Figure 13. From this

$$x(t) = a\mathrm{e}^{-\beta^2 t^2} \tag{7c}$$

has a spectrum – using Fourier transform as treated in Section 4 of **Article 27, Signals in the Frequency Domain, Volume 1**

$$\begin{aligned}\underline{\hat{X}}(\mathrm{j}\omega) &= a\int_{t=-\infty}^{+\infty} \mathrm{e}^{-\beta^2 t^2}\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t \\ &= a\mathrm{e}^{-\frac{\omega^2}{4\beta^2}}\int_{-\infty}^{+\infty} \mathrm{e}^{-(\beta t-\mathrm{j}\omega 2\beta)^2}\,\mathrm{d}t \\ &= \frac{2a}{\beta}\mathrm{e}^{-\frac{\omega^2}{4\beta^2}}\int_0^{\infty} \mathrm{e}^{-x^2}\,\mathrm{d}x = \frac{a\sqrt{\pi}}{\beta}\mathrm{e}^{-\frac{\omega^2}{4\beta^2}}\end{aligned} \tag{7d}$$

This is also a bell-shaped spectrum having no side maximums and therefore leading to signals with a smaller required frequency range. Other time truncation windows

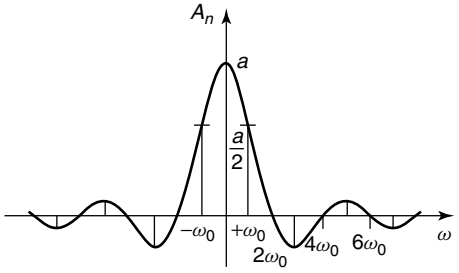


Figure 12. Spectrum of the periodic rectangular signal.

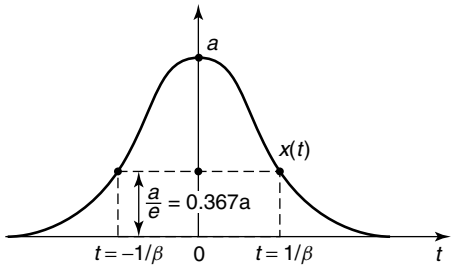


Figure 13. Bell-shaped pulse.

and their advantages are treated in (Johnson, 1989) or (Oppenheim and Schafer, 1989).

The method of periodograms may be useful here (Oppenheim and Schafer, 1989).

8 EXAMPLES OF SIGNAL PROCESSING

To illustrate the main methods of signal processing using the results of the investigations of this article, consider first the integration of a signal $x(t)$.

The first approximation is a rectangular one described by the algorithm

$$y(t) = y(t - T_A) + Tx(t) \tag{8a}$$

Using the discrete Fourier, Laplace, or Z-transform, the frequency response or Z-transform is

$$G_{\mathrm{R1}}(\mathrm{j}\omega) = T\frac{1}{1-\mathrm{e}^{-\mathrm{j}\omega T}} \tag{8b}$$

$$G_{\mathrm{R1}}(s) = T\frac{1}{1-\mathrm{e}^{-sT}} \tag{8c}$$

$$G_{\mathrm{R1}}(z) = T\frac{z}{z-1} \tag{8d}$$

The next better approximation is a trapezoidal one with the algorithm

$$y(t) = y(t - T) + \frac{T}{2}[x(t) + x(t - T)] \tag{8e}$$

leading to the frequency functions

$$G_{\mathrm{R2}}(s) = \frac{T}{2}\frac{1+\mathrm{e}^{-sT}}{1-\mathrm{e}^{-sT}} \tag{8f}$$

$$G_{\mathrm{R2}}(z) = \frac{T}{2}\frac{z+1}{z-1} \tag{8g}$$

$$G_{\mathrm{R2}}(\mathrm{j}\omega) = \frac{T}{2}\frac{1+\mathrm{e}^{-\mathrm{j}\omega T}}{1+\mathrm{e}^{-\mathrm{j}\omega T}} \tag{8h}$$

The frequency response of the ideal integration is

$$G(\mathrm{j}\omega) = \frac{1}{\mathrm{j}\omega} \tag{8i}$$

As Figure 14 shows, the trapezoidal approximation is better because here only an amplitude error and no phase error appears. On the other hand, we have to pay for this advantage by using more processing time!

The mean-square error in using the results of **Article 50, Errors in Signal Systems, Volume 1** for the

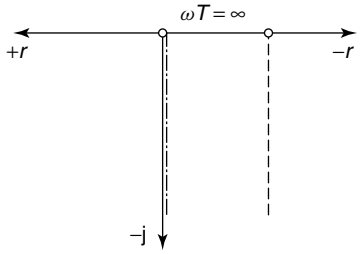


Figure 14. Frequency responses of integration algorithms; —— ideal; - - - - rectangular.; —·—·—· trapezoidal.

rectangular integration and a model signal $S(x) = S_o/(1 + \omega/\omega_0)$ (Woschni, 1996a,b) is given by

$$\overline{\varepsilon^2} = 2\int_0^{\frac{x}{T}} \frac{S_o}{1+\left(\frac{\omega}{\omega_0}\right)} \left|\frac{1}{j\omega} - \frac{T}{1-e^{-j\omega T}}\right|^2 d\omega \qquad (8j)$$

Another example is the prediction problem. The ideal frequency response reads as follows:

$$G_R(s) = e^{sT},\ G_R(z) = z,\ G_R(j\omega) = e^{+j\omega T} \qquad (9a)$$

The first approximation is the linear continuation of $x(t)$, for example, the algorithm is

$$y(t) = x(t) + x(t) - x(t - T_s) = 2x(t) - x(t - T) \qquad (9b)$$

The frequency functions are, using the same method as before

$$G_{R1}(j\omega) = 2 - e^{-j\omega T} \qquad (9c)$$

$$G_{R1}(s) = 2 - e^{-sT} \qquad (9d)$$

$$G_{R1}(z) = 2 - z^{-1} \qquad (9e)$$

The next better approximation includes the two last values with the algorithm

$$y(t) = 3x(t) - 3x(t - T) + x(t - 2T) \qquad (9f)$$

leading to the frequency functions

$$G_{R2}(s) = 3 - 3e^{-sT} + e^{-2sT} \qquad (9g)$$

$$G_{R2}(z) = 3 - 3z^{-1} + z^{-2} \qquad (9h)$$

$$G_{R2}(j\omega) = 3 - 3e^{-j\omega T}s + e^{-j2\omega T} \qquad (9i)$$

Figure 15 shows the results where the general principle is evident. In general, the better the algorithm, that is, the smaller the error – the longer the time needed for processing. Owing to the development of ever faster microelectronic circuitry having increased processing performance good approximations are possible with smaller errors.

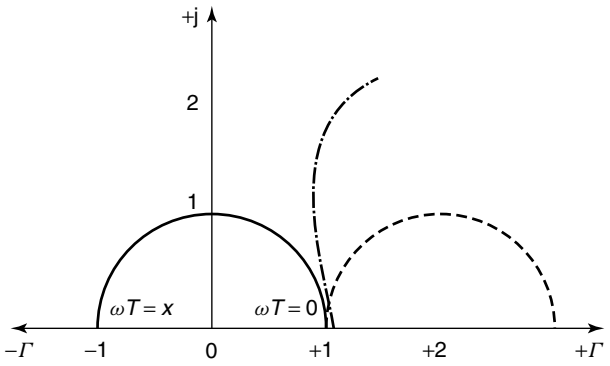


Figure 15. Frequency responses of prediction algorithms; —— ideal; - - - - first approx.; —·—·—· second approx.

Finally, it may be hinted at the limitation of the frequency range from $f = 0$ to $f = f_s/2$ due to the Sampling theorem. Furthermore, the processing error is a function of the word length s in bits of the code used. Between the signal-to-noise ratio S/N in dB and the word length s in bits there exist the general relation (Woschni, 1988)

$$S/N = 20 \log 1/2^s = -20\, s/^2 \log 10 = 6\, s \qquad (9j).$$

REFERENCES

Cooley, J.W. and Tuckey, J.W. (1965) An Algorithm for the Machine Calculation of Complex Fourier Series. *Mathematical Computation*, **19**, 279–301.

Hart, H., Lotze, W. and Woschni, E.-G. (1997) *Meßgenauigkeit*, Oldenboug-Verlag, Munchen 5 (p. 112).

Johnson, J.R. (1989) *Introduction to Digital Signal Processing*, Prentice Hall, Englewood Cliffs, NJ.

Oppenheim, A.V. and Schafer, R.W. (1989) *Discrete-Time Signal Processing*, Prentice Hall, Englewood Cliffs, NJ.

Thrane, N. (1979) The Discrete Fourier Transform and FFT Analysers, *Bruel and Kjaer Technical Review*, **1**, 3–25.

Woschni, E.-G. (1988) *Informations Technik*, Huthig, Heidelberg.

Woschni, E.-G. (1996a) Signal Processing after Sampling – Today's Problems and Solutions in Future, *Proceedings of ISMTII'96*, Hayama Kanagawa Pref., Japan.

Woschni, E.-G. (1996b) Signal Processing after Sampling – Today's Problems and Solutions in Future, *Proceedings of ISMTII'96*, Hayama Kanagawa Pref., Japan.

Woschni, E.-G. (2001) Measurement in Robotics – Trends due to the Development of Microelectronics Proceedings, *10th*

International Workshop on Robotics RAAD 2001, Techn. Universitat, Wien.

FURTHER READING

Buck, J.R., Daniel, M.M. and Singer, A.C. (2001) *Computer Explorations in Signals and Systems Using MATLAB*, 2nd edn, Prentice Hall, NJ.

Grover, D. and Deller, J. (1998) *Digital Signal Processing and the Microcontroller*, Prentice Hall PTR, NJ.

Hayes, M.H. (1998) *Schaum's Outline of Digital Signal Processing*, McGraw-Hill Trade, New York.

Karu, Z.Z. (1995) *Signals and Systems Made Ridiculously Simple*, ZiZi Press, Huntsville, AL.

Miller, M.J. (1982) Discrete Signals and Frequency Spectra, in *Handbook of Measurement Science*, Vol. 1 (ed. P.H. Sydenham), Wiley, Chichester.

Williamson, D. (1999) *Discrete-Time Signal Processing: An Algebraic Approach*, Springer-Verlag, Berlin.

Woschni, E.-G. (1992) The Importance of Estimation and Approximation Methods in System Theory. *Cybernetics and Systems*, **23**, 335–343.

Ziemer, R.E., Tranter, W.H. and Fannin, D.R (1998) *Signals and Systems: Continuous and Discrete*, Prentice Hall, NJ.

32: Geometrical Signal Representations

Eugen Georg Woschni
Technical University Chemnitz, Chemnitz, Germany

1 Representation of Signals in Euclidean Space 182
2 Representation of Signals in Non-Euclidean Space 183
3 Hamming Distance 183
4 Representation with Codegraphs 184
References 184
Further Reading 185

1 REPRESENTATION OF SIGNALS IN EUCLIDEAN SPACE

Signals with n components $x_1, \ldots, x_n$ can be represented by a signal vector $\mathbf{x}$ in the n-dimensional space:

$$\mathbf{x} = (x_1, x_2, x_3, \ldots, x_n) \tag{1a}$$

where the end point of the vector determines the corresponding signal.

For analog signals, use is made of Euclidean space, wherein the Pythagorean theorem applies. Thus, for the magnitude of the vector, also called the *norm* in geometry, one obtains

$$\|\mathbf{x}\| = \left(\sum_{r=1}^{n} x_r^2 \right)^{\frac{1}{2}} \tag{1b}$$

The various components are then the projections onto the different axes of the n-dimensional space, which is called the *signal space*.

The notion of the distance d between two or more signals is of great practical importance. This distance results as a norm of the difference between two signal vectors $\mathbf{x}$ and $\mathbf{y}$ as follows:

$$\mathrm{d}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \left(\sum_{v=1}^{n} |x_v - y_v|^2 \right)^{\frac{1}{2}} \tag{1c}$$

The scalar product

$$\mathbf{x} \times \mathbf{y} = \sum_{v=1}^{n} x_v y_v \tag{2a}$$

can be used to write the angle α between the two vectors in the following way:

$$\cos \alpha = \frac{\mathbf{xy}}{\|\mathbf{x}\| \|\mathbf{y}\|} \tag{2b}$$

Figure 1 explains the relations.

For continuous analog signals $x(t)$, which are defined in the range $a \leq t < b$, one can accordingly indicate a norm:

$$\|\mathbf{x}\| = \left(\int_a^b x^2(t)\, \mathrm{d}t \right)^{\frac{1}{2}} \tag{3a}$$

Physically, it represents the square root of the energy of the signal. A special Hilbert space is thus defined (Blumenthal, 1961). The distance between two signals corresponds to the root mean square error

$$\mathrm{d}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \left(\int_a^b |x(t) - y(t)|^2\, \mathrm{d}t \right)^{\frac{1}{2}} \tag{3b}$$

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

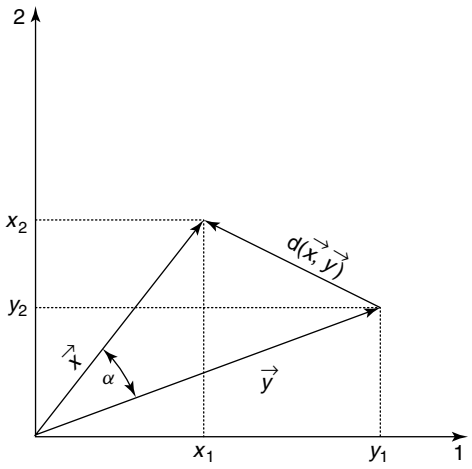


Figure 1. Two-dimensional model of the signal representation in Euclidean space.

It is often used in measurement technology and cybernetics as a measure for the error – *see* **Article 50, Errors in Signal Systems, Volume 1**.

2 REPRESENTATION OF SIGNALS IN NON-EUCLIDEAN SPACE

The above representation in the Euclidean space is suitable for analog signals. Use is made of a representation in non-Euclidean signal space for discrete signals whose importance is constantly increasing as digital systems replace evermore the analog ones. In this space, the distance between two vectors **x, y** is defined to be the sum of the differences of the individual components (Blumenthal, 1961)

$$\mathrm{d}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \sum_{v=1}^{n} |x_v - y_v| \qquad (4a)$$

according to the norm of this space:

$$\|\mathbf{x}\| = \sum_{v=1}^{n} x_v \qquad (4b)$$

The most important discrete signals are binary signals, that is, signals in which the individual components can assume the values 1 and 0 only. For such signals, the signal space constitutes an n-dimensional hypercube having the edge length 1, where the various edges are occupied by possible signals only.

Figure 2 shows such signal words with one, two, and three bits. Obviously, the representation in non-Euclidean space with the norm according to (4b) has the advantage that the distance d indicates the number of digits by which two signal words differ from each other. Similarly defined minimum distance in a signal alphabet is called the Hamming distance and constitutes an important characteristic value for the investigation into a system's sensitivity to noise (Peterson, 1962). To investigate distances between signals, use is also made of distance matrices as shown in the next section. In the analog-to-digital conversion of signals – *see* **Article 132, Data Acquisition Systems (DAS) in General, Volume 3** and **Article 139, Analog-to-Digital (A/D) Converters, Volume 3**, a transform between the corresponding signal spaces takes place.

3 HAMMING DISTANCE

The distance between two code words in Figure 1 is the minimum length that can be taken only along the axis x_1, x_2, x_3. Another possibility to obtain the distances uses the so-called distance matrix as demonstrated for the signals of Figure 2 in Table 1.

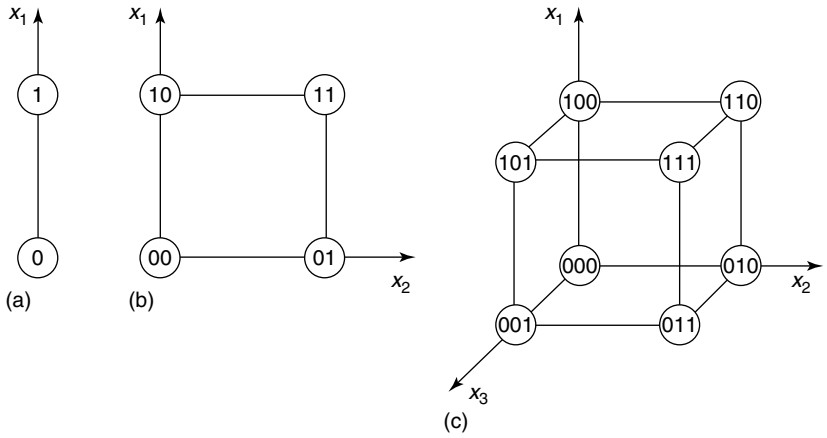


Figure 2. Representation of a binary signal in the signal space: (a) one bit; (b) two bits; (c) three bits.

Table 1. Distance matrices for the signals of Figure 2. a) $n = 1$; b) $n = 2$; c) $n = 3$.

a)

	1	0
	0	1
0	1	0

b)

	00	10	01	11
00	0	1	1	2
10	1	0	2	1
01	1	2	0	1
11	2	1	1	0

c)

	000	001	010	011	100	101	110	111
000	0	1	1	2	1	2	2	3
011	1	0	2	1	2	1	3	2
010	1	2	0	1	2	3	1	2
011	2	1	1	0	3	2	2	1
100	1	2	2	3	0	1	1	2
101	2	1	3	2	1	0	2	1
110	2	3	1	2	1	2	0	1
111	3	2	2	1	2	1	1	0

Because the distance d between two words is equal to the number of bits by which the two words differ from each other, the number of the possible code words between them is $d - 1$. If in an alphabet not all possible words are used, the minimal distance in this alphabet is called the *Hamming distance* $d_{\min}$ or d_{Hamming}. This means, for instance, for $d_{\min} = 2$, that between each word of the alphabet, one not used word is situated. From this fact it follows that 1-bit error leads to a not used word. So bit errors of first order f_{e1} can be noticed. If there are two vacant words between two used words – it means that $d_{\min} = 2$ – errors of second order f_{e2} can be noticed and so on, leading to the general relation between the order n of noticed errors and the Hamming distance

$$f_{en} = d_{\min} - 1 \tag{5}$$

Similar reflections lead to the relation between the order n of correctable errors f_{cn} and the Hamming distance $d_{\min}$. For instance, with $d_{\min} = 2$ in Figure 2, only the words 000 and 111 are used. So errors of first and second order lead to unused words and can be noticed due to (5). But the probability that the first-order error word belongs to the original word is very much higher than for the second-error word. Therefore, a correction can be done to correct the words with first-order error, the order of correction is 1. From this reflection there follows the general relation

$$f_{cn} = \frac{(d_{\min} - 1)}{2} \tag{6}$$

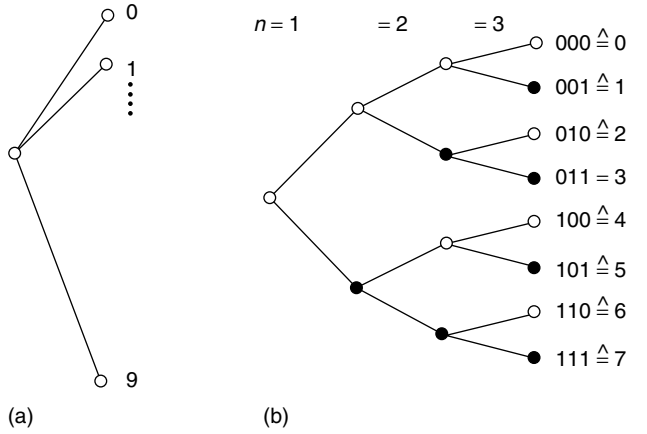


Figure 3. Representation as a codegraph: (a) signal with $m = 10$ possible states and (b) binary signal with the length $n = 3$.

Equation (6) – the application of the principle of an additional redundancy (Woschni, 1988; Shannon, 1948; Goldman, 1953) – is the key to designing data transmission lines with a very small error probability although the original transmission line is of low performance.

4 REPRESENTATION WITH CODEGRAPHS

For an easily understandable explanation of the signal structure, the codegraph may be used. On the basis of the results of the graph theory, the m possible states (symbols) of a signal are shown as an assembly according to Figure 3(a). Figure 3(b) shows the application of this type of representation to a binary signal having three digits. Where the individual code words are of the same length, they can be separated one from the other by counting. No characters are required for the separation of the individual words; the code is irreducible. Obviously, this is the case if the end points of the codegraph are occupied by code words only.

For more details, especially concerning code theory and its application to measurement, *see* **Article 33, Coding Theory and its Application to Measurement, Volume 1**.

REFERENCES

Blumenthal, L. (1961) *A Modern View of Geometry*, Freeman, San Francisco, CL.

Goldman, S. (1953) *Information Theory*, Prentice Hall, New York.

Peterson, W.W. (1962) *Error Correcting Codes*, MIT Press, Cambridge.

Shannon, C.E. (1948) A Mathematical Theory of Communication. *Bell System Technical Journal*, **27**, 379–423.

Woschni, E.-G. (1988) *Informations Technik*, Huthig, Heidelberg.

FURTHER READING

Hamming, R.W. (1986) *Coding and Information Theory*, Prentice Hall, NJ.

Hsu, H.P. and Hse, H.P. (1995) *Schaum's Outline of Signals and Systems*, McGraw-Hill Trade.

Karu, Z.Z. (1995) *Signals and Systems Made Ridiculously Simple*, ZiZi Press, Huntsville, AL.

Oppenheim, A.V., Willsky, A.S. and Hamid, S. (1996) *Signals and Systems*, Prentice Hall, NJ.

Ziemer, R.E., Tranter, W.H. and Fannin, D.R (1998) *Signals and Systems: Continuous and Discrete*, Prentice Hall, NJ.

33: Coding Theory and its Application to Measurement

Eugen Georg Woschni
Technical University Chemnitz, Chemnitz, Germany

1 Coders and Decoders in Measurement Systems 186
2 Coding Theory 186
3 Binary Codes Used in Measurement 187
4 Absolute Optical Encoders and Codes Used 188
5 Incremental Encoders and LSB 189
6 Error Correction by Redundant Codes 189
7 An Example of Error Correction Code Application 190
References 190
Further Reading 190

1 CODERS AND DECODERS IN MEASUREMENT SYSTEMS

The task of measurement as a part of communication is the transmission of a message from an information source to a receiver, as shown in Figure 1. The output of this information source may be a digital or a continuous signal. In measurement, it is the unknown quantity that is to be measured: the output is generated by a random mechanism having a probabilistic nature. Otherwise, the signal would be completely known and there would be no need to obtain the output by means of measurement. The signal coming from the information source is the input signal of the encoder, modulator, or transmitter. If the signal is digital, the term *encoder* is used, while for analog signals the term *modulator* is used. This subsystem is treated in **Article 34, Modulation Theory, Volume 1**. The overall function of this communication system is to transform the signal where there are several information sources by multiplexing them onto the same transmission channel, or to make the signal reasonably immune to unwanted disturbances.

The modulated or coded signal is transmitted by the channel or processed by the processor. This channel may be a microwave or uhf relay link, a wire or cable transmission, or an optical fiber, as are commonly used in measurement systems. Noise arising in the channel is considered to be additive.

In measurement, the signal is usually processed by a microprocessor containing a memory. As in the case of an analog system – still favored in some measurement applications – the behavior of this system may be described by a transfer function provided the computer program is a linear one (Woschni, 1988).

The next link in the serial chain is the decoder or demodulator, whose task is to reconstruct an estimate of the original signal as correctly as possible. In the sense of describing the signal as a vector in the signal space (*see* **Article 32, Geometrical Signal Representations, Volume 1**), it means that the distance between the output and input signals should be kept to a minimum. Many similarities link this problem to the problems of character recognition (Finkelstein, 1976).

2 CODING THEORY

As treated in Section 1 of this article, the first subsystem of a communication system is the encoder, converting the

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

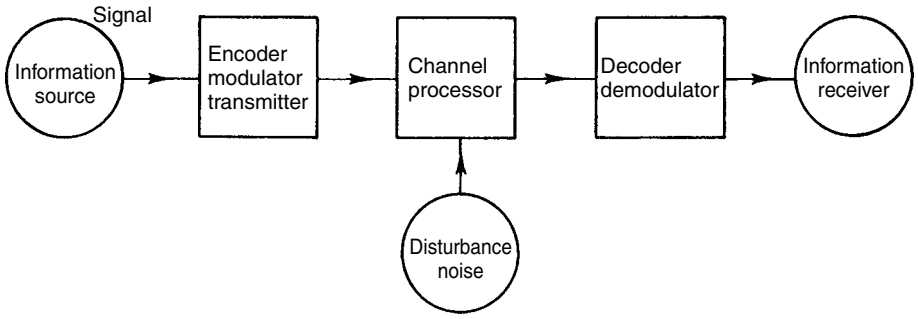


Figure 1. Communication system.

input signal into a series of code words. The task of this coding is the adaptation (interfacing) of the information source to the channel or processor.

In communication systems, the redundancy-diminishing (optimal) source coding, having the purpose of economizing the time needed for communication, often plays an important role. In measurements, however, security of the message against disturbances is the most needed criterion. Here, therefore, error-detecting or error-correcting codes are applied (Peterson, 1962).

For the representation of codes, geometrical descriptions or codegraphs are used in the n-dimensional space (*see* **Article 32, Geometrical Signal Representations, Volume 1**). Many coding and decoding theorems exist. The decoding theorem deals with the problem of identifying a code word by the receiver. For this purpose, the decoder compares the incoming code words with the words of the code alphabet deciding which code word the transmitter has sent. In the case in which the end of a code word is not marked by a special symbol, only the endpoints of the codegraph may be filled with a code word, otherwise a part of a code word would be another code word. The equation that guarantees this is

$$\sum_{k=1}^{K} M^{-l_k} \leq 1 \tag{1a}$$

where M is the number of symbols and l_k is the length of the kth code word. In (1a), the equality sign represents the most advantageous case without code redundancy. Otherwise, the factor c in the equation

$$c \sum_{k=1}^{K} M^{-l_k} = 1 \tag{1b}$$

is a measure of the code redundancy.

This decoding theorem is of great importance in measurement because codes with redundancy are often used, see Sections 6 and 7 of this article. The theorem of optimal coding plays a great role in redundancy-diminishing coding used in communication, giving an equation for the optimum length of the source code words l_k:

$$-\frac{\mathrm{lb}\, p_k}{\mathrm{lb} M} \leq l_k < -\frac{\mathrm{lb}\, p_k}{\mathrm{lb} M} + 1 \tag{1c}$$

where p_k is the probability of the appearance of the kth code word. Some important codes, including those applied in measurement, are now considered.

3 BINARY CODES USED IN MEASUREMENT

There are a lot of binary codes used in several fields of measurement. The data-transmission error-correcting codes are treated in Section 6 of this article.

The simplest code is the counting code, mostly seen in decimal counting. Today, this easily learnable code is displaced in machines that use the binary-coded decimal notation because of the smaller number of bits. As an example, in Figure 2, the 1-out-of-10 code is presented. For manual data coding (data input), a particular form of the binary code

$$\begin{aligned} Z_{\mathrm{bin}} &= A_n 2^n + A_{n-1} 2^{n-1} + \cdots + A_1 2^1 + A_0 2^0 \\ &= A_n A_{n-1} \cdots A_1 A_0 \end{aligned} \tag{2a}$$

the binary-coded decimal system or BCD code, is used. Here, digit-by-digit, the decimal number is converted into

	9	8	7	6	5	4	3	2	1	0
0	0	0	0	0	0	0	0	0	0	1
1	0	0	0	0	0	0	0	0	1	0
2	0	0	0	0	0	0	0	1	0	0
3	0	0	0	0	0	0	1	0	0	0
4	0	0	0	0	0	1	0	0	0	0
5	0	0	0	0	1	0	0	0	0	0
6	0	0	0	1	0	0	0	0	0	0
7	0	0	1	0	0	0	0	0	0	0
8	0	1	0	0	0	0	0	0	0	0
9	1	0	0	0	0	0	0	0	0	0

Figure 2. 1-out-of-10 code.

Table 1. BCD code in binary and excess-three notation.

Decimal code	Binary-coded decimal code		Excess-three notation	
	First tetrad	Second tetrad $x_3x_2x_1x_0$	First tetrad	Second tetrad $y_3y_2y_1y_0$
0	0000	0000	0000	0011
1	0000	0001	0000	0100
2	0000	0010	0000	0101
3	0000	0011	0000	0110
4	0000	0100	0000	0111
5	0000	0101	0000	1000
6	0000	0110	0000	1001
7	0000	0111	0000	1010
8	0000	1000	0000	1011
9	0000	1001	0000	1100
10	0001	0000	0100	0011
11	0001	0001	0100	0100
20	0010	0000	0101	0011
50	0101	0000	1000	0011
51	0101	0001	1000	0100
76	0111	0110	1010	1001
99	1001	1001	1100	1100

the binary code. For each digit, four bits, the so-called tetrad, is necessary (Table 1).

In binary notation, the numbers complement necessary for subtraction in a computer, sometimes leads to a nonexistent code word. The complement of 3 (=0011), for instance, is 1100, which does not exist in the BCD code (Table 1). This disadvantage is avoided by use of the notation of the BCD code, also presented in Table 1. To each decimal number, 3 is added. This code is well known as the excess-three code or the Stiebitz code.

As an example, the conversion of the excess-three code to the binary code is now considered. With the relationships between x and y shown, Table 1 yields in Boolean algebra form

$$x_0 = \bar{y}_0 \tag{2b}$$

$$x_1 = (y_0 \times y_1) + (\bar{y}_0 \times y_1) \tag{2c}$$

$$x_2 = (\bar{y}_1 \times \bar{y}_2) + (\bar{y}_0 \times \bar{y}_2) + (y_0 \times y_1 \times y_2) \tag{2d}$$

$$x_3 = (y_2 \times y_3) + (y_0 \times y_1 \times y_3) \tag{2e}$$

where

$\times$ is the logical AND operation

$+$ is the logical OR operation

The realization of these logic equations leads to the digital logic circuit of Figure 3.

Of importance in measuring systems are the Gray code and the improved variation of this code, the Glixon code

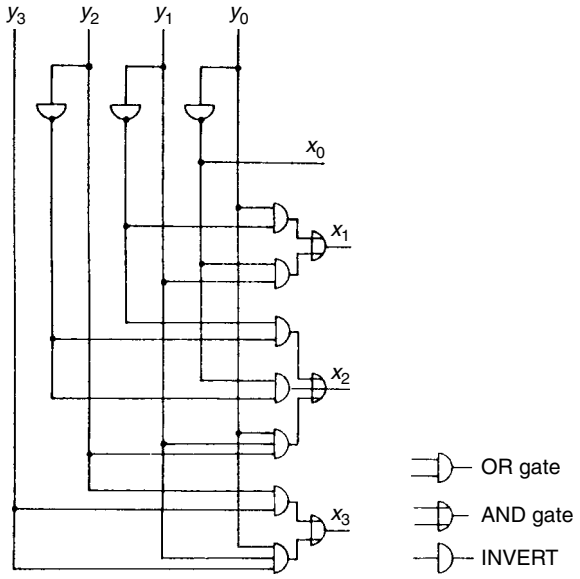


Figure 3. Circuit for conversion of the excess-three code to the binary code.

used especially with optical absolute encoders as described in the next section.

In data transmission, the ISO-CCIT code No. 5 or the teletype CCITT code No. 3 are used in combination with error-correcting codes as treated in Section 6.

4 ABSOLUTE OPTICAL ENCODERS AND CODES USED

There exist two types of encoders – the absolute and the incremental encoders (Carstens, 1993; Pallas-Areny and Webster, 2001). With the absolute encoders, each input has a unique output, while, with the incremental codes, only the difference between two inputs is given in the output (see also next section).

The principle of an optical encoder is seen in the digital shaft encoder demonstrated in Figure 4. It is used to transform the rotational angle directly into a computer recognizable signal. The same principle may be used with a linear encoding scale to transduce length. For them to work correctly, it must be emphasized that the sampling theorem has to be fulfilled in operation. Otherwise, aliasing errors occur, because normally it is not possible to realize the necessary low pass before sampling.

To minimize the errors when counting from one segment to the other, it is appropriate to have a code that differs only in one digit step from one segment to the next.

This group of codes is the reflected codes, arising by counting at first forward, that is, from 0 to 9, and then backward from 19 to 10. One of this group is the Gray code.

Table 2. Formation of the Gray code.

Decimal number	0	1	2	3	4	5	6	7
Binary number	000	001	010	011	100	101	110	111
Shifted binary number	000	001	010	011	100	101	110	111
Σ	0000	00011	0110	0101	1100	1111	1010	1001
Gray code	000	001	011	010	110	111	101	100

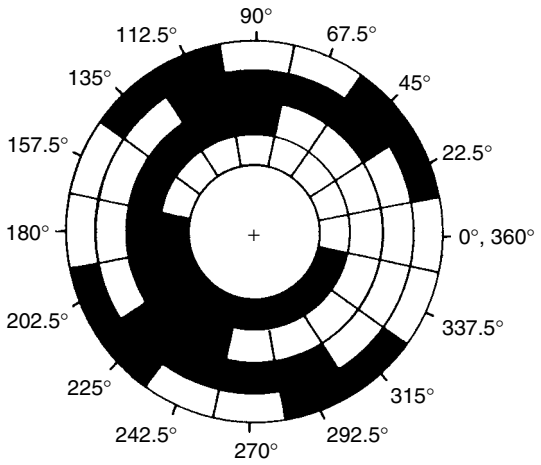


Figure 4. Coding disk (Gray code): code is read radially from outside inward.

Table 3. Glixon code.

Decimal number	Glixon code
0	0000
1	0001
2	0011
3	0010
4	0110
5	0111
6	0101
7	0100
8	1100
9	1000

It is obtained from the binary code as shown in Table 2. The advantage of this code is that any two code words following each other always have unit distance between them, that is, the two code words differ in one digit only. For this reason, this code is often used in measurement for encoding disks or linear encoding scales (Figure 4). The disadvantage of a distance greater than 1 occurring between 9 and 10 is avoided by using the improved Glixon code (Table 3).

5 INCREMENTAL ENCODERS AND LSB

Less expensive to make is the incremental encoder shown in Figure 5. Here, identical lines are counted as they pass

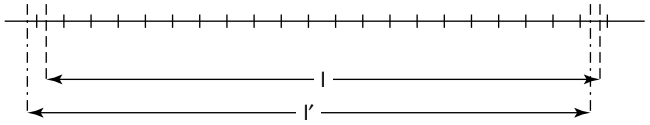


Figure 5. Incremental encoder for coding the length l.

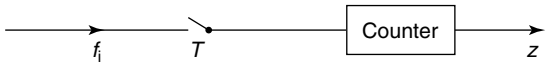


Figure 6. Measuring of frequency f or time T by counting z.

the detector. So counting is made relative to the datum of the starting position, but the starting datum itself is not known. If a count is lost or gained due to noise and so on, the output will be in error from that point on. These, however, have been made to be very reliable in application.

The same principle can be used either to code the length as shown in Figure 5 or the angle shown in Figure 4.

If a pulse frequency f_i is switched on at time T as shown in Figure 6, either the time or the frequency can be coded in the counter z. The relation to get z is

$$z = f_i T \tag{3a}$$

Because it is possible that one more line or pulse is counted as Figure 5 demonstrates – the so-called least significant bit (LSB), the relative error ε is (Woschni, 1988)

$$\varepsilon = \frac{1}{f_i T} \tag{3b}$$

More details are given in Carstens (1993) and in Pallas-Areny and Webster (2001).

6 ERROR CORRECTION BY REDUNDANT CODES

During data input, transmission, or processing procedures, errors may arise, as has already been mentioned in the design of the Gray code in the previous section. Error-detecting and error-correcting codes having additional code redundancy have been designed. The Hamming distance d_{min}, that is, the minimum distance between two code words

of an alphabet, has to be maintained (Peterson, 1962), for error-detecting codes with the degree f_d for errors to be detected,

$$d_{min} = f_d + 1 \tag{4a}$$

and for error-correcting codes, if f_c is the degree of error to be corrected

$$d_{min} = 2f_c + 1 \tag{4b}$$

If correction to the degree $f_c^* < f_c$ only is used, it is possible to detect additional errors up to degree f_d^*:

$$f_d^* = d_{min} - 2f_c^* - 1 \tag{4c}$$

Details concerning the Hamming distance as the procedure of error detection and error correction in the signal space are treated in **Article 32, Geometrical Signal Representations, Volume 1**.

A very simple, but often used method for data input, is the addition of parity bits or parity check. This additional bit is chosen in such a way that the sum of all digits (the so-called weight of the code) is either an even or an odd number. Table 4 presents the BCD code with parity check, detecting all errors with odd weight (1, 3, ...).

Other and more complicated error-detecting or error-correcting codes are the selector code (w-out-of-n code), the Hamming codes produced by feedback shift registers (Hamming, 1986), the recurrent or cyclic code (Peterson, 1962), and codes with block protection (Heise and Quattrocchi, 1995).

If the transmission is bidirectional, it is possible to correct wrong data blocks by means of retransmission (Sydenham, Hancock and Thorn, 1989). A typical example is given in the next section.

Table 4. BCD code with parity check.

First tetrad	Second tetrad	Parity check	Decimal number	Σ1
0000	0000	1	0	1
0000	0001	0	1	1
0000	0010	0	2	1
0000	0011	1	3	3
0000	0100	0	4	1
0000	0101	1	5	3
0000	1001	1	9	3
0001	0000	0	10	1
0001	0001	1	11	3
0010	0000	0	20	1
0101	0000	1	50	3
1001	1001	1	99	5

7 AN EXAMPLE OF ERROR CORRECTION CODE APPLICATION

As an example of error correction by redundant codes, a digital data transmission with a bit error rate (BER) of 10^{-2} is assumed: this corresponds to a normal telephone, hard wire line. The line should have a capacity of 2400 bps, a small value by modern speeds of communication.

We are first interested in the number of bit errors per year that arise in such a link. With 365 days, there will be 756 864 000 bit errors per year!

Now we ask what is the Hamming distance d_H necessary to get no more than one bit error in one year. If the bit error rate is to be 10^{-10}, there will still be 756 864 bit errors per year. The necessary bit error rate has to be 10^{-12} leading to 0.07564 bit errors in one year. From (4b) it follows that $f_c = 5$ a Hamming distance of $d_H = 2 \times 5 + 1 = 11$.

If the capacity of the line were to be 48 kbps, the bit error rate necessary would be 10^{-14} because the value 10^{-12} would lead to $20 \times 0.07564 = 1513$ bit errors per year. Now f_c has to be six, which means that due to (4b), the Hamming distance has to be 13. The realization of the necessary Hamming distance is made by means of checking digits in addition to the information digits (Peterson, 1962) and (Hamming, 1986).

REFERENCES

Carstens, J.R. (1993) *Electrical Sensors and Transducers*, Regents/Prentice Hall, Englewood Cliffs, NJ.

Finkelstein, L. (1976) Paper CTH 301, Preprints *IMEKO Congress*, Institute of Measurement and Control, London.

Heise, W. and Quattrocchi, P. (1995) *Informations- und Codierungs Theorie*, Springer, Berlin.

Hamming, R.W. (1986) *Coding and Information Theory*, Prentice Hall, NJ.

Pallas-Areny, R. and Webster, J.G. (2001) *Sensors and Signal Conditioning*, Wiley, New York.

Peterson, W.W. (1962) *Error Correcting Codes*, MIT Press, Cambridge.

Woschni, E.-G (1988) *Informations Technik*, Huthig, Heidelberg.

FURTHER READING

Wilson, S (1996) *Digital Modulation and Coding*, Prentice Hall, NJ.

34: Modulation Theory

Eugen Georg Woschni

Technical University Chemnitz, Chemnitz, Germany

1 Survey of Modulation Methods	191
2 Amplitude Modulation	192
3 Frequency or Phase Modulation	194
References	195
Further Reading	195

1 SURVEY OF MODULATION METHODS

A special form of processing of a source signal to suit a transmission channel is that of modulation. Transmission of signals together from several sources all over a single channel can be accomplished using frequency-division or time-division multiplexing systems.

In measurement, the problem of transmitting the output signals of many sensors or transmitters over one line is often solved by means of the time-division multiplexing. This time-sharing method is shown in Figure 1. A switch sequentially connects the several input channels into the single output line. The length of the connection time gives the pulse interval t_p and the time between switching back to the same channel again is period t_o. Pulse modulation, with the interleaving of the different signals, is applied for parallel-to-serial conversion.

In general, a modulator may be interpreted as a controlled system with the carrier signal as one input and the modulation signal $x(t)$ as the control input (Figure 2). In the following text, a survey is first given of the several kinds of modulation.

In analog modulation methods (see Figure 3), one, or several, parameters of the sinusoidal oscillation $u(t)$, termed the *carrier oscillation* having a carrier frequency Ω_0 gives

$$u(t) = \widehat{U} \sin(\Omega_0 t + \phi) = \widehat{U} \sin\theta(t) \tag{1a}$$

Thus, the input signal is caused to vary by the modulation signal $x(t)$. If the amplitude $\widehat{U}$ is altered by the input signal $x(t)$, amplitude modulation (AM) results:

$$\widehat{U} = f(x(t)) \tag{1b}$$

Angle modulation is generated by using $x(t)$ to vary the argument of $u(t)$:

$$\phi(t) = f(x(t)) \tag{1c}$$

As $\theta = \Omega_0 t + \phi$, two kinds of angle modulation can be implemented.

Frequency modulation (FM) occurs when Ω_0 is varied as

$$\Omega(t) = f(x(t)) \tag{1d}$$

Phase modulation (PM) occurs when ϕ is varied as

$$\phi(t) = f(x(t)) \tag{1e}$$

Figure 2 shows the modulation methods mentioned above for a sinusoidal modulation signal with the modulation frequency ω:

$$x(t) = \widehat{X} \sin(\omega t) \tag{1f}$$

Pulse modulation methods use the principle of sampling – *see* **Article 31, Discrete Signal Theory, Volume 1**.

As shown in Figure 4, parameters of a regular pulse sequence are changed by the modulation signal $x(t)$ in one of several ways.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.

© 2005 John Wiley & Sons, Ltd.

ISBN: 0-470-02143-8.

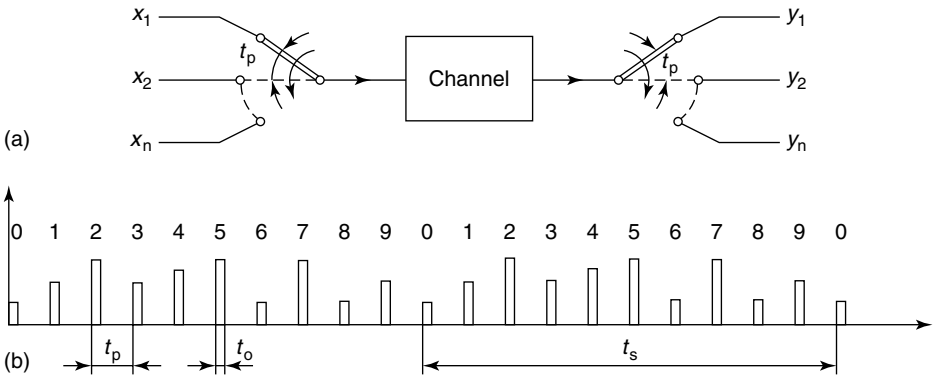


Figure 1. Time-division multiplexing: (a) system and (b) pulse frame (for 10 transmitters).

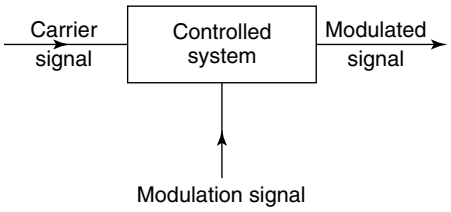


Figure 2. Generalized modulation system.

Variation of the pulse amplitude provides pulse-amplitude modulation (PAM), while with the pulse-duration or pulse-width modulation, the lengths of the pulses correspond to the modulation signal.

A pulse modulation method of great importance, especially in communication because of its good signal-to-noise ratio, is the pulse-code modulation (PCM) (Figure 5).

In the usual case, a pulse-amplitude modulation signal is first generated, which is then converted to a coded pulse sequence using one of the codes of **Article 33, Coding Theory and its Application to Measurement, Volume 1**.

2 AMPLITUDE MODULATION

We now deal with some details of the bandwidth needed and application in measurement.

A sinusoidal modulation signal

$$x(t) = \widehat{X} \sin(\omega t + \varphi) \tag{2a}$$

operating on a carrier $\widehat{U}_0 \sin(\Omega_0 t)$ yields the amplitude-modulated oscillation

$$u(t) = \widehat{U}_0[1 + m \sin(\omega t + \varphi)] \sin(\Omega_0 t) \tag{2b}$$

where the modulation depth is $m = k\widehat{X}$. Representation of (2b) in the frequency domain is (Woschni, 1988;

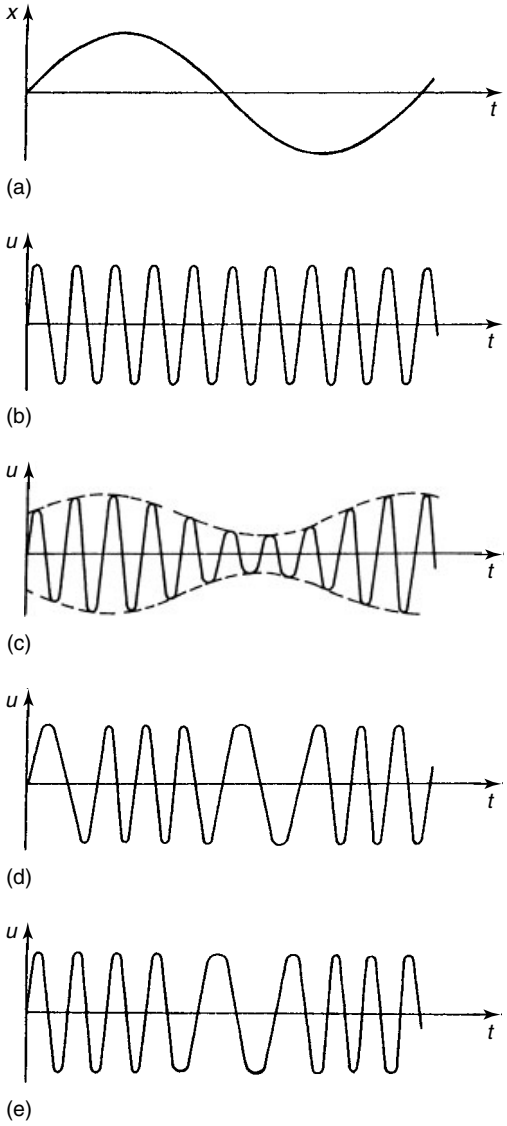


Figure 3. Analog modulation methods: (a) modulation signal $x(t)$; (b) carrier oscillation; (c) amplitude modulation; (d) frequency modulation and (e) phase modulation.

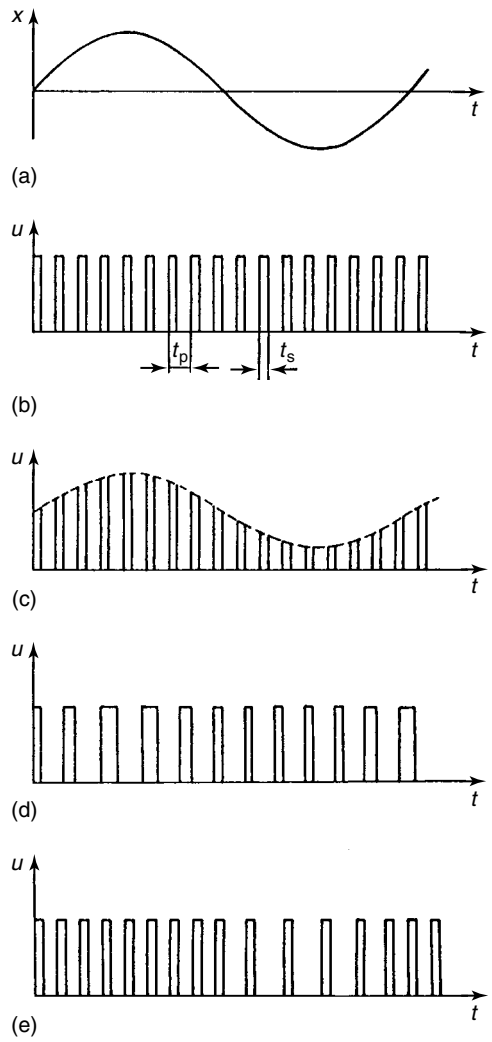


Figure 4. Pulse modulation methods (t_s = sampling time): (a) modulation signal $x(t)$; (b) carrier pulse sequence; (c) pulse-amplitude modulation; (d) pulse-duration or pulse-width modulation and (e) pulse-phase or pulse-position modulation.

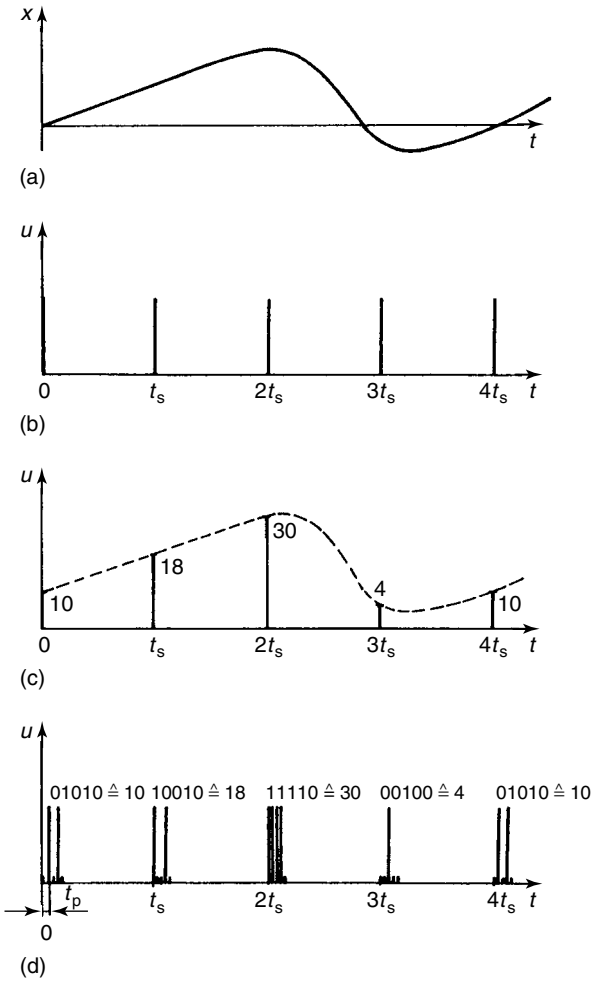


Figure 5. Pulse-code modulation: (a) modulation signal; (b) carrier pulse sequence; (c) pulse-amplitude modulation and (d) pulse-code modulation (binary code).

Wozencraft and Jacobs, 1965)

$$u(t) = \widehat{U}_0 \left\{ \sin(\Omega_0 t) \pm \tfrac{1}{2} m \cos[(\Omega_0 \mp \omega)t \mp \varphi] \right\} \qquad (2c)$$

showing that the resultant is a signal with the carrier frequency and two side frequencies (Figure 6a). The composition of the several spectral frequencies provides the time function, as shown in Figure 6b.

In the general case of an input signal with bandwidth $\omega_1 - \omega_c$ to be modulated, the bandwidth needed by the transmission link is

$$b = 2\,\omega_c \qquad (2d)$$

around the carrier frequency Ω_0. An amplifier, therefore, has to have at least this bandwidth otherwise distortion of the original form of the modulated signal will arise in the later recovered signal (Woschni, 1988).

In measurement, amplitude modulation results at the output of an AC bridge – *see* **Article 126, Electrical Bridge Circuits – Basic Information, Volume 2**, operating with inductive sensors, as presented in Figure 7. To obtain satisfactory dynamic behavior, the condition between the limiting frequency of the measured input ω_c and the carrier frequency Ω_0 needs to be

$$\Omega_0 \geq 5\,\omega_c \qquad (2e)$$

If this condition is not met, it will not be possible to correctly demodulate the amplitude-modulated oscillation.

The condition given by (2e) is very important, for it is the key to avoiding errors in dynamic measurements as shown in detail in **Article 50, Errors in Signal Systems, Volume 1**.

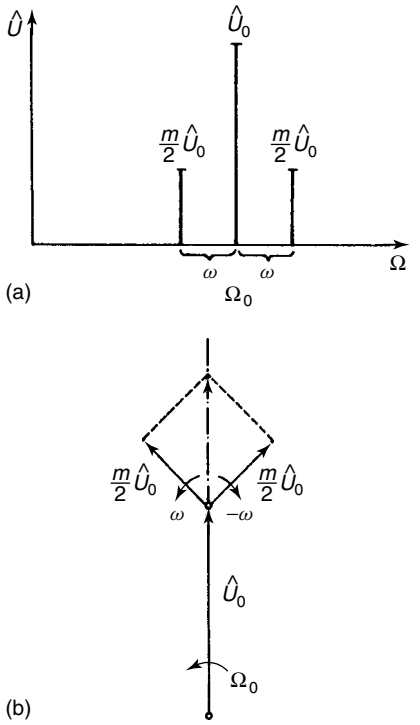


Figure 6. (a) Frequency spectrum of sinusoidal amplitude modulation and (b) phasor representation.

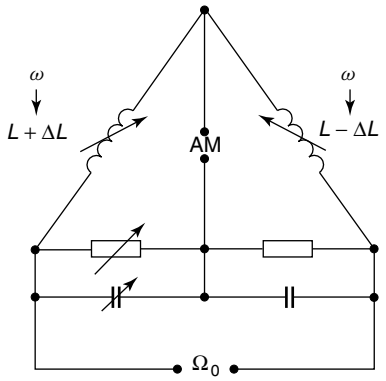


Figure 7. Bridge circuit, delivering an amplitude modulation.

3 FREQUENCY OR PHASE MODULATION

For the operation of capacitive sensors having high sensitivity, and in certain cases for inductive sensors, frequency modulation is used (Figure 8). For sinusoidal variation of the capacitance of the sensor

$$C = C_0 + \Delta C = C_0 \left(1 + \frac{\Delta C}{C_0} \sin(\omega t)\right) \tag{3a}$$

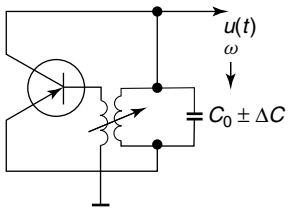


Figure 8. Circuit for the operation of a capacitive sensor to provide frequency-modulated output.

the variation of the natural frequency $\Omega = 1/\sqrt{(LC)}$ is given by (Woschni, 1962)

$$\Omega = \Omega_0 + \Delta\Omega \sin(\omega t) = \frac{1}{\sqrt{(LC_0)}} \left[1 - \frac{\Delta C}{2C_0} \sin(\omega t) + \frac{3}{8} \left(\frac{\Delta C}{C_0}\right)^2 \sin^2(\omega t) - \cdots\right] \tag{3b}$$

Because of the nonlinear characteristic differential capacitance, sensing methods are often used.

The time function yields

$$u(t) = \widehat{U}_0 \sin\left(\int [\Omega_0 + \Delta\Omega \sin(\omega t)]\, \mathrm{d}t\right) = \widehat{U}_0 \sin\left(\Omega_0 t - \frac{\Delta\Omega}{\omega} \cos(\omega t)\right) \tag{4a}$$

The corresponding function for sinusoidal phase modulation is given by

$$u(t) = \widehat{U}_0 \sin[\Omega_0 t + \Delta\phi \sin(\omega t)] \tag{4b}$$

A comparison between both equations shows that

$$\Delta\phi = \frac{\Delta\Omega}{\omega} \tag{4c}$$

Table 1. Relations between frequency and phase modulation.

	Frequency modulation	Phase modulation
Frequency deviation	$\Delta\Omega$	$\Delta\Omega = \Delta\Phi\omega$
Phase deviation	$\Delta\Phi = \dfrac{\Delta\Omega}{\omega}$	$\Delta\Phi$

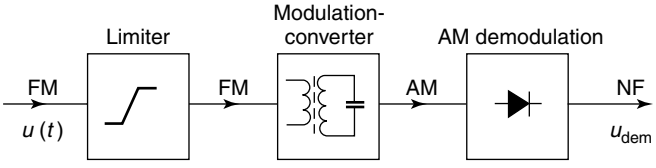


Figure 9. Principle of demodulation of frequency modulation.

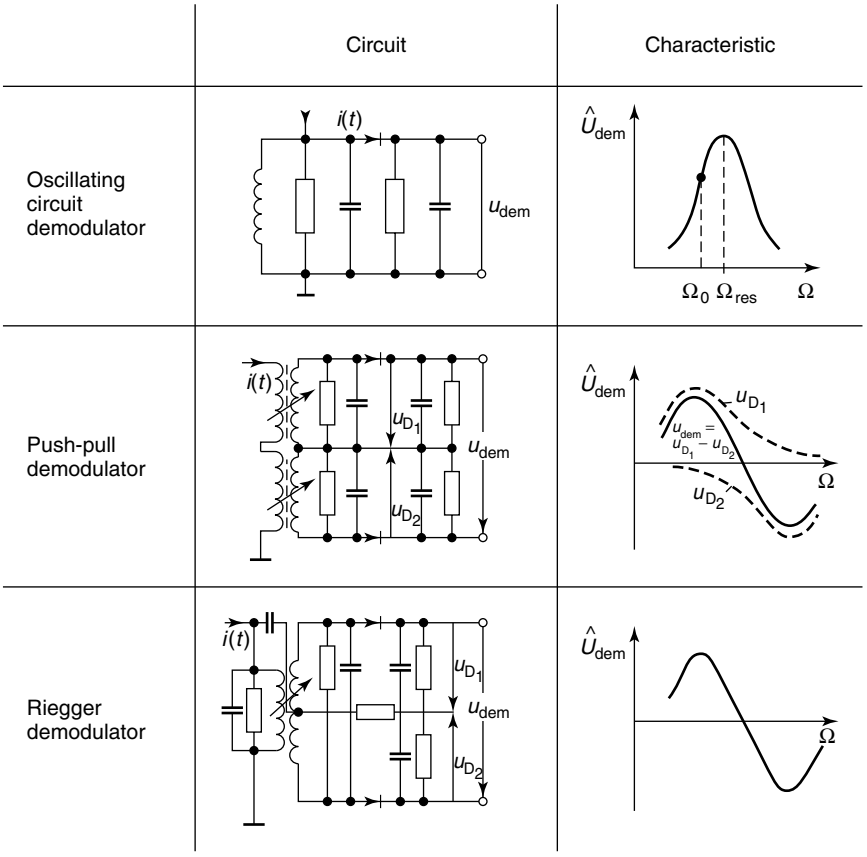


Figure 10. Circuits for realizing FM demodulation.

represents the equivalent phase deviation called the *modulation index*. Therefore, the relationships contained in Table 1, between frequency and phase modulation, are valid (Woschni, 1962). The spectrum of a frequency or phase-modulated signal is derived by means of a series expansion of Bessel functions (Woschni, 1962), leading to a bandwidth necessary for distortion-free transmission given by

$$b \cong 2[\Delta\Omega + k\omega] \tag{4d}$$

with $1 < k < 2$.

The demodulation of a frequency-modulated oscillation is realized as demonstrated in Figure 9. A limiting stage is followed by a modulation converter that converts the frequency modulation into an amplitude modulation, which is rectified by means of a diode.

Some example circuits for FM demodulation are shown in Figure 10 (Woschni, 1962). See also Webster (1999) or Wilson (1996).

REFERENCES

Webster, J.G. (1999) *Measurement, Instrumentation and Sensors Handbook*, CRC Press, Boca Raton, FL.

Wilson, S. (1996) *Digital Modulation and Coding*, Prentice Hall, NJ.

Woschni, E.-G. (1962) *Frequenz Modulation*, Verlag Technik, Berlin.

Woschni, E.-G. (1988) *Informations Technik*, Huthig, Heidelberg.

Wozencraft, J.M. and Jacobs, I.M. (1965) *Principles of Communication Engineering*, Wiley, New York.

FURTHER READING

Haykin, S. (2001) *Communication Systems*, 4th edn Wiley, New York.

35: Systems in the Time Domain

Eugen Georg Woschni
Technical University Chemnitz, Chemnitz, Germany

1 Classification of Technical Systems 199
2 Modeling by Means of Analogies 200
3 Block Diagrams 201
4 Linearization 202
5 Description with Differential Equations 203
6 Simple Harmonic Motion 204
7 State Space Description 204
8 Flowcharting 205
9 General Input Response, Convolution 208
References 209
Further Reading 209

1 CLASSIFICATION OF TECHNICAL SYSTEMS

According to Figure 1, a system can be interpreted as a black box with a family of input variables x_r, which can be regarded as a signal vector **x**, and a family of corresponding output variables y_r, forming a vector **y**. The interior of the black box may consist of several elements being either of electrical, hydraulic, pneumatic, or other energy forms.

The overall behavior of these systems is given by a common mathematical equation of the general form:

$$\mathbf{y} = \mathbf{O}\{\mathbf{x}\} \tag{1}$$

Different systems may have the same overall behavior so this means they follow the relationship of (1). This shows that it is possible for a system to be substituted by another one having the same mathematical equation, for example, by a computer of either analog, or today more likely digital form (an aspect of modeling).

Because of the convenience afforded by a computer to study the behavior of a given system, the role of programming modeling methods plays an important part in the field of cybernetics. Measurement as a part of cybernetics – the general science of feedback – can be comprehended as a mapping of the input signals onto the space of the output signals (Finkelstein, 1977).

Technical systems are characterized as those with active or only passive elements (so-called active or passive systems) or they are described by the number of ports (two-port, three-port, etc.) – *see also* **Article 73, Overview of Electrical and Electronic Technique, Volume 2**.

Table 1 gives a survey of the classification of technical systems. With respect to the difficulties in the treatment of a system, it is of great importance to know whether the system is linear or not, because in the linear case the superposition law is valid. In this article, it is assumed the system is a linear one. Methods of linearization are treated in Section 4.

Another typical characteristic of a system is whether the parameters describing the behavior are functions of time or not. Furthermore, most of the systems used in cybernetics are so-called unidirectional systems, meaning that the parameters of the system are independent of those of the following system. We will only deal with systems that fulfill this assumption. Otherwise, the results of the four-pole theory must be applied (Feldtkeller, 1962).

As an example of practical importance in the field of measurement, the spring–mass–damper system, as illustrated in Figure 2, will be considered. This system makeup is used in numerous sensors, for, among others, in the measurement of length or force. If force F is the input variable

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

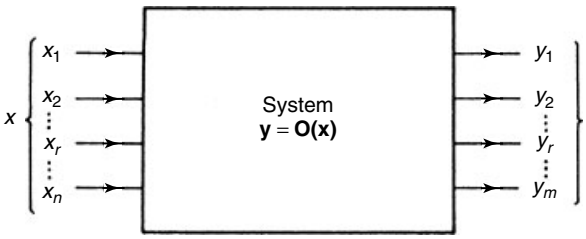


Figure 1. Definition of a system.

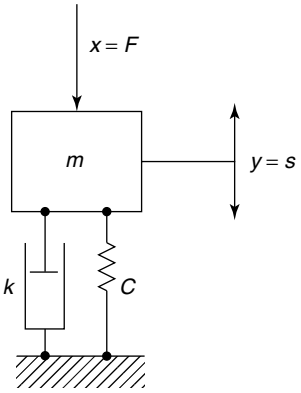


Figure 2. Spring–mass–damper system.

x, and length s is the output variable y, the following special equation corresponding to the general relation (1) is obtained:

$$m\ddot{y} + k\dot{y} + cy = x \tag{2}$$

This second-order differential equation describes the system's dynamic behavior. For more depth of explanation *see* **Article 59, Introduction to the Dynamic Regime of Measurement Systems, Volume 1**; **Article 60, Zero-order System Dynamics, Volume 1**; **Article 61, First-order System Dynamics, Volume 1**; and **Article 62, Second-order System Dynamics, Volume 1**.

2 MODELING BY MEANS OF ANALOGIES

As mentioned before, systems with different interior elements or forms of energy can have the same mathematical relationship between output and input variables. From this fact it follows that a given system can be represented by some other system having the same overall behavior. This modeling opportunity has the advantage that with response investigations of a given system its model system can have the parameters and structures changed by simply programming a computer. Furthermore, it is possible to observe the input and output quantities in a convenient way by means of oscilloscopes, computer displays, or plotters and to change the scale of coordinates or time axes to suit the needs of the investigation. Important methods of modeling are the analogies, application of block diagram representation, and linearization.

Of importance is the fact that mechanical systems, in the same way as pneumatic, hydraulic, and other systems, can be presented by electrical systems, as shown in Figure 3. In the cases illustrated in this arrangement, the following equations are valid:

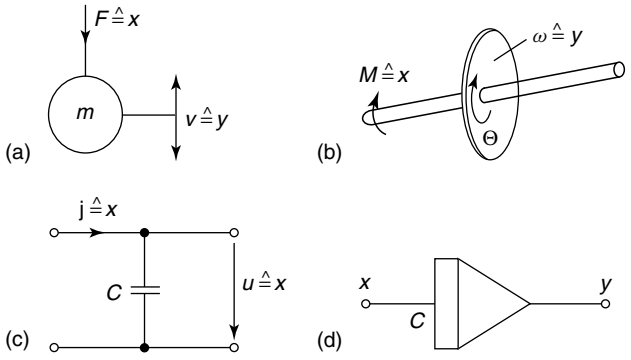


Figure 3. Examples of mechanical and electrical systems: (a) translation, (b) rotation, (c) capacity and (d) computer format.

Table 1. Classification of technical systems.

System	Linear system	Nonlinear system
Mark	Parameters are constant, independent of amplitude Superposition law is valid	Parameters are a function of amplitude Superposition law is not valid
Parameters are time invariant	Examples: most measurement systems with small input levels: amplifiers; filters; transducers	Examples: systems with large input levels: output amplifiers, driver stages; Often linearization is possible
Parameters are functions of time	Examples: controlled amplifiers with multiplicative properties and small input levels: modulators, frequency multipliers, parametric amplifiers	Examples: controlled amplifiers with multiplicative properties and large input levels: modulators, frequency modulators, frequency multipliers

1. In the mechanical example (mass m, velocity v, force F)

$$v = \frac{1}{m}\int F\,\mathrm{d}t \tag{3a}$$

2. In the case of rotary motion (moment of inertia Θ, angular velocity ω, torque M)

$$\omega = \frac{1}{\Theta}\int M\,\mathrm{d}t \tag{3b}$$

3. In the electrical system (voltage u, capacitance C, current i)

$$u = \frac{1}{C}\int i\,\mathrm{d}t \tag{3c}$$

4. In the general case of an analog computer, or more likely today, the digital computer (constant c)

$$y = c\int x\,\mathrm{d}t \tag{3d}$$

Generalizing the dependencies illustrated in Figure 3 realizes the survey of analogies between electrical and mechanical systems shown in Table 2. For the spring–mass–damper system with differential (2) and the representation of Figure 2, the electrical models shown in Figure 4 are obtained. Direct analogy (see Table 2) yields the parallel circuit Figure 4(a) with the equation

$$C\frac{\mathrm{d}u}{\mathrm{d}t} + \frac{1}{R}u + \frac{1}{L}\int u\,\mathrm{d}t = i \tag{4a}$$

while use of indirect analogy provides the series circuit given in Figure 4(b):

$$\frac{1}{C}\int i\,\mathrm{d}t + Ri + L\frac{\mathrm{d}i}{\mathrm{d}t} = u \tag{4b}$$

As can be seen from a comparison of Figures 2 and 4, the indirect analogy is more convenient in this instance because

Table 2. Survey of electromechanical analogies.

Electrical system	Mechanical system			
	Translation		Rotation	
	Direct	Indirect	Direct	Indirect
i	F	v	M	ω
u	v	F	ω	M
R	$\frac{1}{k}$	k	$\frac{1}{k}$	k
C	m	$\frac{1}{m}$	θ	$\frac{1}{\theta}$
L	$\frac{1}{C}$	C	$\frac{1}{C}$	C

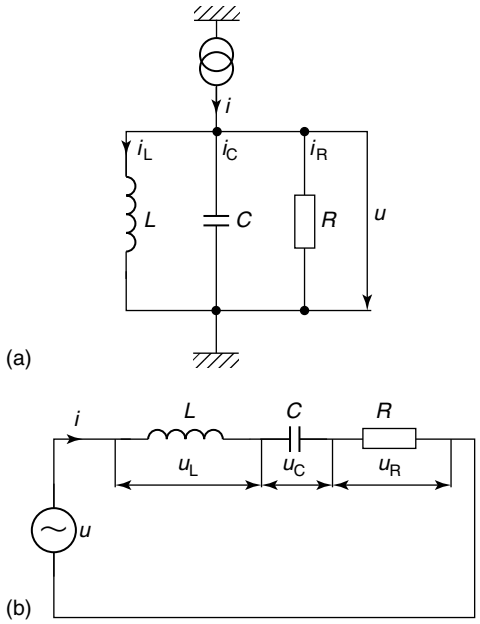


Figure 4. Electrical model of the spring–mass–damper system of Figure 2 (a) parallel circuit and (b) series circuit.

a mechanical parallel circuit is modeled as an electrical parallel circuit.

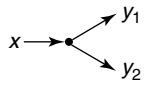
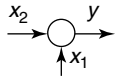
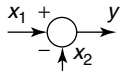
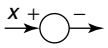
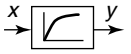
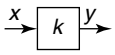
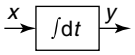
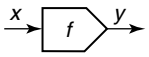
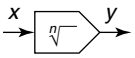
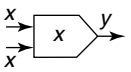
Similar analogies exist between acoustic, hydraulic, and pneumatic systems and to a limited extent to optical systems. Their theory, first ordered in Olson (1943), is paramount in the design and testing of measuring systems. For details, refer to Koenig and Blackwell (1961) and Shearer, Kulakowski and Gardner (1997). The analogies are also the foundation of modern disciplines as mechatronics and micromechanics. Another example is the surface-acoustic wave (SAW) filters.

3 BLOCK DIAGRAMS

Measurement systems are built up from particular combinations of subsystems. Therefore, it is suitable to represent each subsystem by its own block including a symbol indicating the operation the subsystem has to realize. Table 3 contains some of these symbols and signs used for the demonstration of the interconnections between the systems. (It must be pointed out that many standards for such are in use.) It is therefore assumed that the systems are unidirectional and thus it is possible to describe the behavior of such a system by means of an interconnection between several unidirectional systems. Figure 5 illustrates three typical methods of interconnection of systems. The output–input relation of a system is given by the equation

$$y = Gx \tag{5}$$

Table 3. Symbols for block diagrams.

Function	Branch	Summation	Subtraction	Sign inversion	General system
Symbol					
Equation	$y_1 = y_2 = x$	$y = x_1 + x_2$	$y = x_1 - x_2$	$y = -x$	$y = f(x, t)$
Function	Constant factor	Integrator	Nonlinear system	Root calculation	Multiplier
Symbol					
Equation	$y = kx$	$y = \int x\,dt$	$y = f(x)$	$y = \sqrt[n]{x}$	$y = x_1 x_2$

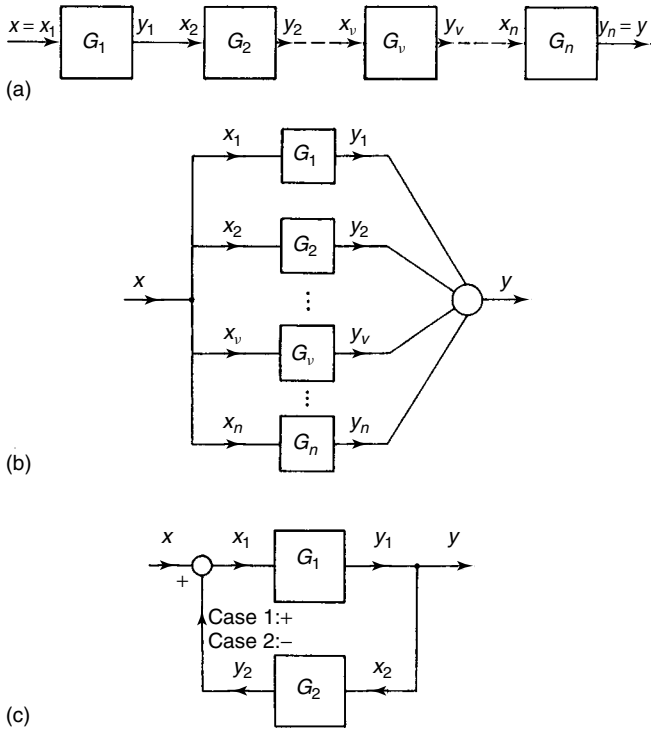


Figure 5. Typical system connections: (a) series circuit, (b) parallel circuit and (c) connection in opposition.

The frequency response G_e for an equivalent system having the same overall behavior as that of the interconnection of the subsystems shown is given by the following:

1. Series circuit (Figure 5a)

$$G_e = \prod_{r=1}^{n} G_r \tag{6a}$$

2. Parallel circuit (Figure 5b)

$$G_e = \sum_{r=1}^{n} G_r \tag{6b}$$

3. In opposition (Figure 5c)

$$G_e = \begin{cases} \dfrac{G_1}{1 - G_1 G_2} & \text{feedforward, oscillator} \quad (6c) \\ \dfrac{G_1}{1 + G_1 G_2} & \text{negative feedback, control} \quad (6d) \end{cases}$$

The product $G_1 G_2$ in (6c, d) describes the frequency response of the open control loop (Levine, 1996; Woschni, 1988).

4 LINEARIZATION

Linear systems are distinguished by the validity of the superposition law. Nonlinear systems are not easily modeled with linear equations and to get round this difficulty a nonlinear system is often linearized in small steps that collectively model the whole excursion region. This enables the advantages of linear systems to be used. The following preliminary conditions must be fulfilled in such a strategy:

1. Only small deviations of the characteristics from the linear path can be used.
2. Only a relatively small drive range of the nonlinear characteristic can be tolerated.

For the linearization, the Taylor series expansion of the nonlinear characteristic $y = f(x)$ at the working point $y_0 = f(x_0)$ is employed. Writing only the deviations from the working point $\Delta x, \Delta y$ yields

$$y - y_0 = \Delta y = \left.\frac{\partial f}{\partial x}\right|_{x_0} \Delta x + \frac{1}{2} \left.\frac{\partial^2 f}{\partial x^2}\right|_{x_0} \Delta x^2 + \frac{1}{6} \left.\frac{\partial^3 f}{\partial x^3}\right|_{x_0} \Delta x^3 + \cdots \tag{7a}$$

In practical measurement technique, the input variable is often the sinusoidal function $x = \hat{X} \sin(\omega t)$.

The proportion of the dominant wave with frequency ω at the output $\widehat{Y}_\omega$, expressed as a ratio with the amplitude of the input $\widehat{X}$, gives

$$\frac{\widehat{Y}_\omega}{\widehat{X}} = \left.\frac{\partial f}{\partial x}\right|_{x_0} \left(1 + \frac{1}{8}\frac{\partial^3 \dfrac{f}{\partial x^3}|_{x_0}}{\partial \dfrac{f}{\partial x}|_{x_0}} \widehat{X}^2\right) \tag{7b}$$

This function is the describing function; it expresses the frequency response of a nonlinear system. Depending on the sign of the third-order differential coefficient, the describing function either increases or decreases with the square of the amplitude of the input as Figure 6 shows. In practice, the case $\partial^3 f/\partial x^3|_{x_0} > 0$ can be troublesome because the amplification factor is increasing with increasing amplitude leading to an unstable oscillating regime (Woschni, 1988; Levine, 1996).

Important in the field of measurement is the rectification effect using a square-wave characteristic created as a rectifier unit that switches between on and off without any state between.

$$\Delta y_0 = \frac{1}{4}\left.\frac{\partial^2 f}{\partial x^2}\right|_{x_0} \widehat{X}^2 \tag{7c}$$

In this case, distortion appears that is described by the distortion factors

$$k_2 = \frac{\widehat{Y}_{2\,\omega}}{\widehat{Y}_\omega} = \frac{1}{4}\left(\frac{\partial^2 \dfrac{f}{\partial x^2}|_{x_0}}{\partial \dfrac{f}{\partial x}|_{x_0}}\right)\widehat{X}, \tag{7d}$$

$$k_3 = \frac{\widehat{Y}_{3\,\omega}}{\widehat{Y}_\omega} = \frac{1}{24}\left(\frac{\partial^3 \dfrac{f}{\partial x^3}|_{x_0}}{\partial \dfrac{f}{\partial x}|_{x_0}}\right)\widehat{X}^2 \tag{7e}$$

5 DESCRIPTION WITH DIFFERENTIAL EQUATIONS

The oldest classical method used in solving problems of system analysis is the method of differential equations. A linear system is described by a linear differential equation of the following form:

$$a'_n y^{(n)} + a_{n-1} y^{(n-1)} + \cdots + a_r y^{(r)} + \cdots + a_2 \ddot{y} + a_1 \dot{y} + a_0 y = b_0 x + b_1 \dot{x} + \cdots + b_m x^{(m)} \tag{8a}$$

This equation may be written as

$$T_n^n y^{(n)} + T_{n-1}^{n-1} y^{(n-1)} + \cdots + T_r^r y^{(r)} + \cdots + T_2^2 \ddot{y} + T_1 \dot{y} + y = G_0 x + \frac{b_1}{a_0}\dot{x} + \cdots + \frac{b_m}{a_0} x^{(m)} \tag{8b}$$

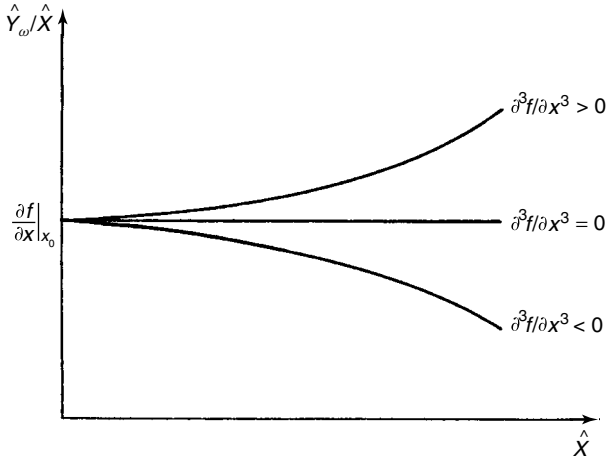


Figure 6. Ratio of the amplitude of the dominant wave of the output Y_ω to the amplitude of the input X (describing function).

where G_0 is the static sensitivity:

$$G_0 = \frac{b_0}{a_0} = \frac{\Delta y}{\Delta x} \tag{8c}$$

G_0 can be measured as an amplification factor by means of a small alteration Δx as stated in (8c). The coefficients

$$T_r = \sqrt[r]{\left(\frac{a_r}{a_0}\right)} \tag{8d}$$

are time constants. For a differential equation of nth order, n time constants are defined, the greatest of which is used to estimate the duration of the transient function. The solution of the differential equation consists of two additive components, a stationary and a dynamic portion:

$$y(t) = y_{\text{st}} + y_{\text{d}}(t) \tag{9a}$$

To solve the homogeneous differential equation leading to the dynamic portion $y_{\text{d}}(t)$, the assumption

$$y_{\text{d}} = C\,\mathrm{e}^{pt} \tag{9b}$$

is used. The zero points of the characteristic equation

$$T_n^n p^n + T_{n-1}^{n-1} p^{p-1} + \cdots + T_2^2 p^2 + T_1 p + 1 = 0 \tag{9c}$$

signify whether the solution is stable or not. In particular, these so-called eigenvalues p_r prove that

1. if the real part is less than zero, that is, $\text{Re}(p_r) < 0$, a stable solution exists;
2. if the real part is greater than zero, that is, $\text{Re}(p_r) > 0$, an unstable solution exists;
3. if the eigenvalues are complex, oscillations with decreasing or increasing amplitudes exist.

With the eigenvalues p_r, the dynamic solution yields (Coddington and Levinson, 1955)

$$y_{\mathrm{d}} = \sum_{r=1}^{n} C_r \exp(p_r t) \tag{9d}$$

If a double root p_0 arises

$$y_{\mathrm{d}} = (C_1 + C_2 t)\exp(p_0 t) \tag{9e}$$

The stationary solution y_{st} is to be found by means of suitable terms satisfying the inhomogeneous differential (8b).

As may be seen from (9d), the eigenvalue p_r corresponds to a time constant $T_r = 1/p_r$, the greatest value of which ($T_{\max}$) is responsible for the duration of the transient process. Because $\mathrm{e}^{-3} = 1/20 = 5\%$ the transient process approximately continues and

$$t_{\mathrm{tr}} = 3T_{\max} \tag{9f}$$

where t_{tr} is the transient time.

6 SIMPLE HARMONIC MOTION

In practice, the special case of a sinusoidal input signal $x(t)$ is very important

$$x(t) = \widehat{X}\mathrm{e}^{\mathrm{j}\omega t} \tag{10a}$$

This test signal is used to gain the frequency response, *see* **Article 36, Systems in the Frequency Domain, Volume 1**. In the steady state regime, which means after a time longer than the transient time since the signal $x(t)$ started, the output signal $y(t)$ will also be a sinusoidal signal but with another amplitude $\widehat{Y}$ and a phase φ

$$y(t) = \widehat{Y}\mathrm{e}^{\mathrm{j}(\omega t+\varphi)} \tag{10b}$$

We use these functions x, y as a statement in the original differential (8b) taking into account that every differentiation means a multiplication with $\mathrm{j}\omega$

$$\begin{aligned}&[T_n^n(\mathrm{j}\omega)^n\widehat{Y}\mathrm{e}^{\mathrm{j}\varphi} + \cdots + T_1\mathrm{j}\omega\widehat{Y}\mathrm{e}^{\mathrm{j}\varphi} + \widehat{Y}\mathrm{e}^{\mathrm{j}\varphi}]\mathrm{e}^{\mathrm{j}\omega t}\\ &= \left[G_0\widehat{X} + \frac{b_1}{a_0}\mathrm{j}\omega\widehat{X} + \cdots + \frac{b_m}{a_0}(\mathrm{j}\omega)^m\widehat{X}\right]\mathrm{e}^{\mathrm{j}\omega t}\end{aligned} \tag{10c}$$

Reducing this equation to the output/input relation, we get the frequency response that is treated in **Article 36, Systems in the Frequency Domain, Volume 1** in detail

$$G(\mathrm{j}\omega) = \frac{\widehat{Y}\mathrm{e}^{\mathrm{j}\varphi}}{\widehat{X}} = \frac{G_0 + \mathrm{j}\omega\dfrac{b_1}{a_0} + \cdots + (\mathrm{j}\omega)^m\dfrac{b_m}{a_0}}{1 + \mathrm{j}\omega T_1 + \cdots + (\mathrm{j}\omega)^n T_n^n} \tag{10d}$$

It may be noted that (10c) is the Fourier transform of the differential (8b) under the supposition of vanishing initial conditions, which means $x(t)_{t=0} = 0$ and so on.

The same is valid under the same suppositions with respect to the Laplace transform, which means $\mathrm{j}\omega \to \mathrm{p}$

$$\begin{aligned}&T_n^n s^n Y(s) + \cdots + T_1 s Y(s)\\ &G_0 X(s) + \frac{b_1}{a_0} s X(s) + \cdots + \frac{b_m}{a_0} s^m X(s)\end{aligned} \tag{10e}$$

leading to the transfer function, *see* **Article 36, Systems in the Frequency Domain, Volume 1**

$$G(s) = \frac{Y(s)}{X(s)} = \frac{G_0 + \dfrac{b_1}{a_0}s + \cdots + \dfrac{b_m}{a_0}s^m}{1 + T_1 s + \cdots + T_n^n s^n} \tag{10f}$$

7 STATE SPACE DESCRIPTION

The basis of the state space description is the classical differential equation discussed above or, in the case of sampled-data systems, the difference equation. The state of a system is described by a set of state variables y_r. The number of these state variables agrees with the degree of the differential equation or in the case of sampled-data systems with the degree of delay. Today, the method is of significant importance because of its suitability for computer simulation of systems.

The nth-order differential equation for such systems

$$T_n^n y^n + T_{n-1}^{n-1} y^{n-1} + \cdots + \dot{y} + y = x \tag{11a}$$

can be transformed into a system of n differential equations of first order:

$$\begin{aligned}y &= y_1\\ \dot{y} &= y_2 = \dot{y}_1\\ \ddot{y} &= y_3 = \dot{y}_2\\ &\vdots\\ y^{(n-1)} &= y_n = \dot{y}_{n-1}\\ y_n &= -\frac{1}{T_n^n}y_1 - \frac{T_1}{T_n^n}y_2 - \frac{T_2^2}{T_n^n}y_3 - \cdots - \frac{T_{n-1}^{n-1}}{T_n^n}y_n\\ &\quad + \frac{1}{T_n^n}x\end{aligned} \tag{11b}$$

This system of equations can be written in the form of a vector differential equation (Zadeh and Desoer, 1963;

Levine, 1996)

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \dot{y}_1 \\ \dot{y}_2 \\ \dot{y}_3 \\ \vdots \\ \dot{y}_n \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ -\dfrac{1}{T_n^n} & -\dfrac{T_1}{T_n^n} & -\dfrac{T_2^2}{T_n^n} & -\dfrac{T_3^3}{T_n^n} & \cdots & -\dfrac{T_{n-1}^{n-1}}{T_n^n} \end{bmatrix}$$

$$\times \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ \dfrac{1}{T_n^n} \end{bmatrix} x \tag{11c}$$

or with the variables **y, T, b**, and **x**:

$$\dot{\mathbf{y}} = \mathbf{T}\mathbf{y} + \mathbf{b}\mathbf{x} \tag{11d}$$

For sampled-data systems (Woschni, 1988; Levine, 1996), the difference equation is

$$\sum_{\mu=0}^{n} a_\mu y(k+\mu) = \sum_{r=0}^{m} b_r x(k+r) \tag{12a}$$

Defining a delay element as

$$D\{y(k)\} = y(k-1) \tag{12b}$$

We obtain also a system of n difference equations of first order

$$Z_\mu(k) = D\Big\{ -a_{(n-\mu)} Z_\mu(k) + Z_{(\mu-1)}(k) + (b_{(n-\mu)} - b_n a_{(n-\mu)}(k)\Big\} \tag{12c}$$

$$Y(k) = Z_1(k) + b_n(k) \tag{12d}$$

This system of equations can be written in matrix form

$$Z(k+1) = \boldsymbol{A} Z(h) + \boldsymbol{B} x(k); \tag{12e}$$

$$Y_k = \boldsymbol{C} z(k) + \boldsymbol{D} x(k) \tag{12f}$$

With the matrices

$$\boldsymbol{A} = \begin{bmatrix} -a_{(n-1)} & 1 & 0 & \ldots 0 \\ -a_{(n-2)} & 0 & 1 & \cdots 0 \\ \vdots & \vdots & \vdots & \vdots \; 1 \\ -a_0 & 0 & 0 & \ldots 0 \end{bmatrix} \tag{12g}$$

$$\boldsymbol{B} = \begin{bmatrix} b_{(k-1)} & \cdots & -b_n a_{(n-1)} \\ b_{(k-2)} & \cdots & -b_n a_{(n-2)} \\ \vdots & \cdots & \vdots \\ b_0 & \cdots & b_n a_0 \end{bmatrix} \tag{12h}$$

$$\boldsymbol{C} = [1, 0 \ldots 0] \tag{12i}$$

$$\boldsymbol{D} = b_n \tag{12j}$$

Consider as an example the spring–mass–damper system of Figure 2 with the differential equation

$$m\ddot{y} + k\dot{y} + cy = x \tag{13a}$$

The vector equation is given by

$$\begin{bmatrix} \dot{y}_1 \\ \dot{y}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\dfrac{c}{m} & -\dfrac{k}{m} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \dfrac{1}{m} \end{bmatrix} x \tag{13b}$$

The state variables are the displacement $y = y_1$ of the mass and the velocity $\dot{y} = y_2$.

8 FLOWCHARTING

Let us start with the simple example of the spring–mass–damper system of Figure 2 and (13a, b).

Figures 7 and 8 respectively show the programming of a computer for modeling this system. The relationships between programming and state space description are easy

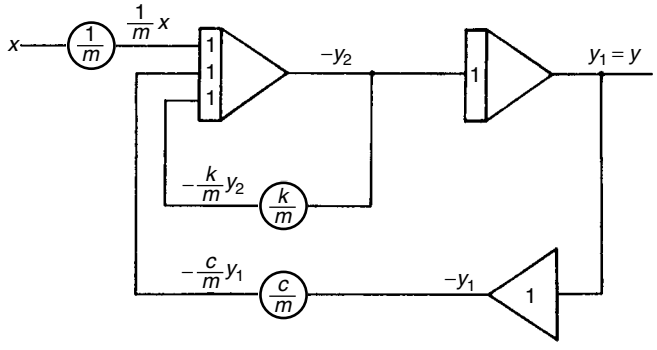


Figure 7. Direct programming of a spring–mass–damper system on a computer.

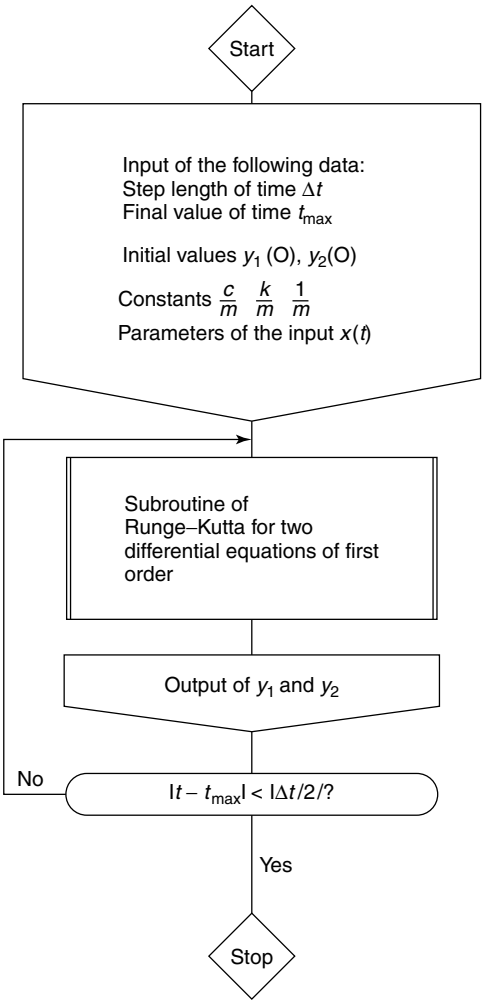


Figure 8. Flowchart for the programming of a spring–mass–damper system on a digital computer.

to recognize, especially due to the equation

$$\dot{z}_2 = \frac{c}{m} z_1 - \frac{k}{m} z_2 + \frac{1}{m} x \qquad (13c)$$

In Figure 9, the path of the five state variables displacement $y = z_1$ and velocity $y' = z_2$ are shown both as a function of time as well as in the state space.

In Figure 10, the programming of a system of nth order is shown as described with the differential (11a) and the state (11c). Here, in the flowchart shown in Figure 8, the subroutine of Runge–Kutta is used to realize n differential equations of first order.

Programming for a sampled-data system, as given by (8c, d) or (8e, f), is presented in Figure 11.

For system description and identification, certain response output functions for test signals at the input are generally used. These are presented here as time-dependent functions but they can be used with any general variable x, spatial distribution being the one used in optical systems.

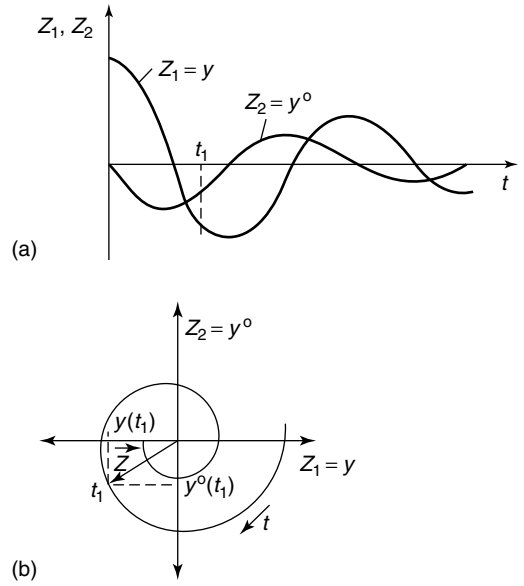


Figure 9. Displacement $y = Z_1$ and velocity $y' = Z_2$ for the spring–mass–damper system (a) as a function of time and (b) in the state space.

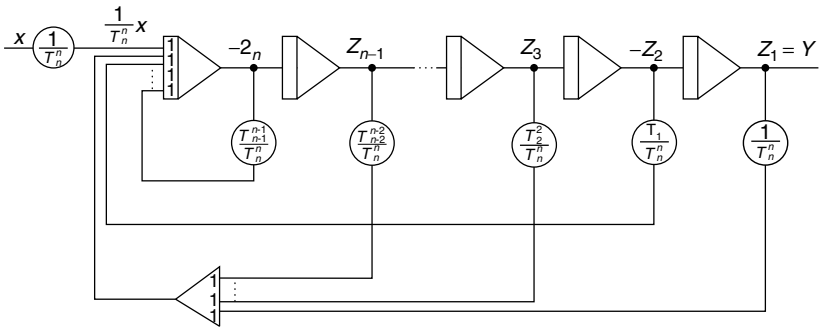


Figure 10. Programming of a system of nth order.

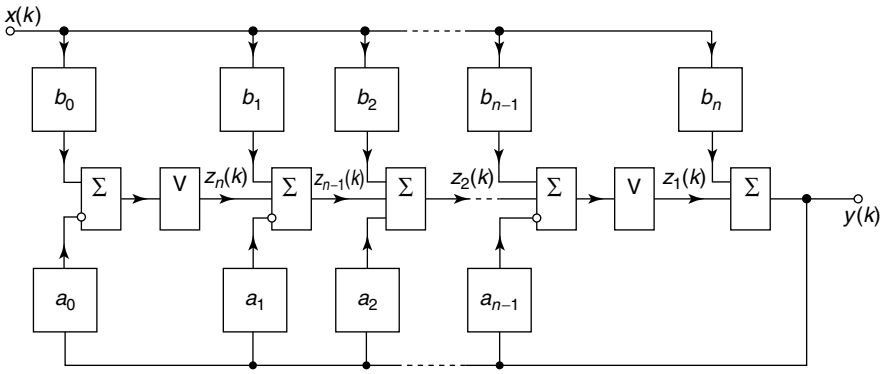


Figure 11. Programming of a sampled-data system.

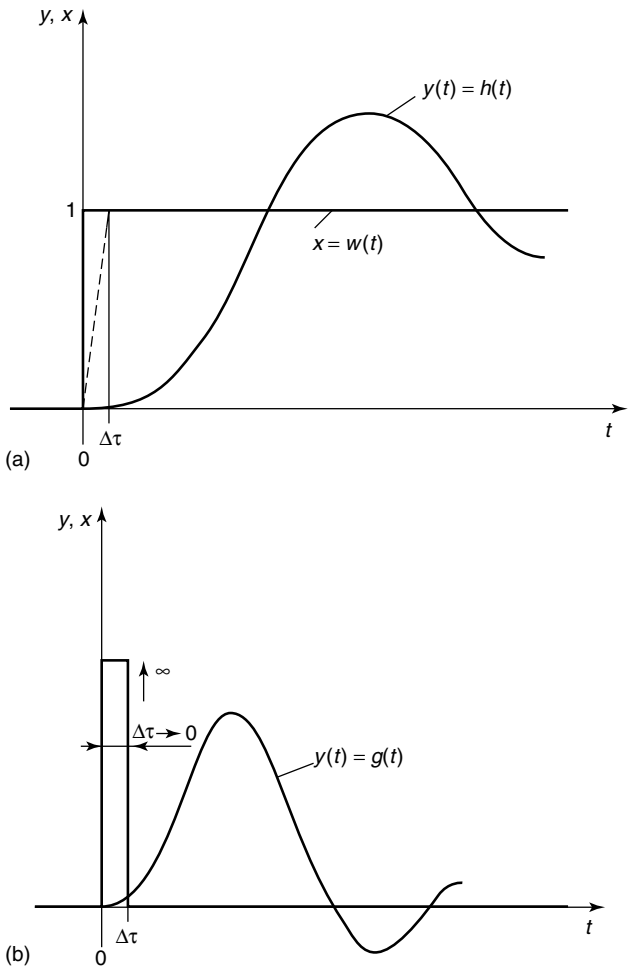


Figure 12. (a) Unit step $w(t)$ and transient response $h(t)$; (b) delta function $\delta(t)$ and weighting function $g(t)$.

1. The response to a step function with a step amplitude of 1 (unit step function, $w(t)$) is the unit step response or transient response $h(t)$ as illustrated in Figure 12(a).
2. The response to a pulse function with an integral value of 1 (Dirac delta function, $\delta(t)$) yields the unit pulse response or weighting function $g(t)$ shown in Figure 12(b).

Because of the linearity of the system, the response function $y(t)$ is to be divided by the step amplitude or the integral value, respectively to get the normalized function. In the case of system identification, if the input function generated by a signal generator is not the ideal function but a function with a rise time $\Delta\tau$ as signified in Figure 12, the condition must be fulfilled that the rise time $\Delta\tau$, or the pulse width $\Delta\tau$, is very much smaller than the transient time t_{tr} of the system (Woschni, 1988).

As can be seen from the comparison of the parts of Figure 12, the pulse function is connected with the step function by means of a differentiation, that is, in the sense of distribution (Gelfand and Schilow, 1960)

$$\frac{\mathrm{d}w(t)}{\mathrm{d}t} = \delta(t) \tag{14a}$$

For linear systems, it is immaterial whether the differentiation is realized at the input or output side of the system, meaning that

$$h(t) = \int g(t)\,\mathrm{d}t \tag{14b}$$

It is only a matter of convenience whether the transient or the weighting functions are used for identification.

For a system with first-order delay, as is used for the approximation of several systems in measurement technology, we gain the differential equation

$$T_1\dot{y} + y = x \tag{14c}$$

with the eigenvalue

$$p_1 = \frac{-1}{T_1} \tag{14d}$$

The stationary solution may be found by means of the assumption

$$y_{st} = c_2$$

for which the total solution yields

$$y = y_{st} + y_d = c_2 + c_1 \exp\left(\frac{-t}{T_1}\right)$$

Using the boundary conditions

$$y|_{t=-\infty} = 0 \quad y|_{t=\infty} = 1$$

we get for the transient response $h(t)$:

$$h(t) = 1 - \exp\left(\frac{-t}{T_1}\right) \tag{14e}$$

The weighting function $g(t)$ follows from (14b):

$$g(t) = \frac{dh(t)}{dt} = \frac{1}{T_1} \exp\left(\frac{-t}{T_1}\right) \tag{14f}$$

The functions $h(t)$ and $g(t)$ are represented in Figure 13. If these functions are obtained experimentally by means of a function generator at the input of the system examined, the time constant T_1 is given by the length of the subtangent (Figure 13). Furthermore, the figure shows the transient time t_{tr} to be nominally three times the time constant T, for $e^{-3} = 1/20$. For more details referring to testing of systems and important examples, *see* **Article 37, Relationship Between Systems in the Time and Frequency Domain, Volume 1**.

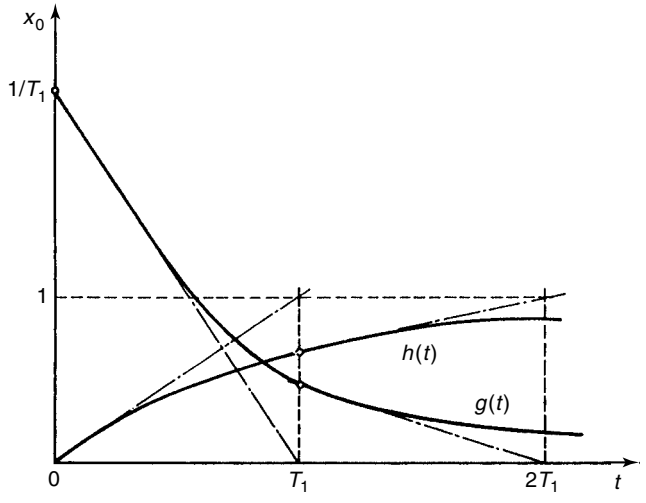


Figure 13. Transient response and weighting function for a first-order system.

9 GENERAL INPUT RESPONSE, CONVOLUTION

Transient response and weighting functions are response functions to special input signals. In the general case, the input function is broken down into a series of weighted Dirac delta functions, which are time delayed as represented in Figure 14. The pulse at the time τ_1 yields the output

$$x(\tau_1)\Delta\tau \times g(t_1 - \tau_1)$$

Because of the linearity of the system the superposition law is valid, which means the entire output is the sum (integral) of all inputs at the time $t - \tau_1 > 0$:

$$\begin{aligned} y(t) &= \int_0^t x(\tau)g(t-\tau)\,d\tau \\ &= \int_0^t x(t-\tau)g(\tau)\,d\tau \\ &= x(t)^* g(t) \end{aligned} \tag{15a}$$

This is the convolution integral, denoted by the sign * for the convolution. The lower limit of the integral may be

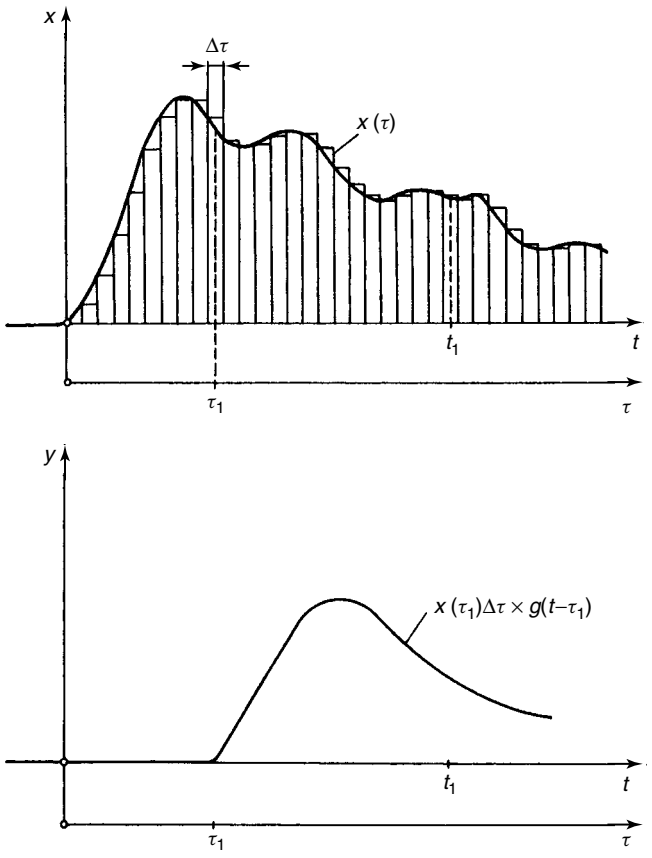


Figure 14. Explanation of the convolution integral.

extended to infinity since the weighting function must be zero before the input is applied. Since the input is zero for $t < 0$, the upper limit may also be extended to infinity, this means (15a) may also be written as

$$y(t) = \int_{-\infty}^{+\infty} x(\tau)g(t-\tau)\,\mathrm{d}\tau = \int_{-\infty}^{+\infty} x(t-\tau)g(\tau)\,\mathrm{d}\tau \tag{15b}$$

Another form of the convolution integral, the Duhamel integral, is obtained by taking into consideration (14b):

$$y(t) = \frac{\mathrm{d}}{\mathrm{d}t}\int_0^t x(\tau)h(t-\tau)\,\mathrm{d}\tau = \frac{\mathrm{d}}{\mathrm{d}t}\int_0^t x(t-\tau)h(\tau)\,\mathrm{d}\tau = \frac{\mathrm{d}}{\mathrm{d}t}[x(t)^* h(t)] \tag{15c}$$

In this equation, the upper and lower limits may also be extended to infinity.

In the field of measurement, convolution is of great importance for system identification (Davies, 1970). The autocorrelation function at the output of a system and the input are related by a double convolution (Woschni, 1988)

$$\psi_{yy}(\tau) = \int_0^\infty \int_0^\infty \psi_{xx}(\tau + \tau_1 - \tau_2)g(\tau_1)g(\tau_2)\,\mathrm{d}\tau_1\,\mathrm{d}\tau_2 \tag{16a}$$

If the autocorrelation function of the input $\psi_{xx}(\tau)$ and the cross-correlation function $\tau_{xy}(\tau)$ are measured, the weighting function of the system can be calculated, for which the following relation is valid (Davies, 1970; Woschni, 1988; Levine, 1996)

$$\psi_{xy}(\tau) = \int_0^\infty g(t)\psi_{xx}(\tau - t)\,\mathrm{d}t \tag{16b}$$

Deconvoluting (16b) gives the required pulse response function $g(t)$; the system may be regarded as being identified. Direct deconvolution techniques using (16b) can be difficult. Therefore, methods in the frequency domain were developed as shown in **Article 36, Systems in the Frequency Domain, Volume 1**. In the special case of white noise with an autocorrelation function $\psi_{xx}(\tau - t) = 2\pi S_0 \delta(\tau - t)$ the deconvolution degenerates to the equation

$$\psi_{xy}(\tau) = 2\pi S_0 g(\tau) \tag{16c}$$

This means that the weighting function corresponds directly to the cross-correlation function (Woschni, 1988).

REFERENCES

Coddington, E.A. and Levinson, N. (1955) *Theory of Ordinary Differential Equations*, McGraw-Hill, New York.

Davies, W.D.T. (1970) *System Identification for Self-Adaptive Control*, Wiley, Chichester.

Feldtkeller, R. (1962) *Einfuhrung in die Vierpoltheorie*, Hirzel, Stuttgart.

Finkelstein, L. (1977) Instrument Science. *Journal of Physics E: Scientific Instruments*, **10**, 566–572.

Gelfand, I.M. and Schilow, G.E. (1960) *Verallgemeinerte Funktionen (Distributionen)*, Vol.1, Deutscher Verlag d. Wiss., Berlin.

Koenig, H.E. and Blackwell, W.A. (1961) *Electromechanical System Theory*, McGraw-Hill, New York.

Levine, W.S. (1996) *The Control Handbook*, CRC Press, New York.

Olson, H.F. (1943) *Dynamic Analogies*, Van Nostrand, London.

Shearer, J.L., Kulakowski, B.T. and Gardner, J.F. (1997) *Dynamic Modeling and Control of Engineering Systems*, Prentice Hall,. Englewood Cliffs, NJ.

Woschni, E.-G. (1988) *Informations Technik*, Huthig, Heidelberg.

Zadeh, L.A. and Desoer, C.A. (1963) *Linear System Theory*, McGraw-Hill, New York.

FURTHER READING

Oppenheim, A.V., Willsky, A.S. and Young, I.T. (1983) *Signals and Systems*, Prentice Hall, Englewood Cliffs, NJ.

36: Systems in the Frequency Domain

Eugen Georg Woschni

Technical University Chemnitz, Chemnitz, Germany

1 Frequency Response 210
2 Logarithmic Characteristics and Bode Plot Representation 211
3 Transfer Function 212
4 Poles and Zeros and Root Locus Representation 213
5 Typical Example: Phase-shift Bridge 215
References 215

1 FREQUENCY RESPONSE

(Note of clarification: Before this article is read, it is necessary to point out that in the literature s is often used instead of p and vice versa. λ is used for Eigen values as well as for p. G and H are often interchanged. There are differences in strict mathematical use, but engineers have found it acceptable to use them interchangeably and it is a matter of personal preference in these explanations.)

Using a sinusoidal input and taking both amplitudes of the output $\widehat{Y}$ normalized to the input $\widehat{X}$, and phase φ as a function of frequency ω follows the frequency response with

$$\text{amplitude characteristic is } |G(\mathrm{j}\omega)| = \frac{\widehat{Y}}{\widehat{X}} \tag{1a}$$

$$\varphi(\omega) \leq y, x \tag{1b}$$

In the complex presentation, the input $x = \widehat{X}\mathrm{e}^{\mathrm{j}\omega t}$ yields the output $y = \widehat{Y}\mathrm{e}^{\mathrm{j}(\omega t+\varphi)} = \widehat{Y}\mathrm{e}^{\mathrm{j}\varphi}\mathrm{e}^{\mathrm{j}\omega t}$, which means that the complex frequency response is

$$\frac{\widehat{Y}}{\widehat{X}}\mathrm{e}^{\mathrm{j}\varphi} = G(\mathrm{j}\omega) = |G(\mathrm{j}\omega)|\mathrm{e}^{\mathrm{j}\varphi(\omega)} \tag{1c}$$

This function is represented in the complex plane as a locus diagram. The complex frequency response may be split into a real and an imaginary part:

$$G(\mathrm{j}\omega) = P(\omega) + \mathrm{j}Q(\omega) \tag{1d}$$

with the relations

$$|G(\mathrm{j}\omega)| = \sqrt{[P^2(\omega) + Q^2(\omega)]} \tag{1e}$$

$$\varphi(\omega) = \arctan\left(\frac{Q(\omega)}{P(\omega)}\right) \tag{1f}$$

Figure 1 shows the representation of the frequency characteristics. If the differential equation is given, it is very convenient to obtain the frequency response by means of the terms

$$x^{(\mu)} = (\mathrm{j}\omega)^{\mu}\widehat{X}\mathrm{e}^{\mathrm{j}\omega t}; \quad y^{(r)} = (\mathrm{j}\omega)^{r}\widehat{Y}\mathrm{e}^{\mathrm{j}\varphi}\mathrm{e}^{\mathrm{j}\omega t} \tag{2a}$$

Substituting the differential equation, *see* **Article 35, Systems in the Time Domain, Volume 1**

$$T_n^n y^{(n)} + \cdots + T_1\dot{y} + y = G_0 x + \frac{b_1}{a_0}\dot{x} + \cdots + \frac{b_m}{a_0}x^{(m)} \tag{2b}$$

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.

© 2005 John Wiley & Sons, Ltd.

ISBN: 0-470-02143-8.

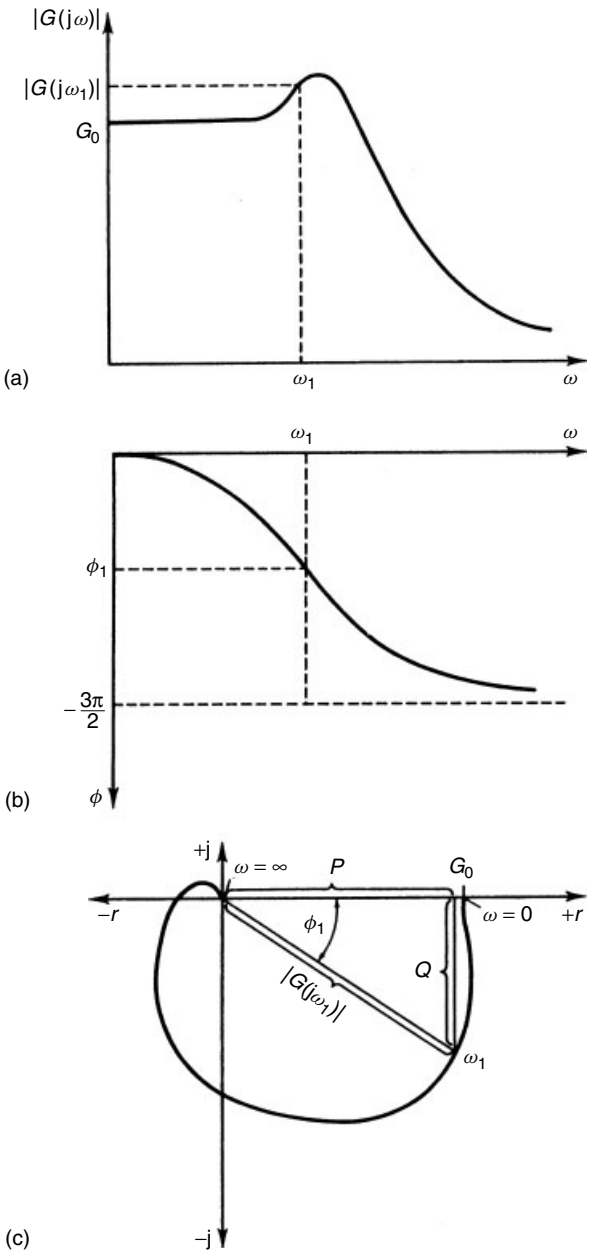


Figure 1. Graphical representation of frequency characteristics: (a) amplitude characteristic; (b) phase characteristic; and (c) locus diagram.

and solving the output–input relation leads to the complex frequency response

$$G(\mathrm{j}\omega) = \frac{\widehat{Y}\mathrm{e}^{\mathrm{j}\varphi}}{\widehat{X}} = \frac{G_0 + \mathrm{j}\omega\left(\frac{b_1}{a_0}\right) + \cdots + (\mathrm{j}\omega)^m\left(\frac{b_m}{a_0}\right)}{1 + \mathrm{j}\omega T_1 + \cdots + (\mathrm{j}\omega)^n T_n^n} \tag{2c}$$

This implies that substitution of nth-order differentiation by $(\mathrm{j}\omega)^n$ and nth-order integration by $(1/\mathrm{j}\omega)^n$ will yield the output–input relation.

The frequency response can be measured, by way of (1a, b), using a sinusoidal test signal. If the input is a stochastic signal with power density $S_{xx}(\omega)$, the output power density $S_{yy}(\omega)$ may be calculated according to Davies (1970) and Woschni (1988)

$$S_{yy}(\omega) = |G(\mathrm{j}\omega)|^2 S_{xx}(\omega) \tag{2d}$$

The double convolution in the time domain corresponds to a multiplication with $|G(\mathrm{j}\omega)|^2$ in the frequency domain (Woschni, 1988; Levine, 1996).

2 LOGARITHMIC CHARACTERISTICS AND BODE PLOT REPRESENTATION

Of great practical importance is the plotting of the amplitude characteristic in a double-logarithmic calibration graph and the phase characteristic with a linear φ-axis and a logarithmic ω-axis, known as *logarithmic frequency characteristics*. In order that a linear scale be used for the y-axis (Bode diagram), the amplitude is generally measured in decibel (dB) units,

$$20\log\left(\frac{\widehat{Y}}{\widehat{X}}\right) = 20\log|G(\mathrm{j}\omega)| \tag{3a}$$

The advantage of this method is illustrated in Figure 2. As treated in **Article 35, Systems in the Time Domain, Volume 1**, the overall behavior of series-connected systems is given by the multiplication of the frequency responses of each of these systems. Because of the logarithmic representation, the multiplication is simplified to a summation, which can easily be realized graphically.

For a system with first-order delay, used for approximation of more complicated systems, from the differential equation – *see* **Article 35, Systems in the Time Domain, Volume 1**.

$$T_1\dot{y} + y = x \tag{3b}$$

it follows that

$$G(\mathrm{j}\omega) = \frac{1}{1 + \mathrm{j}\omega T_1} \tag{3c}$$

$$|G(\mathrm{j}\omega)| = \frac{1}{\sqrt{1 + \omega^2 T_1^2}} \tag{3d}$$

$$\varphi(\omega) = -\arctan(\omega T_1) \tag{3e}$$

These frequency response functions are featured in Figure 3. From Figure 3(b) it can be seen that a critical

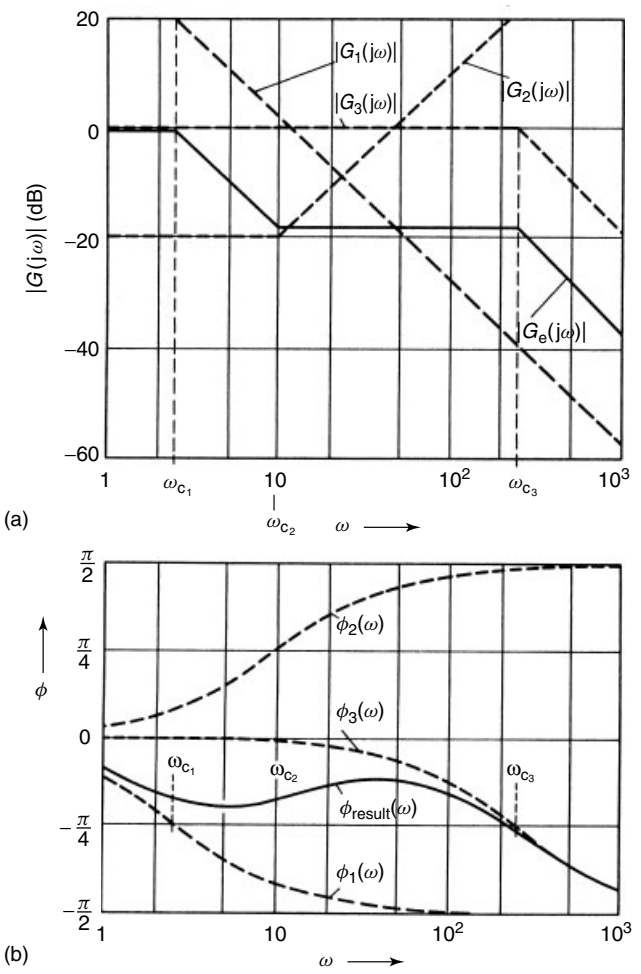


Figure 2. Graphical determination, through addition, of the frequency responses of series-connected systems (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd).

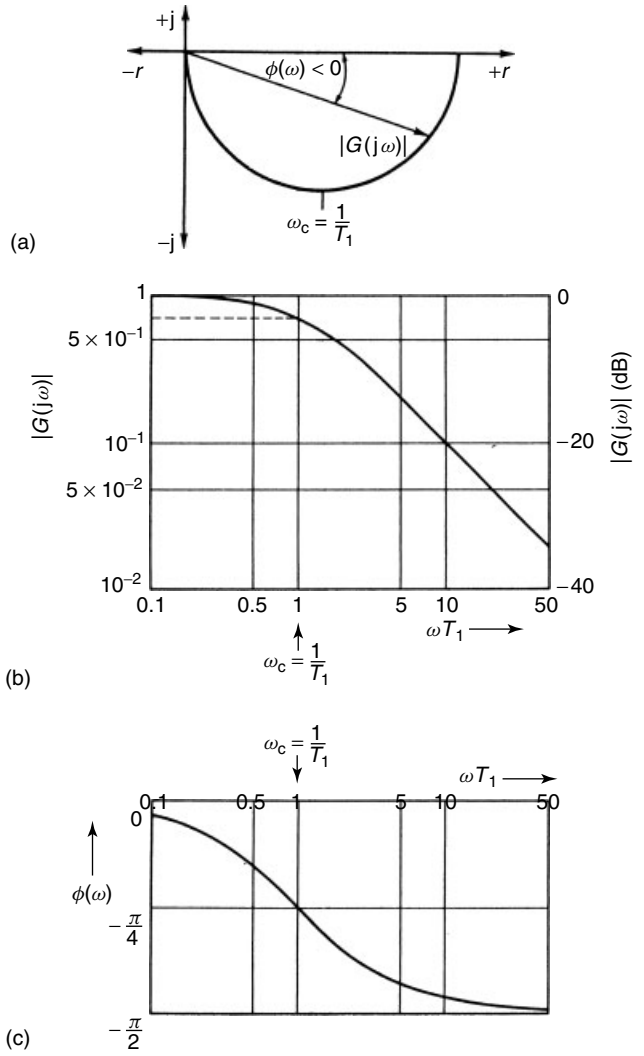


Figure 3. (a) Locus diagram; (b) amplitude characteristic and (c) phase characteristic for a first-order system.

frequency f_c (or ω_c) given by

$$2\pi f_c = \omega_c = \frac{1}{T_1} \tag{3f}$$

can be used to quickly generate approximations of the response behavior (Woschni, 2000).

Important examples of measurement system responses are discussed in **Article 37, Relationship Between Systems in the Time and Frequency Domain, Volume 1**.

3 TRANSFER FUNCTION

A generalization of the complex frequency response arises if the frequency $j\omega$ is extended to the complex frequency

$$p = j\omega + \delta \tag{4a}$$

with the increased constant δ.

This complex frequency is the variable p (or s in some literatures) used with Laplace transformation (*see* **Article 37, Relationship Between Systems in the Time and Frequency Domain, Volume 1**). In the physical sense, p means a harmonic oscillation with a exponentially increasing or decreasing amplitude (Woschni, 1988):

$$x = \hat{X}e^{pt} = \hat{X}e^{\delta t}e^{j\omega t} \tag{4b}$$

If s is represented in the p-plane, the left-hand side of this plane signifies stable solutions ($\delta < 0$) while the right-hand side leads to unstable solutions ($\delta > 0$).

This criterion is used for assessing the stability of systems as treated in **Article 38, Stability Issues, Volume 1**. The complex frequency p used instead of ω results in the transfer function $G(p)$ for the system described by (2c)

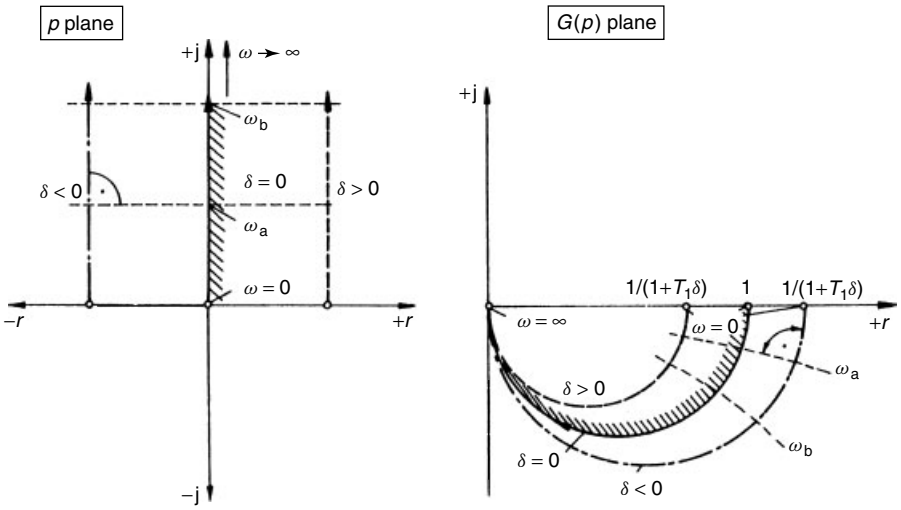


Figure 4. Representations of the transfer function for a first-order system.

with $b_0 = b_0/a_0$:

$$G(p) = \frac{\dfrac{b_0}{a_0} + p\dfrac{b_1}{a_0} + \cdots + p^m\dfrac{b_m}{a_0}}{1 + pT_1 + \cdots + p^n T_n^n} = \frac{\displaystyle\sum_{r=0}^{m} \frac{b_r}{a_0} p^r}{\displaystyle\sum_{\mu=0}^{n} T_\mu^\mu p^\mu} \tag{4c}$$

The $G(p)$ plane is a conformal mapping of the p-plane (Zadeh and Desoer, 1963; Woschni, 1988), which means that the side directions of the curves remain valid and small squares remain squares, as the next example shows. The system with first-order delay with the frequency responses (3c, d) has the transfer function

$$G(p) = \frac{1}{1 + pT_1} \tag{4d}$$

shown in Figure 4 with several values of δ. It can be observed that Figure 3 is a special case included in Figure 4.

4 POLES AND ZEROS AND ROOT LOCUS REPRESENTATION

By means of searching the zero points of both the numerator p_r^* and the divisor p_μ of the fraction (4c) one gains the equivalent product representation, the polynomial equation

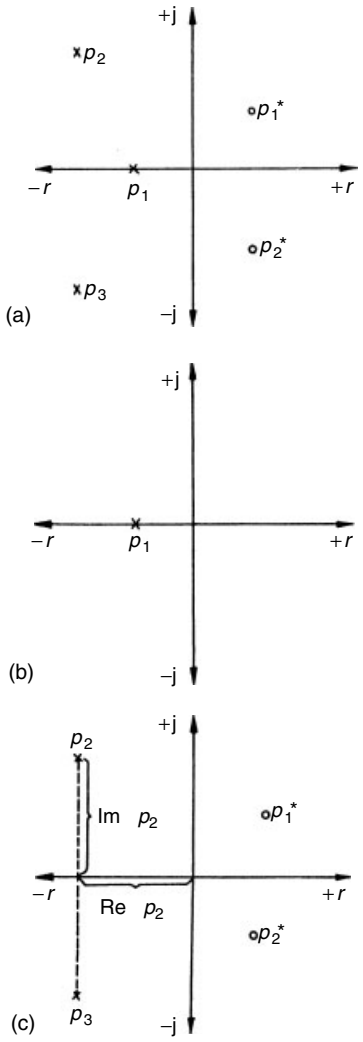


Figure 5. (a) Splitting up the pole-zero plane of a system; (b, c) into series-connected subsystems (x poles; 0 zeros).

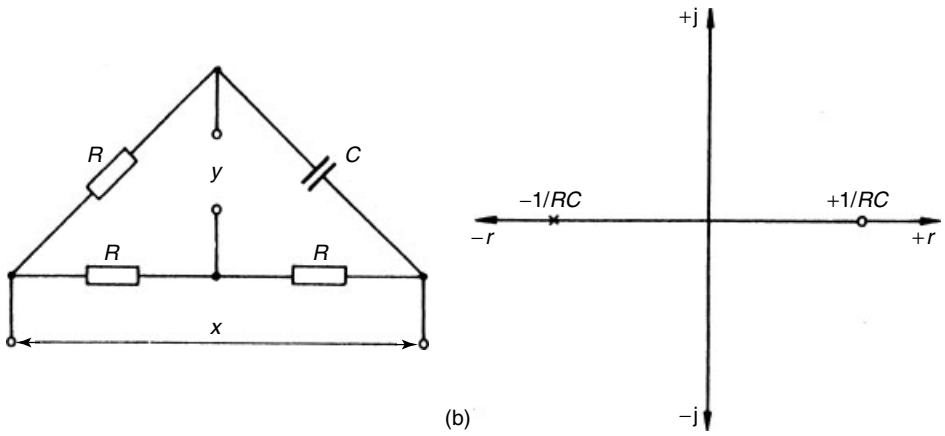


Figure 6. Phase-shift bridge: (a) circuit; and (b) pole-zero plane diagram.

alternative of (4c)

$$G(p) = c\frac{(p-p_1^*)(p-p_2^*)\ldots\ldots(p-p_m^*)}{(p-p_1)(p-p_2)\ldots\ldots(p-p_n)} = c\frac{\prod_{r=1}^{m}(p-p_r^*)}{\prod_{\mu=1}^{n}(p-p_\mu)} \tag{5a}$$

having the poles p_μ and the zeros p_r^* of the transfer function. Thus, the properties of a system are described completely, up to a constant c, by the position of poles and zeros, represented in the pole–zero plane (Figure 5). The poles agree with the Eigen values of the differential equation.

The frequency response can be calculated from the pole–zero plane by

$$G(\mathrm{j}\omega) = c\frac{\prod_{r=1}^{m}(\mathrm{j}\omega - p_r^*)}{\prod_{\mu=1}^{n}(\mathrm{j}\omega - p_\mu)} \tag{5b}$$

Figure 5 shows how, for series connections, the pole–zero representation of a complicated system can be split into a sum of simpler systems. Poles and zeros at the same point cancel each other; this feature called *compensation* is used for the correction of systems by means of additional series-connected correcting elements – *see* **Article 38, Stability Issues, Volume 1**). The position of the poles contains information as to whether the system is stable or not: existence of poles in the right half plane signifies instability, *see* **Article 38, Stability Issues, Volume 1**.

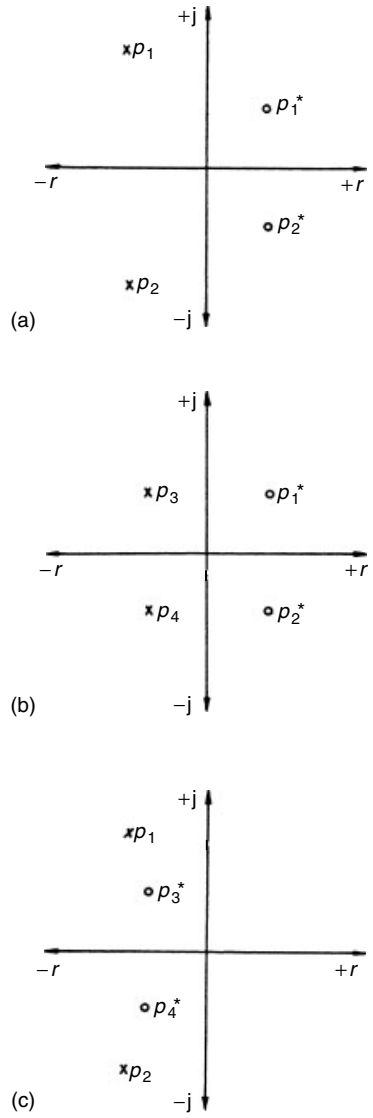


Figure 7. (a) System of Figure 5c; (b) all-pass system; and (c) minimal-phase system.

Of special interest in connection with stability problems is the path of the poles as a function of the parameters of systems. This method is called *root locus representation* and is the basis for system stability investigations (Levine, 1996).

5 TYPICAL EXAMPLE: PHASE-SHIFT BRIDGE

In measurement, the phase-shift bridge (Figure 6) is often used. The transfer function is

$$G(p) = \frac{1}{2}\frac{1 - pCR}{1 + pCR} \tag{6a}$$

and the pole and zero at

$$p_0 = -\frac{1}{RC};\ p_0^* = \frac{1}{RC} \tag{6b}$$

The pole–zero plane (Figure 6) shows a symmetrical position of pole and zero referred to as the imaginary axis. This configuration of poles has, typically, for all-pass circuits, a constant amplitude characteristic and a frequency-dependent phase characteristic as follows directly from (5b). All-pass systems play an important role in the correction of the phase characteristic (Woschni, 1988; Haykin, 2001).

Every system containing zeros in the right half plane can be split into an all-pass system and a so-called minimal-phase system without zeros in the right half plane, as shown in Figure 7 for the system represented in Figure 5c.

REFERENCES

Davies, W.D.T. (1970) *System Identification for Self-Adaptive Control*, Wiley, Chichester.

Haykin, S. (2001) *Communication Systems*, 4th edn Wiley, New York.

Levine, W.S. (1996) *The Control Handbook*, CRC Press, New York; *Proceedings XVI IMEKO World Congress*, Vienna.

Woschni, E.-G. (1988) *Informations Technik*, Huthig, Heidelberg.

Woschni, E.-G. (2000) *Approximations in Measurement – Today still Necessary? Proceedings XVI IMEKO Congress*, Wien, 209–211.

Zadeh, L.A. and Desoer, C.A. (1963) *Linear Systems Theory*, McGraw-Hill, New York.

37: Relationship Between Systems in the Time and Frequency Domain

Eugen Georg Woschni
Technical University Chemnitz, Chemnitz, Germany

1 Laplace Transform in General 216
2 Laplace Transforms of Typical Signals 217
3 Theorems of Laplace Transform 217
4 Important Correspondences of Laplace Transform 217
5 An Important Example 217
References 221
Further Reading 221

1 LAPLACE TRANSFORM IN GENERAL

In **Article 29, Relationship Between Signals in the Time and Frequency Domain, Volume 1**, the relations between signals in the time and frequency domain are treated. In a similar way, the relations in systems may also be described.

The basic idea of the calculation of the time functions, that is, $g(t)$ or $h(t)$, is as follows. Both the spectral function $\widehat{X}(\mathrm{j}\omega)$ of the input and the frequency response of the system $G(\mathrm{j}\omega)$ or generalized $G(p)$ are given. It is then possible to derive the output caused by any of the several sinusoidal input spectral oscillations by multiplying the complex spectral density $\widehat{X}(\mathrm{j}\omega)$ with the frequency response and then applying the summation (integration) of all frequency components. This is valid because the superposition law applies for linear systems.

In system theory, the Laplace transform is preferred to the Fourier transform because it converges more quickly. It can be derived by substituting $\mathrm{j}\omega \rightarrow p$ in the Fourier transform equation (*see* **Article 27, Signals in the Frequency Domain, Volume 1**)

$$F(p) = \int_0^\infty f(t)\mathrm{e}^{-pt}\,\mathrm{d}t = L\{f(t)\} \tag{1a}$$

$$f(t) = \frac{1}{2\pi\mathrm{j}}\int_{c-\mathrm{j}\infty}^{c+\mathrm{j}\infty} F(p)\mathrm{e}^{pt}\,\mathrm{d}p = L^{-1}\{F(p)\} \tag{1b}$$

In system theory, this so-called one-sided Laplace transform is made use of; only the region $t \geq 0$ is of interest. For solving optical problems, the two-dimensional Laplace transform or Fourier transform is applied in the so-called optical transfer function (Goodman, 1968).

The convergence abscissa c in (1b) is chosen in such a way that the poles remain to the left of this abscissa. Owing to the residue theorem (Kaplan, 1962),

$$\int_{c-\mathrm{j}\infty}^{c+\mathrm{j}\infty} F(p)\mathrm{e}^{pt}\,\mathrm{d}p = 2\pi j \sum \mathrm{Re}\, s(p_r) \tag{2a}$$

and for a pole of nth order

$$\mathrm{Re}\, s(p_0) = \frac{1}{(n-1)!}\lim_{p\rightarrow p_0}\frac{\mathrm{d}^{n-1}}{\mathrm{d}p^{n-1}}[F(p)\mathrm{e}^{pt}(p-p_0)^n] \tag{2b}$$

Physical considerations lead to the following relationships between output and input (Woschni, 1988):

$$L\{y(t)\} = L\{x(t)\}G(p)$$
$$y(t) = L^{-1}\{L\{x(t)\}G(p)\} \tag{3a}$$

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

This equation yields for the transfer function $G(p)$

$$G(p) = \frac{L\{y(t)\}}{L\{x(t)\}} = \frac{\widehat{Y}(p)}{\widehat{X}(p)} \tag{3b}$$

Using the Laplace transforms of the unit step $1/p$ and the Dirac function 1 (Table 1) provides, instead of (3b),

$$G(p) = \frac{L\{g(t)\}}{L\{\delta(t)\}} = L\{g(t)\} = \int_0^\infty g(t)\mathrm{e}^{-pt}\,\mathrm{d}t \tag{3c}$$

or

$$G(p) = \frac{L\{h(t)\}}{L\{w(t)\}} = pL\{h(t)\} = p\int_0^\infty h(t)\mathrm{e}^{-pt}\,\mathrm{d}t \tag{3d}$$

and for the calculation of the time functions

$$g(t) = L^{-1}\{G(p)\} = \frac{1}{2\pi\mathrm{j}}\int_{c-\mathrm{j}\infty}^{c+\mathrm{j}\infty} G(p)\mathrm{e}^{pt}\,\mathrm{d}p \tag{3e}$$

$$h(t) = L^{-1}\left\{\frac{G(p)}{p}\right\} = \frac{1}{2\pi\mathrm{j}}\int_{c-\mathrm{j}\infty}^{c+\mathrm{j}\infty} \frac{G(p)}{p}\mathrm{e}^{pt}\,\mathrm{d}p \tag{3f}$$

From (3a) it follows

$$\begin{aligned} y(t) &= L^{-1}\{L\{x(t)\}L\{g(t)\}\} \\ &= \int_0^t x(\tau)g(t-\tau)\,\mathrm{d}\tau = x(t) * g(t) \end{aligned} \tag{3g}$$

or

$$\begin{aligned} y(t) &= L^{-1}\{L\{x(t)\}pL\{h(t)\}\} \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\int_0^t x(t)h(t-\tau)\,\mathrm{d}\tau \\ &= \frac{\mathrm{d}}{\mathrm{d}t}[x(t) * h(t)] \end{aligned} \tag{3h}$$

giving the convolution theorem of the Laplace transform (Table 2). Figure 1 demonstrates this fact. By means of these relations, the deconvolution problem may be solved (Davies, 1970; Woschni, 1988).

$$\begin{aligned} x(t) &= L^{-1}\left\{L\{y(t)\}\frac{1}{G(p)}\right\} \\ &= \int_0^t y(t)g^*(t-\tau)\,\mathrm{d}\tau = x(t) * g^*(t) \end{aligned} \tag{3i}$$

with

$$g^*(t) = L^{-1}\left\{\frac{1}{G(p)}\right\} \tag{3j}$$

Deconvolution often becomes extremely difficult because it is generally not possible to realize the inverse system functions. An alternative approach to the solution, by means of Fourier transforms, is often used (Davies 1970, Woschni 1988).

2 LAPLACE TRANSFORMS OF TYPICAL SIGNALS

In Table 1, typical signals including test signals as pulse, step, and ramp function are put together with the time function and the Laplace transform.

3 THEOREMS OF LAPLACE TRANSFORM

Table 2 shows theorems of Laplace transform. The initial values are often zero because systems are at rest before $t = 0$ and the input signal starts at $t = 0$. In this starting state, the differentiation theorem is very much easier as Table 2 demonstrates

$$L\{f^{(n)}(t)\}\ p^n\ L\{f(t)\} \tag{4}$$

Using this result one gains easily, from the differential equation, the frequency response or transfer function; and vice versa as treated in detail in **Article 36, Systems in the Frequency Domain, Volume 1**.

4 IMPORTANT CORRESPONDENCES OF LAPLACE TRANSFORM

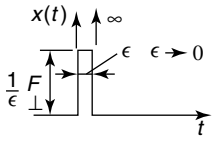
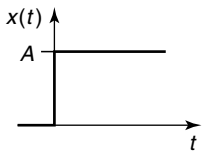
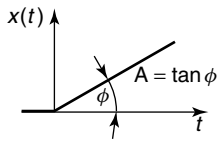
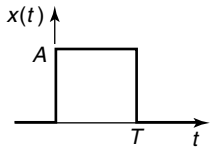
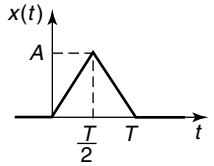
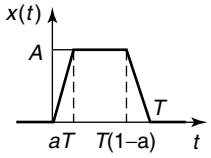
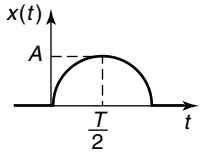
Table 3 contains important correspondences of the Laplace transform to be used to gain the transform of a given time function and vice versa. It is advantageous for it is possible to gain the transforms directly without solving the appropriate equations as shown in the next example.

5 AN IMPORTANT EXAMPLE

As an example, let us deal with a temperature sensor as shown in Figure 2. With the temperature of the environment δ_e and the temperature of the sensor δ_s, the volume V, the surface S, the specific density c, the specific heat or thermal equivalent χ and the heat-transfer coefficient α one obtains the differential equation

$$\begin{aligned} c\chi V\frac{\mathrm{d}\delta_\mathrm{s}}{\mathrm{d}t} + \alpha S(\delta_\mathrm{s} - \delta_\mathrm{e}) &= 0 \\ \frac{c\chi V}{\alpha S}\dot{\delta}_\mathrm{s} + \delta_\mathrm{s} &= \delta_\mathrm{e} \end{aligned} \tag{5a}$$

Table 1. Laplace transforms of test signals.

A periodic test signal function	Equation	Laplace transform
	(Dirac) pulse with integral value $F_\perp$	
	$x(t) = F_\perp \delta(t); \quad \delta(t) = \begin{cases} \infty & \text{for } t = 0 \\ 0 & \text{for } t \neq 0 \end{cases}$	$F_\perp(p) = F_\perp$
	Step function	
	$x(t) = Aw(t); \quad w(t) = \begin{cases} 0 & \text{for } t < 0 \\ 1 & \text{for } t > 0 \end{cases}$	$F_\sqcup(p) = \dfrac{A}{p}$
	Ramp function	
	$x(t) = \begin{cases} 0 & \text{for } t \leq 0 \\ At & \text{for } t \geq 0 \end{cases}$	$E_\angle(p) = \dfrac{A}{p^2}$
	Rectangular pulse	
	$x(t) = \begin{cases} 0 & \text{for } t < 0, t > T \\ A & \text{for } 0 < t < T \end{cases}$	$E_\sqcap(p) = A\dfrac{1 - \mathrm{e}^{-pT}}{p}$
	Triangle pulse	
	$x(t) = \begin{cases} 0 & \text{for } t \leq 0, t \geq T \\ \dfrac{2A}{T}t & \text{for } 0 \leq t \leq \frac{T}{2} \\ 2A\left(1 - \dfrac{t}{T}\right) & \text{for } \dfrac{T}{2} \leq t \leq T \end{cases}$	$E_\wedge(p) = \dfrac{2A}{p}\left(\dfrac{1 - \mathrm{e}^{-p\left(\frac{T}{2}\right)}}{p}\right)^2$
	Trapezium pulse	
	$x(t) = \begin{cases} 0 & \text{for } t \leq 0, t \geq T \\ \dfrac{A}{aT}t & \text{for } 0 \leq t \leq aT \\ A & \text{for } aT \leq t \leq (1-a)T \\ \dfrac{A}{a}\left(1 - \dfrac{t}{T}\right) & \text{for } (1-a)T \leq t \leq T \end{cases}$	$E_\wedge(p) = \dfrac{A}{aT}\dfrac{(1 - \mathrm{e}^{-aTp})(1 - \mathrm{e}^{-(1-a)Tp})}{p^2}$
	Half-sine pulse	
	$x(t) = \begin{cases} 0 & \text{for } t \leq 0, t \geq T \\ A\sin\left(\dfrac{\pi}{T}t\right) & \text{for } 0 \leq t \leq T \end{cases}$	$E_\Omega(p) = \dfrac{A\pi}{p}\dfrac{(1 - \mathrm{e}^{-pT})}{p^2 + \left(\dfrac{\pi}{p}\right)^2}$

Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.

Table 2. Theorems of Laplace transform.

1. Addition theorem

$$L\{f_1(t) + f_2(t)\} = L\{f_1(t)\} + L\{f_2(t)\}$$

2. Multiplication theorem

$$L\{af(t)\} = aL\{f(t)\}$$

3. Shifting theorem

$$\text{for } a > 0 \quad L\{f(t-a)\} = \mathrm{e}^{-pa}F(p)$$

$$\text{or } L\{f(t+a)\} = \mathrm{e}^{pa}\left(F(p) - \int_0^a \mathrm{e}^{-pt} f(t)\,\mathrm{d}t\right)$$

4. Likeness theorem

$$L\{f(at)\} = \frac{1}{a}F\left(\frac{p}{a}\right) \quad \text{if} \quad a > 0$$

5. Attenuation theorem

$$L\{\mathrm{e}^{-at}f(t)\} = F(p+a) \text{ if Re } (p+a) \geq p_0 > 0$$

6. Limit theorem

$$\lim_{t\to\infty} f(t) = \lim_{p\to 0} pF(p); \quad \lim_{t\to 0} f(t) = \lim_{p\to\infty} pF(p)$$

7. Integration theorem

$$L\left\{\int_0^t f(\tau)\,\mathrm{d}\tau\right\} = \frac{1}{p}L\{f(t)\} \text{ if Re } (p) > 0$$

8. Differentiation theorem

$$L\{f^{(n)}(t)\} = p^n L\{f(t)\} - p^{n-1}f(+0) - \cdots - f^{(n-1)}(+0)$$

if the limits

$$\lim_{t\to 0} f(t) = f(+0); \quad \lim_{t\to 0} \dot{f}(t) = \dot{f}(+0); \quad \ldots;$$

$$\lim_{t\to 0} f^{(n-1)}(t) = f^{(n-1)}(+0), \text{ exist}$$

9.–10. Convolution theorem
If the integrals $\int \mathrm{e}^{-pt} f_1(t)\,\mathrm{d}t$ and $\int \mathrm{e}^{-pt} f_2(t)\,\mathrm{d}t$ both are absolutely convergent or at most one absolutely and the other conditionally convergent yields

$$L\{f_1(t)\}L\{f_2(t)\} = L\{f_1(t) * f_2(t)\}$$

with

$$f_1(t) * f_2(t) = \int_0^t f_1(\tau) f_2(t-\tau)\,\mathrm{d}\tau = \int_0^t f_1(t-\tau) f_2(\tau)\,\mathrm{d}\tau$$

Table 3. Laplace transforms of time-dependent functions.

Time function, $f(t)$	Laplace transform, $F(p)$	Time function, $f(t)$	Laplace transform, $F(p)$
0	0	$\frac{t}{2}\sinh(at)$	$\frac{ap}{(p^2-a^2)^2}$
$\delta(t)$	1	$\frac{\sin(at)}{t}$	$\arctan\left(\frac{a}{p}\right)$
$w(t)$	$\frac{1}{p}$		
$g(t)$	$G(p)$	$\frac{2\sinh(at)}{t}$	$\log\left(\frac{p+a}{p-a}\right) = 2\ \text{arc tanh}\left(\frac{a}{p}\right)$
$h(t)$	$G(p)\frac{1}{p}$	$\cos(at+b)$	$\frac{p\cos b - a\sin b}{p^2+a^2}$
$\frac{t^{n-1}}{(n-1)!}$	$\frac{1}{p^n}$	$\cos^2(at)$	$\frac{p^2+2a^2}{p(b^2+4a^2)}$
e^{at}	$\frac{1}{p-a}$	$\cosh^2(at)$	$\frac{p^2+2a^2}{p(p^2-4a^2)}$
$\frac{t^{n-1}}{(n-1!)}\mathrm{e}^{at}$	$\frac{1}{(p-a)^n}$		
$\frac{\mathrm{e}^{at}-\mathrm{e}^{bt}}{a-b}$	$\frac{1}{(p-b)(p-a)}$	$\frac{t}{2}\cos(bt)\sin(at)$	$\arctan\left(\frac{2ap}{p^2-a^2+b^2}\right)$
$\frac{a\mathrm{e}^{at}-b\mathrm{e}^{bt}}{a-b}$	$\frac{p}{(p-b)(p-a)}$	$2\frac{\cos(bt)-\cos(at)}{t}$	$\log\left(\frac{p^2+a^2}{p^2+b^2}\right)$
$\sin(at)$	$\frac{a}{p^2+a^2}$	$2\frac{\cosh(bt)-\cosh(at)}{t}$	$\log\left(\frac{p^2-a^2}{p^2-b^2}\right)$
$t\sin(at)$	$\frac{2ap}{(p^2+a^2)^2}$	$\frac{a\sin(bt)-b\sin(at)}{ab(a^2-b^2)}$	$\frac{1}{(p^2+a^2)(p^2+b^2)}$

Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.

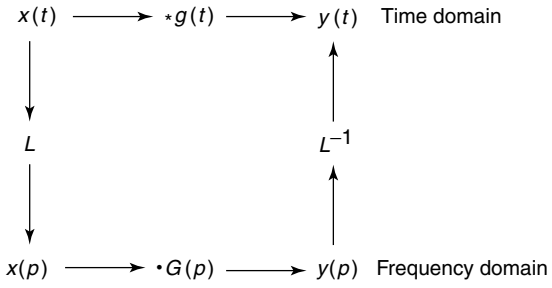


Figure 1. Convolution relationships.

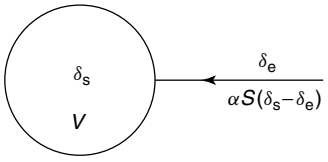


Figure 2. Temperature sensor.

Using (4) the transfer function is

$$G(p) = \frac{1}{1+pT} \text{ with } T = \frac{c\chi V}{\alpha S} \tag{5b}$$

The time functions are obtained using (3e, f)

$$g(t) = \frac{1}{2\pi\mathrm{j}}\int_{c-\mathrm{j}\omega}^{c+\mathrm{j}\omega}\frac{\mathrm{e}^{pt}}{1+pT}\,\mathrm{d}p \tag{5c}$$

$$h(t) = \frac{1}{2\pi\mathrm{j}}\int_{c-\mathrm{j}\omega}^{c+\mathrm{j}\omega}\frac{\mathrm{e}^{pt}}{p(1+pT)}pp \tag{5d}$$

and using the correspondences found in Table 3

$$g(t) = \frac{1}{T}\mathrm{e}^{-\frac{t}{T}} \tag{5e}$$

$$h(t) = 1 - \mathrm{e}^{-\frac{t}{T}} \tag{5f}$$

The time constant is then

$$T = c\chi\frac{V}{\alpha S} \tag{5g}$$

So it follows that the relation of $V/S =$ constant and has to be small for good dynamic behavior, which means

the diameter d of the sensor has to be so small as design conditions allow.

REFERENCES

Davies, W.D.T. (1970) *System Identification for Self-Adaptive Control*, Wiley, Chichester.

Goodman, J.W. (1968) *Introduction of Fourier Optics*, McGraw-Hill, New York.

Kaplan, W. (1962) *Operational Methods for Linear Systems*, Addison-Wesley, Reading, MA.

Woschni, E.-G. (1988) *Informations Technik*, Huthig, Heidelberg.

FURTHER READING

Hsu, H.P. and Hse, H.P. (1995) *Schaum's Outline of Signals and Systems*, McGraw-Hill, New York.

Levine, W.S. (1996) *The Control Handbook*, CRC Press, NJ.

Oppenheim, A.V., Willsky, A.S. and Hamid, S. (1996) *Signals and Systems*, Prentice Hall, NJ.

38: Stability Issues

Eugen Georg Woschni

Technical University Chemnitz, Chemnitz, Germany

1 Concept of Stability 222
2 Stability Criteria 222
3 Root Locus Representation 223
4 Feedback and Feed Forward as an Important Example 224
5 Optimization of Systems 225
6 Deconvolution 226
7 Testing of Systems 226
References 227
Further Reading 228

1 CONCEPT OF STABILITY

Let us deal, as an introduction, with an amplifier having amplification V as shown in Figure 1. If a part of $1/V$ of the output y is coupled back to the input x with the right phase (feedback), the amplifier will produce its own output y without there being an input. The system becomes unstable.

The system is stable if one of the following conditions is met:

1. The eigenvalues of the differential equation p_r exhibit a real part less than zero, that is

$$\mathrm{Re}(p_r) = \delta_r < 0 \tag{1a}$$

2. On the right-hand side of the pole-zero plot there are

no poles of the transfer function; (1b)

3. Decreasing eigenfunctions occur; this means that the weighting function $g(t)$ fulfills the condition

$$\lim_{c\to\infty} \int_0^c |g(t)|\,\mathrm{d}t \le M < \infty \tag{1c}$$

From examination of these stability conditions, certain stability criteria can be applied to test the situation to estimate its degree of stability, for instance, using the Hurwitz and Routh test.

If the following differential equation is given

$$a_n y^{(n)} + a_{n-1} y^{(n-1)} + \cdots + a_0 y = b_0 x + \cdots + b_m x^{(m)}$$

then all coefficients a_r, and the determinants

$$\Delta_\mu = \begin{vmatrix} a_1 & a_0 & 0 & \dots & 0 \\ a_3 & a_2 & a_1 & a_0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{2\mu-1} & a_{2\mu-2} & & \dots & a_\mu \end{vmatrix} \tag{1d}$$

have to be positive, that is, $a_r > 0$, $\Delta_\mu > 0$.

2 STABILITY CRITERIA

For practical applications, graphical methods based on the locus diagram representations are of value. The transfer function $G(p)$ consists of polynomials in the numerator $N(p)$ and the divisor $D(p)$:

$$G(p) = \frac{N(p)}{D(p)} \tag{2a}$$

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.

© 2005 John Wiley & Sons, Ltd.

ISBN: 0-470-02143-8.

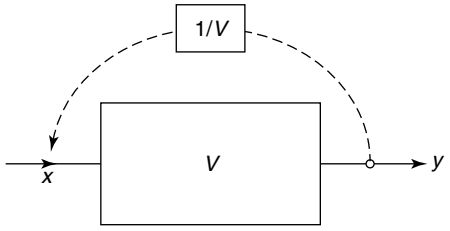


Figure 1. Amplifier with a feedback.

Application of the stability condition (1b) shows that the polynomial of the divisor $D(p)$ is not permitted to have zeros in the right-half plane including those lying upon the imaginary axis (2b) (stability limit).

For testing the stability of a given system, the divisor of the frequency response, that is, $D(\mathrm{j}\omega)$, is drawn on the complex plane (locus diagram) forming a closed curve by inclusion of negative frequencies. Because of the conformal mapping, the unstable field is always the one lying on the right-hand side of the locus curve drawn from $\omega = -\infty$ to $\omega = +\infty$ (see Figure 2). To test for stability, it is necessary to verify where the zero point is situated.

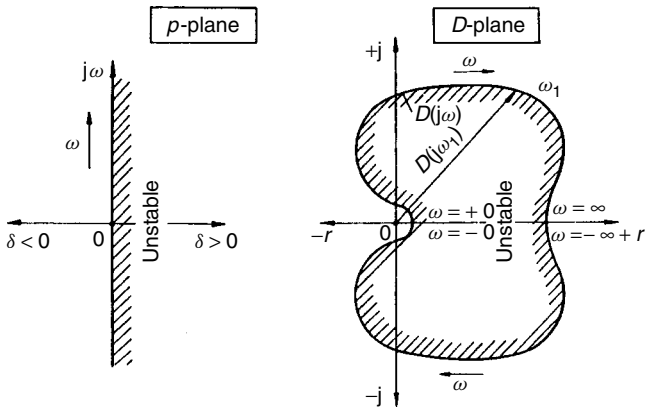


Figure 2. Locus diagram of the divisor $D(p)$.

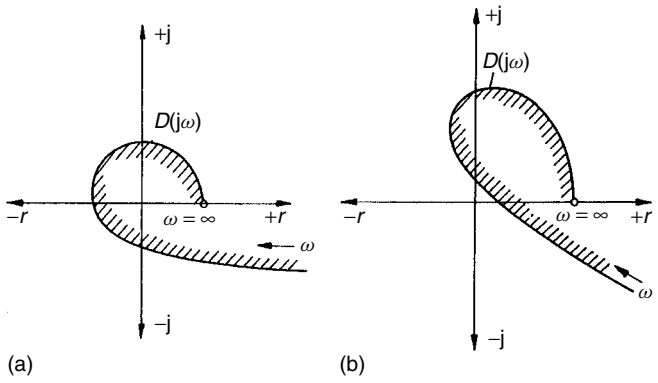


Figure 3. (a) Unstable system and (b) stable system.

If the zero point of the $D(p)$-plane remains left of the locus curve $D(\mathrm{j}\omega)$, passes from $\omega = -\infty$ to $\omega = +\infty$, the system is stable. If not, it is unstable (Figure 3) to some degree.

The procedure is also suitable for developing a statement on the stability margin (see Woschni, 1988; Zadeh and Desoer, 1963 and Section 4 of this article).

3 ROOT LOCUS REPRESENTATION

As described in (1b), the locus of the poles of the transfer function $G(p)$ facilitates the test of the stability of a system. The transfer function

$$G(p) = \frac{\dfrac{b_0}{a_0} + p\dfrac{b_1}{a_0} + \cdots + p^m\dfrac{b_m}{a_0}}{1 + pT_1 + \cdots p^n T_n^n} = \frac{\displaystyle\sum_{\nu=0}^{m}\frac{b_\nu}{a_0}p^\nu}{\displaystyle\sum_{\mu=0}^{n} T_\mu^\mu p^\mu} \tag{3a}$$

may be written in the form (Woschni, 1988; Levine, 1996), *see also* **Article 36, Systems in the Frequency Domain, Volume 1**

$$G(p) = c\frac{(p-p_1^*)(p-p_2^*)\ldots(p-p_m^*)}{(p-p_1)(p-p_2)\ldots(p-p_n)} = c\frac{\displaystyle\prod_{r=1}^{m}(p-p_r^*)}{\displaystyle\prod_{\mu=1}^{n}(p-p_\mu)} \tag{3b}$$

This means the roots of the divisor are the poles of $G(p)$ and thus they are responsible for stability. Stability means that there exist no poles in the right-hand side of the pole-zero plane (Figure 4).

Sampled-data systems with the sampling time T_S yields $z = \mathrm{e}^{pT_S}$, (Woschni, 1988; Levine, 1996)

$$G(z) = c\frac{(z-z_1^*)(z-z_2^*)\ldots(z-z_m^*)}{(z-z_1)(z-z_2)\ldots(z-z_n)} = c\frac{\displaystyle\prod_{\nu=1}^{m}(z-z_\nu^*)}{\displaystyle\prod_{\mu=1}^{n}(z-z_\mu)} \tag{3c}$$

Here, instead of the right-hand side of the pole-zero plane, as with (1b), the region outside the unit circle is the unstable one. So poles have now to be inside this circle to achieve stability as Figure 4 shows.

The main idea of the root locus representation is to allow movement of the poles by varying system parameters when investigating the stability. The method was introduced by Evans (1954) and is applied to find the critical

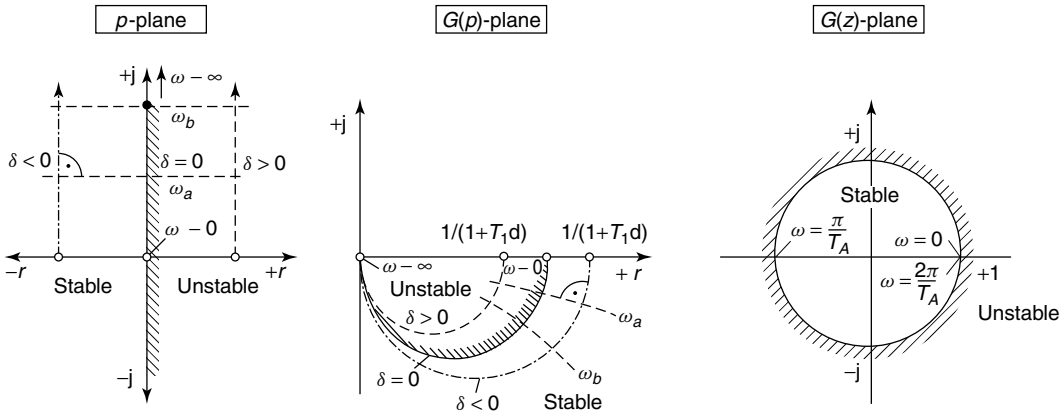


Figure 4. Mapping of the p-plane to the $G(p)$- and the $G(z)$-plane.

amplification V of closed-loop systems (automatic control systems) as dealt with in Section 4.

4 FEEDBACK AND FEED FORWARD AS AN IMPORTANT EXAMPLE

As an example of a commonly met system in instrument systems, consider the feedback system shown in Figure 5. For negative feedback (control situation), the frequency response is given by (*see* Section 3 of **Article 35, Systems in the Time Domain, Volume 1**)

$$G_e(j\omega) = \frac{G_1(j\omega)}{1 + G_1(j\omega)G_2(j\omega)} \tag{4a}$$

The $D(j\omega)$ function is

$$1 + G_1(j\omega)G_2(j\omega) \tag{4b}$$

Instead of testing this function, it is more convenient to verify that the open-loop frequency response

$$G_1(j\omega)G_2(j\omega) \tag{4c}$$

fulfills the condition related to the point '−1' as shown in Figure 6 (Nyquist diagram). The diagram illustrates that an increasing amplification V will take the system from a stable to an unstable state. This fact may be caused by nonlinearity and may lead to unstable oscillatory behavior, (see Section 3 of this article).

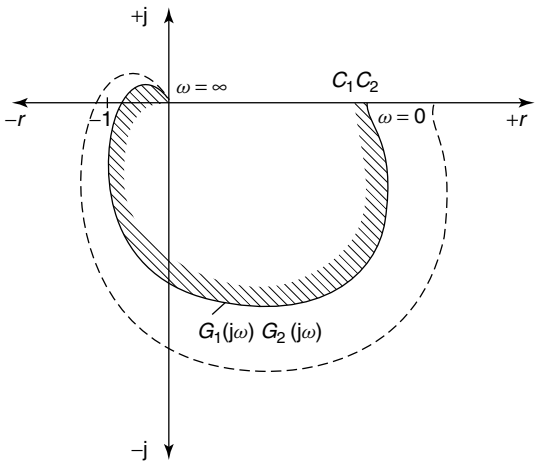


Figure 6. Testing stability of closed-loop systems (control).

A very important device is the operational amplifier shown in Figure 7.

Using (4b) the frequency response is with $|\underline{Z}_1| \ll |\underline{Z}_2|$ and $\underline{V} \to \infty$

$$\frac{\hat{Y}}{\hat{X}} = \underline{V}^* = \frac{\underline{V}}{1 + \dfrac{\underline{V}\,\underline{Z}_1}{\underline{Z}_1 + \underline{Z}_2}} = \frac{1}{\dfrac{1}{\underline{V}} + \dfrac{\underline{Z}_1}{\underline{Z}_2}} = \frac{\underline{Z}_2}{\underline{Z}_1} \tag{4d}$$

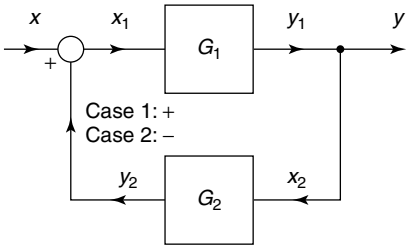


Figure 5. Feedback system.

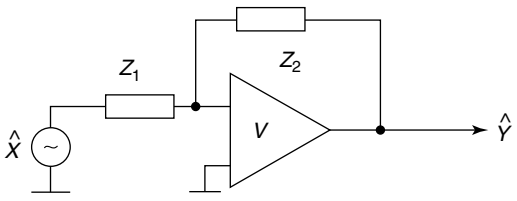


Figure 7. Wired-up operational amplifier.

Table 1. (a) Structurally stable and (b) structurally unstable systems..

a)					
$G_{1,2}(\mathrm{j}\omega)$	c	$\dfrac{c}{1+\mathrm{j}\omega T_1}$	$\dfrac{c}{\mathrm{j}\omega}$	$\dfrac{c}{\mathrm{j}\omega}$	$\dfrac{c}{1+\mathrm{j}\omega T_1}$
$G_{2,1}(\mathrm{j}\omega)$	c	c	c	$\dfrac{c}{1+\mathrm{j}\omega T_2}$	$\dfrac{c}{1+\mathrm{j}\omega T_2}$
b)					
$G_{1,2}(\mathrm{j}\omega)$	$\dfrac{c}{\mathrm{j}\omega}$	$\dfrac{c}{-\omega^2}$	$\dfrac{c}{\mathrm{j}\omega}$	c	
$G_{2,1}(\mathrm{j}\omega)$	$\dfrac{c}{\mathrm{j}\omega}$	c	$\dfrac{c}{\mathrm{j}\omega(1+\mathrm{j}\omega T_2)}$	$\dfrac{c}{1-\omega^2 T_2^2}$	

Source: Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.

With $\underline{Z}_2 = R_2$ and $\underline{Z}_1 = 1/\mathrm{j}\omega C$ it is a differentiation network

$$\underline{V}^x = \mathrm{j}\omega C R_2 \tag{4e}$$

and with $\underline{Z}_2 = (1/\mathrm{j}\omega C)$ and $\underline{Z}_1 = R_1$ an integrator (Miller-integrator, bootstrap-generator)

$$\underline{V}^x = \frac{1}{\mathrm{j}\omega C R_1} \tag{4f}$$

Table 1 also compares structurally stable and unstable systems.

5 OPTIMIZATION OF SYSTEMS

Optimization is treated here specifically with respect to system theory. Both dynamic errors and errors caused by disturbances, for example, noise, are taken into consideration.

The behavior of measuring systems depends upon a number of parameters k_i, for example, time constants T_r, which, at least to a certain extent, may be freely selected. If the parameters of the system cannot be changed, an optimization can be effected by a following correction (also called *compensation*) system or by a correction system connected in opposition, the parameters of such a system being adjustable. In the commonly found microprocessor-based applications, the behavior of the system is set up by suitable programming.

The fundamental idea is shown for a one-dimensional problem in Figure 8. The difference between the output variables of the ideal optimum system (model) and the real system, with a following correction system, is formed and assessed according to the computing instruction defined by the optimization criterion. The parameters k_i of the correction system are set so that the real system resembles the optimum system as far as possible.

Distinction is made between static and dynamic optimization (Bellman, 1961) according to the definition of the performance criterion. If the system is dimensioned such that a performance function depending directly on the parameters

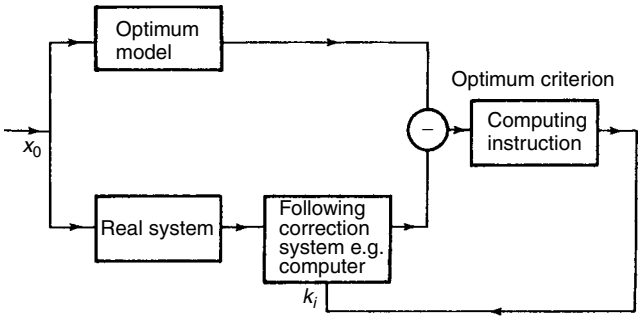


Figure 8. Principle of optimization.

k_i becomes optimum, we have a static optimization. In dynamic optimization, on the other hand, a process $x(t)$ is sought such that a performance function depending on a changing functional process assumes an extremal value.

The previous considerations imply that the central need of optimization is the creation of a suitable optimization criterion.

In most cases, the criterion of the mean-square error, as shown in Figure 9, is used. If $G_{\mathrm{real}}(\mathrm{j}\omega)G_{\mathrm{corr}}(\mathrm{j}\omega)$ is the frequency response of the system corrected by a following correction system $G_{\mathrm{corr}}(\mathrm{j}\omega)$ as shown in Figure 8, we obtain the mean-square error (Woschni, 1988)

$$\begin{aligned}\overline{\varepsilon^2} &= \int_0^\infty S_{xx}(\omega)|G_{\mathrm{id}}(\mathrm{j}\omega) - G_{\mathrm{real}}(\mathrm{j}\omega)G_{\mathrm{corr}}(\mathrm{j}\omega)|^2\,\mathrm{d}\omega \\ &\quad + \int_0^\infty S_{zz}(\omega)|G_{\mathrm{corr}}(\mathrm{j}\omega)|^2\,\mathrm{d}\omega \\ &= \overline{\rho^2} + P_z(5)\end{aligned}$$

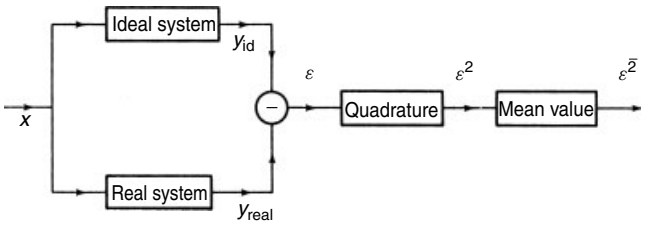


Figure 9. Mean-square error generation.

with the assumption having been made that the disturbances occur prior to the correction computer.

6 DECONVOLUTION

In **Article 35, Systems in the Time Domain, Volume 1**, the convolution integral is introduced to obtain the output $y(t)$ if the input $x(t)$ and the weighting function of the system $g(t)$ are known (Levine, 1996; Woschni, 1988).

$$y(t) = \int_0^t g(t)x(t-\tau)\,\mathrm{d}\tau \tag{6a}$$

The inverse problem is of great importance in measuring systems and is known as deconvolution. There the output signal $y(t)$ is measured with the input $x(t)$ or the wanted signal. The behavior of the measuring device, the weighting function $g(t)$, or the transfer function $G(p) = \mathbf{L}\{g(t)\}$ is known. System theory and Laplace transformation rules yields (Woschni, 1988; Levine, 1996)

$$x(t) = \int_0^t g^*(\tau)y(t-\tau)\,\mathrm{d}\tau \tag{6b}$$

With the weighting function $g^*(t)$ of the inverse system

$$g^*(t) = \mathbf{L}^{-1}\left\{\frac{1}{G(p)}\right\} \tag{6c}$$

As introduced in **Article 36, Systems in the Frequency Domain, Volume 1**, the transfer function of the measuring system $G(p)$ may be described by the pole-zero plane plot. Figure 10 shows this plane for a general nonminimal-phase system containing a minimal-phase system and an all-pass system. The inverse system $g^*(t)$ now has poles at the positions where the original system has zeros and vice versa.

Until there are no zero points of the original system in the right-half plane of the pole-zero plane (so-called minimal-phase system), it is possible to realize the inverse system. If the degree of the divisor is greater than the degree of the counter, this is possible as a limiting case. If the original system is a nonminimal-phase kind, the zero points of the all-pass in the right half of the pole-zero plane will be poles of the inverse system. This means that the system will be unstable and a realization of the deconvolution is not possible. In this case, it is only possible to fulfill the amplitude conditions of (6c) and to use an additional all-pass unit to gain a constant and increasing delay time.

7 TESTING OF SYSTEMS

Testing of systems is the task of system identification.

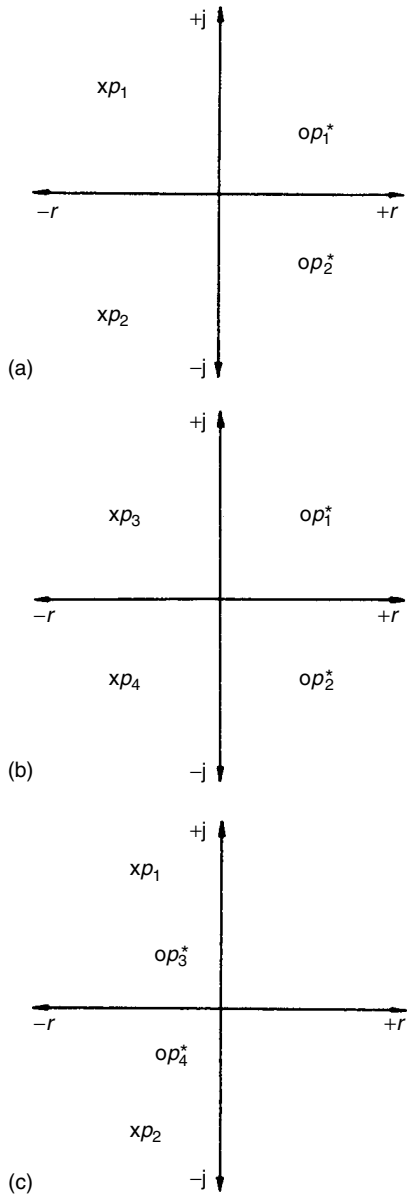


Figure 10. (a) Nonminimal-phase system (b) an all-pass system; and (c) a minimal-phase system.

Because of the relationships between performance in time and frequency domains, it is possible to calculate either of these functions if the other one is measured.

Often, it is the availability of test equipment that decides which of these characteristic functions is to be found. For instance, to test the frequency response of a temperature sensor, it is easier to rapidly change its environment (impulse stimulus giving a time domain response) than it is to apply cyclically varying temperature (frequency domain).

Testing of systems is performed by means of test signals produced by a test generator, recording the corresponding

Table 2. Survey of test signals.

Characteristic function	Input test signal	Output function	Characteristic values
Differential equation	Not specified	Not specified	Time constants T_r
In the frequency domain	Harmonic oscillation $x = \widehat{X}e^{j\omega t}$	Frequency response $G(j\omega)$; Transfer function $G(p)$	Critical frequency $f_c; \omega_c$
In the time domain	Unit step function $w(t)$ Unit pulse = Dirac delta function $\delta(t)$	Transient response $h(t)$ Weighting function $g(t)$	Transient time t_r, dead time t_d, delay time t_l, compensating time t_c, overshoot Δx_0
Stochastic functions	White noise $S(\omega) = \text{constant}$	Cross-correlation function $\psi_{xy}(\tau) \triangleq g(\tau)$	Transient time t_{tr} Correlation time t_{corr}

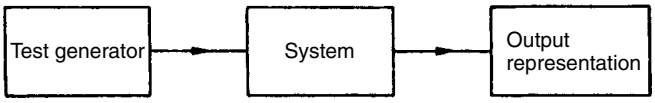


Figure 11. Testing of systems.

output signals as illustrated in Figure 11. Table 2 gives a survey of the test signals used, the output signals, and the characteristic values used for approximations. Today, the process of obtaining the characteristic functions is usually automated, making use of digitally programmable function generators that are controlled by microcomputers, to form the appropriate input signals.

In particular, the frequency response $G(j\omega)$ is obtained by measuring both phase angle and proportion of the amplitudes of output to a given sinusoidal input.

Before recording the true values, it is necessary to wait until the steady state solution appears, that is, $t > t_{tr} = 1/(2f_c)$.

If the measuring device itself has a nonideal frequency response $G_x(j\omega)$, $G_y(j\omega)$, the real frequency response $G(j\omega)$ may be calculated from the incorrect value $G^*(j\omega)$ by means of the relation (Woschni, 1972)

$$G(j\omega) = G^*(j\omega)\frac{G_x(j\omega)}{G_y(j\omega)} \tag{7a}$$

Other principles of calibration make use of comparison systems or reciprocity principles for systems with reversible operation (Woschni, 1972).

If the characteristic functions in the time domain, $g(t)$ of $h(t)$, are to be found, a problem arises in that the input signals are not the ideal ones, as shown in Figure 12 for a nonideal step function $w^*(t)$. Instead of the real transient response $h(t)$, $h^*(t)$ is used:

$$h^*(t) = \frac{d}{dt}\int_0^t w^*(\tau)h(t-\tau)\,d\tau \tag{7b}$$

By means of a deconvolution, it is possible to gain the real transient response where $w^*(t)$ is known (Woschni, 1972).

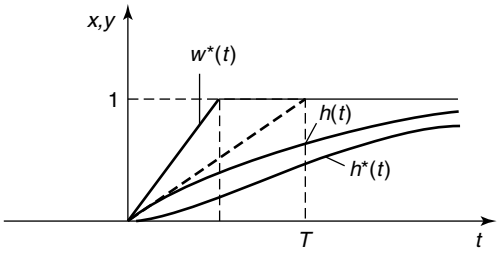


Figure 12. Testing with a nonideal step function $w^*(t)$.

In practice, it is usually sufficient that the rise time of the step function is smaller than one magnitude of the transient time of the system to be tested.

Of practical importance, especially for self-adaptive systems, are those methods of identification that use the noise as the input of a system (Davies, 1970). The weighting function $g(t)$ may be calculated by deconvoluting the relation

$$\psi_{xy}(\tau) = \int_0^\infty g(t)\psi_{xx}(t-\tau)\,dt \tag{7c}$$

with $\psi_{xx}(t)$ autocorrelation function of the input and $\psi_{xy}(t)$ cross-correlation function.

Because the deconvolution technique becomes difficult, methods in the frequency domain or special noise sources are used (Davies, 1970). Use of white noise yields

$$\psi_{xy}(\tau) = \text{constant} \times g(\tau) \tag{7d}$$

which means that the weighting function becomes a constant multiplied by the cross-correlation function (Woschni, 1988). For more details, including errors arising from use of a nonideal correlation function, and the reduction of errors see Davies (1970).

REFERENCES

Bellman, R. (1961) *Adaptive Control Processes*, Princeton University Press, Princeton, NJ.

Davies, W.D.T. (1970) *System Identification for Self-Adaptive Control*, Wiley, Chichester.

Evans, W.R. (1954) *Control System Dynamics*, McGraw-Hill, New York.

Levine, W.S. (1996) *The Control Handbook*, CRC Press, New York.

Woschni, E.-G. (1972) *Meßdynamik*, Hirzel, Leipzig.

Woschni, E.-G. (1988) *Informations Technik*, Huthig, Heidelberg.

Zadeh, L.A. and Desoer, C.A. (1963) *Linear System Theory*, McGraw-Hill, New York.

FURTHER READING

Hsu, H.P. and Hse, H.P. (1995) *Schaum's Outline of Signals and Systems*, McGraw-Hill, New York.

Oppenheim, A.V., Willsky, A.S. and Hamid, S. (1996) *Signals and Systems*, Prentice Hall, NJ.

39: Characteristics of Data, Information, Knowledge, and Wisdom

Timothy Lindsay John Ferris
University of South Australia, Mawson Lakes, South Australia, Australia

1 Definition of Data and its Characteristics 231
2 Definition of Information and its Characteristics 231
3 Definition of Knowledge and its Characteristics 232
4 Definition of Wisdom and its Characteristics 233
5 Relationship between Data, Information, Knowledge, and Wisdom 233
Related Articles 234
References 234

1 DEFINITION OF DATA AND ITS CHARACTERISTICS

Data in a measurement system is the raw form of what comes out of the measuring equipment. Measurement apparatus is applied to the phenomenon, which is observed and produces whatever representation of the observed phenomenon is consistent with the transformation effected by the instrument. A measurement system has two essential elements involved in the process of production of data; being the observed phenomenon and the measuring apparatus. Data is the result of the application of the apparatus to the phenomenon. The observed phenomenon has some state, which follows from the fact that there is some reality present, and that state is observed by means of the application of a standard instrument of a kind, sensitive to variations of some kind in the observed phenomenon.

The explanation above is necessarily abstract, because it intends to consider the broad range of possible scenarios, in which the phenomenon may be either a tangible attribute of tangible objects, such as mass, or some intangible attribute, such as intelligence. In any case, the instrument applied must be of a suitable kind to distinguish manifestations of the phenomenon in such a manner as to effect a mapping to a scale used for description of the phenomenon. This relationship is illustrated in Figure 1, which shows in symbolic form the relation of the observed phenomenon, X, which is an attribute of something, usually a tangible object, which may or may not itself be tangible. The measuring instrument transforms this to a parameter, Π, which is the data achieved by the measurement process. The transformation linking the phenomenon to the parameters is called Ψ, to symbolize that this transformation is a physical transformation.

The domain of the mapping, Ψ, is the possible value of the attribute manifestation in the observed, X; the range is the mapping of that domain into the data space of the possible parameter set, Π. A data set is produced when instruments are correctly applied to the observed. The data may take many forms but it is an image of the observed reduced to the dimensions detected by the instrument used. The image is 'flatter' than the observed because it could be investigated with respect to multiple attributes, but the image is the result of observation with respect to only a subset of the whole. This section is based on (Ferris 1997a, b, Section 8.3).

2 DEFINITION OF INFORMATION AND ITS CHARACTERISTICS

Information is a step up on data in complexity and value. Measurement data are the symbols representing the

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

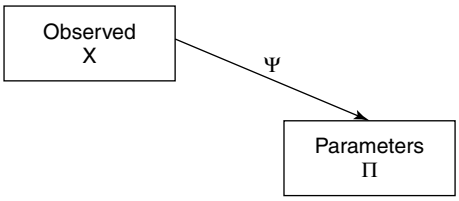


Figure 1. Illustration of the relation of the observed phenomenon, X, which is observed using a measurement instrument performing the transformation, Ψ, to produce the data, parameters, Π.

observations, whereas measurement information is the representation of the observation, which is embodied in the data (Crouch and Sydenham, 1994). Data is not information, but data contains information, which is the significance of the data, and is the vehicle for shifting information around. Depending on the representation scheme, the symbols may or may not appear to correspond with their significance. Information must be conveyed symbolically because significance cannot be directly transferred (Ayer, 1969; Jolley, 1973).

The distinction between data and information can be usefully drawn by considering the work done in quantification of 'information' in different disciplines. Let the measure of information developed by Shannon in communication theory be I_{Sh} (Shannon, 1964) and that developed by Dretske in philosophy be I_{Dr} (Dretske, 1981). In this analysis, I_{Sh} is found:

$$I_{Sh} = k \times \ln P_f \tag{1}$$

where $k \in \Re$ is a constant

P_f is the number of possible message forms

and I_{Dr} is found:

$$I_{Dr} = k \times \ln P_m \tag{2}$$

where $k \in \Re$ is a constant

P_m is the number of possible message meanings.

Both these measures of information quantity are important in measurement theory since instruments produce a representation of the observed in the form of parameter data, which must be communicated within the system, and information is carried by that data as a means of communicating something of substance about the observed.

I_{Sh} relates to the channel capacity required to convey the measurement data in a digital form. The number of possible message forms is the total number of possible different messages that could be conveyed using the number of digits in the message representation. The quantity of data I_{Sh} is dependant on such measurement system characteristics as the range and resolution of the instrument, which together define the number of states of the observed phenomenon that the measurement apparatus is capable of distinguishing. However, the representation of the data demands an integer number of bits, and so the minimum possible value of I_{Sh} is rounded up to the nearest integer above the number of resolvable states of the observed.

I_{Dr} pertains to the resolution of meaning effected by the message and so relates to the number of distinct cases that can be represented by the data in the system. The number of resolvable states of the observed is the number of possible meanings in the data received. Therefore, certainty of a particular state reduces the uncertainty about the state of the observed by defining which one of P_m possible states is present. The consequence is that I_{Dr} is as defined above.

The above discussion makes it clear that $I_{Sh} \geq I_{Dr}$. This is proved in Ferris (1997b, Th 3.3).

3 DEFINITION OF KNOWLEDGE AND ITS CHARACTERISTICS

Knowledge is a more complex and valuable entity than information, but is dependant on information. Information is not knowledge but is the foundation upon which knowledge is built. Information is transformed into knowledge when integrated with other information about the observed phenomenon, which enables the observer to increase understanding of the instance of the observed phenomenon through the observations that have been made.

The concept of knowledge is commonly associated with communication effected through natural language and this is a useful point to develop an understanding of knowledge in a measurement situation. Natural language communicates knowledge, that it has meaning, through the existence of many definitions, relating to words and the implications of grammatical constructions; conventions, relating to the interpretation of the combination of linguistic expression and context; and intention, relating to the purpose for which a particular speaker assembles a particular combination of words to assemble a means of communication of a thought (Burge, 1975). In a natural language communication situation, knowledge arises because the listener hears the discourse and interprets it in the light of the combination of the known definitions, conventions, and perceived intentions of the speaker.

To derive knowledge about the observed phenomenon is a common, and worthy, purpose of measurement activities. But, in the measurement situation, the data and the information that the data contains arises in a quite different situation than natural language. Measurement information arises

from nature, not intention, since measurement information is a statement of the state of a particular phenomenon at the time of observation (Butcharov, 1964). Therefore, interpretation of measurement information to produce knowledge of the observed is not affected by the factors used in interpretation of natural language, these being definitions, conventions, and intentions.

To obtain knowledge from measurement information requires that information existing because of the nature and structure of the world be assigned significance by the interpreter of the information. Situations, such as measurement, in which information exists independent of intention are fundamentally different than cases where information exists as a result of intention (Lewis, 1946; Brodbeck, 1968). Where information arises out of nature, the interpretive approach of perception, in which the information is recognized as containing evidence that the observed phenomenon is in a state evidencing one of some finite number of meaningful categories, is required. Perception is a process of recognizing that an observed phenomenon corresponds to a class of phenomenon that the perceiver knows and so the case observation is linked into a meaningful structure of knowledge of the world, and so the observed state of the observed become the basis of knowledge of the contextual significance of the state of the observed.

4 DEFINITION OF WISDOM AND ITS CHARACTERISTICS

Wisdom is the ability to judge the appropriateness of action and to behave in the correct manner in a situation. Wisdom has been recognized and valued since ancient times, having formed one of the traditional branches of philosophy. Some ancient philosophies were essentially embodiments of wisdom, with examples of this kind of writing found in Confucius and parts of the Old Testament, amongst other sources.

In measurement systems, wisdom relates to the taking of appropriate action on the basis of the observations made through the sensors. This occurs in two ways, the automatic feedback control system and systems in which measurements contribute to decisions.

In an automatic control system, illustrated in Figure 2, there is a controlled variable, the output, which is observed using a sensor. The signal produced by the sensor is used as an input to the primary system to adjust the output. The intention is to have a system in which the actual output approximates as closely as possible, at all times, the desired output.

In the case of a feedback control system, 'wisdom' is defined in the control loop design as the control input

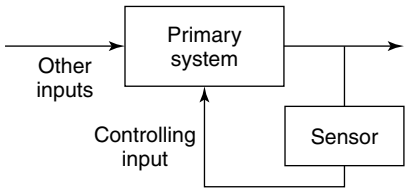


Figure 2. A generalized automatic feedback control system block diagram.

required to adjust the behavior of the primary system so that the output becomes the desired output. Determination of the control input is nontrivial because the control system is a dynamic system, which normally involves complicating factors such as time delays, with the result that a simplistic adjustment of the control input is liable to introduce significant problems such as overshoot, undershoot, or oscillation of the output relative to the desired output value.

In a measurement-based decision system, the contribution of the measurement system is to provide knowledge that classifies the observed phenomenon in a way that enables a decision to take appropriate action. The action may be taken by either a person or a machine. In a measurement-based decision system, the action taken is not a feedback control action of seeking to adjust the controlled output, but rather is an action of an apparently unrelated kind that is an appropriate action to take on the basis of the measurement observations. For example, in medicine, a patient may be measured in one way or multiple ways, resulting in knowledge of the state of the patient, embodied in a diagnosis. The action of wisdom, performed by the doctor, is a particular therapeutic maneuver, performed to correct the situation evidenced by the state of the patient, but not as a mechanistic adjustment of a control input that adjusts the observable output to lead to the conclusion that the patient's state is satisfactory.

5 RELATIONSHIP BETWEEN DATA, INFORMATION, KNOWLEDGE, AND WISDOM

Data is the lowest level of entity that is used in a system for processing matters of meaning concerning things. Data can be transformed into information when the data is organized into a form that distinguishes between resolvably different cases of the observed phenomenon. Knowledge is gained when information from one or more sources is integrated and analyzed to produce a coherent understanding of the matter of observation. Wisdom concerns the performance of appropriate action based on integration of various sources of knowledge. Wisdom normally involves analysis of items,

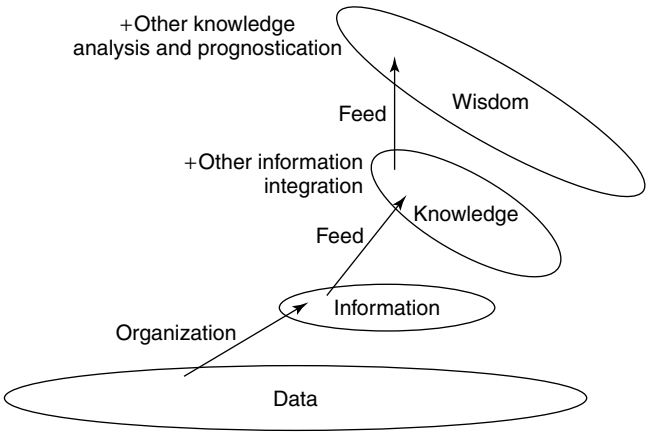


Figure 3. A pictorial representation of the relationship between data, information, knowledge, and wisdom.

probably many, of knowledge, and some kind of modeling of the situation in order to provide prognostication of the outcome resulting from various possible actions. The ideas of this paragraph are expressed in Figure 3.

Each of the elements, data, information, knowledge, and wisdom is a member in a hierarchy of both complexity, order, and value. The relative sizes of the elements in Figure 3 is also significant. Following the relation stated at the end of the Section 2 that discusses information, it is, in general, true that the quantity of information is less than or equal to the quantity of data in a measurement system. Knowledge, if quantified, would be less than or equal to the quantity of information because knowledge concerns categorization of the observation into meaningful categories of distinct. The quantity of wisdom is the smallest because there is always one best solution, although in many situations it may be unclear what the best solution is, and one rational and justifiable solution is selected from a small group of candidate actions instead.

The entities are situated in a stack of increasing order, because each entity is an expression of a more ordered arrangement of representation of the state of the observed phenomenon and the appropriate action arising from that state. The order is found in the fact that the entities each express the understanding of the observed phenomenon in greater simplicity of representation of the observed.

The complexity of representation decreases as one progresses upwards in the stack of elements. However, the complexity of the elements themselves increases as one advances up the stack because the simplicity of representation results from increased integration of various factors into the representation of the observed effected at each level.

In summary, a measurement process is used to collect data about the observed phenomenon; that data is organized as information about the observed and leads to the knowledge of the state of the observed phenomenon when integrated with other information about the observed; and the knowledge, when integrated with other knowledge about the observed and its context provides a basis for the exercise of wisdom.

RELATED ARTICLES

Article 13, Signals, Information and Knowledge, and Meaning, Volume 1; **Article 14, Hierarchical Aspects of Measurement Systems, Volume 1**; **Article 24, Principles of Epistemology as Related to Measurement, Volume 1**.

REFERENCES

Ayer, A.J. (1969) *Metaphysics and Common Sense*, Macmillan, London.

Brodbeck, M. (1968) Meaning and Action, in *The Philosophy of Science*, (ed. P.H. Nidditch), Oxford University, London, (pp. 97–120).

Burge, T. (1975) On Knowledge and Convention. *The Philosophical Review*, **84**(2), 249–255.

Butcharov, P. (1964) Knowledge of Meanings and Knowledge of the World. *Philosophy*, **39**(148), 145–160.

Crouch, V. and Sydenham, P. (1994) Relationship between T&E and M&I', in *Australasian Instrumentation and Measurement Test and Evaluation Conference*, Adelaide (pp. 1–6).

Dretske, F.I. (1981) *Knowledge and the Flow of Information*, MIT, Cambridge.

Ferris, T.L.J. (1997a) The Concept of Leap in Measurement Interpretation. *Measurement Journal of the International Measurement Confederation*, **21**(4), 137–146.

Ferris, T.L.J. (1997b) *Foundation for Medical Diagnosis and Measurement*, PhD thesis, University of South Australia.

Jolley, J.L. (1973) *The Fabric of Knowledge*, Gerald Duckworth, London.

Lewis, C.I. (1946) *An Analysis of Knowledge and Valuation*, Open Court, La Salle.

Shannon, C.E. (1964) *The Mathematical Theory of Communication*, University of Illinois, Urbana (pp. 29–125).

40: Sources of Information

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 How Information Propagates from Discovery 235
2 Information Flow from Sources to Depositories 235
3 Problems of Scale 236
4 Enter the Internet 237
Related Articles 238
Reference 238

1 HOW INFORMATION PROPAGATES FROM DISCOVERY

The world contains an infinite amount of latent information – that residing within systems, waiting to be established. The creation of available information begins with people who convert some of the latent information into a recordable form that suits the storage classification then in vogue. Figure 1 depicts how information passes around a circuit in which the information that is useful to measurement finds its way back to be the basis of the design of more measuring instruments, which in turn aids the flow of more information.

2 INFORMATION FLOW FROM SOURCES TO DEPOSITORIES

Figure 2 provides an appreciation of the flow of information through the many kinds and places where information may be found. Scholarship and experimental observation lead to latent information being made available; the first stage is termed *primary information*. This may occur as internal reports within organizations, in the personal records of workers, or in primary published literature. At times, books, often regarded as secondary sources, will contain primary material.

Within a few months of being established (weeks in some cases), primary information, of the publicly open kind, will often be published in primary journals where it is available through library services. Material published in primary journals is usually subject to expert refereeing that helps maintain, but cannot guarantee, the quality of reported material.

Primary information is also available in student theses (dissertations), patent specifications, and in some conference proceedings. A considerable part of it is held in company and defense secret locations where it is not openly available. It is also to be found in the proceedings of conferences, a source that may not be as well refereed as reputable primary journals; but it is usually available sooner.

There are no universal rules to regulate what is published, or where it is published. Primary information often also appears in part in the next class, the secondary outlets. Academic researchers work with the moral duty to publish their findings, but the same does not apply to many others. Because something cannot be located in open depositories is no guarantee that it has not been externalized from the latent state.

Primary information is, for a variety of reasons, not always presented in the form that is required, so it gets processed in many ways. It may need to be collected together to form a work on a common topic, or be reexpressed in a more acceptable form for others to make use of it. In this class are secondary journals, reviews, books, encyclopedias, dictionaries, videotapes, DVDs, CDROMs, the Internet, audiotapes, and films.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

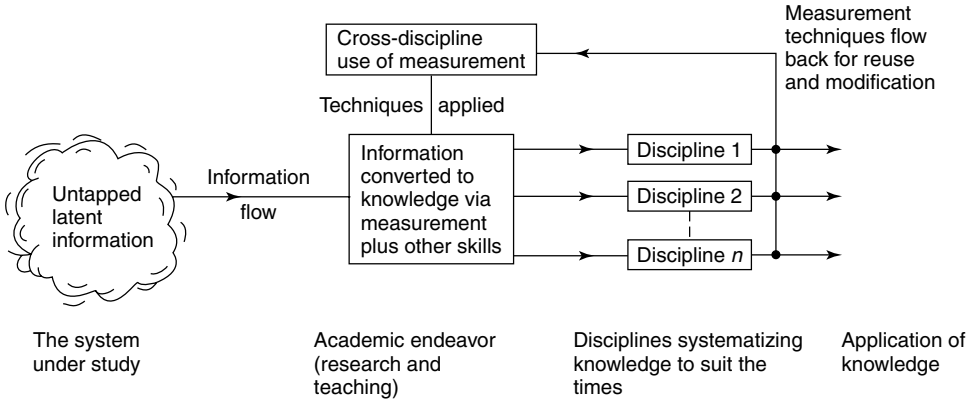


Figure 1. Extraction of latent information leads to new measuring instruments.

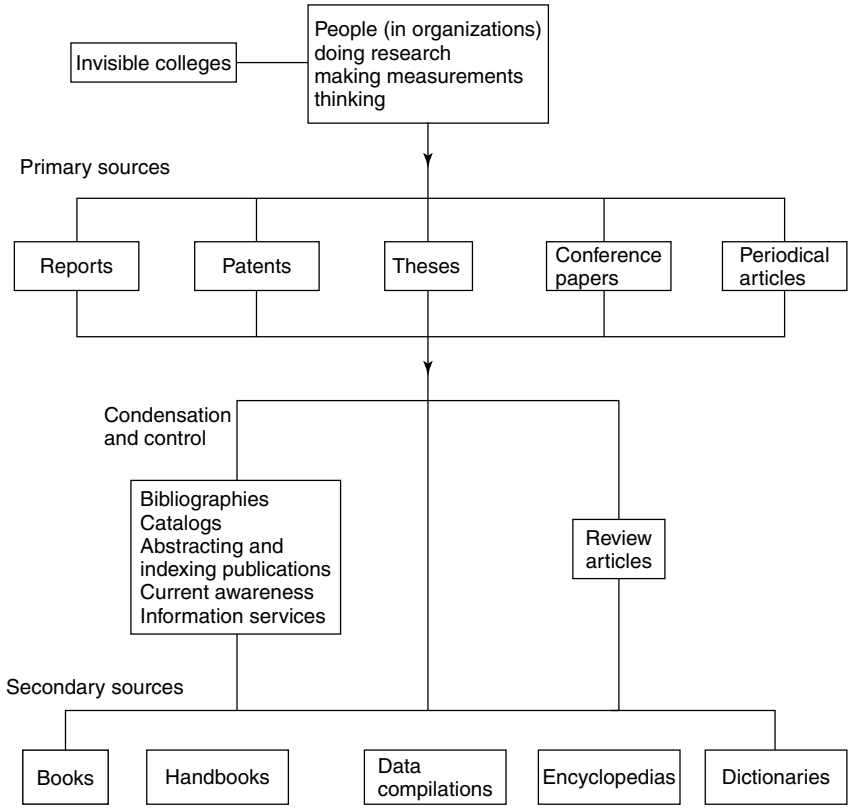


Figure 2. Flow of information through recorded, retrievable systems of documentation. (Reproduced from Melton (1978) with permission of Institute of Physics.)

Secondary information is usually easier to apply, for it distills the essence from the primary reports, leaving out the detail that does not stand the test of time or that is of little bearing to its use.

A major kind of published secondary source is that comprising the reference material that aids retrieval of primary material – bibliographies, abstracts, catalogues, information services.

Reviews and bibliographies are formed from secondary information but their construction, classification, and presentation can present a primary viewpoint, often being original in nature.

3 PROBLEMS OF SCALE

The rate at which information is published is almost incomprehensible. By 1900, the Royal Society in London had to abandon its attempts to catalog all of the scientific publications from 1800 to 1900 – the Author Index alone was

19 volumes in length! In 1974, an estimated 80 000 regular scientific and technical journals in a total of 150 000 of all kinds of journals were being published. Abstracting periodicals then numbered 1500, in a total of 3500. An estimate of that time suggested there were 300 000 books published in a year with over 3000 million books in print. By the end of the twentieth century, this increased exponentially to vast numbers. The rate is not slowing.

The diverse nature of measurement means that information needed could rest almost anywhere in the vast system of documentation, if it exists at all. Well-devised retrieval is essential to glean information in an efficient manner.

4 ENTER THE INTERNET

To provide a more reliable method of communication information, the United States created what has today become known as the Internet or simply the Web. This has grown over the past decade into a vast network of millions of computers that are all connected together via the servers of Internet Service Providers (ISP). The Web is now predominantly used by civilians all over the globe.

These computers offer information on a virtually infinite range of topics. To find the required information, the user makes use of search engines to mine down into the vast resource to select those instances that appear relevant. The user then opens them to see if they are what is needed; many are not, and much time can be wasted in seeking what is not there!

Search engines currently only allow users to locate material on topics of their choice by using symbol searches for words and strings in text files. Being developed, and more widely available soon, will assuredly be the facility to search images for contextual graphical constructs; this uses ontologies for topics, these being large sets of relevant rules about the structure and content of images of the supported domain. For example, the textile industry is adopting this method as a means to find patterns from large numbers of different designs now stored as images in computers. Software applications are available, which use this method for searching images. The development of the ontologies needed is a costly business. A typical ontology will contain over 100 000 rules.

The availability of the Internet can be likened in the time span of man to the epoch when the printing press of the 1400s made information available to many more people of that time. Then, recorded information was only available to a privileged few in painstakingly produced single copies of books of the monks, and in the art of religions such as stained glass windows and paintings.

The Internet has made information available to just about every person on the Earth at a comparatively affordable cost and in almost any geographical location.

On the surface it would seem that the Internet gives its millions of subscribers all they need to know for free. Searches will indeed often discover considerable material on a topic of one's choice. The reality is, however, that the freely available information may not have the veracity required, its time stamp is not always clear, and it rarely provides a sufficient depth of knowledge to allow serious decisions to be taken from it.

Publishers of the traditional technical books have placed considerable emphasis on offering book material online via their highly developed Web sites. These sources of knowledge can be relied upon better and do give organized material that is refereed and recorded in a highly professional and concentrated manner. To provide this service, they must charge fees by subscription or, along the way, as a 'pay as you use' service.

Benefits of having information stored in the electronic, or e-file form, as compared to the traditional hard copy form, are many:

- Access is to any part of it with immensely faster find times for specific knowledge than by using books.
- Hugely larger sets of information can be made available at virtually any location.
- Parts can be selected and reused to form new sets of knowledge (where this is permitted).
- Upgrading the information is a matter of editing files; no reprinting of all of the information with its changes is needed.
- Foundational knowledge will become clear and settle down to allow expansion from sound foundations and not be rejigged for each new edition of a book.

Measurementor, www.measurementor.com, Sydenham (2004), is an example of this form of information delivery. Some of its statistics illustrate the problem that will be faced by those needing up-to-date and definitive knowledge.

Creating the basic knowledge base listing for it showed that the sound set of information needed by a measurement system designer would involve over 4000 one-page items and that even this set of information could only serve satisfactorily as an entry gateway to their fuller needs. It has been estimated by this writer that at least 1000 books would be needed to provide a deep library of information for a profession. All of the titles could not be economically created by one publishing source alone. Internet and gateway products, like Measurementor, can provide this degree of access, for it allows thousands of commercial providers to offer material that sits in their specialty area.

Several Web sites give useful information on measuring systems. They are useful but too often only get the user acquainted with the topics. They also are not coordinated with respect to their information, nor structured to any scheme that stands much scrutiny.

GlobalSpec, http://www.globalspec.com/, claimed to be the world's largest database of technical products and services, provides rapid location and access to a vast number of items – 9000 suppliers catalogues, 850 000 product families and 45 million parts. Many are supported with explanations given in its tutorials and application notes of its Technical Library. The aim of the service is to sell products and services to users on behalf of its clients.

The Sensors magazine Web site http://www.sensorsmag.com/ provides a Literature Showcase for many companies and gives access to the past articles of the Sensors magazine. Search facilities are provided to assist in finding material.

Many manufacturers of measuring systems and instruments offer information assistance via their own Web site. An example of the upper end of this type of information support is that of National Instruments http://www.ni.com/. In their offerings is a service called 'Instrupedia' – a short encyclopedia of measurement that will be found to be useful at times.

The impact of the Internet is already clearly evident. It seems reasonable to suggest that within the next decade, e-sources of information will become routinely linked to parts of all people's lives, especially in support of their professional duties. Already the law and medical professions have taken up Internet use to give their practitioners rapid and full access to their particular bodies of knowledge. Others will follow as the cost of time and time taken to locate information becomes a major cost factor in such areas as teaching, design, and procurement fields.

RELATED ARTICLES

Article 1, Sophistication of Measurement and its Body of Knowledge, Volume 1; **Article 2, Organization of Instrument Science, Volume 1**; **Article 5, Humans in the Real World, Volume 1**; **Article 41, Terminology and Classification of Measurement Systems, Volume 1**; **Article 42, Information Databases of Relevance to Measurement, Volume 1**.

REFERENCE

Sydenham, P.H. (2004), Developing a Measuring Systems Body of Knowledge, *Proc.10th IMEKO TC7 International Symposium*, June 30–July 2, Saint-Petersburg, Russia.

41: Terminology and Classification of Measurement Systems

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Terminology Used in Measurement 239
2 List of Key Terms Relevant to Measurement 239
3 List of Inspec Terms Relevant to Measurement 240
4 Classes of Publications Relevant to Measurement 241
5 Library Classification Schemes 241
6 List of Dewey Classification Codes Where Measurement Material is Located 242
7 List of Journals Relevant to Measurement 245
Related Articles 245

1 TERMINOLOGY USED IN MEASUREMENT

Information retrieval by browsing through original material makes use of many recognition features – words, illustrations, mathematical formulas – but this method is invariably too slow for searching the bulk of the literature.

Realistic retrieval procedures are almost exclusively based on the principle of tracing linguistic terms used by authors in their titles, abstracts, and index keys. Consequently, the choice of terms written into abstracts and used in searching is of paramount importance.

The diversity and uncoordinated multiplicity of the sources of generation (and regeneration!) of instrument concepts has brought with it a highly varied and inconsistent nomenclature. Indeed, some terms used have multiple meanings. It is only in recent times that standardization of terms has begun to be accepted; they are not, as yet, widely used nor broadly developed. Data bases relevant to measurement knowledge are discussed in **Article 42, Information Databases of Relevance to Measurement, Volume 1**.

The wide and often incorrect use of synonymous terms makes literature retrieval difficult. Apparently obvious terms often lead nowhere – metrology, measurement, sensor, and instrumentation in particular; they are too commonly used. The name *instrument science*, for example, might be indexed using at least 10 alternatives, such as measurement systems, measurement physics or measurement science, and so on. The fundamental term *metrology* also has at least two different meanings and it often gets confused with 'meteorology'.

Terms describing the activity of the process of measurement exist in great abundance. Figure 1 lists some of these. It is easy to see how many different terms are used for describing the same thing.

2 LIST OF KEY TERMS RELEVANT TO MEASUREMENT

Measurement, in many specific disciplines, has its own name. Figure 2 lists a compilation of these. Thus, the literature of measurement is spread over the whole body of knowledge. Use of a large-size dictionary usually gives little assistance in separating them, for it is their technical

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

Accounting	Effect	Prospection
Accredit	Evaluation	Quality
Adjustment	Expectation	Rate making
Administer	Experiment	Recognize
A measure of	Exploration	Resolution
Analyse	Features	Sensing
Appraisal	Group	Standardize
Appreciation	Incidence	Style
Arbitrate	Indication	Taxonomy
Assay	Influence	Tests
Assessment	Information	Value
Choice	Inspection	Weighing up
Classification	Interpretation	Worth
Classify	Judge	
Composition of	Logistic	
Criticism	Management	
Criticize	Material constants	
Decide	Measurement	
Decision	Mensuration	
Design	Metricate	
Detection	Monitoring	
Determining	Number	
Diagnosis	Performance	
Distribution	Precision	
Document	Properties	

Figure 1. Terms describing the activities of the process of measurement.

Accounting	Measurement systems engineering
Anthropometrics	Measurement technology
Archaeometry	Mensuration
Assay	Metallography
Automatic testing	Metrology
Biometry	Nuclear engineering
Cybernetics	Pathology
Decision theory	Posology
Diagnostics	Psychrometrics
Econometrics	Quality control
Epidemiology	Quantity surveying
Ergonomics	Radiometry
Faunistics	Scientific instrumentation
Geophysics	Seismology
Instrument physics	Sensitometry
Instrument-science	Sensory physiology
Instrument technology	Stoichiometry
Management	Symptomatology
Man–machine-systems	Taxonomy
Measurement engineering	Teleology
Measurement physics	Time and motion study
Measurement-science	Value engineering

Figure 2. Some names of disciplines involved with measurement.

usage that is of interest, and dictionaries are not good at defining things in those terms.

3 LIST OF INSPEC TERMS RELEVANT TO MEASUREMENT

Use of databases, in particular, requires careful choice of terms in the search profile. In some of them, only terms given in a standard thesaurus are allowable inputs. Figure 3 lists a selection of over 650 (of a total 10 000) terms that

Approximation theory	Non-electric sensing devices
Artificial intelligence	Nuclear instrumentation
Automatic testing	Oceanographic techniques
Character recognition	Optical systems design
Characteristics measurement	Optical instruments
Computerized instrumentation	Parameter estimation
Cybernetics	Patient diagnostics
Data acquisition	Performance index
Data processing	Physical instrumentation control
Data reduction and analysis	Physiological models
Digital instrumentation	Predictor–corrector methods
Display instrumentation	Probes
Dynamic testing	Production testing
Education	Program testing
Electric variables measurement	Quality control
Environmental testing	Quantization
Error analysis	Radiation monitoring
Error detection	Recorders
Fundamental law tests	Relay protection
Geophysical equipment	Research and development management
Height measurement	Reviews (see also published lists of bibliographies with author indexes)
History	Robots
Impulse testing	Seismology
Information science	Self-organising systems
Instrumentation	Sensitivity analysis
Instruments (53 kinds fall into this)	Sensory aids
Laboratory apparatus and techniques	Signal processing
Logic testing	Social and behavioural sciences
Machine testing	Spatial variables measurement
Magnetic variable measurement	Standardization
Man-machine systems	Student laboratory apparatus
Materials testing	Switchgear testing
Measurement (29 specific variables are listed: all appear elsewhere). This is also used instead of term metrology	Telemetering systems
Measurement standards	Testing
	Thesauri
Measurement systems	Transducers
Measurement theory	Units (measurement)
Mechanical testing	Value engineering
Medical diagnostic equipment	Water pollution and detection
Meteorological instruments	
Noise measurement	
Nomenclature and symbols	

Figure 3. Selected terms related to measurement in the INSPEC thesaurus.

relate to the activity of measurement these being contained in the thesaurus of the INSPEC™ database.

4 CLASSES OF PUBLICATIONS RELEVANT TO MEASUREMENT

Consider now the necessary classification of a measurement classification from the author's and reader's viewpoints. It is found that numerous possibilities exist for deciding the key list of index terms.

In general, authors contributing to measurement align more with their field of application or discipline than with sources publishing fundamentals of measurement science. Often, they see the instrument content as insignificant and not worth abstracting in their report. Many facets of a report might not be cited in the few keywords used.

The reality is that the report on a topic including measurement, although given several descriptors, should only be published in one journal. Ethics suggest this is the correct policy, but there is a case for multiple releases in different types of publications so that the contribution can be found by those who need it, as they approach the material from different perspectives.

As an example, consider a hypothetical report containing, as part only, the calibration of a novel, laser-based, alignment system used, by the XYZ Company, to test the military load-carrying capacity of an historic concrete bridge while the bridge is conveying a mobile nuclear reactor. Just where the report should be published is an open question.

A report can be placed into an outlet that emphasizes, with an example of each, its

- principle used (Hall effect)
- philosophical thinking (impact on the philosophy of science and technology)
- application area (manufacturing)
- device (SQUID)
- technology (MEMS)
- theory (sensitivity analysis)
- proprietary aspect (XYZ new product)
- fashionable topic (fuzzy systems)
- industry (user survey), and so on.

Many primary journals publish measurement contributions. Those more obviously dedicated to measurement are many. Figure 4 gives a list of measurement-related journals that could be chosen for publication of the above report. It also gives the rationale for the choice of each.

Other aspects might be that it relates to include transportation, power generation, environmental issues, civil engineering, environmental safety, and many more.

Authors should pay attention to the need to prepare several papers on a major topic, each on its key aspects and for a different readership.

5 LIBRARY CLASSIFICATION SCHEMES

Now consider classification from the viewpoint of location in a library collection. The location of book literature is identified in a library by the classification code-numbering

Example: Calibration of a novel, laser-based, alignment system used, by the XYZ Company, to test the military load-carrying capacityof an historic concrete bridge while the bridge is conveying a mobile nuclear reactor.

Class basis chosen	Example of typical journal for publication
Physical principle used (interferometer)	Applied Optics
Discipline based on a device used (laser)	Laser Focus
Contemporary nature (novel and timely)	New Scientist
Military relevance	Classified report of a military organization
Discipline of use (surveying)	Photogrammetric Engineering
Parameter of measurement (measurand is related to strain)	Journal of Strain Analysis
Measurement principle used (angle measuring interferometer)	Measurement Science and Technology
Constructional material used (concrete)	Magazine of Concrete Research
Standardization (calibration)	Newsletter of National Conference of Standards Laboratories
Design implications	Journal of Elasticity
Testing	International Journal of Non-destructive Testing
Instrumentation detail	Instrumentation and Control Systems
Engineering heritage	Transactions of the Newcomen Society
Trade journal	XYZ Affairs (hypothetical name)

Figure 4. Classes and example journals in which an instrument might be reported.

scheme used. Placement and retrieval are based on an allocated code number to each title, not by the use of a set of key terms, as is the case in computer retrieval of journal literature abstract information.

Several schemes of library location classification are in use. Predominant systems in use are

1. Dewey Decimal Classification;
2. Library of Congress (LC);
3. Universal Decimal Classification (UDC).

Other schemes used are local subject code allotment and even author order, but only in some cases. Subject retrieval by that alone is virtually impossible in this last case. Electronic catalogs have made it easier to find material, for they can search in words in the title, extended keywords, and abstract where used.

Each of the established systems has its prime-listing directory – such as the Dewey Decimal Classification and Relative Index.

The UDC system is fundamentally supported by an extensive set of British Standards Institution standards documents, which enable great detail in subject identification. For instance, one code number is specifically for the cards used to print weight in automatic weighing machines! In practice, some libraries using the UDC system make use of a much abridged version.

UDC and Dewey are somewhat alike, but not identical. The LC system is quite different in code allocations. Equivalence between LC and Dewey codes is published.

The diverse and ubiquitous nature of measurement often makes classification of such material a difficult decision. It is publishers who assign a class call code to a book, that being part of their publishing support – but those that assign them are seldom experts on measurement literature. Most acquisition librarians will make use of this call number, for they also will not be expert enough to challenge the one allocated.

It is, therefore, quite possible for a book to be placed in one of many alternatives – it is not unheard of for a second edition of a book to be coded differently than the first!

6 LIST OF DEWEY CLASSIFICATION CODES WHERE MEASUREMENT MATERIAL IS LOCATED

The lack of structure of the knowledge of measurement science is reflected in library holdings: material will be found diffused throughout many classes. An impression of the situation is obtained by the study of the selected list of subjects and code numbers given in Figure 5 that is taken from the Dewey Classification. Overall, the system contains over 30 000 named code numbers of which at least some 700 are measurement related.

For most effective searching, the full classification would need consulting – a matter of many days of work! Library catalogs will have an alphabetical subject index in their entries but this will seldom provide the detail sought. The wide choice of measurement topics requires searching on the basis of a detailed study of the handbook of the classification system – but even then a book might not be in the class to which a user thinks it should be allocated.

It is appropriate here to explain the development of the Dewey Decimal System, for this has a strong bearing on current day problems with its use in class engineering and technology titles.

The Dewey system was devised in the late nineteenth century when technical publishing was in its infancy. Its structure is a reminder of who did the intellectual thinking at that time. Its construction did not cater to the subsequent explosion of technical knowledge. For this reason, subjects that often began as quite a reasonable and small subset of knowledge have had to be allocated a relatively extended code length, thus going deep into the numbering system.

For that reason, already long code numbers had to be greatly extended to cope with what are now major subsets of knowledge. For example, the 620.0–629.0 series was devoted to engineering and allied operations. This required electrical measurements to have a long call number of 621.381043.

All this contrasts with the nineteenth-century awareness of time measurement that allocated the code 529.78 to that topic. It can be seen that length of the call number is not an indicator of depth of specialist topics.

Many aspects of modern measurement are not placed in the Dewey Decimal Classification where they could now be expected to be in a library. For example, telegraphy (384.1) is placed in the Social Science 300's group. Holography (774) appears in the Arts 700's group along with games theory at 795.01. Computing has all but taken over class 1.0, originally assigned to knowledge.

In some, but not all, groups, there is a specific code for relevant instrumentation for that application. An example is that of flow properties and their measurement (532.053). This is not at all standard practice – many principles or application areas of measurement do not possess such a specific measurement aspect call number.

Occasionally, to add yet more difficulty, a new volume of the classification index itself is published. This requires some changes to codes (deletion, placement elsewhere, new code numbers). This produces only minor changes in the location of titles in the established libraries as it generally affects only books cataloged after that event.

Dewey Decimal code	Classification volume entry
001.422	Statistical method (formerly 311.2)
001.51	Communication theory
001.53	Cybernetics
001.533	Self-organizing systems
001.6	Data processing
003	Systems
016	Bibliographies and catalogs of specific disciplines and subjects. Note: Other codes used such as —.0 of notation
119	Number and quantity
120	Knowledge, cause, purpose, man
121	Epistemology
152	Physiological and experimental psychology
169	Analogy
330	Economics
339.3	Measures of national income
364.12	Criminal investigation (Detection)
368.011	Rates and rate making (Insurance)
371.26	Educational tests and measurements
378.16	Educational measurement and student placement
384.1	Telegraphy
384.9	Visual signaling
389	Metrology and standardization
501	Philosophy and theory
502.8	Techniques, apparatus, equipment materials
511.43	Error theory
512.7	Number theory
515.42	Theory of measure and integration
519.52	Theory of sampling
519.86	Quality control and other statistical adjustments
522.2–522.5	Instruments (astronomy)
526.3	Geodetic surveying
529.75	Time systems and standards
530.16	Measurement theory
530.7	Instrumentation (for measurement, control, recording)
530.8	Physical units, dimensions, constants
531.382	Elastic constants and their measurement
532.053	Flow properties and their measurement
534.4	Measurements, analysis and synthesis of sound
535.33	Optical instruments (see also specific functions class) Microscopes formerly 578.1
536.5	Temperature (536.51–536.54 Measurement)
539.77	Detection and measurement of particles and radioactivity
542.3	Measuring apparatus (laboratory-chemistry)
543.07	Instrumentation (chemistry)
544	Qualitative analysis
545	Quantitative analysis
548.1	Geometrical crystallography
551.4	Geomorphology (.4607 is Deep sea surveys and exploration)
573.6	Anthropometry
574.028	(no key title given but biological instrumentation here see 610.28)
578.4	Use of microscopes
591.18	Movements senses, control mechanisms (Zoology)
610.28	Medical instrumentation (see also 574.028)
611.8	Nervous system and sense organs
612.8	Neurophysiology and sensory physiology (extensive sub grouping)
614.7	Environmental sanitation and comfort (614.71–614.77 cover pollution but no specific class for measurements)
615.19018	Assay methods (pharmacy)
616.07	Pathology (includes diagnoses and prognoses)
617.05	Surgery by instrument and technique
617.752	Optical work

Figure 5. Sample list from the likely Dewey Decimal codes that relate to measurement.

Dewey Decimal code	Classification volume entry
617.89	Audiology
617.9178	Surgical instruments, apparatus, material
620.0042	Engineering analysis, synthesis, design
620.0044	Testing and measurement (engineering)
620.1127	Non-destructive testing of materials (subgroups on radiographic, tracer, ultrasonic, magnetic methods)
620.32	Measurements (of mechanical vibration)
620.72	Systems analysis, synthesis, design
621.313	Generating machinery and converters (includes maintenance and testing)
621.37	Electrical testing and measurement
621.372	Units and standards of measurement
621.373	Recording meters
621.374	Measurement of electric quantities (various subgroups exist)
621.379	Measurement of non-electrical quantities
621.381043	Measurements (electronic)
621.38137	Testing measurements standardization (Classifier instructed to place measurements in group for application)
621.381548	Testing and measuring devices and their use (Electronic)
621.383	Specific instruments and apparatus of wire telegraphy
621.38417	Measurements and standardization (radio)
621.3843	Specific instruments and apparatus of radio telegraphy
621.3887	Measurements and standardization (television)
621.3895	Underwater devices (subgroups include hydrophics and Sonar)
621.4831	Reactor physics (includes testing of physical phenomena occurring within reactors)
621.4835	Operation, control, safety measures
621.756	Inspection technology (in factory operations engineering)
621.902	Machine tools (including numerical control and other automation techniques)
621.994	Measuring tools
622.15	Geophysical exploration (several subgroups)
622.8	Mine health and safety
623.71	Intelligence and reconnaissance topography (military engineering)
623.819	Design tests (nautical)
623.863	Nautical instruments
624.1513	Soil mechanics (includes tests)
624.17720–624.1779	Specific structural elements (includes strength tests)
625.794	Traffic control equipment
627.81	General principles (Dams and reservoirs) (includes surveying)
628.177	Measurement of consumption (sanitary engineering)
629.045	Navigation (celestial in 527)
629.1345–629.1346	Aircraft tests, measurements, standards, maintenance, repair (see subgroups)
629.135	Aircraft instrumentation and systems (many subgroups)
629.273	Panel instrumentation (vehicles)
629.4775	Control of temperature, humidity, air supply and pressure
634.9285	Mensuration (forestry)
637.127	Quality and purity determinations (cows milk)
640.73	Consumer education (includes guides to quality and value of products and services)
657.48	Analytical (Financial accounting) (Measurement of profitability, of financial strength, of income, of liquidity, of flow of funds)
658.28	Equipment for safety and comfort (includes noise control)
658.516	Standardization (equipment, procedures, in general management)
658.562	Standards and specifications (form, size, dimensions, quality, materials, performance, including standardization formerly in 658.16)
658.568	Inspection (includes statistical methods of quality control)
658.834	Consumer research (studies of consumer preferences, attitudes, motivations, behaviour)

Figure 5. (*Continued*).

Dewey Decimal code	Classification volume entry
660.283	Process equipment (chemical instruments, apparatus, machinery)
662.622	Properties, tests, analysis (of coal)
664.07	Tests, analysis, quality controls (food—for texture, taste, odor, colour, contaminants)
665.0288	Tests analyses, quality controls (oils, fats waxes, greases)
666.13	Tests, analyses, quality controls (ceramics)
669.95	Metallography
674.13	Properties and their tests (lumber)
675.29	Properties, tests, quality controls (leather, fur)
676.121	Properties, tests, quality controls (Pulp)
676.27	Properties, tests, quality controls (paper and paper products)
681	Precision instruments and other devices (several groups)
697.9315	Psychrometrics
697.9322	Temperature controls
771.37	Focusing and exposure apparatus
774	Holography
781.91	Musical instruments
788.971	Mechanical and electrical reproducers
913.0310285	Interpretation of remains (includes dating, use of data processing in ancient studies)

Figure 5. (*Continued*).

Australian Journal of Instrumentation and Control
British Communications and Electronics
Control
Control Engineering
Electronic Engineering
Engineers Digest
Flow Measurement and Instrumentation
IEE Proceedings – Science, Measurement and Technology
IEEE Control Systems Magazine
IEEE Instrumentation and Measurement Magazine
IEEE Sensors Journal
IEEE Transactions on Instrumentation and Measurement
IEEE Transactions on Ultrasonics, Ferroelectronics and Frequency Control
ISA Transactions (Instrumentation, Systems and Automation Society)
Journal of Applied Measurement
Journal of Optical Society of America A: Optics Image Science and Vision
Journal of Optical Society of America B: Optical Physics
Journal of Optics A: Pure and Applied Optics
Laser Focus World
Measurement
Measurement and Control
Measurement Science and Technology (formerly Journal of Physics E: Scientific Instruments)
Metrologia
Nature
NCSL Newsletter
New Scientist
Physics Today
Review of Scientific Instruments
Science
Sensors and Actuators A: Physical
Sensors and Actuators B: Chemical
Sensor Review
Strain

Figure 6. Some journals relevant to measuring systems.

7 LIST OF JOURNALS RELEVANT TO MEASUREMENT

Figure 6 is a selected list of journals of major relevance to measurement. These cover application areas such as medicine, agriculture, space, and manufacturing, and will also invariably contain useful material on measuring systems for their domain.

RELATED ARTICLES

Article 1, Sophistication of Measurement and its Body of Knowledge, Volume 1; **Article 2, Organization of Instrument Science, Volume 1**; **Article 5, Humans in the Real World, Volume 1**; **Article 42, Information Databases of Relevance to Measurement, Volume 1**.

42: Information Databases of Relevance to Measurement

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Commonly Used Information Sources for Measurement 246
2 INSPEC Abstracting Service 248
3 NCSLI Services 248
4 National Laboratories 248
5 NIST Services 248
6 IMEKO Book Survey 248
7 Personal Lists 249
Related Articles 249
Reference 249

1 COMMONLY USED INFORMATION SOURCES FOR MEASUREMENT

There exist some, but not many, Internet-based services that provide valuable information on measuring systems as their main line of support. Major organizations that offer lists, guides, and relevant databases, INSPEC, National Conference of Standards Laboratories International (NCSL) and National Institute for Science and Technology (NIST), are covered in later sections.

Global Quality Services www.GlobalQualityServices.com offers a search engine in the field of metrology. It indexes a collection of Web sites in the field of metrology.

IMEKO (*International Measurement Confederation*) http://www.imeko.org is a nongovernmental federation of 34 national member organizations individually concerned with the advancement of measurement technology. Its objectives are the promotion of international interchange of scientific and technical information in the field of measurement and instrumentation and the enhancement of international cooperation among scientists and engineers from research and industry. It was founded in 1958. It offers information on the many meetings of its Technical Committees but does not provide detailed database information about measurement.

NI (*National Instruments*), http://www.ni.com provides a range of support services. One information service offered is a database of some 2000 instrument solutions; another is its Instrupedia, an Encyclopedia of Instrumentation provided on CD-ROM. The glossary part is provided online.

Ei COMPENDEX http://edina.ac.uk/compendex is a comprehensive interdisciplinary engineering information database having over six million summaries of journal articles and conference proceedings and 220 000 new additions every year.

It uses thesaurus-style indexing. Citations are arranged by subject and numbered sequentially; abstracts are provided. Each entry is indexed by article title, author name, and affiliation, journal or proceeding source, date of publication, and page length. Access to this database is via personal or shared accounts. It covers mechanical, civil, environmental, electrical, structural, process, materials science, solid-state physics and superconductivity, bioengineering, energy, chemical, optics, air and water pollution, solid-waste management, hazardous waste, road transportation, transportation safety, manufacturing, quality control, and engineering management issues. Thus, while it does not have a category for measurement, considerable content will be found on that topic by careful searching.

GIDEP (*Government-Industry Data Exchange Program*) http://www.gidep.corona.navy.mil is a cooperative activity between government and industry participants seeking

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

to reduce expenditures of resources by sharing technical information essential during the research, design, development, production, and operational phases of the life cycle of systems, facilities, and equipment. Its content contains metrology data, such as material constants. Utilization of GIDEP data can materially improve the total quality and reliability of systems and components during the acquisition and logistics phases of the life cycle, and reduce costs in the development and manufacture of complex systems and equipment. Again, it is not set up to support measurement as such but can be useful.

ISI® (*Science Citation Index*) http://www.isinet.com/isi/products/citation/sci provides access to current and retrospective bibliographic information, author abstracts, and cited references found in 3700 of the world's leading scholarly science and technical journals covering more than a hundred disciplines. The Science Citation Index Expanded™ format and the online version, SciSearch®, cover more than 5800 journals.

BIOSIS http://www.biosis.org has the objective of facilitating understanding of the living world by helping researchers, educators, students, and others in accessing information relevant to the life sciences. It is an entire life sciences file corresponding to Biological Abstracts (BA) and Bio Research Index (BioI). It includes, from 1969, biomedical material and is sourced from over 9000 serials.

Biology Browser www.biologybrowser.org is an interactive portal designed for the life sciences community. It creates a dynamic environment that connects users with valuable information through interactive discussion forums, quality-controlled Web links, science news, and other free resources.

CHEMCON http://www.chemcon.net is a global network of chemical conferences in the field of chemical control regulations.

CDI http://www.virtualref.com Comprehensive Dissertation Index (not to be confused with another CDI, the Center for Defense Information) contains all dissertations at academic doctoral level granted in the United States and in over 125 other-country institutions. It covers from 1861 and grows at over 500 citations a month. Since 1976, coverage has been expanded to include European institutions, and later other institutions worldwide – but their coverage is rather patchy.

IFI CLAIMS http://www.ificlaims.com is a 45-year old, multifaceted, database holding data on US patents. It is offered as being a trusted and comprehensive source of US patent information.

CPI Conference Papers http://library.dialog.com has provided, since 1973, a Conference Papers Index (CPI) that contains over 800 000 entries of papers given at over 1500 major conferences. It provides citations to sessions presented at major scientific meetings around the world. Subject emphasis since 1995 has been in the life sciences, environmental sciences, and the aquatic sciences, while older material also covers physics, engineering, and materials science. Information is derived from final programs, abstracts booklets, and published proceedings, as well as from questionnaire responses.

ENVIROLINE http://library.dialog.com covers the world's environment related information. It provides indexing and abstracting coverage of more than 1000 international primary and secondary publications reporting on all aspects of the environment. These publications highlight such fields as management, technology, planning, law, political science, economics, geology, biology, and chemistry as they relate to environmental issues. It corresponds to the print material *Environment Abstracts*.

GEOARCHIVE http://library.dialog.com is a comprehensive database covering all types of information sources in geosciences, hydrosciences, and the environmental sciences. The criteria for inclusion in GeoArchive are that the source should be publicly available and have relevant information content, even if the reference is to a small news item in a magazine. It provides, from 1969, international coverage of over 5000 serials, books from over 2000 publishers, geological maps, and doctoral dissertations. It is indexed using *Geosaurus*, the hierarchical thesaurus that classifies references by subject, geography, and stratigraphy. It includes an alphabetical index of around 30 000 terms available from the producer. The printed publications *Geotitles, Hydrotitles, Geoscience Documentation*, and the *Bibliography of Economic Geology* are derived from GeoArchive.

ISMEC Mechanical Engineering Abstracts http://www.cas.org/ONLINE/DBSS/ismecss.html is a bibliographic database (from 1973) containing in-depth coverage of worldwide literature on mechanical engineering, engineering management, and production engineering. Theoretical perspectives and applications for current research results are included. Records contain bibliographic information, controlled and uncontrolled terms, and abstracts. A specific topic area is Measurement and Control.

NTIS (*National Technical Information Service*) http://www.ntis.gov is the central source for US government scientific, technical, and business information. It leads to titles on research and development activities published by over 200 US Federal Agencies. Audiovisual materials, computer data-files, and software are included.

WPI (*World Patent Index*) http://www.derwent.com files (partially, since 1963) patent specifications issued by the major industrial nations. It corresponds to CPI and WPI and grows at over 250 000 new inventions per year.

These few examples illustrate the various kinds of services that are available. Many may need to be searched

to locate measurement information; there seems to be no easier method than to patiently work through them.

2 INSPEC ABSTRACTING SERVICE

INSPEC http://www.iee.org/Publish/INSPEC is produced by the Institution of Electrical Engineers. It is a major English-language bibliographic information service providing access to the world's scientific and technical literature in physics, electrical engineering, electronics, communications, control engineering, computers and computing, and information technology. The INSPEC database was formed in 1969, being based on the *Science Abstracts* service provided by the Institution of Electrical Engineers since 1898. It provides Current Awareness, New Product Information, Technological Forecasting, Competitive Intelligence, and Patent-Related Searching. Over 3400 scientific and technical journals and some 2000 conference proceedings, as well as numerous books, reports, and dissertations are scanned each year for relevant articles to abstract and index for inclusion in the Database. It contains over 7 million bibliographic records and is growing at the rate of 350 000 records each year. It also publishes an INSPEC Thesaurus of terms that contains over 10 000 terms, of which at least 800 are relevant to a search for an instrument-related subject. As they are diffused throughout the whole set, considerable time is needed to properly choose suitable key terms. For an explanation of the difficulties of finding measuring systems literature *see* **Article 41, Terminology and Classification of Measurement Systems, Volume 1**.

3 NCSLI SERVICES

NCSLI (National Conference of Standards Laboratories, International) http://www.ncsli.org was established in 1961 as a professional association for individuals engaged in all spheres of international measurement science. It provides real-time professional career support and advancement opportunities and sponsors an annual technical Annual Workshop & Symposium. Topics covered by the site include implementing national and international standards, achieving laboratory accreditation, new measurement technology, advances in measurement disciplines, laboratory management procedures and skills, equipment management, workforce training, and new instrumentation. In particular http://www.ncsli.org/resources/index.cfm is an internal location for some useful resources. These include an Acronym and Abbreviation Glossary, Procedures Database, Training Information Directory and NCSLI Archives.

4 NATIONAL LABORATORIES

Technological countries have at least one of these carrying the responsibility for developing and maintaining the prime physical standards. They additionally offer staff expertise through consultancies and by participation in industry clubs.

The National Physical Laboratory (NPL), UK, http://www.npl.co.uk is an example. Their bibliography of staff publications can be useful at times but this site does not appear to offer lists and guides, as does the next of the NIST.

5 NIST SERVICES

The National Institute for Science and Technology, NIST, US – formerly the National Bureau of Standards (NBS) – http://www.nist.gov offers a vast array of pull-up information services. One of regular value is a guide to the use of SI units.

6 IMEKO BOOK SURVEY

The practical problems of establishing the existence of book material on measurement science and technology were recognized in 1978 by the Higher Education Committee (TC-1) of the International Measurement Confederation (IMEKO).

It was decided that the Committee should establish a list of titles for free distribution to member nations of IMEKO, and any other interested user. The bibliography was completed in 1980 (Sydenham, 1980).

Initially it was thought that the list would be quite short and that key articles in the serial literature could also be included. Entries were to be for works in the four official IMEKO languages – English, French, German, and Russian.

It was soon established that computer searching, at that time, was not productive for book location. The best approach was found to be to select entries from a decade collection of publishers' catalogs and promotional mailings, from committee members' personal book collections, and from visits to the catalogs of libraries that were closely related to measurement interests.

It was found that individual contributor's book lists rarely overlapped each other, yet each person was convinced theirs was in touch with the bulk of the related book literature. The final list includes over 800 titles up to that date. Books related to, but not directly concerned with, measurement and instrumentation (such as control theory, systems design) were not included. It is, unfortunately not yet available in electronic form.

Sections of the listing (by author only), cover

1. Fundamental concepts (theory of measurement for instance);
2. Units, standards, calibration;
3. Measurement uncertainty, results processing, errors;
4. Instrumentation practice (in 10 groups);
5. Instrument system design, construction, and evaluation;
6. Education and training;
7. Miscellaneous (mainly historical).

An entry is included only once in the listing.

The problems of deciding what are significant papers in the journal literature was addressed and found to be too problematic to produce. The truly vast quantity of papers available was realized when the first 18 instrument-related terms (of a possible 650 from the INSPEC thesaurus!) were combined with the term 'review' and tested on a short file length of INSPEC. It produced over 800 abstracts, most of which appeared to be valid entries for a bibliography. Clearly, the user must go direct to the databases to generate citations, sorting them online to yield those of individual relevance.

7 PERSONAL LISTS

Browsing, current awareness services, and other sources such as reprint requests and conference attendance, provide the individual with a steady influx of quite specific information. When the number of entries reaches a few hundred, it becomes necessary to use a computer database to record them. For this to be useful, it in turn requires ordered storage.

Traditional personal methods involving card indexes have given way to the use of personal computers, especially to the application of relational databases. Setting up citation references for general use is not easy for some 100 attributes are needed to define the paper and book reference in all of the ways required. Publishers ask for so many variations that it is not easy to accommodate them in a simple database. ENDNOTES is a database set up for processing bibliographies. There is no doubt that making satisfactory records of citation lists needs diligence and care to detail as the sources are found.

The tediousness of rekeying reference items into the required form could well do with increased use of standard formats for bibliographic entries.

Users with personal lists are encouraged to mount them in an enduring Web site and maybe then advise Measure-Mentor of that location for inclusion in its link.

RELATED ARTICLES

Article 1, Sophistication of Measurement and its Body of Knowledge, Volume 1; **Article 2, Organization of Instrument Science, Volume 1**; **Article 5, Humans in the Real World, Volume 1**; **Article 41, Terminology and Classification of Measurement Systems, Volume 1**.

REFERENCE

Sydenham, P.H. (ed.) (1980) A Working List of Books Published on Measurement Science and Technology in the Physical Sciences, *International Measurement Confederation IMEKO*, Applied Physics Dept, Technische Hogeschool Delft, Delft.

43: Units

Brian W. Petley
National Physical Laboratory, Teddington, UK

1 Units 255
2 The Definitions of the SI Base Units 258
3 Traceability 259
4 Other Unit Systems 259
Further Reading 260

1 UNITS

All measurement is made in terms of comparison with a like quantity. If we live in isolation, the latter quantity may be anything that we choose. However, we might lose the latter and if we wish to communicate the result with ourselves or with other people possibly in another place at a later date, then it is obvious that it is useful, even essential, that we all use the same reference quantity or can relate our more transient reference quantity to it. Thus, one might well use one's foot or a convenient stick when spacing plants in the garden, but would convert the spacing to meter if one were writing a book on planting.

The general arrangements by which this agreement is obtained are illustrated in Figure 1. The SI or Systéme International des Unités (Table 1) has been set up to achieve this universality and all dimensioned measurements may be related to the units of this system.

The English language has some 80 000 words, but with measurement we try to have the minimum possible 'vocabulary'. Initially, it was intended that all SI measurements would be expressed in terms of six basic units but it was decided that a seventh should be added in order to take account of the needs of chemistry and related fields (Table 2).

Many of our measurements are expressed in terms of jawbreaking combinations of these basic units. It soon became apparent that it was impracticable to have a 'measurement vocabulary' of only seven base units and the derived units were also agreed, the number of these being kept to a minimum (Table 3).

The current International Vocabulary on Metrology (VIM) defines a unit as 'a scalar quantity defined and adopted by convention, with which other quantities of the same kind are compared in order to express their magnitudes relative to that quantity'.

The following notes are added:

1. Units are designated by conventionally assigned names and symbols.
2. Units of quantities of the same dimension may be designated by the same name and symbol even when the quantities are not of the same kind. For example, the joule per kelvin, J/K, is the name and symbol of both a unit of heat capacity and a unit of entropy.
3. Units of dimensionless quantities are numbers. In some cases, these numbers are given special names, for example, radian and steradian.
4. If there is a risk of confusion, the term 'unit of measurement' should be used.

Units that are named after people take the lowercase for the name of the SI unit and the uppercase for the symbol; thus the ampere, symbol A, is the SI base unit of electric current.

1.1 SI conventions

The value of a quantity Q is given by the product of the numerical value of Q, $\{Q\}$ and the unit or units of Q, $[Q]$, thus:

$$Q = \{Q\} \times [Q] \tag{1}$$

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

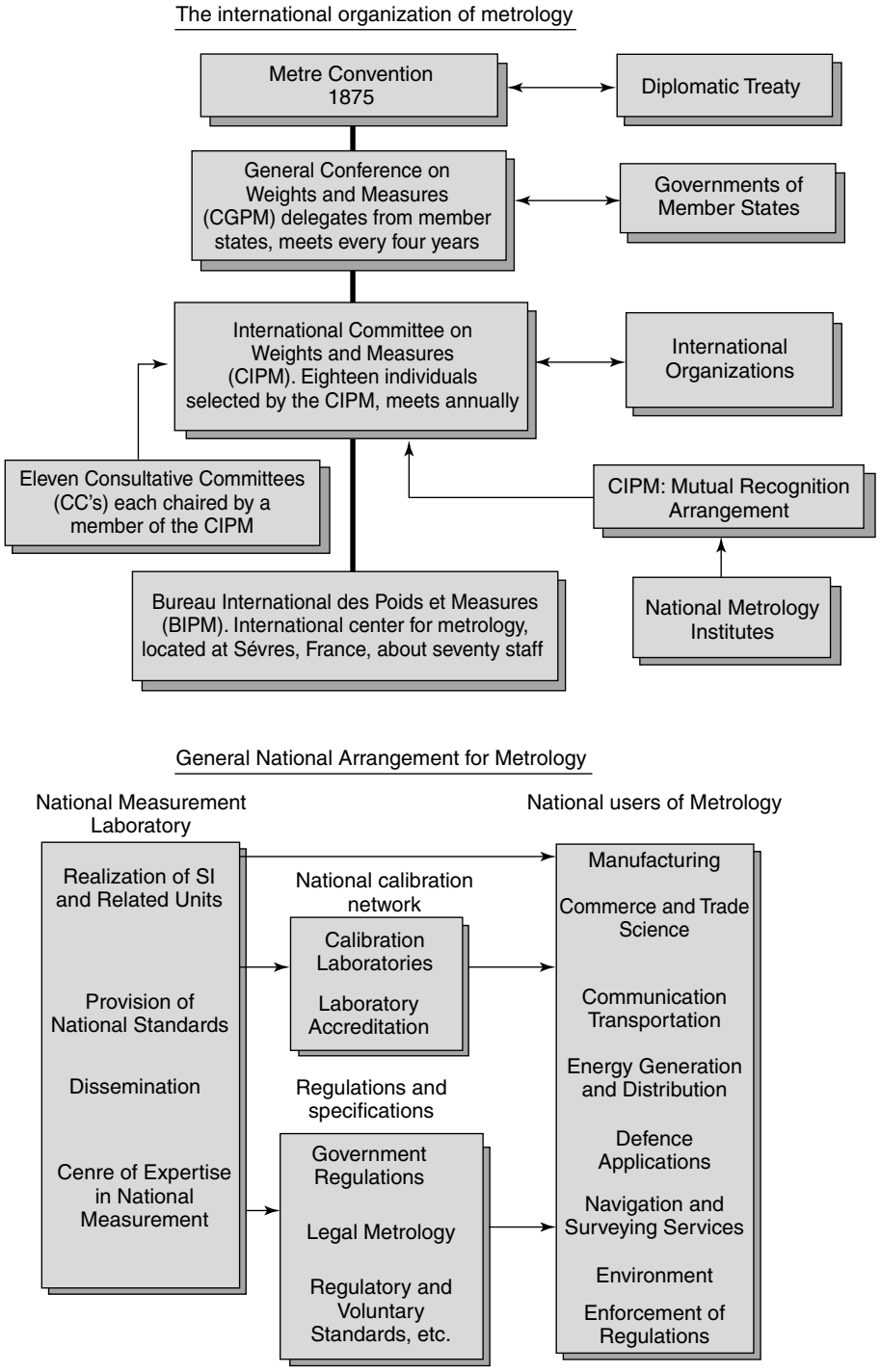


Figure 1. The international organization of metrology.

The convention is that italic font is used for the symbols for quantities, and upright Roman font for units.

1.2 Coherence

The agreed derived units had to satisfy the criterion that whatever correctly dimensioned combination of base and derived values was used, the numerical value remained the same. This excluded popular units such as the liter, degree, minute, hour, day, and so on. However, certain units such as these were in such widespread use that it was agreed that certain noncoherent units could be used alongside the SI units.

Table 1. The composition of the International System of Units (SI) and prefixes: 2002.

Type of unit	Number
SI base units	7
SI named derived units	21
Named SI prefixes	20
Experimentally determined[a]	3
Units outside the SI (day, minute, etc.) (off system units)[a]	10

[a]Units approved for use alongside the SI.

Table 2. The SI base units.

Base quantity	SI base unit name	Symbol
length	metre (meter in USA)	m
mass	kilogram	kg
time	second	s
electric current	ampere	A
thermodynamic temperature	kelvin	K
amount of substance	mole	mol
luminous substance	candela	cd

1.3 The SI prefixes

The 11th General Conference on Weights and Measures (CGPM) authorized certain prefixes to be used in association with SI units. The initial list has since been extended to that given in Table 4. Any of these may be attached to any SI unit but they must not be cascaded. For example, 10^{-6} kg = 1 mg (1 milligram), but not 1 μkg.

1.4 Non-SI units (i) accepted for use, (ii) values determined experimentally, and (iii) used currently (temporarily) with the International System

The International Committee on Weights and Measures (CIPM) (1969) recognized that users would wish to employ the SI together with units that are not part of it but are important and widely used, listed three categories of non-SI units: units to be maintained; to be tolerated temporarily; and to be avoided. In 1996 the CIPM agreed on a new classification of non-SI units: units accepted for use with the SI, Table 5; and units accepted for use with the SI whose values are obtained experimentally, Table 6.

Table 5 lists non-SI units that are accepted for use with the SI. It includes units that are in continuous everyday use, in particular, the traditional units of time and of angle, together with a few other units, which have assumed increasing technical importance.

It has proved useful to have some experimentally derived units in fairly common use as well, and these are shown

Table 3. The SI derived units with special names.

Quantity	Name	Symbol	Expression in terms of:	
			Other units	SI base units
angle	radian	r		$m \times m^{-1}$
solid angle	steradian	sr		$m^2 \times m^{-2}$
frequency	hertz	Hz		s^{-1}
force	newton	N		$m \times kg \times s^{-2}$
pressure, stress	pascal	P	$N\ m^{-2}$	$m^{-1} \times kg \times s^{-2}$
energy, work, quantity of heat	joule	J		$m^2 \times kg \times s^{-2}$
power, radiant flux	watt	W		$m^2 \times kg \times s^{-3}$
electric charge, quantity of electricity	coulomb	C		$s \times A$
electric potential, potential difference, electromotive force	volt	V	W/A	$m^2 \times kg \times s^{-3} \times A^{-1}$
capacitance	farad	F	C/V	$m^{-2} \times kg^{-1} \times s^4 \times A^2$
electric resistance	ohm	Ω	V/A	$m^2 \times kg \times s^{-3} \times A^{-2}$
electric conductance	siemens	S	$A \times V$	$m^{-2} \times kg^{-1} \times s^3 \times A^2$
magnetic flux	weber	Wb	$V \times s$	$m^2 \times kg \times s^{-2} \times A^{-1}$
magnetic flux density	tesla	T	Wb/m^2	$kg \times s^{-2} \times A^{-1}$
inductance	henry	H	Wb/A	$m^2 \times kg \times s^{-2} \times A^{-2}$
Celsius temperature	degree Celsius	°C		K
luminous flux	lumen	lm	$cd \times sr$	$cd \times m^2 \times m^{-2} = cd$
illuminance	lux	lx	$m^{-2} \times cd \times r$	$m^{-2} \times cd$
activity (referred to a radionuclide)	becquerel	Bq		s^{-1}
absorbed dose, specific energy (imparted), karma dose equivalent, ambient dose equivalent, directional dose equivalent, personal dose equivalent, organ equivalent dose	gray	Gy		$m^2 s^{-1}$
catalytic activity	katal	kat		$s^{-1} \times mol$

Table 4. The range of SI prefixes.

Factor	Name	Symbol	Factor	Name	Symbol
10^{24}	yotta	Y	10^{-1}	deci	d
10^{21}	zetta	Z	10^{-2}	cento	c
10^{18}	exa	E	10^{-3}	milli	m
10^{15}	peta	P	10^{-6}	micro	μ
10^{12}	tera	T	10^{-9}	nano	n
10^{9}	giga	G	10^{-12}	pico	p
10^{6}	mega	M	10^{-15}	femto	f
10^{3}	kilo	k	10^{-18}	atto	a
10^{2}	hecto	h	10^{-21}	zepto	z
10^{1}	deca	da	10^{-24}	yocto	y

Table 5. Non-SI units accepted for use with the International System.

Name	Symbol	Value in SI units
minute	min	$1\,\text{min} = 60\,\text{s}$
hour	h	$1\,\text{h} = 60\,\text{min} = 86\,400\,\text{s}$
day	d	$1\,\text{d} = 24\,\text{h} = 86\,400\,\text{s}$
degree	°	$1^\circ = (\pi/60)\,\text{rad}$
minute	′	$1' = (1/60)^\circ\pi$
second	″	$1'' = (1/60)' = (\pi/6400)\,\text{rad}$
liter	l, L	$1\,\text{L} = 1\,\text{dm}^3 = 10^{-3}\,\text{m}^3$
metric ton	t	$1\,\text{t} = 10^3\,\text{kg}$
neper	Np	$1\,\text{Np} = 1$
bel	B	$1\,\text{B} = (1/2)\ln 10\,\text{Np}$

Note: Other units which are currently approved for use alongside the SI in order to satisfy the needs of commercial, legal and specialized scientific use include: nautical mile (knot), are (a), hectare (ha), bar (bar), angstrom (Å), and barn (b).

Table 6. Non-SI units accepted for use with the International System, whose values are obtained experimentally.

Name	Symbol	Value in SI units
electronvolt	eV	$1\,\text{eV} = 1.60217653(14) \times 10^{-19}\,\text{J}$
unified atomic mass unit	u	$1\,\text{u} = 1.66053886(28) \times 10^{-27}\,\text{kg}$
astronomical unit	ua	$1\,\text{ua} = 1.49597870691(30) \times 10^{11}\,\text{m}$

Note: The latest CODATA values may be obtained from the National Institute for Science and Technology (NIST) and International Bureau of Weights and Measures (BIPM) Websites: http://physlab.nist.gov/cuu/Constants/Citations/Search.html, http://www.ibipm..fr/enus/, or http://physics.nist.gov/cuu/Constants/Table/allascii.txt

in Table 6. The electron volt is probably the most frequently encountered of these via its extensive use in measuring X-ray energies. The unified mass unit is particularly important in chemistry and serves as the reference quantity for atomic weights and atomic masses. The astronomical unit serves as the basic length reference unit in astronomy, particularly within the solar system, and larger scale units such as the parsec are subsequently related to it.

2 THE DEFINITIONS OF THE SI BASE UNITS

The definitions of the SI units are carefully framed to reflect the increasingly stringent requirements of modern science and technology. Although the names and symbols for the SI units are intended to be constant, the definitions are modified in the light of progress in order to increase the stability and invariance of the SI. Each time that the definition is changed care is taken to see that the new unit lies within the reproducibility of the old. The definitions of the SI units were originally derived from concepts related to properties of the earth and these were soon transferred to artifact units such as the prototype meter and kilogram. One can readily imagine that the reproducibility of the meter was ultimately limited by the width of the engraved markings as well as by the stability of the platinum–iridium bar.

2.1 Unit of length (meter, symbol m)

The meter is the length of the path traveled by light in a time interval of 1/299 792 458 of a second.

2.2 Unit of mass (kilogram, symbol kg)

The kilogram is the unit of mass. It is equal to the mass of the international prototype of the kilogram.

2.3 Unit of time (second, symbol s)

The second is the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium 133 atom.

2.4 Unit of electric current (ampere, symbol A)

The ampere is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible cross section, and placed 1 m apart in vacuum, would produce between these conductors a force equal to 2×10^{-7} newton per metre of length.

2.5 Unit of thermodynamic temperature (kelvin, symbol K)

The Kelvin, unit of thermodynamic temperature, is the fraction 1/273.16 of the thermodynamic temperature of the triple point of water.

Because of the way that thermodynamic scales used to be defined, it remains common practice to express a thermodynamic temperature T, in terms of the difference from the reference temperature $T_o = 273.15$, the ice point. This temperature is called the Celsius temperature, symbol t, and is defined by the equation

$$t = T - T_o \tag{2}$$

The numerical value of a Celsius temperature, expressed in degrees Celsius, is given by

$$\frac{t}{^\circ\mathrm{C}} = \frac{T}{\mathrm{K}} - 273.15 \tag{3}$$

2.6 Unit of amount of substance (mole, symbol mol)

The definition of the mole comprises two parts:

1. The mole is the amount of substance of a system, which contains as many elementary entities as there are atoms in 0.012 kg of carbon 12; its symbol is the 'mol'.
2. When the mole is used, the elementary entities must be specified and they may be atoms, ions, electrons, other particles, or specified groups of such particles.

In this definition, it is understood that unbound atoms of carbon, at rest and in their ground state are referred to.

2.7 Unit of luminous intensity (candela, symbol cd)

The candela is the luminous intensity, in a given direction, of a source that has emitted monochromatic radiation of frequency 540 1012 Hz and has a radiant intensity in that direction of 1/683 watt per steradian.

3 TRACEABILITY

Traceability to the SI units involved is the desired characteristic of any measurement. Thus, if one measures a length with a cheap plastic ruler, one often finds that the lengths so measured will differ from those made with a steel ruler because the markings on the plastic ruler are not quite at the correct separations. One should always assume, as a matter of routine, that all other measuring equipment, analog or digital, electronic or mechanical, and so on, may similarly be in error at some level of accuracy (including nonlinearities). Thus, one's local measurement units will almost certainly differ slightly from the correct SI values. Of course, even the national and international representations of the units will be slightly in error.

It follows from this that all measurements must be made with devices that have been calibrated in some way against nationally or internationally maintained SI standard representations of the SI units, or other agreed and known units to the appropriate degree of accuracy. This has given rise to the concept of traceability. The current official definition of traceability is given in the International Vocabulary on Metrology as follows:

> The property of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons having a stated uncertainty.

4 OTHER UNIT SYSTEMS

There are other unit systems in common use. These have a fixed relationship with the SI, such as the cgs and fps systems, although the size and spelling between the British and the US versions of the derived units in the latter (for example, the US gallon and the Imperial gallon) may well differ.

There are other systems that are useful for scientific purposes whose relationship with the SI must be determined experimentally. Notable examples of these are the System of Astronomical units and the atomic system of units, the system of natural units, the Planck units, and so on. These systems are based on the fundamental physical constants. The accuracy of our knowledge of these has been rapidly

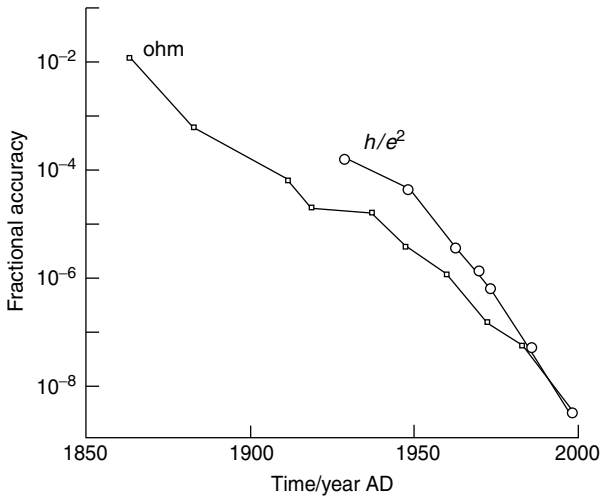


Figure 2. Illustrating how the accuracy of the fine structure constant and/or h/e^2 caught up with that of the realization of the ohm: the ohm may now be maintained in terms of an internationally agreed value for the Klitzing constant.

catching up with the accuracy with which the SI unit(s) may be realized. See, for example, Figure 2.

FURTHER READING

The definitive sources for the current SI, including the definitions of the SI units, are contained in *The International System of Units (SI)* (7th edn, BIPM, 1998): http://www.bipm.fr/enus/3_SI also published as NIST Special Publication 330 (NIST, 2000): http://physics.nist.gov/cuu/Units/index.html and *The International System of Units, Supplement 2000*, BIPM, Sèvres (2000): http://www.bipm.sfr/pdf/si-supplement2000.pdf.

Crovini, L. and Quinn, T.J. (eds) (1989) *Metrology at the Frontiers of Physics and Technology: Proceedings of the International School of Physics "Enrico Fermi," Course CX*, 27 June–7 July, Elsevier Science and Technology Books, Amsterdam.

Cutler, P.H. and Lucas, A.A. (eds) (1983) *Quantum Metrology and Fundamental Physical Constants*, *NATO ASI Series. B: Physics*, Plenum, New York.

Ferro, A., Giacomo, M.P. and Leschiutta, S. (eds) (1976) *Metrology and Fundamental Constants: Proceedings of the International School of Physics "Enrico Fermi, Course LXVIII*, Academic Press, Amsterdam.

Mills, I.M. (1997) The Language of Science. *Metrologia*, **34**(1), 101–109.

Mohr, P.J. and Taylor, B.N. (2000) *Reviews of Modern Physics* **72**, 351. A complete listing of the current CODATA recommended values for the fundamental constants is available at http://physlab.nist.gov/cuu/Constants/Citations/Search.html.

Petley, B.W. (ed.) (1986) *The Fundamental Constants and the Frontier of Measurement*, Adam Hilger, London.

Quinn, T.J., Leschiutta, S. and Tavella, P. (eds) (2000) Recent advances in metrology and fundamental constants, *Proceedings of the International School of Physics "Enrico Fermi" Course CXLVI*, 25 July–4 August, IOS Press, Amsterdam.

Roche, J.J. (1998) *The Mathematics of Measurement*, Athlone, London.

44: Types of Paper Standards and their Purpose

Halit Eren
Curtin University of Technology, Perth, Western Australia, Australia

1 Introduction 261
2 Published Standards 262
3 Standards Organizations Relevant to Measurement 263
4 Links Between Standards Organizations 263
5 Development of Published Standards 264
6 Obtaining Standards 265
7 Use of Standards in Support of Design and Operation 266
8 Example of Development of a Published Standard 267
9 Summary of Benefits of Published Standards 268
Further Reading 268

1 INTRODUCTION

The term standard has many definitions. The Oxford English dictionary defines *standard* as '**noun 1** a level of quality or attainment. **2** a required or agreed level of quality or attainment. **3** something used as a measure, norm, or model in comparative evaluations. **adjective 1** used or accepted as normal or average. **2** (of a size, measure, etc.) regularly used or produced'.

The measurement standards, on the other hand, are those devices, artifacts, procedures, instrument, systems, protocols, or processes that are used to define measurement units and on which all echelon measurements depend on the fundamental physical standards.

From the above definitions, the word 'standard' implies comparative evaluations, which can be physical quantities and/or processes and procedures. Consequently, there are several kinds of standards. In addition to the *fundamental physical measurement standards*, there are standards of *practice or protocols* that can be defined as documents describing the operations and processes to achieve unified results.

This article largely deals with the *practice and protocol* standards, which can loosely be termed as *published standards*.

Both physical and published measurement and metrology standards play a fundamental role in shaping the efficiency of domestic and global economies. Published standards are important as the documents that provide textual and illustrative information on what and how measurements should be done, and how they are done.

They are necessary for worldwide progress in trade liberalization, research and development activities, interpenetration of different technological sectors, and also for ensuring consistency in emerging technologies. Consequently, because of the national and international implications, there are many institutions that are responsible to investigate, develop, determine, and maintain the relevant standards to support worldwide scientific and industrial activities.

However, national and international authorities are subject to various internal and external forces, differences of expertise opinions, and commercial interests during the development of the standards that may not always end up with the ideal situations. This, unfortunately, may result in a confusing number of standards with different versions and interpretations for a specific artifact or measurement. In addition, publications grow over time as new technological developments take place, and the standards authorities and procedures change.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

In this article, an introductory explanation on the nature, scope, and generation of the published standards is given with some examples.

The process of developing the published standards is explained, and guidance is given on how to go about finding the relevant standards for a specific application. After having read this article, the reader should be able to appreciate the importance of the standards in the design and application of measuring systems and should have some informed idea on the selection of most applicable standards for his/her needs.

2 PUBLISHED STANDARDS

Published standards are documents describing the operations and processes that must be performed to achieve a particular task. The National Standards Policy Advisory Committee (http://www.nist.gov) describes a standard as 'a prescribed set of rules, conditions, or requirements concerning definitions of terms: classification of components; specification of materials, performance, or operations; delineation of procedures; or measurement of quantity and quality in describing materials, products, systems, services, or practices'.

It is important to realize that measurement standards are developed by various national and international bodies, sometimes operating totally independently; hence there may be different versions for the same subject. Examples of national and international standards bodies are the International Organization of Standardization (ISO), the International Electrotechnical Commission (IEC), the American National Standards Institute (ANSI), and so on. These organizations have multiple internal departments, committees, subcommittees, and working groups to support their activities.

Common to all organizations, in the initial stages of developing published standards, general terminology, vocabulary, and definitions concerning standardization and related activities are determined. This acts as an agent to develop a mutual understanding amongst the members, groups, and concerned parties at individual, governmental and nongovernmental levels. The aim is that these standards are prepared in such a manner that they can be understood, used, and referred to easily. To stimulate this, standards organizations have strict guidelines to assist technical groups and committees in drafting and producing the standards.

The typical preparation of a standard includes two main stages:

- *Preliminary work* that includes overall analysis, identification of target groups, determination of subject delimitation, finding the relevant sources, determining the affecting concepts, choosing the languages, working out the schedules, and so on.
- *Working procedure* that includes collecting and recording terminological data, establishing the term list, determining the concept fields and concept systems, formulating definitions, establishing and structuring the entries, order of entries, indexes, graphic representation, bibliography, and so on.

Given that standards on a specific topic exist, determining the relevant one for a specific application can be tackled by adopting a top-down approach. That is, the standards can be classified under general headings, such as the following

- generalities, terminology, standardization, documentation
- mathematics, natural sciences
- environmental, health protection, safety
- metrology and measurements
- mechanical systems and components
- health care and related technology
- electrical, electronics, telecommunications engineering
- road vehicles, railways, aircraft and space vehicle engineering
- civil engineering
- domestic and commercial equipment, entertainment, sports
- mining, metallurgy, petroleum and related technologies

and others.

Each general standard contains many discipline-based standards. For example, metrology and measurement standards may have the following

- metrology and measurement including measuring instruments, preferred numbers, standard measures, general aspects of reference materials, and so on
- measurement of volume, mass, density, viscosity
- measurement of time, velocity, acceleration, angular velocity
- acoustics and acoustic measurements
- optics and optical measurements
- thermodynamics and temperature measurements
- electricity, magnetism, electrical, and magnetic measurements
- radiation measurements

and others.

Discipline-based standards contain specific standards. For example, electrical, and magnetic measurements may comprise the following

- electrical engineering in general including voltages, general electrical terminology, electrical documentation, electrical tables, safety, fire-hazard testing
- electromagnetic compatibility

- insulating materials and fluids
- galvanic cells and batteries
- power transmission and distribution networks
- resistors, inductors, capacitors, semiconductor devices
- integrated circuits, microelectronics, logic, and analog microstructures
- radio-communications, mobile services, integrated services digital network (ISDN)
- fiber optic communications, telecontrol, telemetering
- television and radio broadcasting

and others.

These specific standards can further be divided into individual standards. Consider electromagnetic compatibility (EMC) as an example, which may have the following standards:

- electromagnetic compatibility in general
- emission
- immunity, and so on.

Application areas of specific standards may differ greatly. Continuing with the EMC, there are standards for the following

- wheelchairs and motorized scooters
- robots
- earth moving machinery
- space systems
- agricultural and forestry machinery
- lifts, escalators, conveyors, road vehicles, and so on.

More information on EMC measurement will be provided in the examples; see Section 8.

3 STANDARDS ORGANIZATIONS RELEVANT TO MEASUREMENT

Standards are formulated by standards development organizations (SDOs). These organizations can be government agencies, professional societies, international or regional organizations, private companies, or trade associations.

World standard organizations can be categorized in four major groups as national, international, consortium, and industry organizations, as illustrated in Figure 1.

Each organization has various branches that are responsible for a particular area. As an example, Figure 2 illustrates some important branches of the American National Standards Institute.

A list of some of the national and international organizations concerned with published measurement standards are given in Table 1.

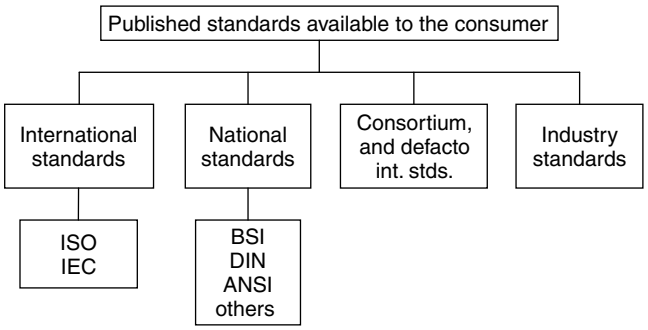


Figure 1. World standards organizations.

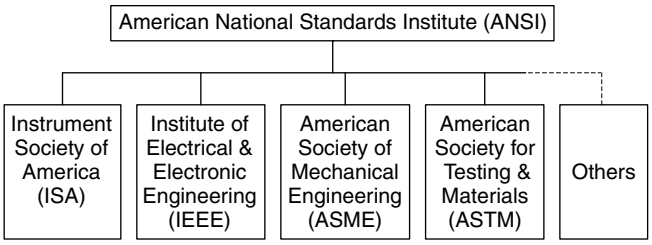


Figure 2. Branches of American National Standards Institute.

Each organization has several specialist groups that encompass a broad range of scientific disciplines, such as electromagnetic, temperature, chemical, mechanical, radiation, optic standards, and so on.

4 LINKS BETWEEN STANDARDS ORGANIZATIONS

Standards are important for scientific, research and development activities and for international trade; consequently, there are continuous links and cooperation between national, regional, and international organizations as listed in Table 1.

For example, in the case of ISO, there are many affiliated standards organizations.

Affiliated organizations largely represent countries as

1. *full members* that are entitled to participate and exercise full voting rights on any technical committee and policy committee of ISO;
2. *correspondent members* that do not take an active part in the technical and policy development work, but are entitled to be kept fully informed about their work of interest; and
3. *subscriber members* that represent the countries with small economies.

Table 1. List of some of the national and international standard organizations.

Acronym	Organization name and web address
ANSI	American National Standards Institute (http://www.ansi.org/)
AFNOR	Association française de normalization (http://www.afnor.fr/)
BSI	British Standards Institution (http://www.bsi-global.com/)
CEI	Comitato Elettrotecnico Italiano (http://www.ceiuni.it/)
CEN	Comité Européen de Normalisation (http://www.cenorm.be/)
CENELEC	Comité Européen de Normalisation Electrotechnique (http://www.cenelec.be/)
CSSN	China Standards Information Center (http://www.cssn.net.cn/)
DIN	Deutsches Institut für Normung (http://www2.din.de/)
IBN	The Belgian Institution for Standardization (http://www.ibn.be/)
IEC	International Electrotechnical Commission (http://www.iec.ch/)
ISO	International Organization for Standardization (http://www.iso.org/)
JISC	Japan Industrial Standards Committee (http://www.jisc.go.jp/)
KATS	Korean Agency for Technology and Standards (http://www.ats.go.kr/)
NIST	American National Standards and Technology (http://www.standards.com.au/)
ON	Austrian Standards Institute (http://www.on-norm.at/)
SAA	Standards Australia (http://www.standards.com.au/)
SCC	Standards Council of Canada (http://www.scc.ca/)
SII	The Standards Institution of Israel (http://www.sii.org.il/)
SNZ	Standards New Zealand (http://www.standards.co.nz/)
TSE	Türk Standardlari Enstitüsü (http://www.tse.org.tr/)

The membership fees of the subscriber members are reduced but they are allowed to be represented in some committees.

Currently, the ISO has about 96 full member countries including Australia, Germany, France, UK, and USA; 36 correspondent members; and 14 subscriber members. More information on this can be found in the ISO website (http://www.iso.org).

Apart from the international organizations that are made up with many voting and nonvoting members, numerous organizations maintain links between standard institutions and organizations. Examples of some of these organizations are as follows:

- *The Asia-Pacific Legal Metrology Forum* (APLMF) is a grouping of legal metrology authorities in the Asia-Pacific Economic Cooperation (APEC) economies and other economies on the Pacific Rim. The objective of this forum is the development of legal metrology and the promotion of free and open trade in the region through the harmonization and removal of technical or administrative barriers.
- *Bureau International des Poids et Mesures* (BIPM) that ensures worldwide uniformity of measurements and their traceability to the International System of Units (SI).
- *World Trade Organization* (WTO), an international agency whose purpose is to help trade flow as smoothly as possible in a system based on rules, to settle trade disputes between governments, and to organize trade negotiation.

Once a published standard is developed by an organization, depending on its applicability that standard may be adapted by other organizations after some modifications.

In addition, many standards organizations collaborate as consortiums to develop and implement particular standards, as in the case of field-bus systems – *see* **Article 172, Field-bus Systems, Volume 3** – for industrial digital information transmission applications.

5 DEVELOPMENT OF PUBLISHED STANDARDS

Standards development organizations (SDOs) undertake the task of producing relevant standards in their expertise areas. There are many different national and international SDOs, for example, in the electrical engineering field, some of the organizations in the USA are as follows:

- Department of Energy(DOE)
- Electronic Industries Association (EIA)
- International Electrochemical Commission (IEC)
- Institute of Electrical and Electronic Engineers Standards Association (IEEE)
- National Electrical Manufacturers Association (NEMA)
- National Institute of Standards and Technology (NIST).

Similar arrangements apply for other disciplines.

At the individual organization level, they have different ways of developing published standards depending on their

structures and sizes. Here, the ISO will be given as a typical standards body; more information can be found on the ISO website (http://www.iso.org/) and on the websites of many other organizations. The ISO operates with the following committees and groups:

- *Policy development committees* that are established by the General Assembly to draft policies. The committees are open to all full member and correspondent members, and they directly report to the council. Policy development committees have several branches such as CASCO (Committee on conformity assessment), COPOLCO (Committee on consumer policy), and so on.
- *Council* that operates the ISO. It consists of 18 elected representatives of full members and some permanent officers.
- *Council standing committees* that advise the Council on policy and strategic matters and raise any related issues.
- *Technical management board* that examines proposals, monitors the work of technical committees, and establishes and dissolves technical advisory groups.
- *Technical advisory groups* (TAGs) that advise the Board on basic, sectoral and cross-sectoral coordination, coherent planning, and the needs for new work. TAG 4 is responsible for metrology matters.
- *Technical committees* that are responsible for a particular area such as TC 12 for quantities, units, symbols, conversion factors; TC 108 for mechanical vibration and shock; and so on.
- *Ad hoc advisory groups* that recommend standardizations.

The standards to be internationalized can be brought to the attention of the ISO from a country or industry or business through one or more of the member organizations.

Members propose the new work item to the ISO as a whole. If accepted, the work item is assigned to an existing technical committee.

Technical committees are composed of experts appointed from industrial, technical, and business sectors. These experts may be joined by others with relevant knowledge, such as representatives of government agencies, testing laboratories, consumer associations, environmentalists, and so on. Generally, experts are chosen from the national delegations nominated by the member institutes of the countries concerned.

The ISO rules state that member institutions are expected to take account of the views of a broad range of interested parties and other stakeholders so that acceptance of the proposal becomes a consolidated and national consensus.

National delegations of experts of a technical committee meet to discuss and debate a new or amended standard until a consensus is reached on a draft agreement. The draft agreement is then circulated as a Draft International Standard (DIS) to ISO's members for comments and balloting.

Members are responsible for providing public-review procedures for making draft standards known and available to interested parties as well as to the general public. Once the feedback is received from members, the draft standard is formulated in its final form to be voted by the members.

If voting is in favor, the document, with eventual modifications, is circulated to the ISO members as a Final Draft International Standard (FDIS). If that vote is positive again, the document is then published as an International Standard.

Similar processes take place in other organizations, for example, in the case of International Electrotechnical Commission (IEC), there are seven stages to develop a full standard document (http://www.iec.org/):

1. Proposal Stage – New work item proposal (NP)
2. Preparatory Stage – Working draft(s) (WD)
3. Committee Stage – Committee drafts (CD)
4. Inquiry Stage – Inquiry draft (DIS, CDV)
5. Approval Stage (FDIS – Final draft international standard)
6. Publication Stage (IEC)
7. If an FDIS fails, the Technical Report 2 (TR2) is published and revised after 2 years.

TR2 status means publication of work is still under technical development, or where for any other reason there is a future but not an immediate possibility of agreement, on an International Standard.

6 OBTAINING STANDARDS

There are several hundred thousand standards and technical regulations in the world containing special requirements for a particular country or region. Finding information on the specific standards can be a very difficult task.

Nevertheless, once the relevant standard is identified for a particular application, the standards can be obtained in paper or electronic forms from various sources. The institutions of the country where the standards are applied can supply all the relevant information. Alternatively, one can purchase the standard directly from the issuing body or through private suppliers. Most of the universities, institutions of higher learning, and public libraries keep the current and the most-used standards. If the standards are not available in a particular library, it is possible to get them through the interlibrary loan services.

Nowadays, the standards can be purchased online from standards institution or other providers using their catalogues. Some of the web sites that provide full-text standards are given below, but the list is not exhaustive.

- ETSI, European Telecommunications Standards Institute, offers direct access to publicly available ETSI telecommunications standards (http://www.etsi.org/).
- IAEA, International Atomic Energy Agency, offers most recent IAEA publications online (http://www.iaea.org/).
- !IEEE Xplore!, Institute of Electrical and Electronics Engineers, offers standards for members (http://ieeexplore.ieee.org/).
- NISO, National Information Standards Organization, offers free full-text access to NISO information standards (http://www.niso.org/).
- ANS, American National Standards Institute, offers a range of standards for purchasing (http://www.ansi.org/).
- Global Engineering Documents offers national and international standards (http://global.ihs.com/)
- HIS, Information Handling Services, offers most engineering standards (http://www.ihs.com/).
- WSSN, World Standards Services Networks, offers national and international standards (http://www.wssn.net/).
- NSSN, National Standards Services Networks, offers national and international standards (http://www.nssn.org/).

As an example, the WSSN is a network of publicly accessible Web servers of standards organizations around the world. This server contains links to national, regional, and international standardization bodies. It also provides access to other national and international organizations that develop standards in their specialized subject areas.

If web browsers are selected to find standards, search can be made by keyword entry, such as EMC, temperature, noise, and so on; or by standard number, such as TP1, AB4, and so on; or by complete title or portion of a title, such as motors and generators, electrical equipment, and so on; or by date if the publication dates are known.

7 USE OF STANDARDS IN SUPPORT OF DESIGN AND OPERATION

Standards in measurement and metrology largely define guidelines for testing methodology, accuracy, and reporting criteria for consistency and accuracy. From guidelines, the necessary instrumentation, procedures, and control processes can be determined and used appropriately for the design, construction, operation, maintenance, installations, safety, and other related areas.

Use of standards in the design and operations offer many advantages:

- improves the communication between designers and operators
- provides practical application of expert knowledge
- represents accumulated experience, thus avoiding ground-up starts
- enhances operability and interoperability, thus ensuring compatibility
- sets a benchmark for global applications
- sets the recognized level of quality
- increases market access and user acceptance
- improves communication between peers in industry and other concerned parties
- provides immediate access to technical resources
- provides an understanding on the future trends
- provides consistency in products
- provides traceability
- facilitates interchangeability of products and quality assurance
- ensures safety
- facilitates migration of products for integrated systems.

Because of the advantages listed above, standards are used for many purposes, such as the following:

1. To identify and plan the measurement activities. Once the data is collected, it is processed in a prescribed manner so that it can be easily understood by other parties.
2. To select appropriate instruments and experimental setups to yield credible and reliable results.
3. To design new products and improve the existing ones.
4. To comply with the consumer/customer expectations.
5. To meet the legal requirements and counteract possible challenges, and so on.

As mentioned earlier, one problem with using standards is that there may be too many of them for a particular application; hence selection of the appropriate one may be difficult. Usually, customers require the use of a particular standard that suits his/her requirement within that country. In the case of international applications, the recommended ones by the standards authorities need to be selected and complied with.

8 EXAMPLE OF DEVELOPMENT OF A PUBLISHED STANDARD

In this section, electromagnetic compatibility (EMC) will be given as an example of the types of standards and the selection of the appropriate ones.

Electromagnetic radiation is a form of energy that can propagate through a medium. It can be generated intentionally or unintentionally that can interface with electronic systems, instruments, and other working equipment, thus degrading their performance.

This electromagnetic interference (EMI) may be generated by almost all electrically operating devices. Owing to increase in man-made EMI generated around the globe, allowable limits as well as measurement techniques on the radio frequency (RF) noise have been set at national and international levels.

The Federal Communication Commission (FCC), International Electrotechnical Commission (IEC), and the military are some of the bodies setting standards on the EMI.

In the USA, the FCC and military are two regulating bodies governing the EMC standards.

In Europe, each country has its own EMC governing body as well as its own standards. Recently, European countries adopted the IEC 1000 as the common EMI/EMC standard.

The frequency spectrum over which EMI can be generated spans from DC to γ-rays (10^{21} Hz) and beyond. The spectrum for use in EMC mainly covers from a few Hertz to 40 GHz. Lower frequencies are a main concern for the biological, while higher frequencies cause serious concern for communications, instruments, most electronic devices, and military applications.

Supposing that we want to do an EMC test and we are beginners on the subject, we will face typical information briefly listed in Table 2.

Table 2. A brief list of standards concerning EMC.

EN 50065-1:2001	Signaling on low-voltage electrical installations
EN 55015:2001	Radio disturbance of electrical lighting and similar equipment
EN 55020:2002	Immunity of broadcast receivers and associated equipment
EN 50083-2:2001	Cabled networks for television, sound signals, interactive services
EN 5009022:1997	Home and building electronic systems
EN 60669-2-1:2000	Switches for household and similar fixed electrical installations
EN 501304:1996	Alarm systems
EN 50148:1996	Electronic taximeters
EN 50199:1998	EMC – Arc welding equipment
EN 50227:1999	Control circuit devices and switching elements
EN 50263:2000	EMC – Measuring relays and protection equipment
EN 50270:1999	Electrical apparatus for combustible gases, toxic gases, or oxygen
EN 55011:2000	Industrial, scientific, and medical RF equipment
EN 55014-1:2001	Household appliances, electric tools, and similar apparatus
EN 60204-31:1998	Safety of machinery
EN 604391:1999	Low-voltage switch gear and control gear assemblies
EN 60945:2002	Maritime navigation and radio communication equipment
EN 60947-5-1:1998	Low-voltage switch gear and control gear
EN 61000-6-4:2001	Generic – Emission standard for industrial environments
EN 611312:2000	Programmable controllers
AS/NZS 3652:1998	Arc welding equipment
AS/NZS 4251:1999	EMC – Residential, commercial, and light industry
AS/NZS 4251: 1999	EMC – Industrial environments
AS 62040.2	Uninterruptible power systems
CSA C108:1977	Electromagnetic interference measuring instrument
CSA-C108.6-M91	Industrial, scientific, and medical RF equipment
IEEE 299:1997	Measuring the effectiveness of the shielding enclosure
IEEE 475:2000	Measurement for field-disturbance sensors, 300 MHz to 40 GHz
IEEE C37.100:1992	Definitions for power switchgear
ISO 7637-1:2002	Road vehicles
ISO 13766:1999	Earth-moving machinery
MIL-HDBK-1857	Grounding, bonding, and shielding design practices
MIL-HDBK-241B	Design guide for EMI reduction in power supplies
MIL-STD-469B	Radar engineering interface requirements
MIL-STD-1541A	EMC requirements for space systems
SAE ARP 958D	EMI measurement antennas – Standard calibration method

This list in Table 2 is by no means exhaustive. In fact, we can find over 500 entries on the subject. However, we can see in the list that some of the standards are duplicated, probably originating from the same source. We also notice that standards organizations specialize in some areas, which mostly concerns their area of expertise.

9 SUMMARY OF BENEFITS OF PUBLISHED STANDARDS

Both physical and published measurement and metrology standards are important in shaping the efficiency of domestic and global economies. They are necessary for worldwide progress in trade liberalization, research and development activities, interpenetration of different technological sectors, and also for ensuring consistency in emerging technologies.

Many national and international standards organizations are maintained by governments and private enterprises.

Many standards exist because of the diversity of organizations and accumulation of the documented material over time. However, a user can select the appropriate standards for a particular application easily by carefully identifying the area of concern.

The use of standards in design and applications offer many advantages such as providing expert knowledge, setting level of quality and consistency, ensuring safety, and so on.

FURTHER READING

American National Standards Institute, ANSI, (2004) (http://www.ansi.org/).

ASTM dictionary of engineering, science, & technology, 9th edn, (2000) TA402.A86 2000 Engineering Library – Dictionaries, West Conshohocken, PA.

British Standards Institution, BSI, (2004) (http://www.bsi-global.com/).

Dyer S.A. (ed.) (2001) *Instruments, Survey of Instrumentation and Measurement*, Wiley, New York.

Eren, H. (2004) *Electronic Portable Instruments-Design and Applications*, CRC Press, LLC, Boca Raton, FL.

IEEE Xplore, (2004) (http://ieeexplore.ieee.org/).

International Electrotechnical Commission, IEC, (2004) (http://www.iec.org/).

International Organization for Standardization, ISO, (2004) (http://www.iso.org/).

National Information Standards Organization, NISO, (2004) (http://www.niso.org/).

National Standards Services Networks, NSSN, (2004) (http://www.nssn.org/).

Standards Council of Canada, SCC, (2004) (http://www.scc.ca/).

Standards Library for Measurement and Control: Guidelines for Quality, Safety, and Productivity, (2001) Research Triangle Park, NC, ISBN: 1556175310.

Techstreet Thomson, standards catalogs, (2004) (http://www.techstreet.com/).

Webster, J.G. (ed.) (1999) *The Measurements, Instrumentation and Sensors Handbook*, CRC and IEEE Press, New York, ISBN: 0-8493-8347-1.

World Standards Services Networks, WSSN, (2004) (http://www.wssn.net/).

45: Calibration Process

Halit Eren
Curtin University of Technology, Perth, Western Australia, Australia

1 Definition of Calibration 271
2 Benefits of Calibration 272
3 Calibration Personnel 273
4 Calibration Laboratory Practice 273
5 Physical Features to be Calibrated 274
6 Labels Used in Calibration 274
7 Measurement Assurance Planning (MAPS) 275
8 Costs of Calibration 275
9 Trends in Calibrations and Case Studies 276
10 Literature on Calibration 277
Further Reading 277

1 DEFINITION OF CALIBRATION

Calibration of instruments and processes is essential for checking their performances against known standards. This provides consistency in readings and reduces errors, thus validating the measurements universally. The calibration procedure involves comparison of the instrument against primary or secondary standards. In some cases, it may be sufficient to calibrate a device against another one with a known accuracy. After the calibration of a device or a process, future operation is considered to be error bound for a given period of time under similar operational conditions.

The process of calibration is carried out in a hierarchical order. At the highest level, the primary reference standard is assigned a value by direct comparison with the reference base of SI units. The primary standards are designated and widely acknowledged as having the highest metrological quantities that have values without reference to other standards of the same quantity. In the second level, the secondary reference standards are calibrated by comparing with primary standards of the same quantity using a high precision comparator and making appropriate corrections. In the third level, working standards are routinely used to calibrate instruments and processes against the secondary reference standards or their representatives. More is provided about physical standards in **Article 43, Units, Volume 1**.

The most important element in calibration is the relationship between a single measurement and the reference base for the unit of measurement: the reference base is the prime source of authority. The base units of measurement are the Le Systeme International d'Unites (SI) units maintained in the Bureau International des Poids et Mesures, Paris. These are kilogram for mass, meter for length, second for time, candela for luminous intensity, kelvin for thermodynamics, ampere for current, and mole for amount of substance. Other reference bases such as newton for force, hertz for frequency, and so on, are derived from the base units and maintained by national standards.

Recently, with the wide applications of digital systems, many intelligent instruments can make self-calibrations – **Article 160, Smart Sensor System Features, Volume 3**. In these cases, postmeasurement corrections are made and the magnitudes of various errors are stored in the memory to be recalled and used in the laboratory and field applications. A new trend is that certain calibrations can be conducted over the Internet by entering the appropriate Websites of manufacturers or calibration authorities – *see* **Article 47, Internet Calibration, Volume 1**.

The Oxford English Dictionary defines the term calibrate as '**1** mark (a gauge or instrument) with a standard scale of readings. **2** compare the readings of (an instrument) with

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

those of a standard. **3** adjust (experimental results) to take external factors into account or to allow comparison with other data.'

Calibration can be defined as 'a set of operations to establish the relationship between values of quantities indicated by measuring instruments and systems under specified conditions.'

Therefore, calibration is a process that assigns values to the response of an instrument relative to reference standards or to a designated measurement process. Calibration aims to eliminate or reduce bias in the measurement system relative to the reference base in accordance to a specific algorithm. The bias may be introduced by the instrument itself and remain constant independent of the direction of measurement. Alternatively, bias may be caused by linear drift over the time of measurement.

Some of the important definitions related to calibration are as follows

Calibration procedure: 'A documented, verified, and validated process that describes a set of operations in accordance to a given method.'

Calibration provider: 'Laboratory or facility including personnel that performs calibration in an established location.'

Errors: 'Result of a measurement minus the true value of the measurand.'

Reference standard: 'Measurement standard having the highest metrological quality available in an organization.'

Uncertainty of measurement: 'Dispersion of values that can be attributed to the measurand.'

In accordance with the above definitions, a calibration can serve two main purposes:

1. If the instrument reads in the same units as the reference standards, the purpose of calibration is to eliminate biases and minimize the instrumental errors.
2. If the instrument reads in different units than the reference standards, the calibration serves the purpose of converting the instrument readings to the units of interest.

In both cases, most calibrations are likely to have uncertainties.

Since most calibrations are likely to have uncertainties in the calibrated values, they need appropriate evaluations.

Uncertainties can be evaluated as type A evaluations and type B evaluations. The type A evaluation applies to both random error and bias and uses statistical techniques. Type B evaluations can apply to both random error and bias too, but the calculation of the uncertainty component is not based on a statistical analysis of data but on factors such as experience, scientific judgment, scant data, use of different laboratories, and so on. The ISO guidelines, such as ISO 11095, give guidance on how to assess, correct, and calculate uncertainties.

2 BENEFITS OF CALIBRATION

Calibration is a process of testing and comparing the errors of measurement instruments and processes with accepted standards in order to detect and correct variations in performance. Therefore, calibration assures that devices and processes meet expected performance specifications within universally acceptable levels and accuracy. Hence, calibration has the following benefits:

- It determines whether measurements made before the calibration were valid.
- It gives confidence that the future measurements will be accurate.
- It assures consistency and compatibility with those made elsewhere.
- It leads to repeatability and reproducibility assessments of the instruments and processes.
- It provides confidence that products meet their specifications, thus reducing legal liability – *see* **Article 25, Introduction to Signals in Physical Systems, Volume 1**.
- Without calibration, the product quality may be poor, thus opening up legal challenges and high failure rates of the products, thus increasing costs.
- It increases efficiency by ensuring that measurements are correct.
- In the process industry, calibration of devices assures that the processes are well controlled and that the products meet expected specifications.
- It leads to documentation of performance of instruments and processes to meet quality standards such as ISO 9000, ISO 1400, and QS-9000.
- Frequent calibrations can provide a graphical view of the equipment uncertainty over time, thus leading to reliability of performance. This gives in-service life analysis; hence, depreciation and replacements can be predicted in an informed manner.
- Measurements made within international standards promotes global acceptance, thus increasing competitiveness.
- It helps convenient implementation of related regulations and legislation that govern the use of equipment in a particular application.
- As the technology changes, the regulations and legislation of test and measuring instruments change continually, and calibration helps compliance validity of measurements and processes under changing conditions.

- In some cases, calibration can be used as a gain; the value multiplied by some input can produce a scaled output.

3 CALIBRATION PERSONNEL

The calibration is conducted by the use of appropriate devices representing the reference standards that are deployed by trained personnel. Appropriate devices are kept as secondary standards and working standards by national authorities and companies themselves. The calibration process involves set rules, regulations, and processes.

During the calibration, the calibrating personnel make the following assumptions:

1. Handling procedures are consistent.
2. Test item and reference standard respond in the same manner in the test environment.
3. Random errors associated with the measurement are independent.
4. Multiple measurements form a distribution curve with the same standard deviation.
5. The test item and reference standards are stable during the measurement.
6. The bias of the test item and reference standard are the same, and hence it can be eliminated by taking the difference between the two readings.
7. Once the calibrations are carried out, ideally the difference is zero or can be explained by appropriate charts and statistical relationships.
8. Calibration is repeatable, yielding the same results.

Skilled calibration personnel are essential to the precision of the calibration. Calibration personnel may need to have appropriate clothing (static free clothing, gloves, face masks, etc.) for safety as well as to avoid the possibility of contamination. Although fully trained and experienced, it is possible for different operators to produce measurements differing in signs and magnitudes. To overcome this problem, measurements by different operators can be plotted and compared. Another solution may be to maintain separate calibration curves by the same operator. This may not be a problem in automated calibrations.

4 CALIBRATION LABORATORY PRACTICE

Calibration is conducted by organizations that perform tests and or calibrations in permanent, temporary, or remote locations. These organizations may have several laboratories (called *calibration laboratories*) for calibrating different instruments and processes.

Some organizations cover a broad spectrum of technologies within a wide variety of industries such as communication, defense, medical, aerospace, automotive, and chemical industries.

Laboratories are accredited by authorities in accordance with guidelines such as the ISO Guide 58.

Accreditation is a formal recognition that a particular laboratory is competent to conduct specific tests and/or calibrations.

Calibration of an instrument is intended to eliminate or reduce bias in its readings over a continuous range of values. For this purpose, a reference standard with known values is measured by the instrument and the functional relationship between the values of the standards and the corresponding measurements is established.

A successful calibration procedure requires the following basic steps:

- selection of an appropriate reference standard with known values covering the range of interest;
- application of the instrument or the reference standard;
- conducting calibration curves (i.e. least-squares fit) to establish the relationship between the measured and known values of the reference standard;
- correction of measurements using calibration curves;
- preparation of the appropriate documentation of the calibration procedure, results, analysis, and interpretation of results for the client.

Once a procedure is adapted, the calibration relies on the instrument continuing to respond consistently and in the same way during the calibrations. If the system drifts or takes unpredictable deviations, the calibrated values may not be easily corrected for bias and errors, thus degrading the accuracy of the measurements.

During the calibration process, the readings of the test item are compared with the reference standards such that values can be assigned on the basis of the known values of the reference standard. The reference standards can be: resistors, length standards, voltage standards, and so on.

However, in many situations, it may be impossible to achieve a perfect calibration because of instrument and measurement biases, and uncontrollable random errors. Mathematically: Ideal value = Measured value + Bias + Error

Similarly, a reference value may be subject to bias and error: Ideal reference value = Reference value + Bias + Error

This leads to a deficiency in the calibration: Deficiency = Ideal measured value − Ideal reference value

Owing to randomness of errors, this deficiency may not be zero; therefore, the calibration measurements have to be repeated many times. Then, the statistical techniques can be applied to work out the calibration curves, average readings, and standard deviations.

The process of collecting data for creating the calibration curve is critical to the success of the calibration program. As a rule of thumb, a minimum of five reference standards is required for a linear calibration curve and ten reference standards are sufficient for more complicated calibration models.

Nowadays, most calibrations are conducted using computers to capture and analyze data. Once the results are obtained, software packages can assist in analyzing the information. Most packages use the method of least squares for estimating the coefficients. Some of the packages are capable of performing a weighted fit if the errors of the measurements are not constant over the calibration interval. The software tools provide information such as the coefficients of a calibration curve, standard deviations, residual standard deviation of the fit, and goodness of the fit.

Calibrations of instruments and processes can be made under static or dynamic conditions. If the calibrated variable has multiple inputs and outputs, the input is varied in increments, in increasing and decreasing directions, over a specified range.

The observed output then becomes a function of that single input. For better results, this procedure may be repeated by varying the input, thus developing a family of relationships between the inputs and outputs.

In multivariable situations, the input/output relationship usually demonstrates statistical characteristics. From these characteristics, appropriate calibration curves can be obtained, and statistical techniques can be applied for analysis.

5 PHYSICAL FEATURES TO BE CALIBRATED

The physical features to be calibrated depend on the characteristics of the measuring instrument or process under test.

Calibration is applied in a diverse range of measurement instrument and processes. Some examples follow.

Flow calibration: There are many devices and facilities available for the measurement of liquid, air, or solid flow – *see* **Article 190, Calibration and Standards in Flow Measurement, Volume 3**. Once the method of measurement is determined by an appropriate flowmeter set up, static or dynamic calibrations can be carried out. In the case of static-gravimetric liquid flow, a calibration facility may include a reservoir, a pumping system, a pipeline, flowmeter under test located on the pipeline, collection system, computers and interface, supporting software, and so on. The calibration of flow of fluid through the meter can be determined by collecting the prescribed mass of steady fluid flowing over a measured time interval.

Sensor calibration: Sensors may have to be calibrated after having their data integrated with a signal conditioning system. This process for calibrating the processing part of the system requires injection of a known input signal into the sensor. By observing the output, a correct output scale can be configured for that particular application. If the sensor is used for time varying inputs, the dynamic calibration becomes necessary. In most cases, transient behavior of sensor top step response may be sufficient to assess the dynamic response of the sensor.

Calibration of food products: As food contains many chemical substances, the calibration of devices becomes complex. In the case of honey, for example, the following parameters need to be identified by calibrated instruments and processes: fructose, glucose, turanose, maltose, moisture level, acidity, and so on.

Calibration of images: Calibration is one of the first steps in image processing. For example, astronomical images are calibrated to eliminate the effect of cameras, light pollution, and distortions. Various methods are used to ensure calibrated images by eliminating thermal, readout, and other effects. For thermal effects, the cameras are cooled below certain temperatures and dark frames are used to compensate the noise generated by camera electronics.

6 LABELS USED IN CALIBRATION

Calibration labels are important as they supply information on the efficiency and applicability of instruments. They indicate parameters of instrument, tolerances, and special conditions of use. Calibration labels need to conform to the requirements of ISO/IEC 17025 General Requirements for the Competence of Testing and Calibration Laboratories. ISO 17025 requires the following

- All measurement equipments shall be securely and durably labeled.
- The labels should clearly indicate the name of the calibration laboratory, date of calibration, due date, usage equivalent, and the authorized officer.
- Information on the label must be legible and durable under reasonable use and storage conditions.

- When it is impractical to affix a label directly on an item, the label may be affixed to the instrument container.
- Temperature-resistant seals may be used when necessary.
- Functional labels should contain reference standards.

As an example, labels on radioactivity-measuring instruments contain information on calibration parameters, applicable dose levels, type of radioactivity detected, source used, and so on, as illustrated in Figure 1.

Meters with multiple probes have additional labels for each additional probe containing information such as the following

- date last calibrated and call date
- probe or sensor type
- information on organization and personnel who conducted the calibration
- range of measurement and correction factors for each range
- efficiency of the instrument, and so on.

Calibration labels must be durable and legible. They are made from various materials such as metal plates, soft adhesive aluminum, vinyl overlays, or transparent adhesive tapes. In the case of metal plates and adhesive aluminum, information and markings may be permanently indented for durability. Calibration labels may have the following information:

- calibrate before use
- not calibrated
- obsolete
- do not use
- does not conform
- indication only
- information only
- out of calibration
- reference only
- uncalibrated instrument
- user-calibrated instrument, and so on.

Cal. Date:__/__/__ Cal. By
Cal. Due:__/__/__ _____ Probe type:_____

Range	*C.F. Eff.*	*Isotope*	*Eff.*	*Isotope*
X 0.		C^{14}	%	%

Figure 1. A typical calibration label.

7 MEASUREMENT ASSURANCE PLANNING (MAPS)

Measurement Assurance Planning (MAP) is a group of activities designed for critical evaluation of a group of measurements. The term was generated by the National aeronautics and Space Administration (NASA) and National Institute of Standards and Technology (NIST) for describing the calibration of instruments for their Earth Observing System (EOS) programs. The concept is applied to calibration of EOS satellite, ground-based and airborne instruments where it consists of a series of carefully designed activities in space explorations and measurements.

However, the idea of MAPs has been adapted by many other organizations as a guideline for good practices in measurement and calibration activities under various names such as Volt MAP, Mass MAP, Accelerometer MAP, and so on.

NASA applies MAPs to achieve accurate, precise, and consistent long-term series of radiometric measurement data form multiple instruments and multiple platforms to advance scientific understanding of the earth as a system. It is identified that the success of the project depends on proper calibration of all instruments against a set of recognized physical standards and careful characterization of the instrument performance at the system level.

Other important points would be adhering to good measurement practices and established protocols, intercomparing measurements wherever possible, and establishing traceability for all instruments to the common scale of physical quantities maintained at the national standards laboratories.

8 COSTS OF CALIBRATION

A successful calibration process requires hardware and software, special equipment, and manpower, hence the costs are variable depending intensity of use of these variables. The cost of calibration depends on what is calibrated and who is calibrating by it.

In simple cases where a one-off instrument is involved, cost can be lower than one hundred dollars, but complex cases can cost thousands of dollars.

Calibration cost depends on whether the calibration is carried out on the premises of calibrating laboratories or on the factory floor being outsourced to third parties.

Certification by ISO 10012-1, ISO 9001, MIL-STD 45662A, and MIL-HDBK-52B requires calibration for measuring equipment. In many situations, such as weighting systems calibration, it is a statutory requirement.

One of the major factors for cost is the frequency of calibration of an instrument. Most calibration systems issue

a validity period during which the instrument can be used without concern for major errors and uncertainties. Some organizations use finely worked out methods for determining calibration intervals, while others use conservative calibration intervals barely able to meet the legal demands.

The perception exists that calibration cost can be reduced if the interval can be stretched legitimately. The use of uncalibrated instruments in an organization can be costly as it may affect the product quality and quality of downstream operations.

Standards such as MIL-STD 45662A suggest good calibration intervals. As a rule of thumb, 85 to 95% of all instruments returned for calibration meet the calibration limits. The calibration limits are determined by probability charts of age of instruments and their failure data.

Usually, an instrument must be calibrated if the failure rate increases or functionality deteriorates when compared to other standard instruments.

It is possible that a planned calibration costs, say, $200 as opposed to an unexpected failure costing thousands of dollars. In this respect, a number of different mathematical techniques, such as the Weibull statistics and renewal equations, can be employed to analyze the costs. There are different software tools (e.g. visualSMITH, Calibration Manager, etc.) available for cost analysis and for determining calibration intervals.

9 TRENDS IN CALIBRATIONS AND CASE STUDIES

With the availability of advancing technology, the classical calibration process is changing on at least three fronts, these being the following:

1. electronic calibration
2. e-calibration using Internet and communication techniques
3. intelligent and self-calibrating instruments and sensors.

These techniques are now explained with some examples.

Electronic calibration: Many modern instruments offer features for closed-case calibrations so that electronic calibration can be employed. Electronic calibration is a single connection and one- or two-port calibration technique without disturbing the components inside the case. Once the calibrating equipment, for example, computer, is linked with the device under calibration, appropriate software generates the necessary calibration information. Errors due to gains and offsets of the instrument are corrected mathematically within the instrument processor to obtain the correct measured values. Analog corrections can also be made via the adjustment of the parameters of the digital-to-analog converters. Corrected calibration constants are kept within the nonvolatile memory for permanent use.

As an example of this method, the Agilent™ electronic calibration modules 8719, 8720, and 8722 microwave network analyzers provide a broad frequency range of calibrations from 10 MHz to 67 GHz (http://we.home.agilent.com/USeng/nav/). Similarly, Fluke™ offers electronic calibration facilities for multifunction process, pressure, and temperature instruments (http://www.fluke.com/calibrators/cal.asp).

Remote calibration: A new trend of calibration is emerging in the form of Internet calibration. Undoubtedly, the use of Internet is likely to expand the calibration capabilities without the transportation of equipment from one place to another. Internet calibration technique is supported by web-accessible test procedures and appropriate hardware and software.

As examples of Internet calibration, Fluke™, UK, offers calibration systems for the Fluke 4950 multifunctional instruments (http://www.npl.co.uk/npl/news/press/pr16-2000.html).

Anritsu™ is active in investigating the possibilities of Internet-based calibration using portable OTDR MW9076 (http://www.electronicstalk.com/news/anr/anr163.html). The OTDR is controlled via the Internet with the aid of appropriate software. The software can be controlled through a PC via modem, mobile telephone access, or PCI cards.

Similar efforts are made by NIST to realize Internet-based calibrations particularly aimed for radiation processing industry – see also the work of NPL in UK, as described in **Article 47, Internet Calibration, Volume 1**.

In some cases, calibration algorithms and production automation know-how are available online by combining standard low-cost integrated circuits and Internet access. Online calibration is supported by appropriate software that includes the mathematical models of the interface electronics and optimizes calibration on the basis of this model.

Self-calibrating instruments and sensors: Nowadays, many smart sensors are available in the marketplace, including digital programmable calibration features, as shown in Figure 2. Particularly, sensors complying with IEEE 1451.4 standards provide comprehensive transducer electronic datasheets (TEDs) that contain configuration, scaling, and calibration information necessary to make measurement through mixed mode interface.

Many smart sensors – *see* **Article 160, Smart Sensor System Features, Volume 3** – are capable of calibrating themselves, scaling the incoming data, computing statistics, and communicating with other digital systems on the

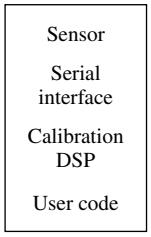


Figure 2. Main features of self-calibrating smart sensors.

network. Most commercially available software permits calibration of smart sensors and upload the new parameters directly to the sensor's circuitry.

In an application of smart sensors (the Atmos™ SSP14 Sensor Signal Processor family) (http://www.sensorsmag.com/isensors/dec00/34/main.shtml), each smart sensor has a memory that is programmed at the factory with a set of default zero and span curves defining the relationships with the physical phenomenon. These default curves represent average sensor output adjusted for the most accurate response at room temperature. Each time calibration is done, the appropriate zero or span curve is adjusted in the vicinity of the calibration.

10 LITERATURE ON CALIBRATION

Literature on calibration is available from various sources. These are:

1. *Manufacturers*: They supply comprehensive information about calibration requirements of their products.
2. *Regulating authorities and standards institutions*: These supply information on calibration requirements of instruments and devices. Calibration can be statutory, particularly where health and safety is important.
3. *Calibration services*: These provide information on calibration processes.
4. *Organizations*: They provide rules and regulations for their equipments for assurance planning.
5. *Books*: These provide information on calibration issues and processes. Details of some of these books can be found in the bibliography.

Many nations and organizations maintain laboratories with the capability of primary functions of calibrating instruments and field measuring systems that are used in everyday operations. Examples of these laboratories are: Standards Council of Canada (SCC), National Institute of Standards and Technology (NIST), National Association of Testing Authorities (NATA) of Australia and the British Calibration Services (BCS), and Australian Standard (AS).

Some important information on calibration can be found in ISO, IEC, IEEE, and national standards publications. Some of these publications are briefly explained below.

ISO 17025 is an international standard for calibration and testing laboratories. It requires labs to demonstrate that they operate a quality system covering processes, documentation, and quality management. The laboratories need to generate technically valid results accounting for the equipment procedures and personnel. Information on ISO 1705 can be found in many books or ISO publications (http://www.fasor.com/iso25/)

BS EN ISO 9000:2000 is a family of standards for international quality management. These standards are globally accepted and applied by many organizations. (www.iso.ch/iso/en/iso9000-14000/)

The military standards MIL-STD-45662A are standards that describe the requirements for creating and maintaining calibration systems for measurement and testing. They are available at http://store.mil-standards.com/.

FURTHER READING

BS EN ISO 9000:2000, (2000) http://www.iso.ch/iso/en/iso9000-14000/iso9000/iso9000index.html.

Eren, H. (2004) *Electronic Portable Instruments-Design and Applications*, CRC Press, LLC, Boca Raton, FL.

Fluke (1994) calibration, *Calibration: Philosophy in Practice*, Fluke Corporation, Everet, WA, http://www.fluke.com/calibrators/cal.asp.

ISO/IEC/EN 17025, http://www.fasor.com/iso25/.

i-sensor *Sensor Interface and Calibration Solutions via the Internet* http://www.sensorsmag.com/isensors/dec00/34/main.shtml, 2004.

Microwave calibration kits, Agilent, http://we.home.agilent.com/USeng/nav/, 2004.

Noizette, J.L., Robert, M. and Riviera, J.M. (1997) Intelligent Sensor Calibration. Complexity, Methodology, and Consequences, *IEEE-IMTC Conference Proceedings*, Vol. 2, New York (pp. 948–952).

Sydenham, P.H., Hancock, N.H. and Thorn, R. (1989) *Introduction to Measurement Science and Engineering*, Wiley, New York.

Webster, J.G. (ed.) (1999) *The Measurement, Instrumentation, and Sensors Handbook*, CRC Press, Boca Raton, FL.

46: Calibration Interval

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Physical Features to be Calibrated 278
2 Calibration Labels 279
3 Calibration Interval Assessment 279
4 Types of Calibration Interval 280
5 Calibration of Automatic Test Equipment (ATE) 280
Related Articles 280
References 280

1 PHYSICAL FEATURES TO BE CALIBRATED

Calibration is the process of giving a measuring system traceable accuracy. The general topic of calibration is covered in **Article 45, Calibration Process, Volume 1**. This article deals with how to determine and set the interval between calibrations.

There exist many physical features of a measuring instrument that may need calibration. Often, the manufacturer's original test certificate will indicate which parameters will need periodic calibration and give a recommendation for what interval to use. The following list provides a checklist of some of the many possible error sources that may need quantifying:

1. zero setting;
2. full-scale value;
3. error of nonlinearity;
4. gain or transduction coefficient;
5. effect of external influence parameters (temperature, relative humidity, atmospheric pressure, magnetic fields, vibration, radio frequency interference, ambient illumination level, and more);
6. power supply variations and noise on supply;
7. response to pulsations on measurand;
8. frequency response;
9. step response and other inputs;
10. maximum safe voltage levels;
11. leakage resistance to ground;
12. level of operation of a protective device;
13. general inspection of instrument integrity;
14. battery voltages;
15. transmission and reflectance levels in optical systems.

Before a new instrument is commissioned, it is first necessary to establish a suitable calibration plan. The instrument should be calibrated before use and proper records created. The group that is responsible will be required to maintain a record system. Often a history card is used, on which is recorded all required technical details about the instrument including calibration information. A calibration procedure should be established in writing and be used each time the item is calibrated.

Where automatic testing equipment (ATE) is used, the system will generally produce a calibration printout, which can act as the certified calibration document. Clearly, the data displayed is selected to suit the need.

It is often convenient to hold a copy of the calibration certificate with the instrument but in such cases the original should still be filed elsewhere in a secure location. In the event of severe damage to the instrument, or its complete loss, the certificate will then be most useful to help phase-in a replacement.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

2 CALIBRATION LABELS

Many different reasons exist for requiring one or more of the many physical parameters listed above to be calibrated. A guide of the types of calibration labels that are in use was published in the *NCSL Newsletter* of July 1978, this being the summary of a study organized by one of the Regional Groups of the NCSL membership. The following list is extracted from that report:

1. standards calibration.
2. calibration: manufacturer's specifications.
3. calibration: limited specifications.
4. calibration: facility specification.
5. calibration: limited facility specification.
6. not usable for acceptance testing.
7. cross-check.
8. operational check.
9. calibration not required.
10. preventive maintenance.
11. calibrate before use.
12. inactive or do not use.
13. reject.
14. reference or primary standard.
15. transfer standard.
16. working standard.

The list was compiled from government, military, and private users. It was thought during that meeting that, in many cases, simplification of labeling systems might be possible.

Associated with every calibration is the cost of making a new calibration, and the cost resulting because of not having done it when it needed to be done. The first cost is relatively easy to decide. It depends upon the instrument complexity, ease of adjustment, degree of automation, calibration procedure, ability of staff, and less obvious factors such as cost of loss of production while it is being calibrated, loan or spares holdings needed to reduce such loss of production, transport costs, and costs of the associated paperwork. Clearly, the cost of making a calibration far exceeds the staff time of those involved. A balance must be made between keeping costs low by employing less calibrations and the risk of waiting too long.

Costing the effect of using an instrument that might be (or might not be!) out of calibration is the intangible factor. Overcalibration is often considered to be – see Gebhart (1980) – the safe path, but in the interests of preserving maximum financial efficiency as well as, perhaps, maximum instrument integrity, it is usually necessary to seek a suitable calibration interval along with guiding rules, which will decide when, in the future, the instrument should be calibrated.

Obviously, the interval will depend upon the equipment's characteristics and the purpose of the instrument. In nuclear power stations, the torque wrenches used to tighten critical nuts, such as those that hold on the reactor vessel top, are recalibrated after a dozen or so pulls: for each nut and its associated pull, there is an individual record of the circumstance. At the other end of the scale is the multimeter of the servicing electrician that may not need to be calibrated after the initial build unless the instrument becomes damaged.

Statements on calibration interval are not easily found. Useful sources could be NCSL (1996), Dietrich (1997), Morris (1998).

3 CALIBRATION INTERVAL ASSESSMENT

A priority problem faced by calibration facilities is establishing suitable calibration intervals to be used. At present, it does not appear that there will ever be developed a system that guarantees minimum cost of calibration and zero loss through not having recalibrated an instrument in time. It is necessary to build up information about an instrument as its life passes, making intelligent informed estimates of need that are gained from this experience and by study of what others have done beforehand. The essential state to target for is an adequate state of satisfaction that a responsible decision has been made and that undue risks are not taken.

To assist those without such experience, the following information is extracted from various issues of the *NCSL Newsletter*:

A suitable paper that describes how a very large organization handles the problem – the US Army – is provided in Westmoreland (1980).

The bibliography of the June 1977 *Newsletter* issue is specifically concerned with calibration interval topics.

During early 1978, the NCSL Calibration Systems Management Committee conducted a US survey on this subject. Their findings are published in the July 1978 issue. It contains the following outline of how a typical member laboratory characterizes its calibration interval system:

1. Initial intervals are established on the basis of the recommendation of others, usually the manufacturer of the instrument.
2. All instruments with the same function (e.g. DMMs, scopes, etc.) are calibrated on the same interval.
3. Intervals are adjusted (lengthened or shortened) periodically (about twice a year) based on analysis of data for the manufacturer and model number group of instruments.

4. Major test systems are either calibrated as a system, or individual instruments within the system are removed and calibrated separately.
5. A maximum interval (between one and two years) is imposed, but there is no minimum interval.
6. The interval system is intended to provide either at least 85% or at least 95% in tolerance of instruments at recalibration.

An update of the NCSL work has been published, but is out of print, NCSL (1996). As ISO9000 is so widely used to assist quality operations, Morris (1998) may be of use.

4 TYPES OF CALIBRATION INTERVAL

Calibration intervals are often initially based upon one of two procedures, these being 'recall after a given calendar time period' or, the less favored method being based on 'instrument usage time'. These are discussed in Greb (1973). Once established, the period is varied in accordance with experience. The impact of the smart instrument generation containing self-calibration of a kind adds a new dimension to the task.

The recall periods of a large number of commercial instrument products were surveyed and reported in the May 1976 issue, the topic being followed up in Gebhart (1980).

It is not easy to provide any simple rule to adopt. Recall for calibration periods vary from 1 to 36 months, varying considerably from user to user for the same product. Some instruments have a calibration period as little as only 10 days.

It might be considered that the better the quality of an instrument the longer can be the calibration interval. Quality in this case is defined to be a factor of recalibration interval. An extensive study (Greb, 1976a,b) provides detailed comment on whether such simplistic assumptions are correct. The results of the study are expressed in terms of the intolerance rates (ITR) that occur.

On the quality of precision instrumentation, in the more general sense of the term, the reader is referred to Moss (1978), in which is summarized the experience of the Aerospace and Guidance Metrology Center, US, about the acceptance of new equipment from industry. It is stated there that a considerable amount of incoming equipment was rejected – 21% over one period involving 62 contracts. Of those failing to meet specifications, 43% were adjustable to specification but many were either condemned or returned for other, less serious, reasons. The author states that some improvement was evident at the time of study but that 1 in 6 complex facilities did not comply and 1 in 10 needed action beyond adjustment. These findings would probably not surprise many purchasers of instrumentation. A similar situation appears to still exist.

5 CALIBRATION OF AUTOMATIC TEST EQUIPMENT (ATE)

Where the number of instruments and degree of variation and type arises and the number of calibrations is high, a case may exist where automation of many of the test processes becomes economically viable. Remembering that calibration and testing work is an overhead cost, it is usually necessary to produce a cost-benefit analysis, which can clearly show that there is an economic case to support introduction of automated test equipment (ATE). A case study (Pearson 1980) provides one group's experience of getting started. The topic of calibration in testing is given some content in O'Connor (2001).

RELATED ARTICLES

Article 45, Calibration Process, Volume 1; **Article 47, Internet Calibration, Volume 1**; **Article 48, Common Sources of Errors in Measurement Systems, Volume 1**; **Article 49, General Characterization of Systematic and Stochastic Errors, Volume 1**; **Article 55, Uncertainty Determination, Volume 1**.

REFERENCES

Dietrich, C.F. (1997) *Uncertainty, Calibration and Probability: The Statistics of Scientific and Industrial*, Adam Hilger, Bristol.

Gebhart, C. (1980) Recall Period Pilot Program. *NCSL Newsletter*, **20**(1), 14–18.

Greb, D.J. (1973) Calibration on Basis of Instrument Usage. *NCSL Newsletter*, **13**(3), 26–27.

Greb, D.J. (1976a) Calibration Intervals and Instrument Quality – Introduction. *NCSL Newsletter*, **16**(3), 26–32.

Greb, D.J. (1976b) Calibration Intervals and Instrument Quality – Parts II and III. *NCSL Newsletter*, **16**(4), 29–35.

Morris, A.S. (1998) *Measurement and Calibration Requirements for Quality Assurance to ISO 9000*, Wiley, New York.

Moss, C. (1978) Quality of Precision Measuring Equipment. *NCSL Newsletter*, **18**(1), 23–24.

NCSL (1996) *Establishment and Adjustment of Calibration Intervals: Rp-1*, National Conference of Standards Laboratories.

O'Connor, D.T. (2001) *Test Engineering*, Wiley, Chichester.

Pearson, T.A. (1980) Case Study in Automated Test Equipment Use. *NCSL Newsletter*, **20**(1), 19–20.

Westmoreland, F.G. (1980) Intervals and Calibration System Management. *NCSL Newsletter*, **20**(1), 10–13.

47: Internet Calibration

Richard A. Dudley
National Physical Laboratory, Teddington, UK

1 Use of Internet to Conduct Calibration 281
2 Type of Measurand Suited to Internet Calibration 281
3 Procedures for Carrying Out an Internet Calibration 282
4 Example of an Internet Calibration Service 282
References 283

1 USE OF INTERNET TO CONDUCT CALIBRATION

Regular instrument calibration is an essential part of today's quality driven measurement environment, and a traceable link to a national or international standard should be present. Achieving traceability requires a laboratory to periodically send their standards to be calibrated at a National Measurement Institute (NMI), acquiring a certificate and correction values. The standards are measured under carefully controlled conditions at the NMI, but there is no guarantee that these conditions will be reproduced when the standards are used at the remote laboratory. Furthermore, in some cases, the value of the standards can be affected by transport leading to an uncertainty component, which is difficult to assess.

The downtime experienced by laboratories fulfilling calibration schedules can be extremely disruptive and costly while the equipment is sent away. On the equipment's return, system checks, paperwork, and the update of soft or hard calibration figures compounds the delays. Implementation of remote calibrations using the Internet as a data transmission medium has emerged as a solution to all of the transportation, environmental, downtime, and cost issues present in current calibration schemes. Additional benefits emerge in the dissemination of measurement techniques and good practice equally to all laboratories using an Internet service.

The use of the Internet to assist in metrology has taken a number of forms. In the simplest case, the Internet acts as a means of exchanging information between two remote sites with human operators at each end using video or interactive links (Filipski, 1999).

The second stage of development was remote monitoring of sensors (Lee and Schneeman, 1999 and O'Dowd *et al*., 1997).

However, services are emerging that are truly remote calibrations. These allow a laboratory to log on to a site, perform an accredited measurement and create a full uncertainty evaluation instantly (Dudley and Ridler, 2000, 2001).

Services that offer online standard or instrument calibration have become known as *Internet Calibration Services* or iCals.

2 TYPE OF MEASURAND SUITED TO INTERNET CALIBRATION

Clearly, Internet calibration is not applicable to all metrology areas, because physical constraints or human interaction are essential for some measurements.

However, measurements that rely on a standard issue instrument, on artifact, or are distinctive in the software developed as part of the calibration process, are open to Internet implementation. The implementation of remote calibration via the Internet varies considerably with each

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

measurement example but some general rules are that the system must have one of the following:

- Stable calibration artifact with easily verified performance check.
- Easily transported measurement artifact from which the calibration is derived.
- Unique software for calibration and uncertainty calculation.

3 PROCEDURES FOR CARRYING OUT AN INTERNET CALIBRATION

A laboratory undertaking an Internet-based calibration observes a number of advantages during the measurement process compared to a conventional method. Once a computer has been connected to the measurement system, typically through GPIB (General Purpose Interface Bus) or RS-232 ports, the laboratory only needs to log on to their calibration service provider.

Once connected to the system, the measurement system, standards, and instrumentation are known by the iCal service. It instructs the operator, in the correct order, to perform the measurements.

Effectively, the iCal system provides the operator with a soft procedure to perform the measurements assuming, of course, the operator is skilled and has been accredited with using the system.

Once the measurement procedure is completed, the iCal system generates the required final data with uncertainties ascertained from the measured and database information, providing the operator with a certificate if applicable.

During the measurement process, there is potential for the iCal system to provide measurement assistance screens with video or procedural details for specific parts of the process. In no way does an iCal system 'de-skill' the calibration operator; instead, it provides the maximum assistance to ensure an accurate and efficient calibration.

In addition, the iCal service provides the administration with any measurement procedural changes through new international standards and ensures all groups are following common guidelines. Finally, data warehousing and historical records are automatically kept on the iCal service Web site and can be accessed and checked with minimal demands on the service users.

4 EXAMPLE OF AN INTERNET CALIBRATION SERVICE

The first true iCal service was developed by the UK's National Physical Laboratory (NPL), (Dudley and Ridler, 2000, 2001). It combines the technology of remote monitoring, remote control, and NMI calibration techniques for application in a microwave frequency measurement system called *vector network analysers* (VNA).

VNAs provide a swept frequency measurement of the transmission and reflection coefficients for an electrical network. Calibration is performed using instrument firmware and a set of standard devices, all of which are assumed to be ideal and are available as standard items from the VNA manufacture.

Correction of the measurement data to that of the NMI comes via precision verification artifacts, air-spaced transmission lines, attenuators and terminations, whose properties change little over time. An external computer, running a specially designed code, controls the calibration and evaluates the uncertainty in measurement. The extension of control across the Internet is ideal for this particular system.

A calibration procedure under non-Internet conditions requires the client to send the calibration laboratory their precision verification artifacts for periodic calibration. The NMI returns the verification device plus certificate and correction data, Figure 1(a). The client must then implement their measurements using the verification information and introduce their own uncertainty budget. It should be noted that typically a 1 to 18 GHz calibration for a VNA results in several thousand correction values being generated. The majority of VNA firmware does not allow inclusion of these correction values and a client will usually have to resort to creation of proprietary control software or use a spreadsheet to correct measured data using the calibration information.

The Internet version, while not removing the need for physical movement between sites of the instrument's standard reference, does simplify the process, Figure 1(b).

When the NMI receives the client standard, all correction factors are stored in an online database. The standard is returned, but now all a client needs to do to measure a device with traceability is to connect their control PC to the service Web pages.

While online, the client enters the required measurement parameters and is offered options based on the knowledge the NMI has about the client's equipment. From this point, the entire measurement process is controlled by NMI and the need for clients to assess their own uncertainty budget is removed.

There exists great potential for many new applications of Internet calibration and metrology. Applications are being investigated in the medical and optical networks markets. The optical communications network provides a method of a real standard transfer for wavelength measurements.

If telecommunications networks were all-optical – they are presently a combination of optical and electrical – it would be possible for an NMI to distribute a wavelength

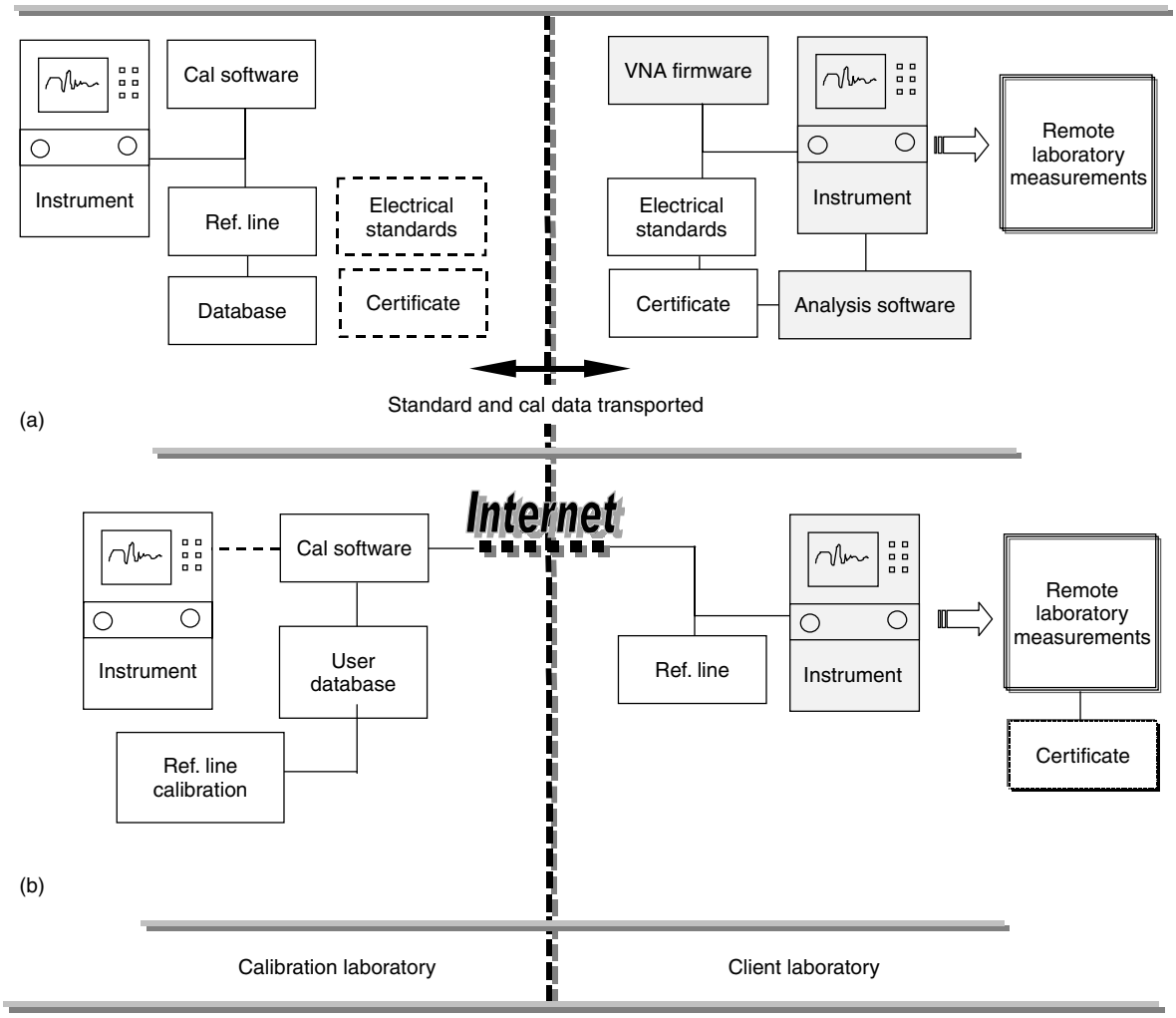


Figure 1. (a) Basic components of Internet calibration system: artifacts, instrument and control firmware, producing a certificate for the remote laboratories' artifacts. (b) The Internet system removes the need for traveling verification artifacts by allowing direct access to the calibration software and client historical database.

standard into the network creating a 'real' international standard artifact.

Other possibilities such as frequency, noise, and possibly time stamping could be placed on the back of the wavelength standard for added value. Futher information is found on http://www.internetcalibrations.com.

REFERENCES

Dudley, R. and Ridler, N. (2000) Internet-based Calibrations of Electrical Quantities at the UK's National Physical Laboratory, in *Proceedings of the NCSL Workshop and Symposium 2000 (National Conference of Standards Laboratories)*, 16–20 July Westin Harbour Castle, Toronto, Ontario, Canada.

Dudley, R. and Ridler, N. (2001) Internet Calibration Direct to National Measurement Standards for Automatic Network Analysers, in *IMTC 2001 Proceedings of the 18th IEEE Instrumentation and Measurement Technology Conference*, 21–23 May, Budapest.

Filipski, P. (1999) SIMnet – A Collaborative Tool for Metrology in the Americas, in *Proceedings of 16th IEEE Instrumentation and Measurement Conference*, Vol. 2. May.

Lee, K. and Schneeman, R. (1999) Distributed Measurement and Control Based on the IEEE 1451 Smart Transducer Interface Standards, in *Proceedings of 16th IEEE Instrumentation and Measurement Conference*, Vol. 2. May.

O'Dowd, R., Maxwell, D., Farrell, T. and Dunne, J. (1997) Remote Characterization of Optoelectronic Devices Over the Internet, in *Proceedings of 4th Optical Fibre Measurement Conference*, October (pp. 155–158).

48: Common Sources of Errors in Measurement Systems

Dietrich Hofmann
Steinbeis Transfer Centre for Quality Assurance & Quality Measurement, Jena, Germany

1 Definition of Errors in Measurement Systems 289
2 How Errors Arise in Measurement Systems 290
3 Terms Used to Describe Errors 290
4 Types of Errors in Defined Classes 290
5 List of Error Sources in Measurements 292
6 Standards on Error Description 292
7 Uncertainties of Measurements 293
8 Glossary of Terms for Errors and Uncertainties 294
Reference 294

1 DEFINITION OF ERRORS IN MEASUREMENT SYSTEMS

A number of crucial definitions are first needed to ensure clarity of discussion. These are as follows.

Measurement error is defined as the difference between the distorted information and the undistorted information about a measured product, expressed in its measurands. In short, an error is defined as real (untrue, wrong, false, no go) value at the output of a measurement system minus ideal (true, good, right, go) value at the input of a measurement system according to (1):

$$\Delta x = x_r - x_i \tag{1}$$

where Δx is the error of measurement, x_r is the real untrue measurement value, and x_i is the ideal true measurement value.

A *measurement* is an experimental process to acquire new knowledge about a product.
A measurement is the realization of planned actions for the quantitative comparison of a measurand with a unit.
A *measurand* is the physical quantity that is to be measured. Generally, a measurand is the input of a measurement system.
Measuring is a process for ascertaining or determining the quantity of a product by application of some object of known size or capacity, or by comparison with some fixed unit – see www.iso.ch/9000e/2000rev_8e.htm.
A *product* is the result of a process after ISO 9000:2000, point 3.4.2 – see http://www.iso.ch.
Four generic product categories exist as follows: service, software, hardware, processed material.
Many products comprise elements belonging to different generic product categories. Whether the product is then called service, software, hardware, or processed material depends on the dominant element. A *process* is a set of interrelated or interacting activities that transform inputs into outputs (ISO 9000:2000, point 3.4.1) – see http://www.iso.ch. Inputs to a process are generally outputs to other processes.
A *measurement process* is a set of operations used to determine the value of a quantity (ISO 9000:2000, point 3.10.2) – see http://www.iso.ch.
Measuring equipment is the measuring instrument, software, measurement standard, reference material or auxiliary apparatus, or a combination, necessary to realize a measurement process (ISO 9000:2000, point 3.10.4) – see http://www.iso.ch.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

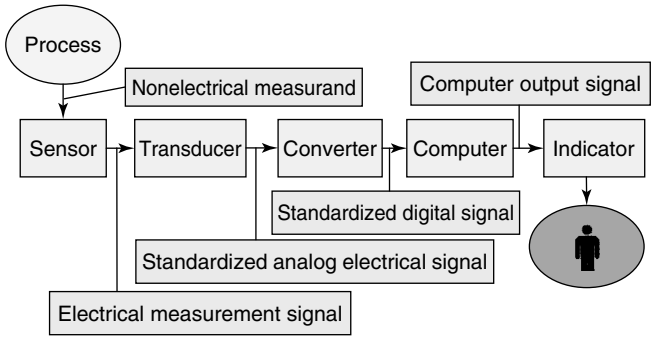


Figure 1. Measurement chain.

A *measurement system* or measuring system is the totality of all measuring equipments and auxiliary means to obtain a measuring result.
A *measurement signal* is the quantity in a measuring instrument or measuring equipment, which is unequivocally related to the measurand.
A *measuring chain* (Figure 1) is the structure of elements of a measuring equipment or measuring system, which channels the transfer of a measurement signal from the input of the measurand to the output of the measured value.
A *measured value* is a value that is unequivocally related to the measurand at the output of a measuring chain.

2 HOW ERRORS ARISE IN MEASUREMENT SYSTEMS

A measurement under ideal conditions has no errors. Real measurement results, however, will always contain measurement errors of varying magnitudes. A systematic (clearly defined process) and systemic (all encompassing) approach is needed to identify every source of error that can arise in a given measuring system. It is then necessary to decide their magnitude and impact on the prevailing operational conditions.

Measurement system errors can only be defined in relation to the solution of a real specific measurement task.

If the errors of measurement systems given in technical documentation are specified, then one has to decide how that information relates to which

- measurand
- input
- elements of the measurement system
- auxiliary means
- measurement method
- output
- kind of reading
- environmental conditions.

If the measurement system has the general structure given in Figure 1, the following errors may appear for a general measurement task:

- input error
- sensor error
- signal transmission error 1
- transducer error
- signal transmission error 2
- converter error
- signal transmission error 3
- computer error
- signal transmission error 4
- indication error.

3 TERMS USED TO DESCRIBE ERRORS

Globalization of the economy is strongly driving international standardization and accreditation programs, for example, ISO 9000 and ISO 17025, are the means of assuring quality control of components fabricated throughout the world.

Accrediting organizations such as American Association for Laboratory Accreditation (A2LA) experienced double digit annual growth in the past decade, particularly in the dimensional metrology field (classified under mechanical).

Most of these standards and accreditation programs require traceable measurements and hence measurement uncertainty statements – see www.mel.nist.gov/proj/sfnmi.htm.

Recent developments in metrology use

- the International System of Units (SI),
- the Guide to the Expression of Uncertainty in Measurement (GUM), and
- the International Vocabulary of Basic and General Terms in Metrology (VIM),

as a starting point for the application of terms to describe errors in the frame of uncertainties (see Figure 2).

4 TYPES OF ERRORS IN DEFINED CLASSES

Systematic error (bias) is a permanent deflection in the same direction from the true value. It can be corrected. Bias and long-term variability are controlled by monitoring measurements against a check standard over time.

Random error is a short-term scattering of values around a mean value. It cannot be corrected on an individual measurement basis. Random errors are expressed by statistical methods.

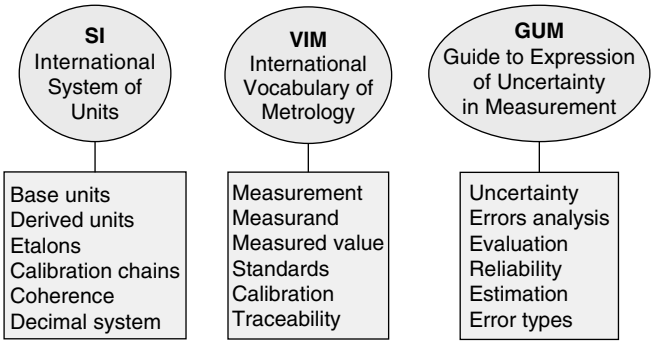


Figure 2. Main tasks of SI (http://physics.nist.gov/cuu/units), VIM (http://www.abnt.org.br/ISO_DGuide_99999(E).pdf), and GUM (http://www.gum.dk/home.html).

A fundamental source for theory and application of statistical methods is the NIST/SEMATECH Engineering Statistics Handbook (*NIST/SEMATECH e-Handbook of Statistical Method* – see http://www.itl.nist.gov/div898/handbook/, 26.03.2003).

It describes, and uses in detail, the so-called EDA method. Exploratory data analysis (EDA) is an approach for data analysis that employs a variety of techniques (mostly graphical) to

1. maximize insight into a data set,
2. uncover underlying structure,
3. extract important variables,
4. detect outliers and anomalies,
5. test underlying assumptions,
6. develop parsimonious models, and
7. determine optimal factor settings.

Most EDA techniques are graphical in nature with a few quantitative techniques included. The reason for the heavy reliance on graphics is that graphics gives the analysts unparalleled power to reveal the structural secrets of data and to be always ready to gain some new, often unsuspected, insight into the measured data in combination with the inherent pattern-recognition capabilities of man – see http://www.itl.nist.gov/div898/handbook/eda/section1/eda11.htm.

Most of the graphical techniques outlined in Figure 3 are demonstrated in the case studies given in the above-mentioned Web site.

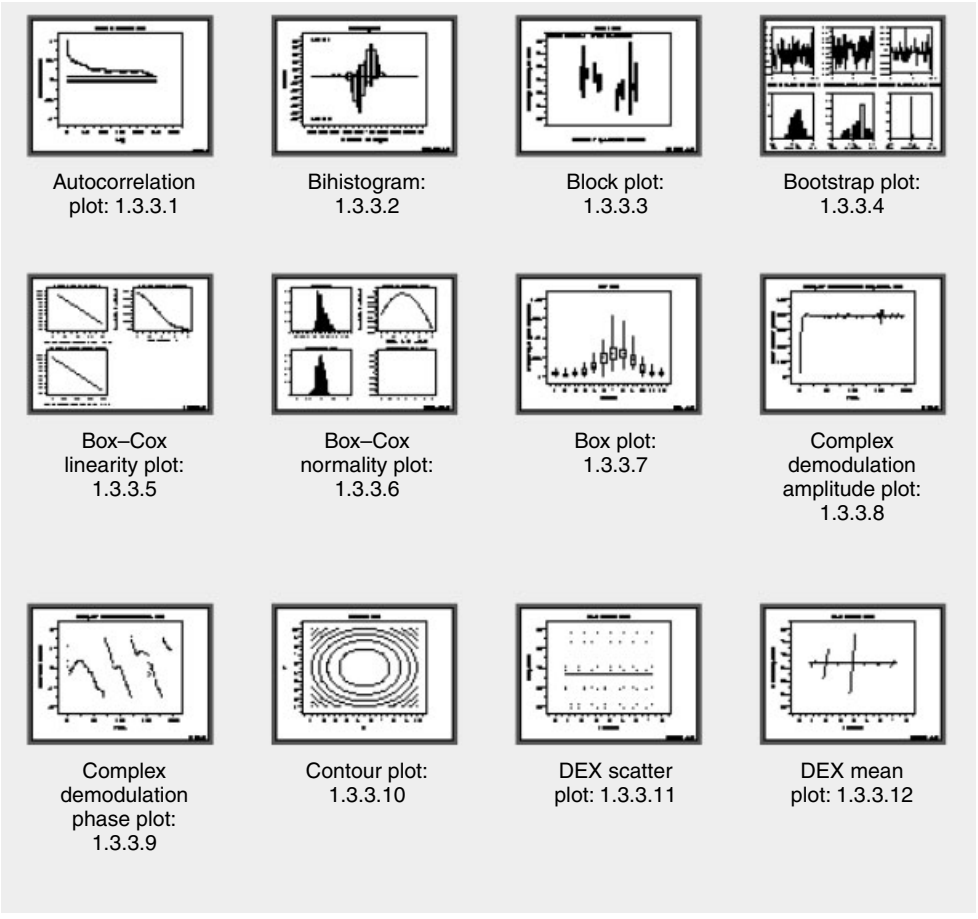


Figure 3. Graphical techniques in alphabetical order (excerpt) (http://www.itl.nist.gov/div898/handbook/eda/section3/eda33.htm).

5 LIST OF ERROR SOURCES IN MEASUREMENTS

Systematic errors or bias are repeatable errors existing with the specified source; these can be adjusted out or compensated for.

The terms 'bias' and 'systematic error' have the same meaning. Bias is defined in the International Vocabulary of Metrology, VIM, as the difference between the measurement result and its unknown 'true value'. It can often be estimated and/or eliminated by calibration to a reference standard – see http://www.itl.nist.gov/div898/handbook/mpc/section7/mpc7.htm#VIM.

To investigate sources of systematic errors, a general checklist of error sources in measurement should be used, which has been collected by specialists working in the field concerned. The main sources are

5.1 Lack of gauge resolution

Resolution better called (but rarely done so) *discrimination* is the ability of the measurement system to detect and faithfully indicate small enough changes in the characteristic of the measurement result – see http://www.itl.nist.gov/div898/handbook/mpc/section4/mpc451.htm.

5.2 Lack of linearity

A test of linearity starts by establishing a plot of the measured values versus corresponding values of the reference standards. This obtains an indication of whether or not the points fall on a straight line with slope equal to 1, which indicates linearity (proportional variation) – see http://www.itl.nist.gov/div898/handbook/mpc/section4/mpc452.htm.

Nonlinearities of gauges can be caused by the following facts:

- gauge is not properly calibrated at the lower and upper ends of the operating range,
- errors in the values at the maximum or minimum range,
- worn gauge,
- internal design problems (in, say the electronic units of the gauge).

5.3 Drift

Drift is defined as a slow change in the response of a gauge.

Short-term drift is frequently caused by heat buildup in the instrument during the time of measurement. Long-term drift is usually not a problem for measurements with short calibration cycles – see http://www.itl.nist.gov/div898/handbook/mpc/section4/mpc453.htm.

5.4 Hysteresis

Hysteresis is a retardation of the effect when the forces acting upon a body are changed (as in viscosity or internal friction); for example, a lagging in the values of resulting magnetization in a magnetic material (as iron) because of a changing magnetizing force. Hysteresis represents the history dependence of a physical system under real environmental conditions – see http://www.lassp.cornell.edu/sethna/hysteresis/WhatIsHysteresis.html.

Specific devices will posses their own set of additional error sources. A checklist needs to be developed and matured. The following is an example of such a list.

Hall Effect measurement error checklist – see http://www.eeel.nist.gov/812/errs.htm.

1. Are the probes or wires making good contact to the sample?
2. Are the contact I-V characteristics linear?
3. Is any contact much higher in the resistance than the others?
4. Do the voltages reach equilibrium quickly after current reversal?
5. Is there visible damage (cracks, especially around the contacts)?
6. Is the sample being used in the dark?
7. Is the sample temperature uniform?
8. Are dissimilar wiring materials used resulting in large temperature gradients across the wiring?

6 STANDARDS ON ERROR DESCRIPTION

Standards, when they are part of a single and coherent set of standards, promote market efficiency and expansion, foster international trade, encourage competition and lower barriers to market entry, diffuse new technologies, protect consumers against unsafe or substandard products, and provide trust and reliability – see http://www.iso.ch/iso/en/commcentre/events/2002/standardsforservices.html.

On the other hand, people talk about raising standards when they perceive slackness in the ropes of control, when they see a sloppiness infiltrating the verities of life, when they begin to be fearful about life-diminishing certainties. Talk of standards is to talk about conservation, about protecting the past in its imagined superiority and security, and defending the future through strong leadership – see http://epaa.asu.edu/epaa/v6n10/c6.htm.

Comprehensive and precise terminologies have an extraordinary importance to the industry. The following are examples:

ISO/TR 7178:1983
Liquid flow measurement in open channels – velocity-area methods – investigation of total error.
ISO 8655-6:2002
Piston-operated volumetric apparatus – Part 6: Gravimetric methods for the determination of measurement error.
ISO/IEC 10021-7:1997/Amd 1:1998
Security error diagnostic codes.
ISO/IEC 10166-2:1991/Amd 1:1995
Minor enhancement for additional error.
ISO 12142:2001
Electronic imaging – media error monitoring and reporting techniques for verification of stored data on optical digital data disks.
ISO 7509:2000
Plastic piping systems – glass-reinforced thermosetting plastic (GRP) pipes – determination of time for failure under sustained internal pressure.
ISO 10365:1992
Adhesives – designation of main failure pattern.
ISO 11782-1:1998
Corrosion of metals and alloys – corrosion fatigue testing – Part 1: Cycles to failure testing.
ISO 13953:2001
Polyethylene (PE) pipes and fittings – determination of the tensile strength and failure mode of test pieces from a butt-fused joint.

7 UNCERTAINTIES OF MEASUREMENTS

Uncertainty is a measure of the 'goodness' of a result. Without such a measure, it is impossible to judge the fitness of the value as a basis for making decisions relating to health, safety, commerce, or scientific excellence – see http://www.itl.nist.gov/div898/handbook/mpc/section5/mpc5.htm.

Uncertainty of measurement is a parameter associated with the result of a measurement that characterizes the dispersion of the values that could reasonably be expected – see http://physics.nist.gov/cuu/Uncertainty/glossary.html.

In this model, it is understood that the result of the measurement is the best estimate of the value of the measurand, and that all components of uncertainty, including those arising from systematic effects, contribute to the dispersion. Typical examples for systematic effects are uncertainties of

- systematic errors that generally can be corrected either numerically or technically,
- measurement standards that generally are negligibly small in comparison with the uncertainty of the measurement system in total.

Standard deviation is the generally used parameter for expressing the value of uncertainty. It is defined as the point of inflection in the Gaussian normal distribution – see Figure 4 – of randomly scattering values.
Evaluation of uncertainty is an ongoing process that can consume time and resources. It can also require the services of someone who is familiar with statistical data analysis techniques – see http://www.itl.nist.gov/div898/handbook/mpc/section5/mpc51.htm.
Measurement result is the best estimate of the value of the measurand.
Standard uncertainty in the measurement result is expressed with a standard deviation – see http://physics.nist.gov/cuu/Uncertainty/glossary.html.
Combined standard uncertainty in the measurement result has to be calculated by the error propagation law for statistical characteristics.
Type A uncertainty in the measurement result is expressed as a deviation evaluated by the method of statistical analysis from a series of observations.
Type B uncertainty in the measurement result is expressed as a deviation evaluated by methods other than the statistical analyses of series of observations.
Expanded uncertainty in the measurement result is expressed as the quantity of an interval that is large enough to include that fraction of the distribution function of the measured values that might be reasonably attributed to the measurand.

A special kind of expanded uncertainty is the confidence interval of a Gaussian normal distribution – see http://mathworld.wolfram.com/GaussianDistribution.html – where the factor $a > 1$.

Coverage factor is a numerical factor used as a multiplier or the combined standard uncertainty in order to obtain

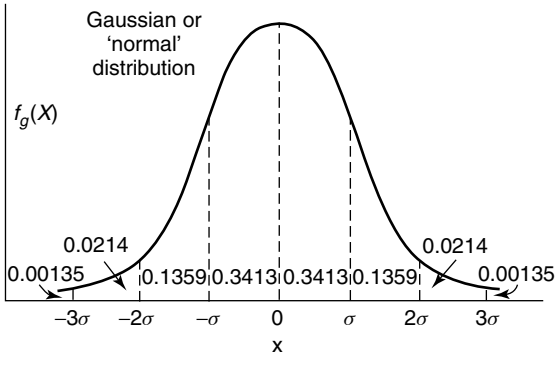


Figure 4. The Gaussian normal distribution (http://hyperphysics.phy-astr.gsu.edu/hbase/math/gaufcn.html).

Table 1. Selected glossary of terms on error and uncertainty (http://www.rit.edu/~vwlsps/uncertainties/uncertaintiespart2.html#glossary).

Term	Brief definition
Absolute error	An error in a quantity, having the same units as the quantity. For $c = (2.95 \pm 0.07)\,\mathrm{m\,s^{-1}}$, the absolute error is $0.07\,\mathrm{m\,s^{-1}}$.
Accuracy	How close a measurement is to being correct. $g = 9.7\,\mathrm{m\,s^{-2}}$ on earth is more accurate than $g = 9.532706\,\mathrm{m\,s^{-2}}$. See *precision*.
Average	When several measurements of a quantity are made, the sum of the measurements divided by the number of measurements.
Average deviation	The average of the absolute value of the differences between each measurement and the average. See *standard deviation*.
Confidence level	The fraction of measurements that can be expected to lie within a given range.
Deviation	A measure of range of measurements from the average. Also called *error or uncertainty*.
Error	A measure of range of measurements from the average. Also called *deviation* or *uncertainty*.
Estimated uncertainty	An uncertainty estimated by the observer based on his or her knowledge of the experiment and the equipment.
Gaussian distribution	The familiar bell-shaped distribution, assumed for random errors. Also called *normal distribution*.
Independent variables	Changing the value of one variable has no effect on other variables. *Error propagation* assumes this independence.
Instrument limit of Error (ILE)	Smallest reading that an observer can make from an instrument. This is generally smaller than the *least count*.
Least count	Size of the smallest division on a scale. Typically the *ILE* equals the least count or 1/2 or 1/5 of the least count.
Precision	The number of significant figures in a measurement. $g = 9.532706\,\mathrm{m\,s^{-2}}$ is more precise than $g = 9.7\,\mathrm{m\,s^{-2}}$.
Propagation of errors	Method of determining an uncertainty in a function of independent variables each with their own uncertainty.
Random error	Deviations from the 'true value' can be equally likely to be higher or lower than the true value. See *systematic error*.
Relative error	The ratio of absolute error to the average, Dx/x. Also called percentage error or fractional uncertainty. See *absolute error*.
Standard deviation	The statistical measure of uncertainty. See *average deviation*.
Systematic error	A situation where all measurements fall above or below the 'true value'. To realize and correct systematic errors is difficult.
Uncertainty	A measure of range of measurements from the average. Also called *deviation* or *error*.

an expanded uncertainty of combined measurements – see http://physics.nist.gov/cuu/Uncertainty/glossary.html.

Problems of uncertainties and error propagation and their solutions are collected in http://www.rit.edu/~vwlsps/uncertainties/Uncertaintiespart2.html#problems/.

The issue of error and uncertainty is covered in more depth in **Article 54, Explanation of Key Error and Uncertainty Concepts and Terms, Volume 1** and other articles of the Subsection 'Errors and Uncertainty'.

8 GLOSSARY OF TERMS FOR ERRORS AND UNCERTAINTIES

A glossary of terms for errors and uncertainties is available that can be used as a checklist for error handling – see http://www.rit.edu/~vwlsps/uncertainties/Uncertainties part2.html#glossary.

Table 1 provides a selected glossary of terms on error and uncertainty from that source.

Terms used in the description of static system performance are also covered in **Article 57, Static Considerations of General Instrumentation, Volume 1** and **Article 58, Description of Accuracy, Linearity, and Drift, Volume 1**. Terms used to cover dynamic performance are dealt with in **Article 59, Introduction to the Dynamic Regime of Measurement Systems, Volume 1**.

REFERENCE

The Web sites cited above are possibly the best sources of information on the terms and concepts for errors and uncertainty. Generally, texts are not sufficiently standardized – often using terms that are now defunct – to act as definitive sources; their use is, as other expressions of what is not an easy topic to be made crisp and accurate.

49: General Characterization of Systematic and Stochastic Errors

Martin Halaj
Slovak University of Technology, Bratislava, Slovak Republic

1 Definition of a Systematic Error 295
2 Nature of a Systematic Error 296
3 Definition of a Random Error 297
4 Nature of a Random Error 298
5 Comparison of Systematic and Random Errors 299
6 Example of a Systematic Error Situation 300
7 Example of Random Error Situation 301
8 Errors and Measurement Result 302
Further Reading 303

1 DEFINITION OF A SYSTEMATIC ERROR

General definition states that error (of measurement) is a result of a measurement minus a true value of the measurand (see VIM, International Vocabulary of Basic and General Terms in Metrology, 3.10).

Since a true value cannot be determined, in practice a conventional true value is used. Measurement error consists of two parts, indicated as systematic and random error.

The same document defines systematic error as a mean that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions minus a true value of the measurand (see VIM 3.14).

Such definition implies that systematic error is equal to overall error minus random error.

Following the definition, let us consider n measurements of the same measurand X, carried out under the same conditions that are designated $x_1, x_2, \ldots, x_n$. Arithmetic mean $\bar{x}$ is the most used estimation of the measurement result

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

If x_{ctv} is the conventional true value of the measurand, absolute systematic error Δx is expressed as

$$\Delta x = \bar{x} - x_{\mathrm{ctv}} \tag{2}$$

Relative systematic error δ_x is calculated as

$$\delta_x = \frac{\Delta x}{x_{\mathrm{cvt}}} \tag{3}$$

Each particular measurement x_i from the series of n measurement is affected by the measurement error. Its systematic and random part is graphically represented in Figure 1.

General interest usually excludes systematic errors from the measurement results. In this situation, the VIM states the so-called uncorrected result of a measurement as being the measurement result before correction for systematic error (see VIM 3.3).

On the other hand, the corrected result of measurement is the result of a measurement after correction for systematic error (see VIM 3.4). Correction is then a negative value of the known part of the systematic error that must be added algebraically to the uncorrected result of measurement.

Like true value, systematic errors, and their causes cannot be completely known. Therefore, only an estimation of the

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

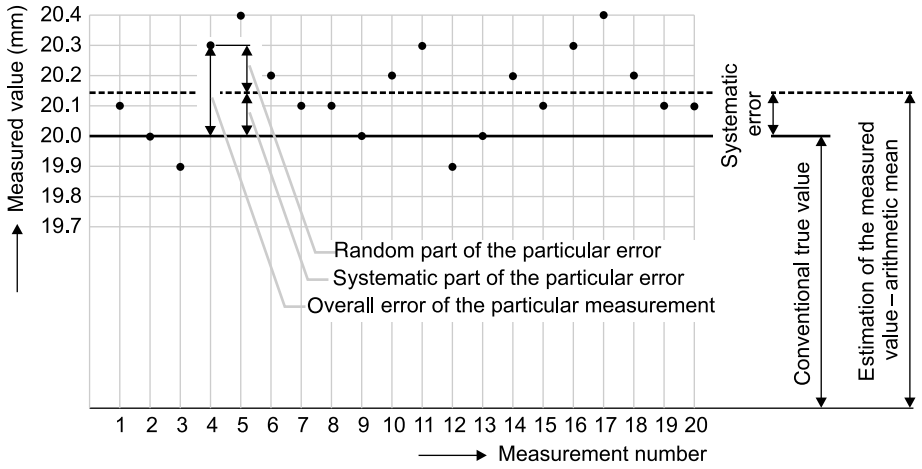


Figure 1. Graphical representation of the systematic error.

systematic error can be expressed. Such an estimation can be done by three ways:

1. analyzing the measurement process;
2. using available information on measuring instruments (they can be obtained from manufacturer's information, calibration certificates, etc.);
3. utilizing experiences from previous measurements.

Not all systematic errors can be estimated. Real systematic error can differ from that being estimated and subsequently being corrected. It appears in the measurement result in the form of the measurement uncertainty, usually being evaluated by the type B method (*see also* **Article 55, Uncertainty Determination, Volume 1**).

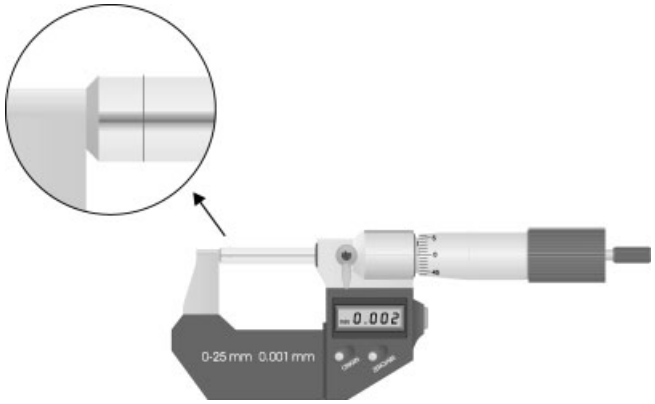


Figure 2. Micrometer with shifted zero position.

2 NATURE OF A SYSTEMATIC ERROR

Systematic error is caused by any factors that systematically affect measurement of the same measurand performed under the same conditions.

Unlike random error, systematic errors tend to be consistently either positive or negative. Because of this, systematic error is sometimes considered to be a bias in the measurement (*see also* **Article 50, Errors in Signal Systems, Volume 1**; and **Article 51, Errors in Digital Signal Systems, Volume 1**).

For instance, if the micrometer has a shifted zero position as shown in Figure 2, all measured dimensions are greater than the correct value. Subsequently, the mean of measured values exceeds the conventional true value, as shown in Figure 1.

As with all errors, systematic errors also arise from four main sources (*see also* **Article 48, Common Sources of Errors in Measurement Systems, Volume 1**; **Article 50, Errors in Signal Systems, Volume 1**; and **Article 51, Errors in Digital Signal Systems, Volume 1**):

1. *Measuring method*: improperly selected methods affect measurement the same way. For indirect measurements, rounded or uncorrected values of constants are often used, for example, $\pi = 3.141$ or improper gravity's acceleration rate g, and so on.
2. *Measuring instrument*: systematic error causes constant difference of the calibration curve from the characteristic curve.
3. *Measuring conditions*: measuring instruments being used in conditions differing from that specified by the manufacturer. A difference in temperature is the most common cause.
4. *Operator*: For example, wrong reading from the line scale. Use of digital indicators avoids this error.

Theoretical division of systematic errors states that two categories of errors exist – additive and multiplicative.

Additive errors are those values that do not change with the measured value, multiplicative errors change their value with change in measured value. The most common additive errors are offset errors and linearity errors; multiplicative error is presented as a gain error. Systematic error of the measuring instrument is manifested in most cases by a combination of those types of errors. Let us describe them in more detail:

1. *Offset error*: (see Figure 3a) the difference, under specified conditions of use, between the actual output value and the specified minimum value of the output range when the input is at the lower range value. Usually is expressed as a percentage of the specified span.
2. *Gain error*: (see Figure 3b) slope of the characteristic curve differs from that of the calibration curve. The absolute value of the gain error depends on the input value; the relative error remains the same.
3. *Linearity error*: (see Figure 3c) the absolute value of the maximum deviation between the calibration curve and the specified straight line. When expressed simply as the linearity error, it is assumed to be independent linearity error.

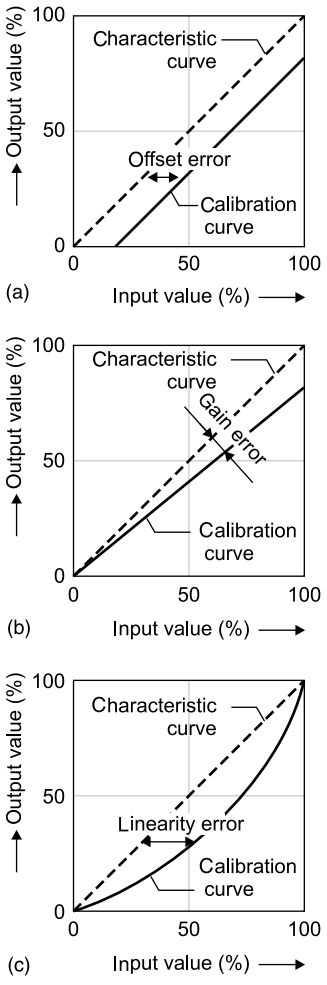


Figure 3. Systematic errors of the measuring instrument.

3 DEFINITION OF A RANDOM ERROR

The VIM defines random error as a result of a measurement minus the mean that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions (see VIM 3.13). Again, such definition implies that random error is equal to the error minus systematic error as shown in Figure 1.

Only a finite number of measurements can be made in practice; therefore, it is possible to determine only an estimate of random error. A finite number of measurements can be plotted in the so-called histogram (see Figure 4), utilizing data as shown in Figure 1.

One should observe the whole range of obtained data and then divide the x-axis into suitable even-width bins. The number of measured values that fit within particular intervals is plotted on the y-axis. This procedure gives the so-called frequency distribution of the measured values.

Let us imagine that a large number of measurements with high accuracy will produce a particular histogram with a smoother shape. For an infinite number of measurements and infinitesimally small intervals (bins), this histogram transforms to a smooth curve that can be described by a theoretical equation and be replaced by a probability distribution function (see Figure 5).

Measurement theory most often assumes that a normal Gaussian distribution, also known as a *bell curve*, can describe the occurrence of any given error (see Figure 6). This assumption is based on the central limit theorem

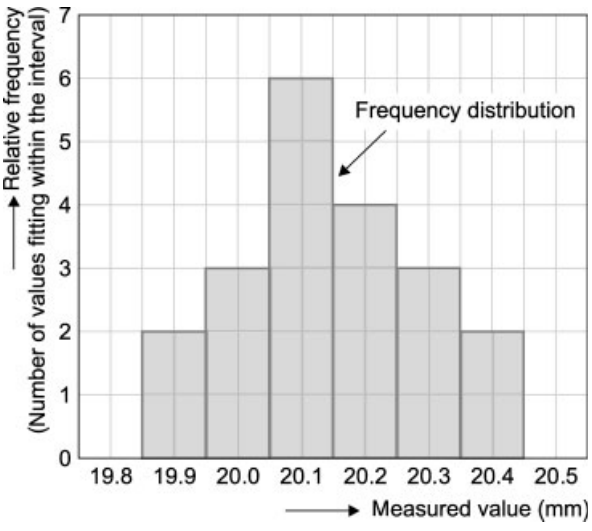


Figure 4. Histogram of the measured data.

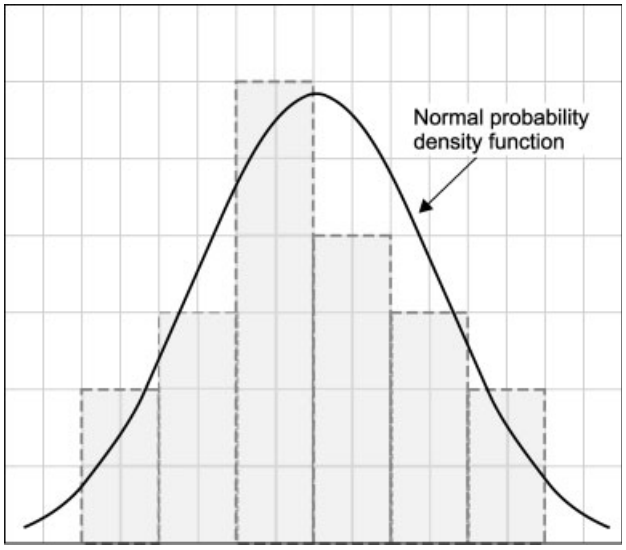


Figure 5. Replacing histogram by the probability distribution function.

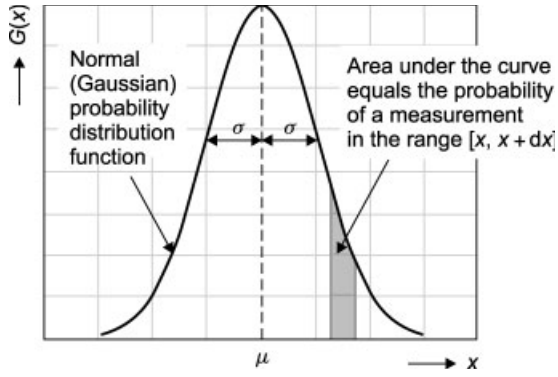


Figure 6. Gaussian distribution.

that does not necessarily always apply; its use must be confirmed by statistical tests.

The probability of finding a measurement in the range $[x,\ x + \mathrm{d}x]$ is equal to the area under the curve in that range. The curve is normalized to have a total area of 1, which is why its amplitude is not also a free parameter. The distribution is symmetric; an error is equally likely to occur in either direction. The equation describing this curve is

$$G(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right] \tag{4}$$

The Gaussian distribution has two independent parameters that completely determine the shape of the curve – the mean μ and the standard deviation σ. The standard deviation σ describes the width of the bell; the mean (x_{m}) lies on the axis of symmetry of the bell (see Figure 6). Both parameters are estimated by statistical methods – *see also* **Article 55, Uncertainty Determination, Volume 1**; **Article 52, Error Models, Error Budgets and their Calculation, Volume 1**; and **Article 53, Calculation and Treatment of Errors, Volume 1**.

4 NATURE OF A RANDOM ERROR

A random error is caused by any factor that randomly affects measurement of the measurand. The important thing about random error is that it does not have any consistent effect on the group of measured data. Instead, it pushes values obtained by measurement up or down randomly.

This means that if we could see all of the random errors in a distribution they would have to add up to 0 – there would be as many negative errors as positive ones. This introduces the important property of random error – it adds variability to the data but does not affect the average performance of the group – the mean (see Figure 9). Because of this, random error is sometimes considered as noise – *see also* **Article 48, Common Sources of Errors in Measurement Systems, Volume 1**; **Article 50, Errors in Signal Systems, Volume 1**; and **Article 51, Errors in Digital Signal Systems, Volume 1**.

The probability that a measured value appears in the intervals $\pm\sigma$, $\pm 2\sigma$, and $\pm 3\sigma$ is equal to the area bounded by the Gaussian curve within particular intervals – see Figure 7.

As stated before, the standard deviation σ describes the width of the bell, as shown in Figure 8. A higher standard

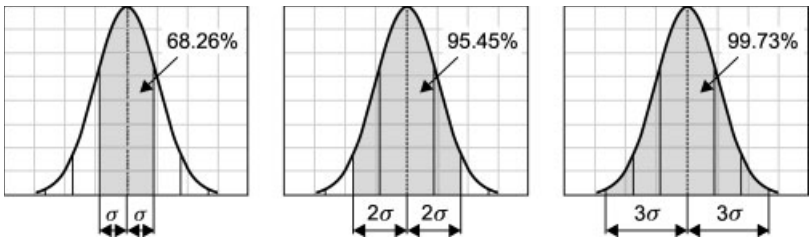


Figure 7. Probability of the presence of measured value within the range.

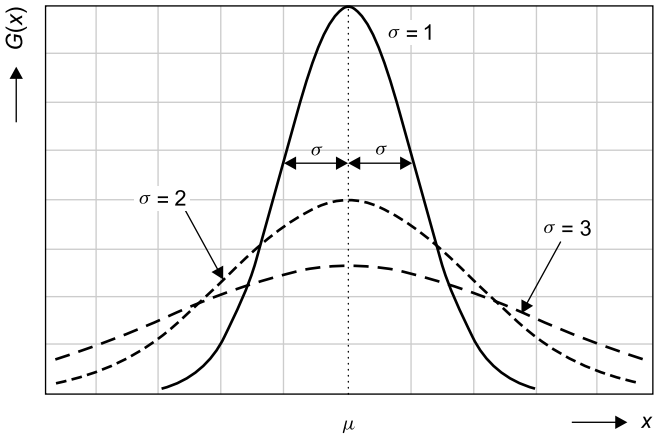


Figure 8. Influence of sigma on the bell curve.

deviation means large errors are more likely to exist. The mean μ (estimated usually as the arithmetic mean $\bar{x}$) lies on the axis of symmetry of the bell.

The mean (also called expectation, expected value) is defined as

$$\mu = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x \tag{5}$$

An estimation of the mean μ for the set of measured values $x_1, x_2, \ldots, x_n$ used in practice is the arithmetic mean $\bar{x}$ is given by

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{6}$$

The variance of the random variable X (measured quantity) is defined as

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,\mathrm{d}x \tag{7}$$

As can be seen from (7), the variance of the random variable is stated as the square of random variable units.

The empirical variance is the estimation of variance for the same set of measured values $x_1, x_2, \ldots, x_n$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{8}$$

The standard deviation, σ, obtained as a second root of the variance, is characteristic of measured data variability. Standard deviation σ is also designated as σ_X^2, $\sigma^2(X)$.

Estimation of the standard deviation for the set of measured data is the empirical standard deviation $x_1, x_2, \ldots, x_n$ defined as

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{9}$$

Variability characteristics are used as a basis for determination of the measurement uncertainty. The Type A method for the uncertainty evaluation, utilizing statistical methods, states that

$$u_{\mathrm{A}} = \sqrt{\frac{s^2}{n}} \tag{10}$$

where u_{A} – standard uncertainty evaluated by the Type A method – (*see also* **Article 55, Uncertainty Determination, Volume 1**).

5 COMPARISON OF SYSTEMATIC AND RANDOM ERRORS

Systematic errors are caused by any factors that systematically affect measurement of the same measurand performed under the same conditions. They tend to be consistently either positive or negative; therefore, they are sometimes considered to be the bias in measurement. Their value remains, under identical conditions, the same, or its changes can be predicted and evaluated.

On the other hand, random errors are caused by any factors that randomly affect measurement. They do not have any consistent effect on the group of measured data. Random errors add variability (estimated by the standard deviation σ) to the data but do not affect the average performance of the group (estimated by the mean μ). Therefore, random error is sometimes considered to be noise.

These facts are documented in Figure 9. Three different situations can occur:

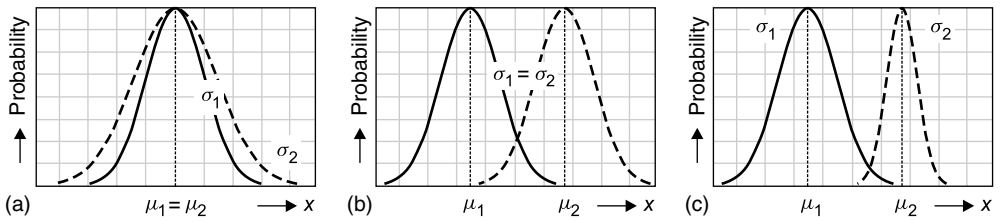


Figure 9. Influence of systematic and random errors.

1. Systematic errors affecting both measurements are the same; therefore, the mean of both data sets is the same. If their standard deviations differ, it points to a strong effect of random errors – see Figure 9(a).
2. Systematic errors differ; therefore, the mean of both data sets differs. Their standard deviations are the same so it points to a strong effect of systematic errors – see Figure 9(b).
3. Both systematic and random errors differ; therefore, the mean and standard deviations of both data sets differ. This is the most general case when both systematic and random errors affect measurements proportionally, as shown in Figure 9(c).

Determination of systematic errors is based on analysis of their sources – *see also* **Article 48, Common Sources of Errors in Measurement Systems, Volume 1**; **Article 50, Errors in Signal Systems, Volume 1**; and **Article 51, Errors in Digital Signal Systems, Volume 1**.

We tend to first determine systematic errors and subsequently either exclude them or correct them. Not excluded systematic errors have an unknown value so that they are estimated. Such estimation contains its own error sometimes called *nonexcluded systematic error*.

For a particular measurement, it has a random character with zero mean and certain variability. Determination and estimation of systematic errors can be a very complicated and difficult process. In practice, we often try to exclude only large systematic errors and keep the rest of them within certain limits.

Random errors are manifested in the case of repeated measurement and can be determined from measured data. They cannot be known so that they cannot be eliminated. We can, however, describe their stochastic behavior. The mean of random errors is zero; their variability is described by their standard deviation, which serves as a basis for evaluating their size.

6 EXAMPLE OF A SYSTEMATIC ERROR SITUATION

Systematic error is caused by any factors that systematically affect measurement under the same conditions. Their value (positive or negative) is either constant or can be calculated according to a known relationship.

As systematic errors can be estimated, their estimations can be subtracted from the measurement result so that the corrected measurement result can be obtained. While those errors are only estimated, and some of them we are not able to estimate, they occur in a measurement result in the form of measurement uncertainty.

All measurement is affected by some systematic errors. The tendency is to find their sources and evaluate their values before taking final measurements.

Systematic errors arise again from four main sources – *see also* **Article 48, Common Sources of Errors in Measurement Systems, Volume 1**; **Article 50, Errors in Signal Systems, Volume 1**; and **Article 51, Errors in Digital Signal Systems, Volume 1**:

1. *Measuring method*: sometimes uncorrected constants are used, for example, gravity's acceleration rate, air density, atmospheric pressure, and so on.
2. *Measuring instrument*: as described before, additive or multiplicative errors occur in measuring instruments. They usually result in (a) offset error, (b) gain error, or (c) linearity error. A combination of all these errors occurs very often.
3. *Measuring conditions*: they can differ from that specified by the manufacturer, standards, internal instruction, and so on. Temperature affects measurement in most cases; besides that, so does humidity, atmospheric pressure, and so on.
4. *Operator*: neglecting to set the measuring chain to zero before measurement, if required, introduces bias in all measurements.

Let us mention only a few examples of the sources of common systematic errors:

1. *Uncorrected constants*. Gravity's acceleration rate on 45° northern latitude, used as normal (standard) gravity's acceleration rate is $g = 9.806\,65\,\mathrm{m/s^2}$, 65° northern latitude is $g = 9.823\,03\,\mathrm{m/s^2}$.
   All measurement where g occurs can be affected if using an incorrect value. For example, measurement of the pressure using a U-tube filled with mercury, where the same difference in levels $h = 0.1\,\mathrm{m}$ is measured on both latitudes, differs by about 22.27 Pa (see Figure 10). This value represents an approximate relative error of 0.17% of the pressure measurement.
2. *Temperature effects*. Using the same example of a U-tube pressure measurement, let us measure the pressure at the same place, having the same difference between two mercury levels $h = 0.1\,\mathrm{m}$. The first measurement is performed at 0 °C, the second one at 20 °C. The difference between the two measured values at the two temperatures is approximately 48.52 Pa, resulting in relative error of 0.36%. Of interest is the fact that using water instead of mercury gives a relative error of 0.8% for the same measurement example.

The value of systematic error can be obtained by calibration – for example, the difference of the real weight mass

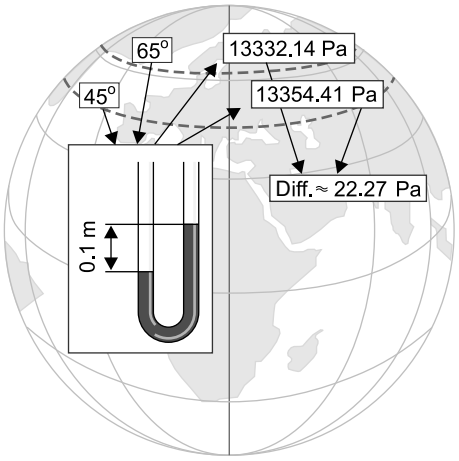


Figure 10. Example of the systematic error situation.

from the nominal value. Uncertainty of the systematic error estimation is also known.

7 EXAMPLE OF RANDOM ERROR SITUATION

The nature of random error implies that those errors are caused by any factors that randomly affect measurement – *see also* **Article 48, Common Sources of Errors in Measurement Systems, Volume 1**; **Article 50, Errors in Signal Systems, Volume 1**; and **Article 51, Errors in Digital Signal Systems, Volume 1**. A wide variety of such factors can occur in everyday life and, therefore, only a few examples of random error sources can be given here, those mainly experienced being as follows:

1. *Measuring method*: improperly selected method and measuring instrument for the intended measurement. For example, a level position of foamed liquid cannot be measured by the ultrasonic ranger; another type of measuring principle must be selected that avoids the effect of foam on measurement results.
2. *Measuring instrument*: wrong or defective parts of measuring chain. Simple examples are defective contact surfaces of micrometer, transient phenomena in electronic parts of the measurement chain, and so on.
3. *Measuring conditions*: the measuring instrument is used in conditions that vary among repeated measurements – temperature, humidity, atmospheric pressure, and so on. For example, the ultrasonic distance sensor is very sensitive to varying measurement conditions.
4. *Operator*: the classic and often seen source of random measurement errors. Usually a wrong reading out from the line scale occurs, followed by a different use of the measuring instrument in the repeated measurement (different positioning of the micrometer on the shaft, different depth of the thermometer in the liquid, etc.).

Figure 11 can be used as a slightly artificial example of the random errors sources when measuring product length with a simple caliper. The measuring method generates random errors by incorrect positioning of the measured object. Some kind of fixture should be used to assure the

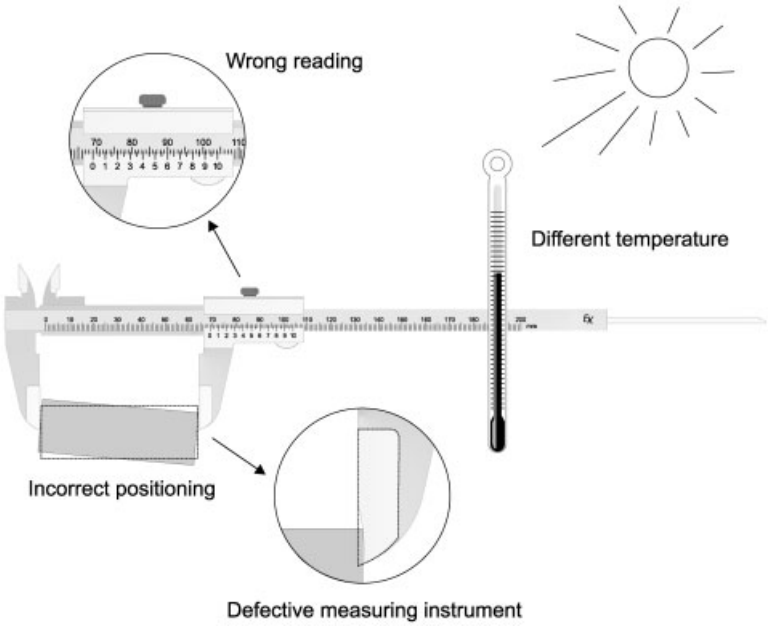


Figure 11. Example of the random error sources.

same position issued for each measured object. Wrong or defective jaws of the caliper contribute to random errors of the measuring instrument.

When positioning the measured object and caliper, different parts of the caliper jaws are in contact with the measured object.

Measuring conditions changing between the individual measurements can affect the measurement results significantly. Temperature changes have a strong influence in cases in which thermal expansion coefficients of the measured object and the measuring instrument differ significantly.

The operator usually contributes to random errors. Reading out the value of scale can depend strongly on the operator's skills and mood. This can be avoided by using a digital display readout device; however, the digital measuring device might introduce random errors caused by resolution of the displaying device.

When measurement is performed once, or only a few times, random error is estimated by a different way – not by the type A method. For example, in the case of single measurement, all errors and their uncertainties are estimated by the type B method. Therefore, uncertainties of the random errors must also be evaluated by the type B method.

8 ERRORS AND MEASUREMENT RESULT

Measurement result is given as a measured value (estimated) minus measurement error. If the errors could be known perfectly, a measurement result as defined above would be the final result. As the errors cannot be determined totally, but only their estimations are known, a measurement result must comprise also of the characteristics of the quality of such estimation. Error estimation is considered as a random quantity with an estimation of its quality expressed as standard deviation. Metrology uses the term measurement uncertainty (*see also* **Article 54, Explanation of Key Error and Uncertainty Concepts and Terms, Volume 1**). Standard measurement uncertainty is equal to standard deviation.

Random errors are estimated as zero and standard deviation of the probability distribution of random errors is equal to standard measurement uncertainty caused by random error. Normal distribution of the random error is assumed in most cases.

Correctness of such an assumption can be verified by application of appropriate statistical tests. If repeated measurements under the same conditions are available, the measurement result will be given as the arithmetic mean of the measured value and the uncertainty will be given as an experimental standard deviation of the arithmetic mean.

If only one measurement is available, standard deviation is determined on the basis of other information on probability measurement of the random error (e.g. on the basis of previous measurements).

Estimations of systematic errors are based on analysis of the measuring process (measuring system, measurement conditions, measuring method, influence of operator, etc.). Uncertainties characterizing those estimations are determined on the basis of process analysis and different information obtained from the manufacturer of the measuring system, calibration certificates, literature, and so on.

The following sample situation occurs often in practice (*see also* **Article 55, Uncertainty Determination, Volume 1**):

1. Limits $\pm z_{\max}$ around the estimation of the systematic error Δ can be determined. Values of the systematic errors fit within this interval. The standard uncertainty $u(\Delta)$ of the systematic error Δ is then calculated as

$$u(\Delta) = \frac{z_{\max}}{k} \tag{11}$$

   where k is the value corresponding to the selected approximation of the probability distribution (see Figure 12).
2. When using a digital measuring instrument, resolution of a displaying device represents one possible error source. Even if the values in the case of repeated measurements do not change, uncertainty is nonzero. Uncertainty estimation is based on assumed uniform distribution

$$u(z) = \frac{\delta(z)}{2\sqrt{3}} = 0.29\delta(z) \tag{12}$$

   where $\delta(z)$ – resolution of the measuring device.

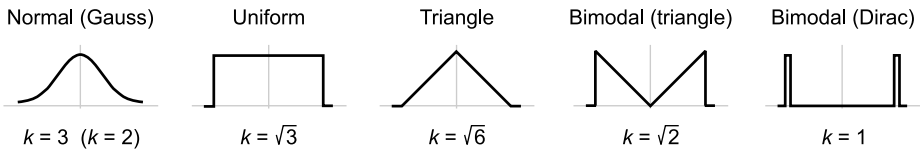


Figure 12. Probability distributions and corresponding k values.

3. When using an analog measuring instrument, the scale division δ_z is used instead of resolution of the displaying device. The uncertainty is calculated using (12).

Measurement result is then stated as

$$x = x^* \Delta \pm U \tag{13}$$

where U – expanded uncertainty is defined as

$$U = ku \tag{14}$$

where k – expansion coefficient – *see also* **Article 55, Uncertainty Determination, Volume 1**.

FURTHER READING

ISO 2602 (1980) *Statistical Interpretation of Test Results – Estimation of the Mean – Confidence Interval*.

ISO 2854 (1976) *Statistical Interpretation of Data – Techniques of Estimation and Tests Relating to Means and Variances*.

ISO 3207 (1975) *Statistical Interpretation of Data – Determination of a Statistical Tolerance Interval*.

ISO 3301 (1975) *Statistical Interpretation of Data – Comparison of Two Means in the Case of Paired Observations*.

ISO 3494 (1976) *Statistical Interpretation of Data – Power of Tests Relating to Means and Variances*.

ISO 3534-1 (1993) *Statistics – Vocabulary and Symbols – Part 1: Probability and General Statistical Terms*.

ISO 3534-2 (1993) *Statistics – Vocabulary and Symbols – Part 2: Statistical Quality Control*.

ISO 3534-3 (1999) *Statistics – Vocabulary and Symbols – Part 3: Design of Experiments*.

ISO 5479 (1997) *Statistical Interpretation of Data – Tests for Departure from the Normal Distribution*.

ISO 8595 (1989) *Interpretation of Statistical Data – Estimation of a Median*.

ISO/TR 13425 (1995) *Guide for the Selection of Statistical Methods in Standardization and Specification*.

ISO 16269-7 (2001) *Statistical Interpretation of Data – Part 7: Median – Estimation and Confidence Intervals*.

EA4/02, EA (1999) *Expression of the Uncertainty of Measurement in Calibration*.

BIPM/IEC/ISO/OIML (1993) *Guide to the Expression of Uncertainty in Measurement* (GUM).

BIPM/IEC/ IFCC/ISO/OIML/IUPAC/IUPAP (1993) *International Vocabulary of Basic and General Terms in Metrology* (VIM).

ISO/IEC Guide 2 (1991) *General Terms and their Definitions Concerning Standardization and Related Activities*.

50: Errors in Signal Systems

Eugen Georg Woschni

Technical University Chemnitz, Chemnitz, Germany

1 First Section Heading	304
2 Errors in Dynamic Measurements	305
3 Typical Wave-shape Distortions	305
4 Statistical Errors	306
5 Errors in Optimal Filtering	309
6 An Example	310
References	313

1 FIRST SECTION HEADING

As Figure 1 shows, the error ε is given by the difference between the output of the real system y_{real} and the ideal system without errors y_{id}

$$\varepsilon = y_{\text{id}} - y_{\text{real}} \tag{1a}$$

Equation (1a) may be interpreted either in the time or the frequency domain. In the time domain, it leads to a time error $\varepsilon(t)$. Of special interest is the mean-square value of this error

$$\overline{\varepsilon^2} = \lim_{t\to\infty} \frac{1}{2T} \int_{-\infty}^{+\infty} \varepsilon^2(t)\,\mathrm{d}t \tag{1b}$$

or in the frequency domain, leading to an amplitude error $|G_\varepsilon(\mathrm{j}\omega)|$

$$|G_\varepsilon(\mathrm{j}\omega)| = |G_{\text{id}}(\mathrm{j}\omega)| - |G_{\text{real}}(\mathrm{j}\omega)| \tag{1c}$$

with a phase error of

$$\varphi_\varepsilon(\mathrm{j}\omega) = \varphi_{\text{id}}(\mathrm{j}\omega) - \varphi_{\text{real}}(\mathrm{j}\omega) \tag{1d}$$

Because of the mean operation, the definition (1b) does not contain the whole information of (1c, d). It is, thus, possible to calculate the mean-square value $\overline{\varepsilon^2}(t)$ if the amplitude and phase errors are known, but the inverse operation is not possible.

As shown in **Article 30, Statistical Signal Representations, Volume 1**, noise signals z lead to statistical errors characterized by the spectral power density $S(\omega)$, or the mean-square value that is $\overline{z^2}$, related to the standard deviation σ by the relation

$$\overline{z^2} = \sigma^2 = \int_{-\infty}^{+\infty} S(\omega)\,\mathrm{d}\omega \tag{1e}$$

Noise is assumed to be generated at the input of the real system $S_\chi(\omega)$, as shown in Figure 1. The mean-square value of the noise at the output is then the statistical error (Woschni, 1988)

$$\overline{z}^2_{\text{output}} = \int_{-\infty}^{+\infty} S(\omega)|G_{\text{real}}(\mathrm{j}\omega)|^2\,\mathrm{d}\omega \tag{1f}$$

The total error $\overline{\varepsilon}^2_{\text{tot}}$ is the sum of both error components when there is no correlation between both error components. Otherwise, correlation has to be taken into consideration. This point is often overlooked.

$$\overline{\varepsilon}^2_{\text{tot}} = \overline{\varepsilon}^2(t) + \overline{z}^2_{\text{output}} \tag{1g}$$

Additionally, there may arise aliasing errors if the sampling theorem is violated. These are investigated in **Article 31, Discrete Signal Theory, Volume 1**.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.

© 2005 John Wiley & Sons, Ltd.

ISBN: 0-470-02143-8.

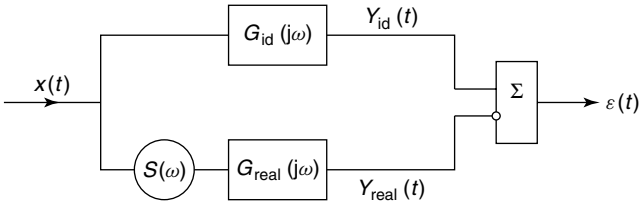


Figure 1. Error definition.

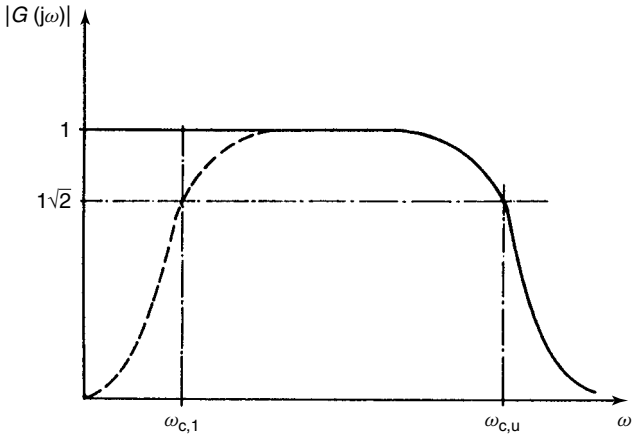


Figure 2. Definition of the critical frequencies.

2 ERRORS IN DYNAMIC MEASUREMENTS

For a description of the dynamic behavior of systems, it is usual and useful to introduce characteristic values that can be gained by means of approximations from the functions in the frequency and time domains (see Figure 2). The amplitude–frequency characteristic, Figure 2, yields the upper and lower critical frequencies $f_{c,u}$ and $f_{c,l}$ or $\omega_{c,u}$ and $\omega_{c,l}$, where $|G(j\omega)|$ decreases to $\sqrt{2} = 0.7 \triangleq 3\,\text{dB}$ of the reference value. From the transient response $h(t)$ (see Figure 4) the important transient time t_{tr}, dead time t_d, delay time t_l, compensation time t_c, and the maximum overshoot Δx_0 are obtained.

The transient time is given approximately as (*see* **Article 35, Systems in the Time Domain, Volume 1**)

$$t_{tr} = 3T_{max} \tag{2a}$$

From the sampling theorem, the transient time and the critical frequency are related by (Woschni, 2000)

$$f_{c,u} = \frac{1}{2} t_{tr} \tag{2b}$$

In measuring systems engineering, these approximate considerations are of importance as convenient ways to estimate performance, for example, as in the selection of suitable measuring instruments for a given task.

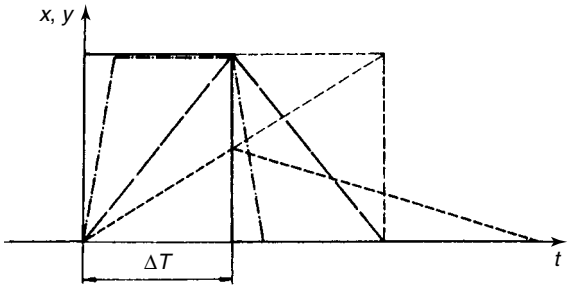


Figure 3. Approximation of the output for the case of a pulse-shaped input – pulse-shaped input $t_{tr} = \Delta T$ - - - -, transient time too long -----, $t_{tr} < \Delta T$ - - - - -.

Consider the following example. Figure 3 illustrates the problem of measuring a pulse-shaped input function (solid line), an approximation for numerous measuring tasks (Woschni, 1972). In this figure, three cases are presented. The long, dashed, triangular shaped response represents the example in which the transient time is equal to the pulse width, that is $t_{tr} = \Delta T$. The pulse height is still correctly indicated, whereas the pulse shape is strongly distorted. In contrast to this particular case, considerable error is introduced when measuring the pulse height if the transient time is too long (short dashes). To permit proper determination of the pulse shape, the transient rise time has to be substantially shorter than the pulse width (chain-dash curve).

The same considerations may also be useful for analyzing measuring errors. If one expects, for instance, a pulse-shaped behavior of the output and receives, as the response, an output variable with a heavily prolonged trailing edge, then an error will be present (short dashes in Figure 3).

It is noteworthy that minor and medium errors in amplitude measurement frequently produce more detrimental effects in practice than do large measuring errors. For errors of some 10%, the machine-made circuit element that has been sized according to this measurement will function for a certain time due to the safety margins, and will often withstand initial testing. After having been produced in series and operated for some time, however, all elements fail at the same point according to the fatigue curve for the number of stress reversals possible up to the failure of the element.

On the other hand, major measuring errors will mostly become evident during the testing period.

3 TYPICAL WAVE-SHAPE DISTORTIONS

Table 1 gives a survey of the more important measuring systems and their characteristic functions in both the time

and frequency domains (Woschni, 1988). The results in this table, in principle at least, may be used to find optimal parameters of a system. For example, the best damping of the spring–mass damping system, typical of a great number of measuring systems, can be seen to be approximately 1 (precise value 0.7). The transfer functions of typical systems, as illustrated in Table 1, allow the user to gain a survey of typical curve distortions and their causes using the methods for approximations given in Section 2 of this article. These considerations lead to the results summarized in Table 2. In these cases, the input function is assumed to be a pulse-shaped curve.

4 STATISTICAL ERRORS

As treated in **Article 30, Statistical Signal Representations, Volume 1**, the sources of noise may be described by the Gaussian distribution density $w(x)$ (Jaffe, 2000) as shown in Figure 4.

With the standard deviation $\sigma = \sqrt{x^2(t)}$ it yields

$$w(x) = \frac{1}{\sqrt{(2\pi)\sigma}} \exp\left(\frac{\exp\left(\frac{-x^2}{2\sigma^2}\right)}{2\sigma^2}\right) \tag{3a}$$

Table 1. Survey of typical systems and their characteristic functions.

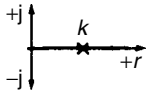
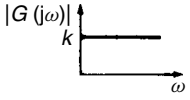
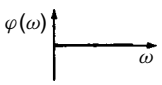
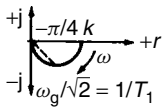
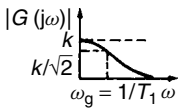
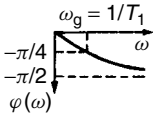
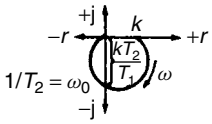
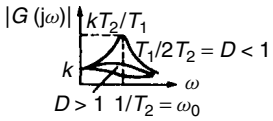
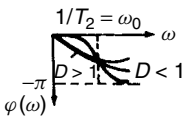
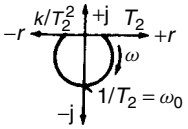
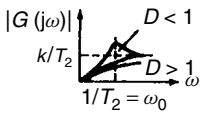
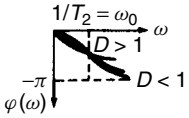
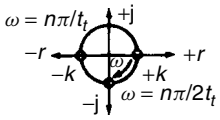
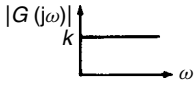
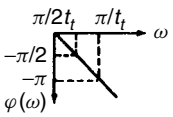
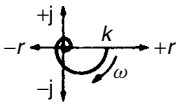
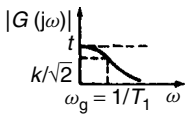
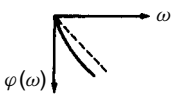
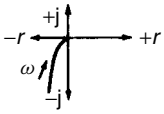
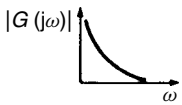
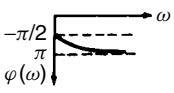
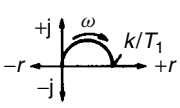
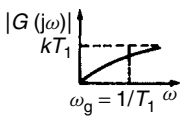
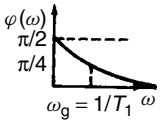
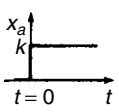
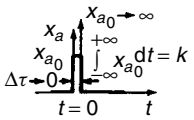
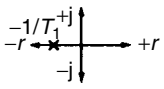
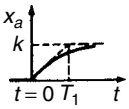
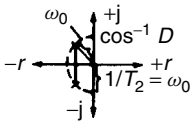
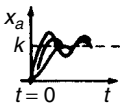
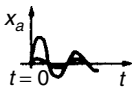
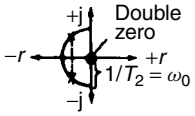
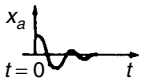
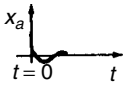
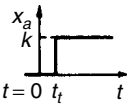
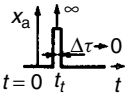
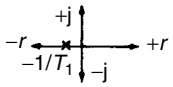
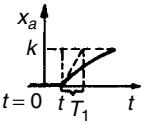
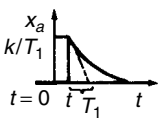
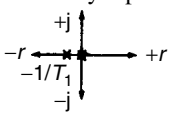
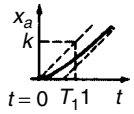
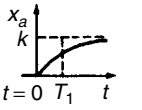
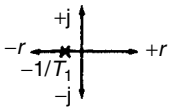
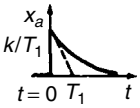
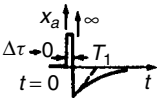
Mathematical formulation	Transfer locus	Amplitude phase curve	
k			
$\frac{k}{1+T_1p}$			
$\frac{k}{1+T_1p+T_2^2p^2}$			
$\frac{-kp^2}{1+T_1p+T_2^2p^2}$			
e^{-t_tp}			
$\frac{ke^{-t_tp}}{1+T_1p}$			
$\frac{k}{p}\frac{1}{1+T_1p}$			
$\frac{kp}{1+T_1p}$			

Table 1. (*continued*).

Poles (×) and zeros (0)	Function of time		Examples for systems with such behavior
	Transfer function $x_a/1 = h(t)$	Weighting function $a_a/1 = g(t)$	
None			Systems with proportional behavior (idealized)
			Temperature-measuring arrangements without protective tube; very heavily damped systems capable of vibrating with proportional behavior (idealized); system with delay of the first order and compensation
			Spring-mass damping system, $D \leq 1$, temperature-measuring arrangements with protective tube; $D < 1$, approximation behavior for vibration systems
			Spring-mass damping system without fixed point $D \leq 1$
Cannot be exactly represented.			System with pure dead time (idealization), caused, e.g., by pipeline, transport path, etc.
			Real systems with dead time, e.g., temperature-measuring arrangement with heat conductivity feed to transducer
Factor $e^{-t_t p}$ cannot be exactly represented.			
			Real integrator
			Real differentiator ($x_a \hat{\approx} \dot{x}_e$)

If this stochastic signal is passing through a linear system (Figure 5), the distribution is not changed. The frequency response of this system $G(\mathrm{j}\omega)$ influences the mean-square value of the noise at the output $P_y = \overline{y^2}(t)$ where $S_x(\omega)$ is the spectral power density of the input signal, as shown in **Article 27, Signals in the Frequency Domain, Volume 1** (Woschni, 1988; Jaffe, 2000)

$$P_y = \int_{-\infty}^{+\infty} S_x(\omega)|G(\mathrm{j}\omega)|^2 \,\mathrm{d}\omega \qquad (3b)$$

An important example is the calibration of a white noise source as shown in Figure 6. To find out the unknown value of S_0, a system of first order $G(\mathrm{j}\omega) = 1/(1 + \mathrm{j}\omega RC)$ is used – a so-called shaping filter – leading to the output power P_y

$$P_y = \int_{-\infty}^{+\infty} S_0 \frac{1}{1 + \omega^2 R^2 C^2} \,\mathrm{d}\omega = S_0\, \omega_0 \pi \qquad (3c)$$

If the output power P_y is measured, the unknown spectral power density S_0 can be calculated

$$S_0 = \frac{P_y}{\omega_0 \pi} \qquad (3d)$$

Table 2. Typical wave-shape distortions and their causes.

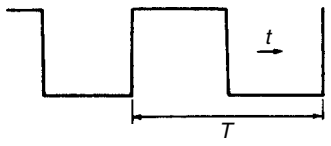
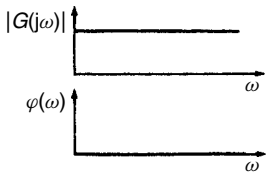
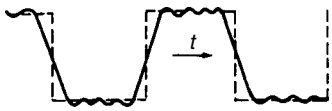
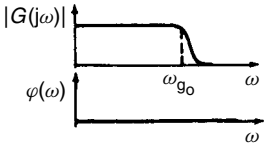
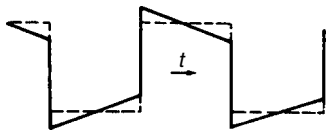
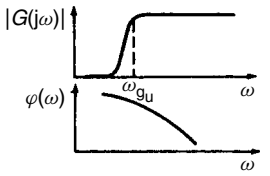
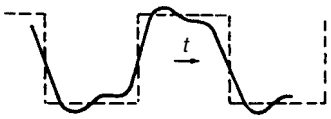
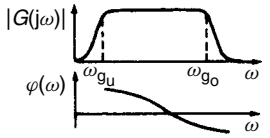
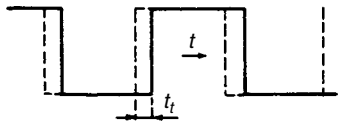
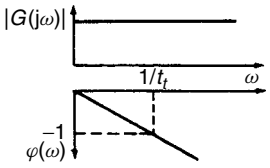
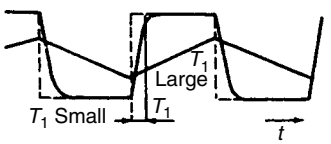
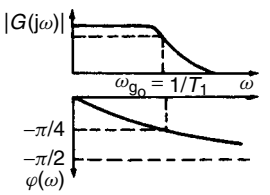
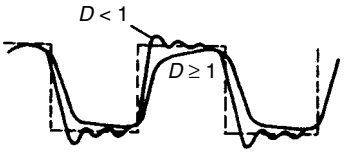
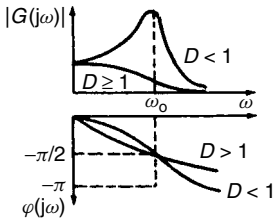
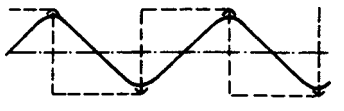
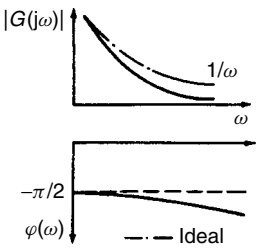
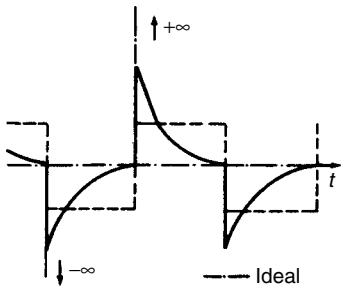
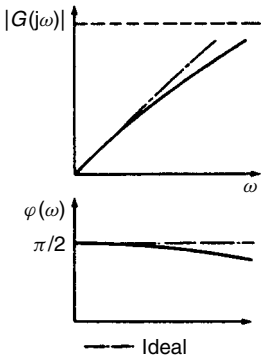
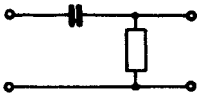
----- Output variable $x_a(t)$ —— Input variable $x_e(t)$ Curve shape	Typical features	Causes: amplitude curve, phase curve $\lvert G(j\omega)\rvert; \varphi(\omega)$	Typical features	Remarks
 1	Sharp corners ⌐ Steel edges ⌐ Straight top ¬		Ideal system, i.e. $f_{g_u} = 0$; $f_{g_o} = \infty$ ideal phase curve	Practically not fully satisfied; only approximately for slow process (T great)
 2	Blurred corners ⌐ Oblique edges ⌐ Straight top ⌒		Upper limiting frequency f_{g_o} too low; lower limiting frequency $f_{g_u} = 0$ ideal phase curve	Short peaks of the input function are not reproduced or are reproduced with wrong height; remedy: to increase upper limiting frequency, e.g. by reducing masses (idealized system)
 3	Sharp corners ⌐ Steep edges ⌐ Inclined top ¬		Lower limiting frequency f_{g_u} too high; upper limiting frequency $f_{g_o} = \infty$	Short peaks are correctly indicated; the error is increasing the faster the process runs; pure static calibration and measurement not possible (e.g. piezoelectric measuring device)
 4	Blurred corners ⌒ Oblique edges ∫ Inclined top ¬		Upper and lower limiting frequencies incorrect	Insufficiencies according to 2 and 3, i.e. short peaks wrongly reproduced; no transfer of static variables
 5	Curve shape correctly reproduced, only entire curve shifted by T_t		Ideal amplitude curve linear phase curve, i.e. pure dead time	With the exception of a time shift, no alteration of the output variable; time shift is often insignificant in measurement engineering; is often detrimental, however, in control engineering; idealized case: with $G(j\omega) =$ constant additional distortions as 2,3,4
 6	Curve consists of pure e-functions with time constant T_1 Oblique edges ∫ Straight top ⌒		Upper limiting frequency too low	Short peaks are not correctly reproduced; heavily damped systems with proportional behavior (idealized)

Table 2. (*continued*).

----- Output variable $x_a(t)$ —— Input variable $x_e(t)$ Curve shape	Typical features	Causes: amplitude curve, phase curve $\lvert G(j\omega)\rvert; \varphi(\omega)$	Typical features	Remarks
 7	For $D > 1$: blurred corners ⌜ Oblique edges ∫ For $D < 1$: oscillation		Upper limiting frequency too low; additionally for $D < 1$: resonance behavior	Short peaks are not correctly reproduced; $D \gtrless 1$: systems with spring–mass damping behavior, e.g. vibration meters $D > 1$: system for temperature measurements with protective tube
 8	Integral curve exhibits no corners at points where x_e jumps integrates correctly at low frequencies only		Integrator shows nonconstant phase rotation by $-\pi/2$; amplitude curve decreases more strongly than with $1/\omega$; has upper limiting frequency for integration	With nonelectric integrators, the influence of the masses is decisive for this effect; upper limiting frequency too low in electric techniques
 9	Differential curve rounded. Differentiates correctly at low frequencies only		Differentiator shows nonconstant phase rotation by $+\pi/2$; amplitude curve not ideal; has an upper limiting frequency f_{g_o} for differentiation	Unavoidable with simple CR component $T_1 = CR$ 

5 ERRORS IN OPTIMAL FILTERING

Gauss introduced the measure of the mean-square error $\overline{\varepsilon^2}$, because this definition leads to closed mathematic solutions as for instance with optimal filtering. As shown in Section 5 of **Article 38, Stability Issues, Volume 1**, the total mean-square error $\overline{\varepsilon^2}$ consists of two components.

As shown in Section 5 of **Article 38, Stability Issues, Volume 1**, a correction or compensation system is series connected with the real system; error can arise due to it having insufficient dynamic behavior $\overline{\rho^2}$ and to the error due to the disturbance P_z. If there is no correlation between signal source and disturbance source, as has been assumed here, the two error components must be added to arrive at the total error.

From 5 the total mean-square error is given by

$$\begin{aligned}\overline{\varepsilon^2} &= \int_0^\infty S_{xx}(\omega)|G_{\mathrm{id}}(\mathrm{j}\omega) - G_{\mathrm{real}}(\mathrm{j}\omega)G_{\mathrm{corr}}(\mathrm{j}\omega)|^2\,\mathrm{d}\omega \\ &\quad + \int_0^\infty S_{zz}(\omega)|G_{\mathrm{corr}}(\mathrm{j}\omega)|^2\,\mathrm{d}\omega \\ &= \overline{\rho^2} + P_z \end{aligned} \tag{4}$$

with the assumption having been made that the disturbances occur in the system prior to the correction computer.

When the error components and total error are plotted against the correction degree a, relationships will always result as presented in Figure 7. The dynamic error component $\overline{\rho^2}$ decreases as a increases and disappears for $a \to \infty$ (ideal correction), whereas the error component due to the disturbance increases with a. Figure 8 provides the

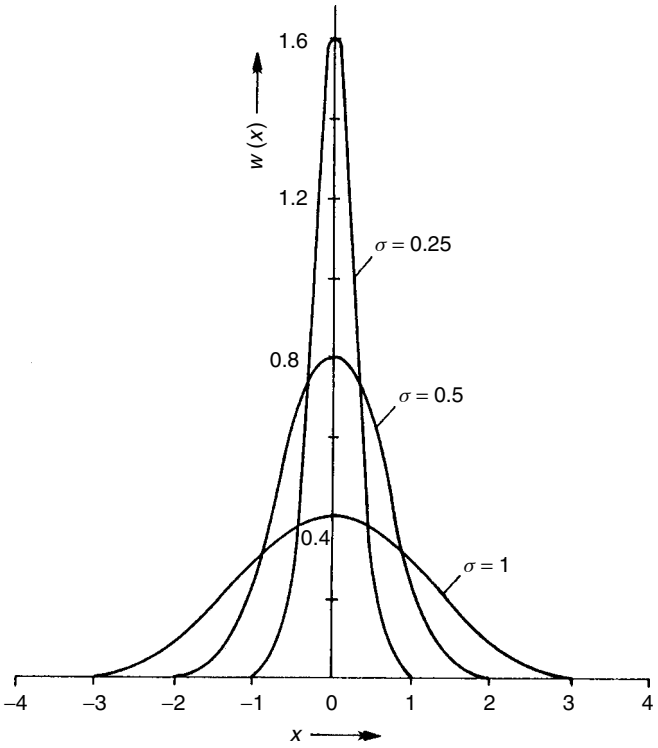


Figure 4. Gaussian distribution density.

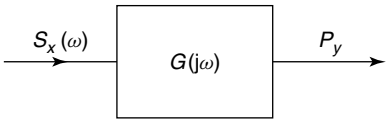


Figure 5. Calculating the noise at the output of a system.

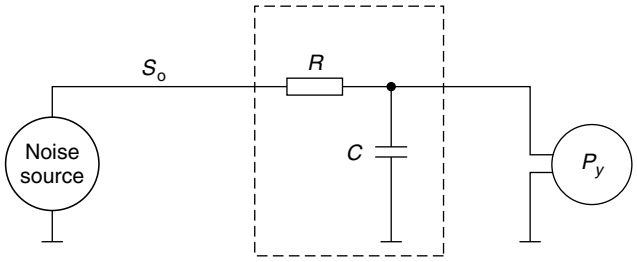


Figure 6. Calibration of a noise source.

physical demonstrative explanation for this. In this figure, the amplitude responses $|G_{\text{real}}(\text{j}\omega)|$ of the uncorrected system as well as of the correction system $|G_{\text{corr}}(\text{j}\omega)|$ are plotted.

To compensate for the decrease of the uncorrected system at frequencies above the limiting frequency $\omega_{\text{c,real}}$, the correction system must raise these frequencies correspondingly up to its limiting frequency $\omega_{\text{c,corr}}$. However, this inevitably causes the spectral frequencies of the disturbances that fall into this range to be raised so that the error component caused by the disturbance becomes the greater the higher the value selected for $\omega_{\text{c,corr}}$. Therefore, there will always be a minimum for the total error $\overline{\varepsilon^2} = \overline{\rho^2} + P_z$, this minimum being deeper and more pronounced the lower the inherent disturbances of the uncorrected system are, the better the dynamic behavior of this system will be. This minimum corresponds to the case of an optimum filter according to Schlitt (1960) and Levine (1996).

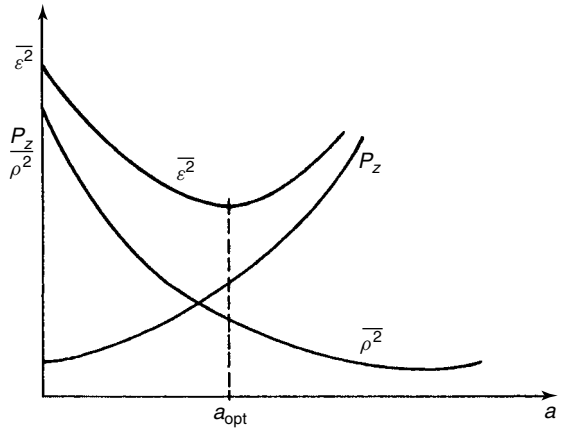


Figure 7. Behavior of the components of the error as a function of the degree of correction α.

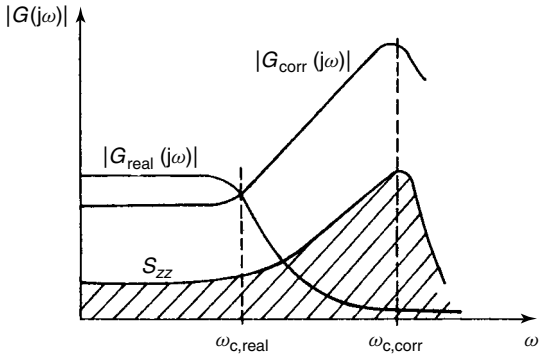


Figure 8. Physical explanation of the increase of the noise in case of correction.

6 AN EXAMPLE

An example considers the first-order system treated in **Article 35, Systems in the Time Domain, Volume 1** and **Article 37, Relationship Between Systems in the Time and Frequency Domain, Volume 1**. The frequency response there is

$$G(\text{j}\omega) = \frac{1}{1 + \text{j}\omega T} \tag{5a}$$

Let S_{x_0} be a white-structure signal applied to the input, let the band be limited to ω_x and let white noise S_{z_0} be the type of interference. Furthermore, no correlation is assumed between the signal and the interference.

The correction program of a series-connected system that can be realized, in this case reads

$$G_{\mathrm{k}}(\mathrm{j}\omega) = \frac{1 + \mathrm{j}\omega T}{1 + \mathrm{j}\omega T_{\mathrm{k}}} \tag{5b}$$

of which the factor

$$\frac{T}{T_{\mathrm{k}}} = \frac{\omega_{\mathrm{c,k}}}{\omega_{\mathrm{c}}} \tag{5c}$$

indicates the improvement of the bandwidth by use of the correction. If a microcomputer is used for the realization of (5b), the limit frequency of the digital conversion, with the sampling time t_{s}, is

$$\omega_{\mathrm{c,s}} = \pi / t_{\mathrm{s}} \tag{5d}$$

Thus, the mean-square error is

$$\overline{\varepsilon^2} = S_{x_0}\left(\int_0^{\omega_{\mathrm{c,s}}} \frac{\omega^2 T_{\mathrm{k}}^2}{1 + \omega^2 T_{\mathrm{k}}^2}\,\mathrm{d}\omega + \omega_x - \omega_{\mathrm{c,s}}\right) + S_{z_0}\int_0^{\omega_{\mathrm{c,s}}} \frac{1 + \omega^2 \mathrm{T}^2}{1 + \omega^2 T_{\mathrm{k}}^2}\mathrm{d}\omega \tag{5e}$$

In order to obtain values that can be compared with the original system and enable an estimate of the efficiency of the correction to be generated, we calculate the error of the system without any correction while assuming a bandwidth limited to $\omega_{\mathrm{c}} = 1/T$:

$$\overline{\varepsilon_0^2} = S_{x_0}\left(\int_0^{\omega_{\mathrm{c}}} \left|1 - \frac{1}{1 + \mathrm{j}\omega T}\right|^2 \mathrm{d}\omega + \int_{\omega_{\mathrm{c}}}^{\omega_x} \mathrm{d}\omega\right) + S_{z_0}\int_0^{\omega_{\mathrm{c}}} \mathrm{d}\omega = S_{x_0}\,\omega_{\mathrm{c}}\left(\frac{\omega_x}{\omega_{\mathrm{c}}} - \frac{\pi}{4} - \frac{S_{z_0}}{S_{x_0}}\right) \tag{5f}$$

To find a more favorable solution in each case, and thus obtain suggestions for synthesis, consider the following two cases: (i) the limiting frequency $\omega_{\mathrm{c,s}}$, that is, the sampling frequency, be adapted to the limiting frequency of the corrected system $\omega_{\mathrm{c,k}}$; and (ii) the limiting frequency $\omega_{\mathrm{c,s}}$ be adapted to the limiting frequency of the signal ω_x:

1. $\omega_{\mathrm{c,s}} = \omega_{\mathrm{c,k}}$. From (5e), for $\omega_{\mathrm{c,s}} = \omega_{\mathrm{c,k}} < \omega_x$, one obtains

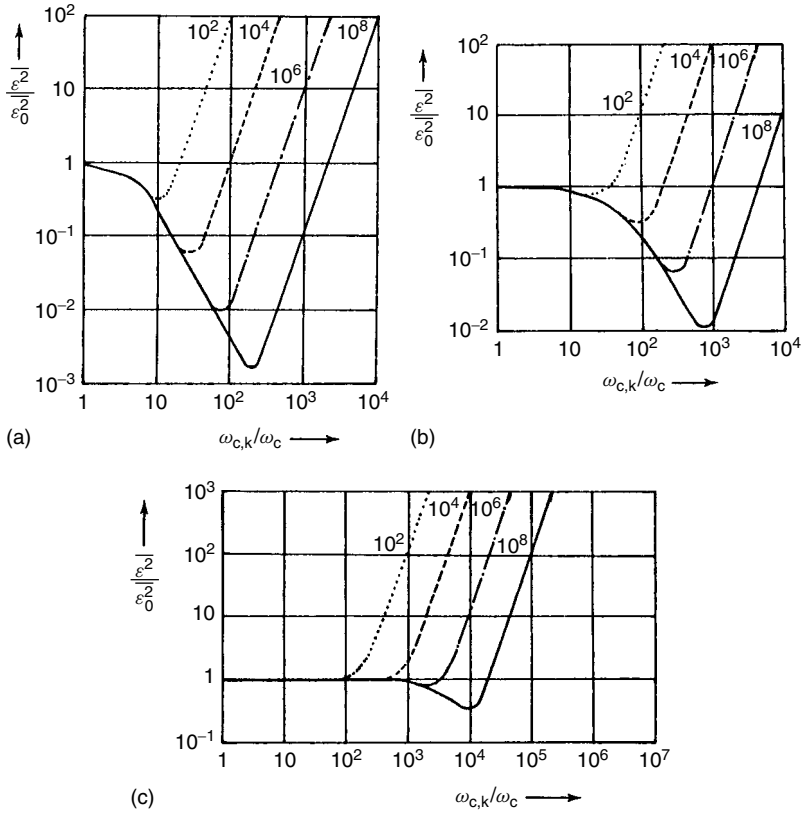


Figure 9. Dependence of relative mean-square error on the bandwidth improved by correction: $S_{x_0}/S_{z_0} = 10^2$; - - - - - $S_{x_0}/S_{z_0} = 10^4$; ·-·-·-·- $S_{x_0}/S_{z_0} = 10^6$; ——— $S_{x_0}/S_{z_0} = 10^8$; (a) $\omega_x/\omega_{\mathrm{c}} = 10$; (b) $\omega_x/\omega_{\mathrm{c}} = 10^2$; (c) $\omega_x/\omega_{\mathrm{c}} = 10^4$.

$$\overline{\varepsilon^2} = S_{x_0}\,\omega_{c,k}\left[\left(1+\frac{S_{z_0}}{S_{x_0}}\frac{\omega_{c,k}^2}{\omega_c^2}\right)\left(1-\frac{\pi}{4}\right) + \frac{\omega_x}{\omega_{c,k}} - 1 + \frac{S_{z_0}}{S_{x_0}}\frac{\pi}{4}\right] \tag{5g}$$

and for $\omega_{c,s} = \omega_{c,k} > \omega_x$, respectively,

$$\overline{\varepsilon^2} = S_{x_0}\,\omega_{c,k}\left[\frac{\omega_x}{\omega_{c,k}} - \arctan\left(\frac{\omega_x}{\omega_{c,k}}\right) + \frac{S_{z_0}}{S_{x_0}}\frac{\omega_{c,k}^2}{\omega_c^2} + \left(1-\frac{\pi}{4}\right) + \frac{S_{z_0}}{S_{x_0}}\frac{\pi}{4}\right] \tag{5h}$$

2.

$$\omega_{c,s} = \omega_x \tag{5i}$$

$$\overline{\varepsilon^2} = S_{x_0}\,\omega_{c,k}\left\{\left(1+\frac{S_{z_0}}{S_{x_0}}\frac{\omega_{c,k}^2}{\omega_c^2}\right) \times\left[\frac{\omega_x}{\omega_{c,k}} - \arctan\left(\frac{\omega_x}{\omega_{c,k}}\right)\right] + \frac{S_{z_0}}{S_{x_0}}\arctan\left(\frac{\omega_x}{\omega_{c,k}}\right)\right\} \tag{5j}$$

Any existing amplification or attenuation in the original system (static transmission factor) may be considered in the usual manner in the signal to noise ratio S_{x_0}/S_{z_0}.

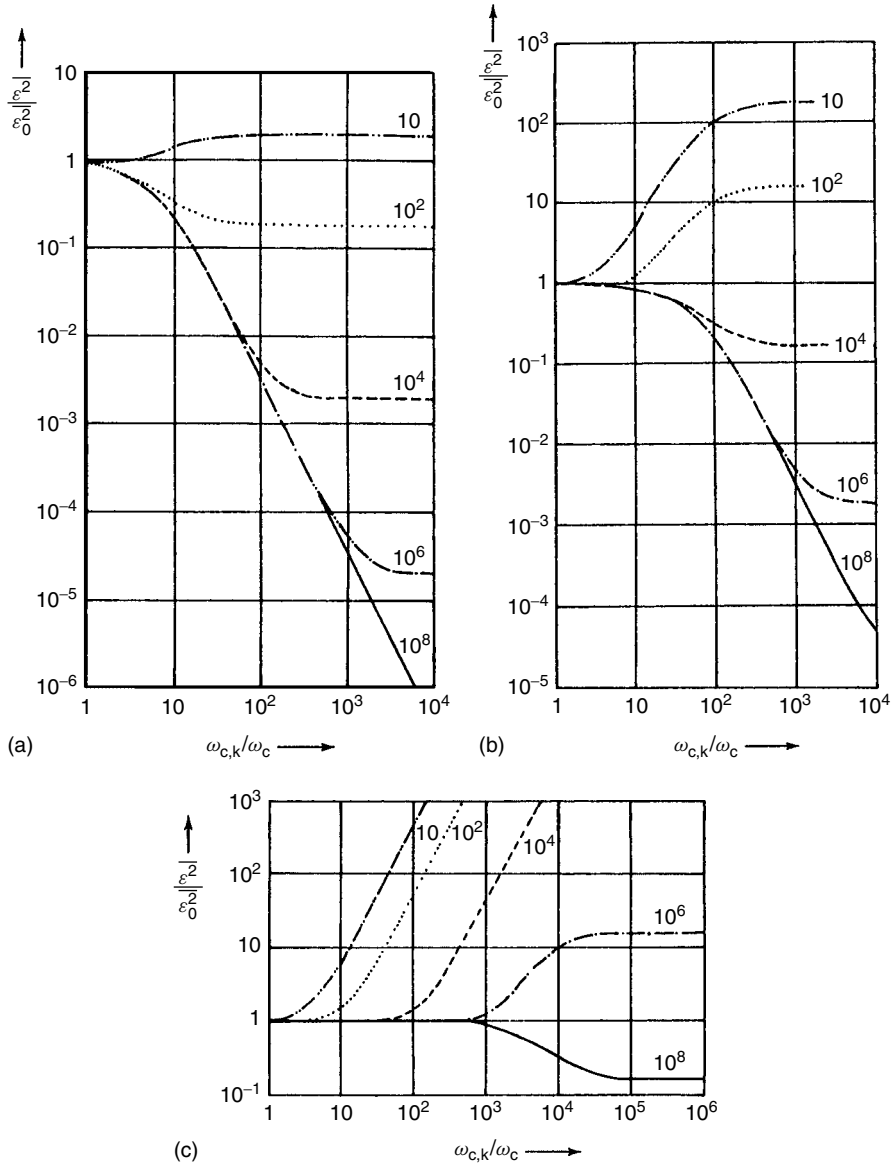


Figure 10. Continuation of Figure 9: ·····$S_{x_0}/S_{z_0} = 10$; · · · ·$S_{x_0}/S_{z_0} = 10^2$; - - - - -$S_{x_0}/S_{z_0} = 10^4$; -·-·-$S_{x_0}/S_{z_0} = 10^6$; ——$S_{x_0}/S_{z_0} = 10^8$; (a) $\omega_x/\omega_c = 10$; (b) $\omega_x/\omega_c = 10^2$; (c) $\omega_x/\omega_c = 10^4$.

In Figure 9(a, b, c), the results obtained from (5g, h) are shown for different values of the signal to noise ratio of 20, 40, 60, and 80 dB, that is for $S_{x_0}/S_{z_0} = 10^2, 10^4, 10^6$, and 10^8, in relation to the bandwidth increase by correction of $\omega_{c,k}/\omega_c$. The parameters of ω_x/ω_c were selected such that dynamically good systems (Figure 9a) as well as dynamically poor systems (Figure 9c) are involved. The values for the mean-square error ε^2 are related to the error of the uncorrected system $\overline{\varepsilon_0^2}$ according to (5f), so as to indicate directly the reduction of the error.

All the diagrams for the adaptation of the bandwidth $\omega_{c,k}$ to the bandwidth of the program for system correction $\omega_{c,s}$ (Figure 9a) reveals a minimum corresponding to the case of an optimum filter: while the dynamic portion of error (first term (5e)) decreases with the rising degree of correction, the interference-dependent portion increases with the rising degree of correction, this being due to the increase in the spectral portions $\omega > \omega_c$. The efficiency of the correction, therefore, will be higher, the dynamically better the system and the smaller the interferences. This substantiates the finding obtained by Woschni (1969) on the basis of physical considerations.

By adapting the bandwidth $\omega_{c,k}$ to that of the input signal (Figure 9b), the results shown in Figure 10(a, b, c) are obtained from (5j). The parameters were selected such that direct comparison with the results represented in Figure 9 is possible. In contrast to Figure 9(a), dependencies are obtained, which tend asymptotically to a limit value. This is because the interference-dependent portion does not rise any more because of band limitation. In dynamically very poor systems, that is, for the case of large values of ω_x/ω_c (Figure 10b, c), the error increases with correction. In these cases, the increase in the fraction of errors caused by interference predominates because of the increase in the high spectral frequencies.

Finally, it should be emphasized that, in practice, further limitations in efficiency occur due to the sensitivity of parameters, which in this investigation, have not been taken into consideration and may arise due to possibly existing nonlinearities (Woschni, 1967), (Bellmann, 1961) and (Levine, 1996).

REFERENCES

Bellman, R. (1961) *Adaptive Control Processes*, Princeton University Press, Princeton, NJ.

Jaffe, R.C. (2000) *Random Signals for Engineers using MATLAB and MATHCARD*, AIP Press, New York.

Levine, W.S. (1996) *The Control Handbook*, CRC Press, New York.

Schlitt, H. (1960) *Systemtheorie fur Regellose Vorgange*, Springer, Berlin.

Woschni, E.-G. (1967) Parameterempfindlichkeit in der Meßtechnik, Dargestellt an Einigen Typischen Beispielen. *Zeitschr. Messen-Steuern*, **10**(4), 124–130.

Woschni, E.-G. (1969) Inwieweit Spielt die Qualität eines Meßgrößenaufnehmers beim Einsatz von On-Line-Rechnern Noch eine Rolle? *Zeitschr. Messen-Steuern*, **12**(10), 384–385.

Woschni, E.-G. (1972) *Meßdynamik*, Hirzel, Leipzig.

Woschni, E.-G. (1988) *Informations Technik*, Huthig, Heidelberg.

Woschni, E.-G. (2000) Approximations in Measurement – Today still Necessary? in *Proceedings XVI IMEKO World Congress*, Wien.

51: Errors in Digital Signal Systems

Luca P. Mari
Università Cattaneo, Castellanza (VA), Italy

1 Uncertainty and Error Sources in Digital Signals 314
2 List of Typical Digital Signal Uncertainties and Errors 315
3 Digital Signal Uncertainties and Errors in Data Acquisition 315
4 Digital Signal Uncertainties and Errors in Data Processing 316
Related Articles 317
Further Reading 317

1 UNCERTAINTY AND ERROR SOURCES IN DIGITAL SIGNALS

Digital systems are adopted today in a broad range of measurement applications. While supported by the current remarkable innovations in microelectronics and the related technologies, the reason for the widespread usage of digital systems in measurement is that they are inherently less prone to error in communication and manipulation.

Traditional measuring systems behave as transducers of measurands to quantities directly perceivable by human beings, such as angular deflections of needles on graduated scales. Their operation can be described in terms of physical transformations for which the interpretation of the physical states as information entities is left to the observer. In this case, any further data processing (leading to the so-called indirect, or derived, measurement) is accomplished by devices external to the measuring system, if not done manually by the observer.

On the other hand, characteristic of the concept of digital coding is the hypothesis that physical signals are simply carriers for univocally recognizable and mutually distinguishable symbols, so that physical transformations of signals are actually modeled and dealt with as formal data processing operations (i.e., mathematical functions) in the symbol domain.

Transition from hardware to software forms for the information medium is the paradigm of this transition. While the information content in analog systems resides entirely as features and states of their hardware, in the case of digital systems, a progressive virtualization of the content can be used that varies from hard-wired logic systems, to microprocessor-based programmable systems, to the so-called virtual instruments, whose operation could be even interpreted as if their characterizing software layer is executed in an ideal hardware subsystem (*see also* **Article 144, DSP Chip Sets, Volume 3**).

Corresponding to each of these levels of abstraction, arise different issues in the metrological characterization of the systems, and in particular in the identification of the typical sources of uncertainty/error, related to both the hardware and the (multiple) software layers.

The combined uncertainty u_Y summarizing the contributions of such multiple sources u_{X_i} depends additively on them, as formalized by the law of propagation of uncertainty, as recommended by the ISO *Guide to the expression of uncertainty in measurement* (GUM) (the simplified version of such a law is shown here, applicable in the case of statistically uncorrelated sources) (*see also* **Article 54, Explanation of Key Error and Uncertainty Concepts and Terms, Volume 1**):

$$u_Y^2 = \sum_{i=1}^{N} \left(\frac{\partial f}{\partial X_i} \right)^2 u_{X_i}^2 \tag{1}$$

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

where f is the function modeling the relationship that links the measurand to its influence quantities. Equation (1) is obtained as a first-order Taylor series approximation of the model function computed in a (N-dimensional) point assumed 'sufficiently closer' to the average values of the quantities X_i and under the hypothesis that f is 'sufficiently linear' in the neighborhood of such a point.

While usually reasonably correct in the case of instrumentation dealing with smoothly varying quantities, these assumptions could become quite critical for digital systems, for then nonlinearities that sometimes are very strong, such as those manifesting as the consequence of bugs in the software, are common.

Given the fundamental requirement to formalize any measurement result by expressing both a measurand value and an estimation of its uncertainty, the usage of digital signals and systems (particularly if under software control) usually implies to trade off flexibility for complexity.

2 LIST OF TYPICAL DIGITAL SIGNAL UNCERTAINTIES AND ERRORS

While in some specific cases digital systems integrally operate on digitally coded entities (e.g. in some cases of counting in which the measurand is inherently discrete), they are also widely used in the measurement of continuously varying quantities so that a preliminary stage of analog-to-digital conversion is implied.

Once such a transduction has been completed, the operations are performed on coded symbols, that is, on a purely algorithmic basis (*see also* **Article 33, Coding Theory and its Application to Measurement, Volume 1**). The results are then fed into a device that acts as an output transducer, which is sometimes required to convert the digital symbols back to analog signals.

As a consequence, a metrological characterization of digital systems involves the analysis of their behavior in reference to three general components, each of them being affected by specific causes of uncertainties/errors.

The input subsystem is set to acquire information on the measurand from the environment and, when needed, converting it into an equivalent digital form.

Its general structure includes a sensor, a signal conditioning component, and an analog-to-digital converter (ADC) that, in PC-based systems, is usually part of a data acquisition card – *see also* **Article 132, Data Acquisition Systems (DAS) in General, Volume 3**; **Article 136, Quantizing Theory Relevant to DAS, Volume 3**.

Digital signals are obtained as the output of such a subsystem; hence, strictly speaking the input subsystem does not contribute to the budget of system uncertainties/errors related to digital signals. On the other hand, ADC characteristics and behavior significantly influence the quality of the generated digital signals (conceptually definable as the degree of correspondence with the originating analog signals and operatively affecting the possibility of reconstructing them from the converted digital signals) – *see also* **Article 138, Sampling Theory Relevant to DAS, Volume 3**

The data processing subsystem is aimed at dealing with digitally coded entities to transform them by means of suitably implemented algorithms and to transfer them to remote devices. Uncertainties/errors can appear in both hardware and software layers because of the presence of physical factors modifying the quantity on which the symbols are coded and the low quality of algorithms (or their implementations) adopted in the processing of such symbols. The latter issue grows in relevance as the software adopted for metrological purposes becomes more and more complex, as is the case with spreadsheets or virtual instruments.

Current developments in this area are particularly important, as witnessed by the emerging applications of pattern recognition, automatic control, and data fusion based on the so-called soft computing paradigm in which techniques such as neural networks and fuzzy logic inference are used to exploit uncertainty and partial information.

The output subsystem is finally aimed at making the processed data available to users and user devices (e.g., actuators of control systems) while possibly converting such data to a corresponding analog form. At this stage, raw data produced by the measuring system must be converted to information that is meaningful to the intended users and useful to them.

The sources of possible uncertainties/errors in the expression of measurement results from the digital signals representing the instrument readings are multiple, all basically related to the mathematical model of the measurement system – *see* **Article 104, Models of the Measurement Process, Volume 2**.

While having been traditionally assigned to a human user, the definition and the metrological qualification of this model is now the main task of the knowledge-based intelligent instruments – *see* **Article 128, Nature and Scope of AI Techniques, Volume 2**.

3 DIGITAL SIGNAL UNCERTAINTIES AND ERRORS IN DATA ACQUISITION

Digitalization of analog signals usually implies their time and amplitude discretization, the two basic parameters qualifying such operations being the sampling rate and the

amplitude resolution (also called *the bit depth*) of quantization, measured in samples per second and bits respectively – *see also* **Article 133, Amplifiers and Filters for DAS, Volume 3**.

Even in the case of an ideal ADC, limitations in the size of the data storage devices and bandwidth of the data transmission channels are sources of errors for the generated digital signals:

The sampling theorem shows that the information conveyed by an analog signal is integrally maintained whenever the signal is sampled at a rate greater than twice its bandwidth (for most applications the time interval between samples is kept constant). The usual technique of low-pass (sometimes band-pass) antialiasing filtering (*see also* **Article 125, Outline of Purpose of Analog Data Filters, Volume 2**) is in fact a trade-off between two systematic errors. Its application helps to avoid aliasing effects but removes any information contained in the cut-off portion of the signal spectrum.

The number of intervals (sometimes called *channels* or *cells*) in which the amplitude range is subdivided in quantization specifies the quantizer resolution, that is, the length of the binary word coding each sample. This establishes the amount of the error introduced by quantization.

In the simplest case of uniform quantization, when all intervals have the same half-width a, each sample of amplitude x is associated with a channel i whose midpoint (dealt with as the reference value to be coded) is c_i: the quantization error is then $x - c_i$ corresponding to a maximum quantization error of ± 0.5 least significant bits (LSBs) and a null average quantization error.

Here again a trade-off is implied: to reduce the quantization error, the bit depth of the code word must be increased. Thus, to enhance the accuracy of conversion by the ADC, its precision must also be increased.

To characterize the actual behavior of a physical ADC, some further parameters have to be taken into account, such as internal and external noise, settling time, short-term repeatability, and long-term stability, offset, linearity of gain, and, in the case in which two or more signals are acquired at the same time, cross talk.

It is usual that the specifications for such parameters are directly given by the ADC manufacturer as the interval $\pm a$ that certainly (i.e. with probability $= 1$) contains the corresponding values/errors.

This is the typical case in which the ISO GUM recommends type B uncertainty evaluations on the basis of uniform probability distributions (*see also* **Article 54, Explanation of Key Error and Uncertainty Concepts and Terms, Volume 1**). The corresponding standard uncertainties are then computed as $a/\sqrt{3}$ and combined by means of (1).

4 DIGITAL SIGNAL UNCERTAINTIES AND ERRORS IN DATA PROCESSING

The simplest kind of data processing is the one performed by systems computing the identity function, that is, producing as their output the same symbols given at their input, as the ideal behavior of an ideal digital transmission channel.

In this case, the presence of errors (generally caused by noise sources external to the channel) is modeled in statistical terms by recognizing that for each input symbol x_i the channel does not deterministically produce an output symbol y_j but a conditional probability distribution $P(y_j|x_i)$ (for binary channels $x_i, y_j \in \{0, 1\}$, and $P(0|x_i) + P(1|x_i) = 1$).

The average value of $-\log_2(P(x_i|y_j))$, called *equivocation* and computed from $P(y_j|x_i)$ by means of the Bayes theorem, represents the average information lost in the transmission process because of errors.

From the channel equivocation $H(X|Y)$ and the source entropy $H(X)$, the channel capacity C is computed as

$$C = \max_X (H(X) - H(X|Y)) \tag{2}$$

a basic informational quantity, measured in bit per symbol (and more usually in bit per second by multiplying it by the rate of symbol transmission over the channel), whose physical grounds are clearly identified in the fundamental relation:

$$C = W \log_2 \left(1 + \frac{S}{N}\right) \tag{3}$$

where W and S/N are the channel bandwidth and signal-to-noise ratio respectively).

In the case in which the information flowing from the source has a rate lower than the capacity C of the channel, several techniques can be adopted to reduce the probability of error at the receiver. All are based on the introduction of redundancies and aimed at either error recognition or correction.

Typical applications of digital signal processing in measurement are digital filtering and DFT/FFT computation (*see also* **Article 143, DFT and FFTs, Volume 3**).

Higher-level operations are now common, for example, in computing statistical parameters as in the case of DC/RMS measurement.

The fundamental parameters qualifying the arithmetic of a processor are its overflow, underflow, and round-off error thresholds.

In the common case of the floating-point number representation (in which numbers are expressed as $(-1)^a \times b \times 10^c$ where $a \in \{0, 1\}$, the mantissa $b \in [1, 10)$ has a fixed

Table 1. Values of machine parameters in IEEE floating-point arithmetic.

Machine parameter	Single precision (32 bits)	Double precision (64 bits)
Machine precision	$2^{-24} \approx 5.96 \times 10^{-8}$	$2^{-53} \approx 1.11 \times 10^{-16}$
Underflow threshold	$2^{-126} \approx 1.18 \times 10^{-38}$	$2^{-1022} \approx 2.23 \times 10^{-308}$
Overflow threshold	$2^{128}(1-\varepsilon) \approx 3.40 \times 10^{38}$	$2^{1024}(1-\varepsilon) \approx 1.79 \times 10^{308}$

number of digits, and the exponent c is an integer spanning between two fixed values, the overflow and the underflow thresholds depend on the maximum positive and negative values of the exponent respectively.

On the other hand, round-off errors depend on the number of digits reserved for the mantissa; they are expressed in terms of the machine precision, a value generally related to the characteristics of not only the processor arithmetic-logic unit (ALU) but also the adopted software platform or compiler. This is an important source of complexity in the metrological qualification of data processing modules.

A commonly implemented reference for the values of these parameters is the IEEE Standard (see Table 1).

The data processing subsystem is usually so complex that, instead of identifying all the relevant sources of uncertainty that would be required to apply (1), a black box solution is sometimes adopted for its metrological qualification. A reference data set is chosen that contains a collection of sampled input data with the corresponding expected output; such input data are fed into the subsystem, and the results are compared with the references. From the analysis of the obtained error, an estimation of the uncertainty of the data processing results is then inferred.

The contribution of all possible hardware faults (and correspondingly the degree of fault tolerance of the system) should also be taken into account for a complete metrological qualification of the system.

RELATED ARTICLES

Article 48, Common Sources of Errors in Measurement Systems, Volume 1; **Article 54, Explanation of Key Error and Uncertainty Concepts and Terms, Volume 1**; **Article 55, Uncertainty Determination, Volume 1**; **Article 132, Data Acquisition Systems (DAS) in General, Volume 3**; **Article 139, Analog-to-Digital (A/D) Converters, Volume 3**.

FURTHER READING

ANSI/IEEE (1985) *IEEE Standard for Binary Floating Point Arithmetic*, Std 754, New York.

ANSI/IEEE (1987) *IEEE Standard for Radix Independent Floating Point Arithmetic*, Std 854, New York.

ISO (1993) *Guide to the Expression of Uncertainty in Measurement*, International Organization for Standardization, Geneva.

Mathworks, manuals related to the MathLab™ software platform (in particular *Data acquisition toolbox – User's guide*, Version 2), downloadable at http://www.mathworks.com.

National Instruments, manuals related to the LabView™ software platform (in particular *LabView user manual*, 7.2000 Edition; *Measurements manual*, 7.2000 Edition; *Data acquisition basics manual*, 1.2000 Edition), downloadable at http://www.ni.com.

National Physical Laboratory, Software Support for Metrology Project, Best Practice Guides, downloadable at http://www.npl.co.uk/ssfm.

Patterson, D.A. and Hennessy, J.L. (1998) *Computer Organization & Design – The Hardware/Software Interface*, Morgan Kaufmann, New York.

52: Error Models, Error Budgets and their Calculation

Rudolf Palenčár
Slovak University of Technology, Bratislava, Slovak Republic

1 Definition of an Error Model 318
2 Error Budget Principles 319
3 Setting Up an Error Budget 320
4 Calculation of an Error Budget 320
5 Evaluation of the Final Error Figures 324
6 Methods for Evaluation of Uncertainties 325
Further Reading 326

1 DEFINITION OF AN ERROR MODEL

Uncertainty determination and presentation is covered in **Article 55, Uncertainty Determination, Volume 1**. This article introduces how errors and uncertainties of measurement are dealt with where many are present when determining a measured value.

The measurement error Δx is defined as a difference between the measured value x_m and true value x_S of the measurand X

$$\Delta x = x_m - x_S \tag{1}$$

If we want to introduce a factor of time, error is defined as

$$\Delta x(t) = x_m(t) - x_S(t) \tag{2}$$

The p domain for linear or quasi-linear measuring systems is

$$\Delta x(p) = x_m(p) - x_S(p) \tag{3}$$

where $x_m(p)$ – the measured value transformed by Laplace transform; $x_S(p)$ – transformed true value.

For simplicity, let us assume that x_S is the value at the input of the measuring system and x_m is the value at the output of the measuring system. Sometimes, value x_S, at the input of the measuring system, is not identical with the value of the measurand. Its inaccessibility and common impact on the measurand and measuring system may cause this. Similarly, the obtained value x_m may not need to be identical with the output value of the measuring system due to such things as signal processing taking place after the data leaves the measuring system.

Error is treated as a time function when quantity X is time dependent, or if the transient processes in the measuring system are still active, as shown in Figure 1.

Measurement error Δx consists of the static error Δx_{st} and the dynamic error Δx_{dy}.

Dynamic error arises when the measurand is time dependent and change occurs before the transient processes are completed or when the measurand is time steady but determination of the value in the time t^* arrives before the transient processes in the measuring system are over.

Measurement error (static and dynamic) consists of the systematic and stochastic part (*see also* **Article 49, General Characterization of Systematic and Stochastic Errors, Volume 1**).

If the transfer function of the measuring system is represented as

$$S(p) = \frac{x_m(p)}{x_S(p)} \tag{4}$$

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

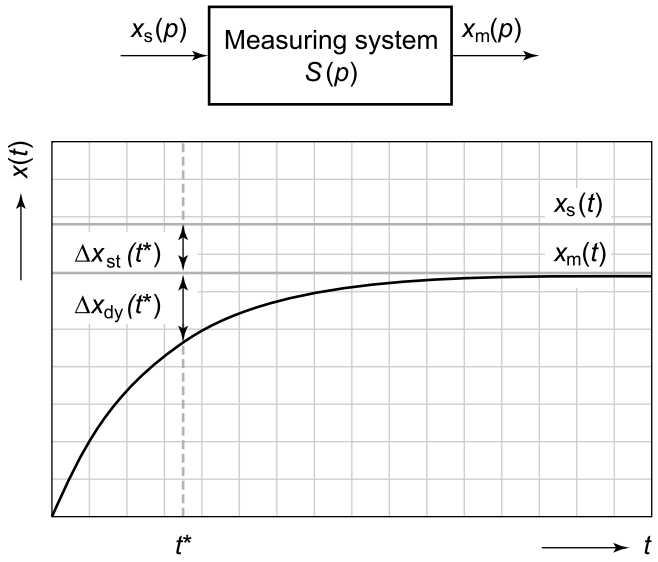


Figure 1. Relationship between the measured and true value of measurand. $x_s(t)$ – path of the input value; $x_m(t)$ – path of the output (measured) value; $\Delta x_{st}(t^*)$ – static measurement error in time t^*; $\Delta x_{dy}(t^*)$ – dynamic measurement error in time t^*.

the error is then

$$\Delta x(p) = \left(1 - \frac{1}{S(p)}\right) x_m(p) \tag{5}$$

The error value in the time domain is determined by the use of the reverse Laplace transformation.

Error Δx is also called an *absolute error*.

The ratio of the absolute error to the true value is called the *relative error*.

In the time domain

$$\Delta x_r(t) = \frac{x_m(t) - x_S(t)}{x_S(t)} = \frac{x_m(t)}{x_S(t)} - 1 \tag{6}$$

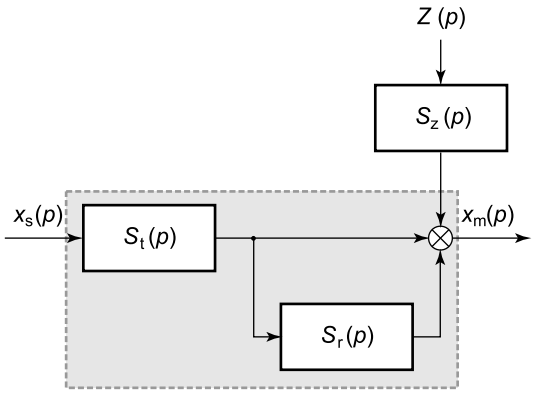


Figure 2. Scheme of the measuring system and the influence of errors. $S_t(p)$ – theoretical transfer function of the measuring system; $S_r(p)$ – transfer function of the inner error; $S_z(p)$ – transfer function of the injected error.

This transforms into the p domain as

$$\Delta x_r(p) = \frac{x_m(p) - x_S(p)}{x_S(p)} = S(p) - 1 = S_r(p) \tag{7}$$

The transfer function $S_r(p)$ characterizes the transfer of the error in the measuring system and does not depend on the value at the system input.

Figure 2 shows schemes of the measuring system with the influence of errors. This figure represents an example for realistic functioning of a measuring system.

2 ERROR BUDGET PRINCIPLES

Absolute measurement error at the output of the measuring system can consist of two parts:

1. Inner error of the measuring system is caused by difference between the theoretical $S_t(p)$ and real transfer function $S(p)$.
2. Error injected into the measuring system from the surrounding by disturbance value z.

Stated errors can comprise of systematic errors with transfer functions $S_S(p)$ and $S_{ZS}(p)$, and random errors with transfer functions $S_R(p)$ and $S_{ZR}(p)$. Figure 3 shows individual cases of the errors acting.

Error sources are associated with the

- measuring system,
- measuring method,
- environment,
- operator.

Details on individual error sources are stated in other articles – *see also* **Article 48, Common Sources of Errors**

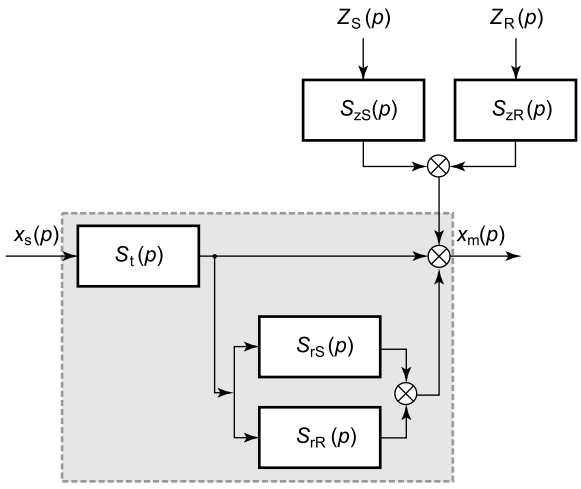


Figure 3. Influence of errors on the measuring system.

in Measurement Systems, Volume 1; **Article 51, Errors in Digital Signal Systems, Volume 1**; **Article 50, Errors in Signal Systems, Volume 1**; and **Article 49, General Characterization of Systematic and Stochastic Errors, Volume 1**.

As stated before, during measurement of the quantity y, individual errors Δx_i arise and the overall error Δy must be determined. It is then necessary to determine for the quantity y, being a function of the quantities $x_1, x_2, \ldots, x_n$ (indirect measurement of the quantity y), that is,

$$y = f(x_1, x_2, \ldots, x_n) \tag{8}$$

Error is determined by the errors propagation law

$$\Delta y = \sum_{i=1}^{n} C_i \Delta x_i \tag{9}$$

where C_i – sensitivity coefficients being obtained as partial derivation of the f function by individual quantities x_i.

Equation (9) was obtained by expanding function (8) into the Taylor series, neglecting members of high order.

For determination of the uncertainty $u(\Delta y)$ of the error Δy (only if its estimation is known) using (9), the uncertainties propagation law is described by the expression (as the standard uncertainty is equal to the standard deviation):

$$u^2(\Delta y) = \sum_{i=1}^{n} C_i^2 u^2(\Delta x_i) + 2 \sum_{i=1}^{n-1} \sum_{j>i}^{n} C_i C_j u(\Delta x_i, \Delta x_j) \tag{10}$$

where $u(\Delta x_i)$ is the standard uncertainty of the error Δx_i; and $u(\Delta x_i, \Delta x_j)$ isthe covariance between errors Δx_i and Δx_j.

Uncertainties $u(\Delta x_i)$ and covariances $u(\Delta x_i, \Delta x_j)$ are determined by the type A and type B evaluation – *see also* **Article 53, Calculation and Treatment of Errors, Volume 1**; **Article 54, Explanation of Key Error and Uncertainty Concepts and Terms, Volume 1**; and **Article 55, Uncertainty Determination, Volume 1**.

If the individual errors Δx_i are not correlated, then

$$u^2(\Delta y) = \sum_{i=1}^{n} C_i^2 u^2(\Delta x_i) \tag{11}$$

The measurement result y obtained by the measurement of the Y quantity will be

$$y = y_{\mathrm{m}} - \Delta y \pm U(\Delta y) \tag{12}$$

where y_{m} is the value obtained from (8) for measured values $x_{\mathrm{m}i}$; and $U(\Delta y)$ is the expanded uncertainty of the error Δy.

Values $x_{\mathrm{m}i}$ can be obtained as

- results of individual measurements,
- arithmetic means or medians or other statistical characteristics of the set of repeated measurements,
- values of physical constants,
- values of different parameters,
- results of previous measurements, and so on,

for estimations of errors when their uncertainties and covariances are known.

3 SETTING UP AN ERROR BUDGET

Basic principles are available that show how individual errors combine and propagate (transfer and combination).

Tables 1, 2, and 3 introduce basic schemes for the transfer of the measurement errors in measuring systems, transfer functions of the measuring system seen from their errors point of view, relative, and absolute errors, and uncertainties of the measuring system.

Similar reasoning leads to the relationships given in Table 4, which shows typical cases for different functions occurring in indirect measurement.

4 CALCULATION OF AN ERROR BUDGET

For the measuring system shown in Figure 3, we can now determine the value of the measured quantity x_{m} in relation to the real quantity x_{s}, as absolute and relative measurement error.

For measured value x_{m}^*, we can determine an estimation Δx^* of the absolute errors Δx and estimation $(\Delta x/x)^*$ of the relative errors $\Delta x/x$, as well as for the uncertainty of estimation of such errors as $u(\Delta x^*)$, respectively $u(\Delta x/x^*)$.

The measurement result will then be an estimation x_{s}^* of the real value x_{s}, obtained as the measured value x_{m}^* corrected for systematic errors (their estimation) and the uncertainty of the estimations of those errors including the uncertainty of random errors.

For a measuring system according to Figure 3, the measured value x_{m} in the p domain is given by:

$$x_{\mathrm{m}} = S_{\mathrm{t}}[1 + S_{\mathrm{S}} + S_{\mathrm{R}}]x_{\mathrm{S}} + S_{Z\mathrm{S}} Z_{\mathrm{S}} + S_{Z\mathrm{R}} Z_{\mathrm{R}} \tag{13}$$

The absolute error will be:

$$\Delta x = [S_{\mathrm{t}}(S_{\mathrm{S}} + S_{\mathrm{R}}) - 1]x_{\mathrm{S}} + S_{Z\mathrm{S}} Z_{\mathrm{S}} + S_{Z\mathrm{R}} Z_{\mathrm{R}} \tag{14}$$

Table 1. Transfer of errors and uncertainties through measuring system with serial sequencing of the measuring instruments.

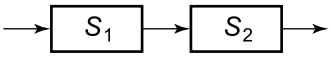
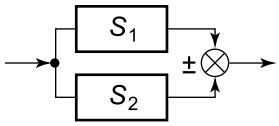
Serial sequencing of the measuring instruments	
	
Transfer function	$S_1 S_2$
Relative error	$\frac{\Delta x_1}{x_1} + \frac{\Delta x_2}{x_2}$
Absolute error	$S_2 \Delta x_1 + S_1 \Delta x_2$
Relative uncertainty	$\sqrt{u^2\left(\frac{\Delta x_1}{x_1}\right) + u^2\left(\frac{\Delta x_2}{x_2}\right) + 2u\left(\frac{\Delta x_1}{x_1}, \frac{\Delta x_2}{x_2}\right)}$
Absolute uncertainty	$\sqrt{S_1^2 u^2(\Delta x_1) + S_2^2 u^2(\Delta x_2) \pm 2 S_1 S_2 u(\Delta x_1, \Delta x_2)}$
Used symbols:	
S_1, S_2	Transfer functions of the measuring instruments
$\Delta x_1, \Delta x_2$	Absolute errors of the measuring instruments
$\frac{\Delta x_1}{x_1}, \frac{\Delta x_2}{x_2}$	Relative errors of the measuring instruments
$u\left(\frac{\Delta x_1}{x_1}\right), u\left(\frac{\Delta x_2}{x_2}\right)$	Relative uncertainties of the measuring instruments (uncertainties of the relative errors of the measuring instruments)
$u(\Delta x_1), u(\Delta x_2)$	Absolute uncertainties of the measuring instruments (uncertainties of the absolute errors of the measuring instruments)
$u\left(\frac{\Delta x_1}{x_1}, \frac{\Delta x_2}{x_2}\right)$	Relative covariance among measuring instruments (covariance among relative errors of the measuring instruments)
$u(\Delta x_1, \Delta x_2)$	Absolute covariance among measuring instruments (covariance among absolute errors of the measuring instruments)

Table 2. Transfer of errors and uncertainties through measuring system with parallel sequencing of the measuring instruments.

Parallel sequencing of the measuring instruments	
Scheme of the measuring system	
Transfer function	$S_1 \pm S_2$
Relative error	$\frac{S_1}{S_1 \pm S_2}\frac{\Delta x_1}{x_1} + \frac{S_2}{S_1 \pm S_2}\frac{\Delta x_2}{x_2}$
Absolute error	$\Delta x_1 + \Delta x_2$
Relative uncertainty	$\sqrt{\frac{S_1^2}{(S_1 \pm S_2)^2}u^2\left(\frac{\Delta x_1}{x_1}\right) + \frac{S_2^2}{(S_1 \pm S_2)^2}u^2\left(\frac{\Delta x_2}{x_2}\right) \pm 2\frac{S_1 S_2}{(S_1 \pm S_2)^2}u\left(\frac{\Delta x_1}{x_1}, \frac{\Delta x_2}{x_2}\right)}$
Absolute uncertainty	$\sqrt{u^2(\Delta x_1) + u^2(\Delta x_2) \pm 2u(\Delta x_1, \Delta x_2)}$

If absolute error is expressed as a function of x_m, then

$$\Delta x = \left[1 - \frac{1}{S_t(1 + S_S + S_R)}\right] x_m + \frac{1}{S_t(1 + S_S + S_R)}(S_{ZS} Z_S + S_{ZR} Z_R) \quad (15)$$

If we know the transfer functions S_S and S_R exactly, as well as the spurious influences and their transfer functions $S_{ZS}Z_S$ and $S_{ZR}Z_R$, we can determine the real value of the measured quantity x_S.

Where we do not know those values exactly, knowing only their estimations as well as the uncertainties of

Table 3. Transfer of errors and uncertainties through measuring system with feedback.

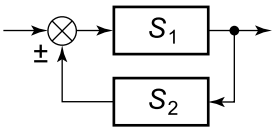
Sequencing of the measuring instruments with feedback	
Scheme of the measuring system	
Transfer function	$\frac{S_1}{1 \pm S_1 S_2}$
Relative error	$\frac{1}{1 \mp S_1 S_2}\frac{\Delta x_1}{x_1} \pm \frac{S_1 S_2}{1 \mp S_1 S_2}\frac{\Delta x_2}{x_2}$
Absolute error	$\frac{S_1}{(1 \mp S_1 S_2)^2}\Delta x_1 \pm \frac{S_1^2 S_2}{(1 \mp S_1 S_2)^2}\Delta x_2$
Relative uncertainty	$\sqrt{\frac{1}{(1 \mp S_1 S_2)^2}u^2\left(\frac{\Delta x_1}{x_1}\right) + \frac{S_1^2 S_2^2}{(1 \mp S_1 S_2)^2}u^2\left(\frac{\Delta x_2}{x_2}\right) \pm 2\frac{S_1 S_2}{(1 \pm S_1 S_2)^2}u\left(\frac{\Delta x_1}{x_1}, \frac{\Delta x_2}{x_2}\right)}$
Absolute uncertainty	$\sqrt{\frac{S_1^2}{(1 \mp S_1 S_2)^4}u^2(\Delta x_1) + \frac{S_1^4 S_2^2}{(1 \mp S_1 S_2)^4}u^2(\Delta x_2) \pm 2\frac{S_1^3 S_2}{(1 \pm S_1 S_2)^2}u(\Delta x_1, \Delta x_2)}$

Table 4. Determination of errors and uncertainties for indirect measurement.

Function	$y = \sum_{i=1}^{n} x_i$
Absolute error	$\Delta y = \sum_{i=1}^{n} \Delta x_i$
Relative error	$\frac{\Delta y}{y} = \sum_{i=1}^{n} \frac{x_i}{x_1 + x_2 + \cdots + x_n}\frac{\Delta x_i}{x_i}$
Absolute uncertainty	$u(\Delta y) = \sqrt{\sum_{i=1}^{n} u^2(\Delta x_i) + 2\sum_{i=1}^{n-1}\sum_{j>i}^{n} u(\Delta x_i, \Delta x_j)}$
Relative uncertainty	$\frac{u(\Delta y)}{y} = \sqrt{\sum_{i=1}^{n} \frac{x_i^2}{(x_1 + x_2 + \cdots + x_n)^2}\frac{u^2(\Delta x_i)}{x_i^2} + 2\sum_{i-1}^{n-1}\sum_{j>i}^{n} \frac{x_i x_j}{(x_1 + x_2 + \cdots + x_n)^2}\frac{u(\Delta x_i, x_j)}{x_i x_j}}$
Function	$y = x_1 - x_2$
Absolute error	$\Delta y = \Delta x_1 - \Delta x_2$
Relative error	$\frac{\Delta y}{y} = \frac{x_1}{x_1 - x_2}\frac{\Delta x_1}{x_1} - \frac{x_2}{x_1 - x_2}\frac{\Delta x_2}{x_2}$
Absolute uncertainty	$u(\Delta y) = \sqrt{u^2(\Delta x_1) + u^2(\Delta x_2) - 2u(\Delta x_1, \Delta x_2)}$
Relative uncertainty	$\frac{u(\Delta y)}{y} = \sqrt{\frac{x_1^2}{(x_1 - x_2)^2}\frac{u^2(\Delta x_1)}{x_1^2} + \frac{x_2^2}{(x_1 - x_2)^2}\frac{u^2(\Delta x_2)}{x_2^2} - 2\frac{x_1 x_2}{(x_1 - x_2)^2}\frac{u(\Delta x_1, \Delta x_2)}{x_1 x_2}}$
Function	$y = x_1 x_2 \cdots x_n$
Absolute error	$\Delta y = \sum_{i=1}^{n} \frac{x_1 x_2 \cdots x_n}{x_i}\Delta x_i$

Table 4. (*continued*).

Relative error	$\dfrac{\Delta y}{y} = \sum_{i=1}^{n} \dfrac{\Delta x_i}{x_i}$
Absolute uncertainty	$u(\Delta y) = \sqrt{\sum_{i=1}^{n} \dfrac{(x_1 x_2 \cdots x_n)^2}{x_i^2} u^2(\Delta x_i) + 2 \sum_{i=1}^{n-1} \sum_{j>i}^{n} \dfrac{(x_1 x_2 \cdots x_n)^2}{x_i x_j} u(\Delta x_i, \Delta x_j)}$
Relative uncertainty	$\dfrac{u(\Delta y)}{y} = \sqrt{\sum_{i=1}^{n} \dfrac{u^2(\Delta x_i)}{x_i^2} + 2 \sum_{i-1}^{n-1} \sum_{j>i}^{n} \dfrac{u(\Delta x_i, x_j)}{x_i x_j}}$
Function	$y = x_1^{l_1} x_2^{l_2} \cdots x_n^{l_n}$
Absolute error	$\Delta y = \sum_{i=1}^{n} \dfrac{l_i x_1^{l_1} x_2^{l_2} \cdots x_n^{l_n}}{x_i} \Delta x_i$
Relative error	$\dfrac{\Delta y}{y} = \sum_{i=1}^{n} l_i \dfrac{\Delta x_i}{x_i}$
Absolute uncertainty	$u(\Delta y) = \sqrt{\sum_{i=1}^{n} \dfrac{l_i^2 (x_1 x_2 \cdots x_n)^2}{x_i^2} u^2(\Delta x_i) + 2 \sum_{i=1}^{n-1} \sum_{j>i}^{n} \dfrac{l_i l_j (x_1 x_2 \cdots x_n)^2}{x_i x_j} u(\Delta x_i, \Delta x_j)}$
Relative uncertainty	$\dfrac{u(\Delta y)}{y} = \sqrt{\sum_{i=1}^{n} l_i^2 \dfrac{u^2(\Delta x_i)}{x_i^2} + 2 \sum_{i-1}^{n-1} \sum_{j>i}^{n} l_i l_j \dfrac{u(\Delta x_i, x_j)}{x_i x_j}}$
Function	$y = \dfrac{x_1 x_2 \cdots x_n}{z_1 z_2 \cdots z_m}$
Absolute error	$\Delta y = \sum_{i=1}^{n} \dfrac{x_1 x_2 \cdots x_n}{x_i z_1 z_2 \cdots z_m} \Delta x_i - \sum_{j=1}^{m} \dfrac{x_1 x_2 \cdots x_n}{z_j z_1 z_2 \cdots z_m} \Delta z_j$
Relative error	$\dfrac{\Delta y}{y} = \sum_{i=1}^{n} \dfrac{\Delta x_i}{x_i} - \sum_{j=1}^{m} \dfrac{\Delta z_j}{z_j}$
Absolute uncertainty	$u(\Delta y) = \sqrt{\sum_{i=1}^{n} \dfrac{(x_1 x_2 \cdots x_n)^2}{x_i^2 (z_1 z_2 \cdots z_m)^2} u^2(\Delta x_i) + \sum_{j=1}^{m} \dfrac{(x_1 x_2 \cdots x_n)^2}{z_i^2 (z_1 z_2 \cdots z_m)^2} u^2(\Delta z_j) + 2 \sum_{i=1}^{n-1} \sum_{j>i}^{n} \dfrac{(x_1 x_2 \cdots x_n)^2}{x_i x_j (z_1 z_2 \cdots z_m)^2} u(\Delta x_i, \Delta x_j) + 2 \sum_{i=1}^{m-1} \sum_{j>i}^{m} \dfrac{(x_1 x_2 \cdots x_n)^2}{z_i z_j (z_1 z_2 \cdots z_m)^2} u(\Delta z_i, \Delta z_j) - \sum_{i=1}^{n} \sum_{j=1}^{m} \dfrac{(x_1 x_2 \cdots x_n)^2}{x_i z_j (z_1 z_2 \cdots z_m)^2} u(\Delta x_i, \Delta z_j)}$
Relative uncertainty	$\dfrac{u(\Delta y)}{y} = \sqrt{\sum_{i=1}^{n} \dfrac{u^2(\Delta x_i)}{x_i^2} + \sum_{j=1}^{m} \dfrac{u^2(\Delta z_j)}{x_j^2} + 2 \sum_{i-1}^{n-1} \sum_{j>i}^{n} \dfrac{u(\Delta x_i, \Delta x_j)}{x_i x_j} + 2 \sum_{i-1}^{m-1} \sum_{j>i}^{m} \dfrac{u(\Delta z_i, \Delta z_j)}{z_i z_j} - \sum_{i-1}^{n} \sum_{j=1}^{m} \dfrac{u(\Delta x_i, \Delta z_j)}{x_i z_j}}$

those estimations, only an estimation of the real measured quantity x_S^* can be determined.

Transfer functions of random errors are estimated as zero, that is, $S_R^* = 0$, $S_{ZR}^* = 0$ and we presume that the uncertainties of those estimations can be determined from measured data; or that we know them.

For measured value x_m^* being valid (assuming $S_R^* = 0$, $S_{ZR}^* = 0$) then:

$$\Delta x^* = \left[1 - \frac{1}{S_t^*(1+S_S^*)}\right] x_m + \frac{S_{ZS}^* Z_S^*}{S_t^*(1+S_S^*)} \tag{16}$$

and the uncertainty of the absolute error is:

$$u^2(\Delta x^*) = \frac{(x_m^* - S_{ZS}^* Z_S^*)^2}{S_t^2(1+S_S^*)^4}[u^2(S_S^*) + u^2(S_R^*)] + \frac{1}{S_t^2(1+S_S^*)^2}[u^2(S_{ZS}^* Z_S^*) + u^2(S_{ZR}^* Z_R^*)] \tag{17}$$

for we do not assume dependence among individual errors.

For error and its uncertainty determined by this procedure in the p domain, we can then extract the error value and its uncertainty using the reverse Laplace transformation.

This is based on the assumption that x_s is the value at the input to the measuring system and value x_m is at the output from the measuring system.

5 EVALUATION OF THE FINAL ERROR FIGURES

If we proceed to use a time domain, the measurement result will be expressed as

$$x = x_m^* - \Delta x^* \pm U(\Delta x^*) \tag{18}$$

where $U(\Delta x^*)$ – expanded uncertainty of the measurement – is determined from the equation

$$U(\Delta x^*) = k u(\Delta x^*) \tag{19}$$

where k – expansion coefficient depending on the probability distribution of the error Δx (value $k = 2$ is often selected in practice).

Example: indirect measurement of the current by measurement of the voltage drop We wish to measure the electric current flowing through a circuit. We will measure the voltage drop on the resistor with a nominal value 1 Ω resistor using a digital voltmeter, as shown in Figure 4.

The surrounding temperature is within $(22 \pm 2)\,^\circ$C. The flowing electric current is expected to be around 50 mA. The calibration certificate of the measuring resistor states that its value for the temperature 22 °C and the current 50 mA is (1–0.0002) Ω and the corresponding expanded uncertainty for the expansion coefficient $k = 2$ is 0.0002 Ω.

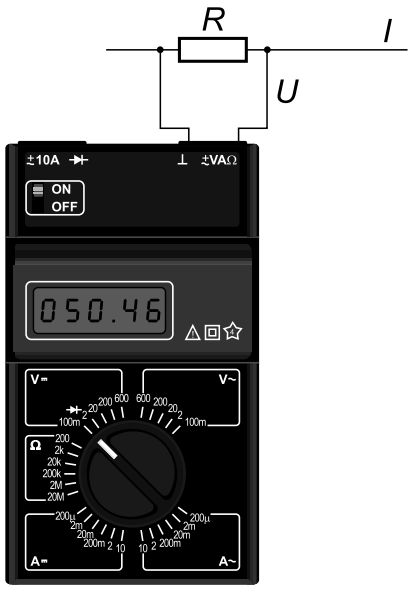


Figure 4. Indirect measurement of electric current I, via use of resistor R to generate voltage U from that current.

The manufacturer states that the digital voltmeter has the maximum permissible error 0.01% of the measured value plus 0.005% of the measuring range for the measuring range 100 mV with the surrounding temperature within 15 to 35 °C.

For the current determination, we use the Ohm's law relationship:

$$I = \frac{U}{R}$$

where I – measured electric current (mA); U – directly measured voltage drop (mV); R – resistance of the measuring resistor (Ω).

We have obtained 10 values by measurement, as listed in Table 5. Measurements were performed under the same conditions.

Following are the estimation results from the measured values:

$$\overline{U} = \frac{1}{10}\sum_{i=1}^{10} U_i = 50.44\,\text{mV}$$

Estimation of the electric current being measured is:

$$\bar{I} = \frac{\overline{U}}{R} = \frac{50.44}{1} = 50.44\,\text{mA}$$

Let us consider only the systematic error of the measuring resistor. This is

$$\Delta I = -\frac{\overline{U}}{R^2}\Delta R = \frac{50.44}{1}0.0002 = -0.01\,\text{mA}$$

Table 5. Values obtained by measurement.

No.	1	2	3	4	5	6	7	8	9	10
U_i (mV)	50.46	50.44	50.45	50.48	50.41	50.49	50.40	50.40	50.45	50.42

Table 6. Balance of errors and uncertainties.

Error ΔX_i	Estimation Δx_i	Standard uncertainty $u(x_i)$	Sensitivity coefficient C_i	Contribution to the standard uncertainty $u_i(y)$
Random	0.00 mV	1.011×10^{-2} mV	1 mA/mV	1.011×10^{-2} mA
Voltmeter ΔU	0.00 mV	0.58×10^{-2} mV	1 mA/mV	0.58×10^{-2} mA
Measuring resistor ΔR	$-0.0002\,\Omega$	$0.0001\,\Omega$	50.46 mA/Ω	0.55×10^{-2} mA
ΔI	−0.01 mA			1.27×10^{-2} mA

Components of the standard uncertainty are as follows:

1. Standard uncertainty of the voltage drop determined by the type A method:

$$u_{\mathrm{A}}(U) = s(\overline{U}) = \sqrt{\frac{1}{10(10-1)}\sum_{i=1}^{10}(U_i - \overline{U})^2}$$
$$= 1.011 \times 10^{-2}\,\mathrm{mV}$$

2. Standard uncertainty of the voltage drop determined by the type B method:

For the voltmeter used and considering the measured value 50.45 mV, the maximum permissible error of the measuring instrument 0.01% of 50.45 mV plus 0.005% of 100 mV, that is, approximately 0.01 mV. Considering the uniform distribution, standard uncertainty is $u_{\mathrm{B}}(U) = 0.01/\sqrt{(3)} = 0.0058$ mV. Current flowing through the voltmeter is neglected. The influence of the temperature that fits within the range $22\,^{\circ}\mathrm{C} \pm 2$ is covered in the basic permissible error of the voltmeter.

3. Standard uncertainty of the measuring resistor:

The expanded uncertainty of the measuring resistor for $k = 2$ is equal to $0.0002\,\Omega$ that matches the standard uncertainty $u(R) = 0.0002/2 = 0.0001\,\Omega$. The influence of the temperature within the range $(22 \pm 2)\ ^{\circ}\mathrm{C}$ on the electrical resistance can be neglected compared to other considered sources.

4. Standard uncertainty of the electric current measurement:

We do not consider any covariances in this case as we measure directly only the quantity of voltage drop. Estimations of the permissible error of the voltmeter and the measuring resistor are considered not to be correlated. Then:

$$u_{\mathrm{I}} = \sqrt{C_U^2 u_{\mathrm{A}}^2(U) + C_U^2 u_{\mathrm{B}}^2(U) + C_{\mathrm{R}}^2 u^2(R)} = 1.27^{-2}\,\mathrm{mA}$$

A summary of the procedure for the uncertainty determination gives the table of the uncertainties balance, see Table 6 and *see also* **Article 54, Explanation of Key Error and Uncertainty Concepts and Terms, Volume 1**; and **Article 55, Uncertainty Determination, Volume 1**.

The final measurement result is

$$I = (50.440 + 0.010 \pm 0.026)\,\mathrm{mA} = (50.450 \pm 0.026)\,\mathrm{mA}$$

6 METHODS FOR EVALUATION OF UNCERTAINTIES

Uncertainties of individual errors are determined either by type A evaluation or by type B evaluation (*see also* **Article 54, Explanation of Key Error and Uncertainty Concepts and Terms, Volume 1**; **Article 55, Uncertainty Determination, Volume 1**; **Article 53, Calculation and Treatment of Errors, Volume 1**).

Type A evaluation is based on statistical analysis of measured data.

Type B evaluation is every evaluation that is not of type A.

For example, in the case of a single measurement, all errors and their uncertainties are estimated by the type B method.

Therefore, uncertainties of the random errors must also be evaluated by the type B method (e.g. utilizing experiences, from data of the manufacturer etc.).

As random errors are generated by different influences acting on the measuring process, it is more accurate to perform repeated measurement for their evaluation.

For n repeated measurements under the same conditions, when a set of values $x^*_{\mathrm{m}1}, x^*_{\mathrm{m}2}, \ldots, x^*_{\mathrm{m}n}$ is available, the measurement result will be given as the arithmetic mean of the measured values $\bar{x}^*_{\mathrm{m}}$ reduced on systematic error.

Overall uncertainty of the random error is evaluated by the type A method according to equation

$$u = \frac{s_x}{\sqrt{n}} = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^{n} (x^*_{mi} - \bar{x}^*_m)^2} \qquad (20)$$

6.1 Covariances

Covariances among the estimations of individual errors determine how much those estimations are influenced by common sources.

Commonly bounded errors contribute to the overall uncertainty more or less than only merging their uncertainties.

Covariances can increase or decrease the resulting overall uncertainty. It depends on their character (whether increasing one error increases another error) and also on the shape of the function that binds them to the output quantity.

Type A evaluation of covariances between estimations of two errors Δx_i and Δx_j is used when n measured values of both quantities $x_{i1}, x_{i2}, \ldots, x_{in}$ and $x_{j1}, x_{j2}, \ldots, x_{jn}$ is available. Then

$$u(\Delta x_i, \Delta x_j) = \frac{1}{n(n-1)} \sum_{k=1}^{n} (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j) \qquad (21)$$

Covariances between two estimations evaluated by the type B method can be determined by studying certificates of the measuring instruments, from literature and so on, or from calculation using the following procedure:

1. For each pair of estimators, one estimates (on the basis of experiences) the correlation coefficient $r(\Delta x_i, \Delta x_j)$ expressing the degree of dependence between estimators.

   Covariance is then obtained from the expression

$$u(\Delta x_i, \Delta x_j) = r(\Delta x_i, \Delta x_j)\, u(\Delta x_i)\, u(\Delta x_j) \qquad (22)$$

2. If two quantities y_1 and y_2 act in indirect measurement and they are defined by the following functions

$$y_1 = f_1(x_1, x_2, \ldots, x_n) \qquad (23)$$

$$y_2 = f_2(x_1, x_2, \ldots, x_n) \qquad (24)$$

the covariance between the quantities y_1 and y_2 is calculated as

$$u(y_1, y_2) = \sum_{i=1}^{n} C_{1i} C_{2i} u^2(x_i) + \sum_{i=1}^{n} \sum_{j \neq i}^{n} C_{1i} C_{2j} u(x_i, x_j) \qquad (25)$$

and for $x_1, x_2, \ldots, x_n$ being commonly independent it is then

$$u(y_1, y_2) = \sum_{i=1}^{n} C_{1i} C_{2i} u^2(x_i) \qquad (26)$$

where C_{1i} respectively C_{1i} – sensitivity coefficients of the function f_1 or f_2 respectively.

Such an approach is used for excluding correlation when a proper measurement model is designed that comprises only independent quantities.

3. If the correlation coefficient cannot be determined and a proper model (excluding correlation) cannot be designed, the maximum influence of correlation to the resulting uncertainty should be determined.

   If $y = f(x_1, x_2)$ then:

$$u^2(y) \leq (|C_1 u(x_1)| + |C_2 u(x_2)|)^2 = C_1^2 u^2(x_1) + C_2^2 u^2(x_2) + 2|C_1 C_2|\, u(x_1)\, u(x_2) \qquad (27)$$

If we do not then have enough information for accurate evaluation of the covariances, and consequently also the resulting uncertainty, the upper limit of the uncertainty can be stated.

FURTHER READING

Books

Hofmann, D. (1986) *Handbuch Meβtechnik und Qualitätssicherung*, 3rd edn, VEB Verlag Technik Berlin.

Sydenham, P.H. (1983) *Handbook of Measurement Science I, II*, John Wiley & Sons, New York.

Standards and guides

ISO 2602:1980 *Statistical Interpretation of Test Results – Estimation of the Mean – Confidence Interval*.

ISO 2854:1976 *Statistical Interpretation of Data – Techniques of Estimation and Tests Relating to Means and Variances*.

ISO 3207:1975 *Statistical Interpretation of Data – Determination of a Statistical Tolerance Interval*.

ISO 3301:1975 *Statistical Interpretation of Data – Comparison of two Means in the Case of Paired Observations*.

ISO 3494:1976 *Statistical Interpretation of Data – Power of Tests Relating to Means and Variances*.

ISO 3534-1:1993 *Statistics – Vocabulary and Symbols – Part 1: Probability and General Statistical Terms*.

ISO 3534-2:1993 *Statistics – Vocabulary and Symbols – Part 2: Statistical Quality Control*.

ISO 3534-3:1999 *Statistics – Vocabulary and Symbols – Part 3: Design of Experiments*.

ISO 5479:1997 *Statistical Interpretation of Data – Tests for Departure From the Normal Distribution*.

ISO 8595:1989 *Interpretation of Statistical Data – Estimation of a Median*.

ISO/TR 13425:1995 *Guide for the Selection of Statistical Methods in Standardization and Specification*.

ISO 16269-7:2001 *Statistical Interpretation of Data – Part 7: Median – Estimation and Confidence Intervals*.

DIN 1319:1985 *Grundbegriffe der Messtechnik. Teil 4: Behandlug von Unsicherheiten bei der Auswertung von Messungen*.

EA-2/04, EA:1997 *Expression of the Uncertainty of Measurement in Calibration*.

Guide to the Expression of Uncertainty in Measurement (GUM). BIPM/IEC/ISO/OIML, 1993.

International Vocabulary of Basic and General Terms in Metrology (VIM). BIPM/IEC/IFCC/ISO/OIML/IUPAC/IUPAP, 1993.

ISO/IEC Guide 2:1991 *General Terms and their Definitions Concerning Standardization and Related Activities*.

53: Calculation and Treatment of Errors

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Errors in Measurement 328
2 Classification of Errors 329
Related Articles 330
References 330

1 ERRORS IN MEASUREMENT

All measurements contain some variability, which obscures the true value of the measurand determination (Galyer and Shotbolt, 1990; Doebelin, 2003; Gregory, 1973; Helfrick and Cooper, 1990; Hofmann, 1982; Rangan, Sarma and Mani, 1983; Taylor, 1982).

The variability in a measurement is called the *error of measurement*, which is defined as follows: The error of a measurement is the discrepancy or difference between the result of a measurement and the absolute true value of the quantity measured.

An error of measurement is of little practical use when considered in isolation. A number of measurements of the same true value, each of which has an associated error of measurement, can be used to provide valuable information regarding the accuracy and repeatability of measurements taken.

Thus, measurement is more than just reading the display on the instrument and recording the data. It involves the process of treating and manipulating measurement data to gain maximum benefit from them by extracting as much information as possible from the data.

Although this process of data reduction causes a loss of information about each specific measurement, an overall benefit is achieved by the result of summarizing the typical characteristics of large amounts of data (Weinberg, 1975). McGhee *et al*. (1996) describes the process of treating measurement data as consisting of the following:

- Assessing the consistency of the data.
- Estimating the uncertainties in the result.
- Predicting the result from theory.
- Comparing the data with physical theories and previous, related experimental work.

It is very well known that no measurement is exact. However, it should be made to an acceptable degree of accuracy (Morris, 1997). It is, therefore, important to specify the uncertainty of the measurement along with the result of the measurement. Representing uncertainty numerically and graphically, as given by McGhee *et al*. (1996), is also described in **Article 55, Uncertainty Determination, Volume 1**. There are a number of different sources for measurement error, such as the measurement method, the measuring instrumentation, or the measurement environment. Therefore, it is important to identify the most significant source of errors so that less significant ones may be neglected if sufficiently small.

There are also a large number of terms used to describe different types of errors, such as systematic errors, random errors, parasitic error, method error, observer error, parallax error, interpolation error, error of indication, repeatability error, rounding error, discrimination error, hysteresis error, response error, datum error, zero error, influence error, temperature error, supplementary error, total error, and so on. The first step to be taken in the reduction of errors is understanding them. Some of these errors are covered here.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

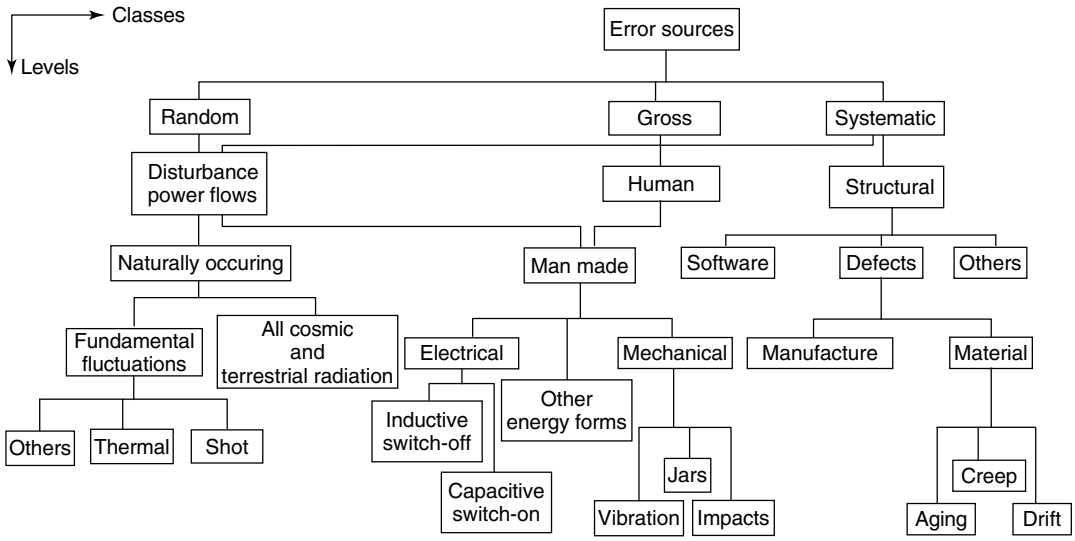


Figure 1. Classification of measurement errors.

2 CLASSIFICATION OF ERRORS

Errors can be classified according to their characteristics. Thus, Figure 1, which shows a summary of the three main groups of error classes, distinguishes between gross errors, systematic errors, and random errors (McGhee *et al*., 1996).

2.1 Gross errors

These are largely human errors. They may be caused by man-made power disturbance flows in any of the COMETMAN energy forms described in **Article 109, Structure and Energy in Sensor Systems, Volume 2**. They also include misreading of instruments leading to incorrect recording of measurement results, which may compound any error due to miscalculation. Other possibilities are incorrect adjustment of an instrument or improper use of an instrument and any computational mistakes. Although gross errors may be difficult to eliminate entirely, they should be anticipated and corrected. Some gross errors can be easily detected but others are difficult to detect.

Gross errors cannot be treated mathematically. They can be avoided only by taking care in the approach and by using error detection routines. For example, errors can be minimized by taking more than one set of readings under the same conditions, or by taking more than one set of readings using different instruments or operators.

In high-precision applications, some measurements need to be taken for the same quantity under the same conditions, by using different operators. In some conditions, the error associated with an operator is included in the estimation of the measurement error.

2.2 Systematic errors

These errors are due to systematic disturbance power flows, caused by human activities, or by structural imperfections or defects in software, inadequacies in manufacturing methods, inadequacies in materials of instrument construction, or some other unidentifiable or unspecified causes.

When a measurand is repeatedly measured under the same conditions, systematic errors always have the same value. In other words although they are constant under constant conditions, they are caused by intrinsic errors arising from long-term variations.

Systematic errors have the property that the next individual measurement error value can be predicted once the error law has been determined.

Systematic errors are divided into instrumental and environmental errors according to their sources. They can also be classified according to their behavior into dynamic and static errors.

- Instrumental errors are defined as imperfections of the measuring instrument. They are inherent in measuring instruments because of their structure. There are many types of instrumental errors, which depend upon the instrument. Instrumental error can be avoided by selecting a suitable instrument for the particular measurement application, calibrating the instrument against a standard, adjusting the instrument before using it, and by applying the estimated factors of correction. Instrumental errors in a measuring instrument can be accounted for, approximately, by replacing it with an instrument of higher accuracy. In this case, the systematic error in

the instrument under test is the difference between the two results.

- Environmental errors are caused by conditions external to the measuring instrument. These conditions include changes in temperature, pressure, humidity, interference, magnetic field, electrostatic field, and many more. If it is possible to control the measuring environment by providing air-conditioning, then calibration accuracy may be improved. Also, if certain components can be hermetically sealed, then errors due to adverse environmental effects may be reduced. When measurements are conducted using appropriate shielding techniques, then errors due to electromagnetic interference can be totally removed.
- Static errors are caused by the limitation of the measuring instrument or the physical law governing its behavior.
- Dynamic errors are caused by the inability of the instrument to respond fast enough to follow the changes in the measurand. However, although systematic errors always exist in any measurement, they can be minimized to a degree suitable to the aimed uncertainty of measurement.
- The terminology for static systems is covered in **Article 58, Description of Accuracy, Linearity, and Drift, Volume 1** and **Article 57, Static Considerations of General Instrumentation, Volume 1**; dynamic description is covered in **Article 61, First-order System Dynamics, Volume 1**; **Article 59, Introduction to the Dynamic Regime of Measurement Systems, Volume 1**; **Article 62, Second-order System Dynamics, Volume 1**; and **Article 60, Zero-order System Dynamics, Volume 1**.

2.3 Random errors

Even when all systematic errors have been accounted for, a residual error will invariably occur. This residual error, due to unspecified or unknown causes, is normally assumed to have a random nature. Random errors are unpredictable in that the next occurring single data value cannot be predicted with certainty, whereas the behavior of the group can arise through irregular chances of random causes normally due to fundamental fluctuations or extraterrestrial generation.

McGhee, Henderson and Sankowski (1986) provide a comprehensive but incomplete list of starting point references. The random sources cause changes in the reading for each repeated measurement under the same conditions. In well-designed experiments, few random errors usually occur, but they become important in high precision measurements.

The only way to compensate for random errors is by taking a large number of readings and by using statistical methods to obtain the best approximation of the true value of the measurand.

In many cases, it is difficult to distinguish between systematic errors and random errors. This can be seen in the summary of the random and systematic errors in measurement and their sources, which is given in Figure 1.

RELATED ARTICLES

Article 48, Common Sources of Errors in Measurement Systems, Volume 1; **Article 49, General Characterization of Systematic and Stochastic Errors, Volume 1**; **Article 55, Uncertainty Determination, Volume 1**.

REFERENCES

Doebelin, E.O. (2003) *Measurement Systems: Application and Design*, 5th edn, McGraw-Hill, New York.

Galyer, J.F.W. and Shotbolt, C.R. (1990) *Metrology for Engineers*, Cassell Publishers, London.

Gregory, B.A. (1973) *An Introduction to Electrical Instrumentation*, The MacMillan Press, London.

Helfrick, A.D. and Cooper, W.D. (1990) *Modern Electronic Instrumentation and Measurement Techniques*, Prentice Hall, NJ.

Hofmann, D. (1982) Measurement Error, Probability, and Information Theory, in *Handbook of Measurement Science: Vol I Theoretical Fundamentals* (ed. P.H. Sydenham), John Wiley & Sons, Chichester.

McGhee, J., Henderson, I.A. and Sankowski, D. (1986) Functions and Structures in Measurement Systems: A Systems Engineering Context for Instrumentation. *Measurement*, **4**(3), 11–119.

McGhee, J., Henderson, I.A., Kulesza, W. and Korczynski, M.J. (1996) *Scientific Metrology*, ISBN 83-904299-9-3, A.C.G.M. LODART, Lodz.

Morris, A.S. (1997) *Measurement and Calibration Requirements*, John Wiley & Sons, Chichester.

Rangan, C.S., Sarma, G.R. and Mani, V.S.V. (1983) *Instrumentation Devices and Systems*, Tata McGraw-Hill Publishing, New Delhi.

Taylor, J.R. (1982) *An Introduction to Error Analysis*, University Science Books, CA.

Weinberg, G.M. (1975) *An Introduction to General Systems Thinking*, John Wiley & Sons, New York.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

54: Explanation of Key Error and Uncertainty Concepts and Terms

Luca P. Mari
Università Cattaneo, Castellanza (VA), Italy

1 Measurement Results and Their Quality 331
2 The Concept of Error 332
3 The Concept of Uncertainty 333
4 Characterizing a Measurement with its Uncertainty 333
5 The Expression of Measurement Results and Their Uncertainty 334
6 The Procedure for Assigning the Measurement Uncertainty: an Example 335
Related Articles 335
Further Reading 335

1 MEASUREMENT RESULTS AND THEIR QUALITY

Measurement is a peculiar means of acquiring and formally expressing information about empirical systems. It is aimed, therefore, at setting up a bridge between the empirical world and the linguistic/symbolic world, the domains of the systems under measurement, and measurement results respectively.

A fundamental evidence is that these two realms exhibit extremely different characteristics. Empirical systems are embedded in the space–time universe, and this generates their space and time dependency: any system is only partially isolated from its environment and its dynamic forces distinguishing between the system itself and its temporal versions, that is, the system states. On the other hand, symbolic entities such as numbers are coextensive with their definitions (in a sense: they *are* their definitions) so that they are always identical to themselves. Paradigmatically, noise exists in the empirical realm, not in the symbolic one; real numbers exist in the symbolic realm, not in the empirical one.

Whenever the two realms interact with each other, as measurement does by means of the mediation of quantities, these diversities (a) require the introduction of a concept of quality of the symbols (in our case, measurement results) chosen as representatives for empirical states and (b) are the cause of several issues affecting such a quality.

The typical operative context of measurement that is presented in Figure 1 shows that the required empirical results ('the output') can be in principle obtained as the transformation of the same empirical states ('the input') by a direct manipulation (an 'empirical procedure') or a transduction to information entities, to be processed and finally transduced back to the empirical realm (an 'informational procedure') shown in Figure 2.

The benefits of informational procedures are commonly recognized (basically due to the fact that it is much easier to deal with symbols than with empirical things), but they depend on the faithfulness of measurement results as representative entities for the corresponding empirical states.

Such faithfulness, and therefore the quality of measurement results, is limited as a consequence of causes related to the following:

- *Model of the system under measurement*: incompleteness, if not faults, in the definition of the measured quantity (the measurand), as in the case of an ill-characterized system dynamics or only a partial identification of the quantities influencing the measurand.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

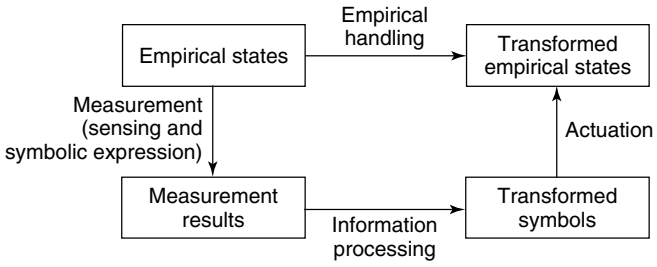


Figure 1. The role of measurement in the relations between empirical and symbolic realms.

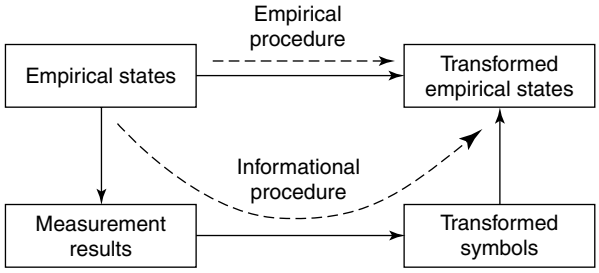


Figure 2. The (possible) equivalence of empirical and informational procedures.

- *Operative accomplishment of the measurement procedure*: poor repeatability, or stability, or selectivity of the adopted measuring system (*see also* **Article 104, Models of the Measurement Process, Volume 2**), if not faults in its usage.

The unavoidable presence of such flaws is the reason requiring us *to state any measurement result by expressing in symbols a measurand value together with an estimation of its deemed quality*.

2 THE CONCEPT OF ERROR

It is a well-known fact that the repeatability of measurements can be increased by

- improving the measuring system in its empirical characteristics,
- reporting the results with a reduced number of significant figures,

that is, by adjusting the sensing device or modifying the symbolic expression respectively (see Figure 3).

The repeatability of a measurement, and in more general terms its quality, is therefore a relative characteristic to be evaluated in reference to the goals for which the operation is performed and the available resources (in epistemological terms, this can be thought of as a confirmation that a concept of absolute, or complete, precision is simply meaningless).

It is amazing in this perspective to note that the indication of the estimated quality of the results became customary in physical measurement only in the late nineteenth century, and, however, several decades after the Theory of Error provided by Gauss at the beginning of that century. A plausible reason for this can be recognized in the commonly (in the past) assumed hypothesis that measurable quantities are characterized by a perfectly precise 'true value'. The choice of adopting the concept of error to model and formalize a less-than-ideal quality of measurements originates from this hypothesis: any discrepancy between the measuring system outputs and the measurand 'true value' should be taken into account as an error, and correspondingly dealt with (*see also* **Article 49, General Characterization of Systematic and Stochastic Errors, Volume 1**).

However,

- an error can be recognized as such only if a corresponding 'right entity' exists;
- errors can be corrected only if their corresponding 'right entities' are known;
- 'true values', that play the role of such 'right entities' in the case of measurement, are in principle unknown (otherwise measurement itself would be useless. . .) and cannot be operatively determined.

These assertions imply that the Theory of Error is grounded on metaphysical, empirically inapplicable bases. Consider the following two statements:

- 'at the instant of the measurement, the system is in a definite state';
- 'at the instant of the measurement the measurand has a definite value'.

Traditionally they would be considered as synonymous, whereas their conceptual distinction is a fundamental fact of metrology: the former represents a basic assumption for measurement (we are not considering here measurement in quantum mechanics), while the latter is epistemically unsustainable and, however, operationally irrelevant. Measurement results are symbolic, and not empirical, entities:

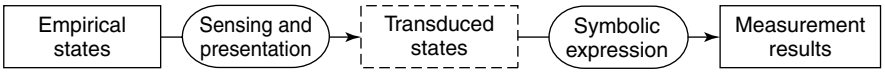


Figure 3. Abstract schematization of a measurement.

what in the measurement is *determined*, and therefore considered preexisting, is the system state, not the measurand value that is instead *assigned* on the basis of the instrument reading and the calibration information.

3 THE CONCEPT OF UNCERTAINTY

The search for a more adequate framework reached a crucial point about 30 years ago, when it was understood that a common approach for modeling and formally expressing a standard parameter describing the quality of measurement results was a condition for establishing strict cooperation among the national calibration laboratories. To build up and maintain mutual confidence between accreditation bodies and to build up and maintain compatibility for their calibration certificates required having the quality of their measurement results evaluated and expressed according to some harmonized protocol. To this goal, the *International Committee for Weights and Measures* (CIPM), started a project together with several international organizations involved in standardization (International Organization for Standardization (ISO), International Engineering Consortium (IEC), International Organization of Legal Metrology (OIML), ...): its final result is the *Guide to the Expression of Uncertainty in Measurement* (GUM), first published in 1993 and later introduced as a Standard by each of such organizations. While originally intended for calibration laboratories, the GUM is presently to be considered as the basis for expressing the results of any measurement performed in accordance with an international standard.

According to the GUM, the uncertainty of a measurement result is 'a parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand'. Apart from this rather classical definition, the most important innovation of the GUM stands in its recognition that the uncertainty of measurement results can be evaluated according to two distinct and complementary methods:

- some uncertainties, designated as 'of type A', are computed as suitable statistics of experimental data, usually obtained as repeated instrument readings;
- some other uncertainties, designated as 'of type B', are instead estimated on the basis of the observer's personal experience and the available a priori information, and therefore express a degree of belief on the possible measurand values.

The recognition that even measurement, an operation traditionally deemed as the paradigm of objective information acquisition, requires the introduction of some subjective evaluation is of capital importance. Therefore, the shift from 'error' to 'uncertainty' is far more than a terminological issue, and witnesses a conceptual transition from an ontological position to an epistemic one: according to the GUM standpoint, to establish the quality of measurement results is an issue related to the *state of knowledge* of the measurer, and therefore 'absolute quality' cannot be reached simply because some *intrinsic uncertainty* is always part of the measurement system.

The possible sources of uncertainty that the GUM itself lists are exemplar in this regard: together with the 'variations in repeated observations of the measurand under apparently identical conditions', the reason usually recognized for random variability, and some causes related to instrumental issues such as 'approximations incorporated in the measurement procedure' and 'instrument resolution or discrimination threshold', the GUM identifies several epistemic sources, among them the incomplete definition of the measurand and the imperfect realization of its definition (*see also* **Article 48, Common Sources of Errors in Measurement Systems, Volume 1**).

4 CHARACTERIZING A MEASUREMENT WITH ITS UNCERTAINTY

To accomplish a measurement process, three distinct activities must be sequentially performed:

- *Acquisition*: by means of a sensing device, the measurand is transduced to a quantity suitable for direct access by the measurer (e.g. the angular position of a needle with respect to a reference scale), possibly through the mediation of an 'intermediate' quantity (a typical role for electrical quantities) to drive processing and presentation devices (see Figure 4).
- *Evaluation*: the access to the transduced quantity (i.e., the instrument reading) concludes the empirical part of the operation; by gathering and processing the available information (the transduced quantity itself, together with everything that is known on the measurement system: the measurand definition and realization, the instrument calibration diagram, the values of relevant influence quantities, etc.), the measurer evaluates the

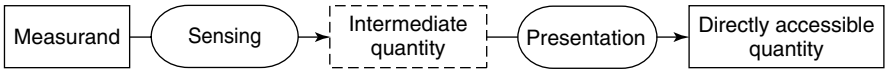


Figure 4. Abstract schematization of the empirical component of a measurement.

measurand value and its uncertainty; this inferential process is based on both objective and subjective information.
- *Expression*: the obtained information is expressed in symbolic form according to an agreed formalization.

It should be noted that the same information could be in principle expressed in different forms for different needs, by adopting, typically, a statistical or a set-theoretical formalization (or some generalization of the latter, as in the case of representations based on fuzzy sets: we will not deal with such generalizations here). Consider the traditional indication, $x \pm y$, that admits two distinct interpretations:

- the measurand value is expressed as the scalar x, with y as its estimated uncertainty;
- as the measurand value, the whole interval $(x - y, x + y)$ is taken, whose half-width, y, expresses the quality (sometimes called *precision*) of such a measurement result.

Neither of them is the 'right one': they should be selected according to the specific application requirements.

The GUM adopts this approach, and while basing its procedure on the first interpretation, recognizes that 'in some commercial, industrial, and regulatory applications, and when health and safety are concerned', it is often necessary to express the measurement results by means of intervals of values.

Measurement results must be therefore assigned according to the goals for which the measurement is performed; they are adequate (and not 'true') if they meet such goals. By suitably formalizing them, the measurer is able to express the available information of both the measurand value and its estimated quality. Quoting the GUM again, no method for evaluating the measurement uncertainty can be a 'substitute for critical thinking, intellectual honesty, and professional skill': indeed 'the quality and utility of the uncertainty quoted for the result of a measurement ultimately depends on the understanding, critical analysis, and integrity of those who contribute to the assignment of its value'.

5 THE EXPRESSION OF MEASUREMENT RESULTS AND THEIR UNCERTAINTY

For both type A and type B evaluation methods, the GUM assumes that measurands (but the same holds for all the quantities involved in the measurement system: influence quantities, correction factors, properties of reference materials, manufacturer or reference data, etc.) can be formalized as random variables, and as such characterized by statistical parameters:

- The measurand value is estimated as the mean value of the random variable; in the case of type A evaluations, for which an experimental population X of n repeated reading data x_i is available, it is computed as

$$m(X) = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

- The uncertainty of the measurand value is estimated as the standard deviation of the measurand value, being itself a random variable; this parameter is termed as the GUM *standard uncertainty* and denoted $u(m(X))$; in the case of type A evaluations it is computed as

$$u(m(X)) = \sqrt{\frac{1}{n}\frac{1}{n-1}\sum_{i=1}^{n}(x_i - m(X))^2} \tag{2}$$

Measurement results can then be reported, for example, as $m_S = 100{,}021\ 47(35)\,\mathrm{g}$, meaning that the evaluated mass m of the system S (whose specification should include the indication of the operative condition in which the measurement has been performed) is 100,021 g with a standard uncertainty of 0,35 mg.

The same couple of values (measurand value, standard uncertainty) is adopted to express measurement results as intervals. To this goal, a *coverage factor* k (typically in the range 2 to 3) is introduced, such that $U(X) = ku(m(X))$, termed *expanded uncertainty*, is adopted as the half-width of the interval representing the measurement result: $[m(X) - U(X), m(X) + U(X)]$, commonly written as $m(X) \pm U(X)$ (if the probability distribution of the random variable is known, this interval can be thought of as a *confidence interval* whose confidence level depends on k).

In the case of *derived measurement*, that is, when the measurand Y is a quantity depending on N input quantities $X_i, i = 1, \ldots, N$:

$$Y = f(X_1, \ldots, X_N) \tag{3}$$

and for each quantity X_i, the estimated value $m(X_i)$ and uncertainty $u(m(X_i))$ are given, the issue arises of how to obtain the corresponding values $m(Y)$ and $u(m(Y))$ for Y.

The measurand value $m(Y)$ is simply obtained by introducing the estimates $m(X_i)$ in the model function f:

$$m(Y) = f(m(X_1), \ldots, m(X_N)) \tag{4}$$

The uncertainty $u(m(Y))$ is instead evaluated by means of the so-called *law of propagation of uncertainty*, that for

statistically noncorrelated quantities is

$$u^2(m(Y)) = \sum_{i=1}^{N} c_i^2 u^2(m(X_i)) \tag{5}$$

where the 'sensitivity coefficients' c_i that define the extent to which Y is influenced by variations of the input quantities X_i are computed as

$$c_i = \frac{\partial f}{\partial X_i} \text{ evaluated at } X_i = m(X_i) \tag{6}$$

In the general case of correlated input quantities (i.e. their covariance $u(m(X_i), m(X_j)) \neq 0$), (5) becomes

$$u^2(m(Y)) = \sum_{i=1}^{N} c_i^2 u^2(m(X_i)) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} c_i c_j u(m(X_i), m(X_j)) \tag{7}$$

in which the combined standard uncertainty of the measurement result $m(Y)$ is computed on the basis of a first-order Taylor series approximation of (3) (*see also* **Article 55, Uncertainty Determination, Volume 1**; **Article 52, Error Models, Error Budgets and their Calculation, Volume 1**).

6 THE PROCEDURE FOR ASSIGNING THE MEASUREMENT UNCERTAINTY: AN EXAMPLE

Measurement uncertainty is a *pragmatic* parameter: its value is not intrinsic to the measurand but is to be established in reference to the specific goals according to which the measurement is performed. No 'true uncertainty' exists, and the preliminary step of a procedure aimed at assigning a value to the uncertainty of a measurand value is therefore to decide a *target uncertainty*, the maximum value of uncertainty compatible with the given goals. In any step of the procedure, if the estimated value is reliably considered less than such a target uncertainty, then the procedure should be stopped with a positive result: the measurand can be evaluated with a satisfying uncertainty, and no further resources are required to refine the procedure. On the other hand, whenever the estimated uncertainty becomes greater than the target uncertainty, the procedure must be definitely stopped with a negative outcome, conveying the information that better measurements are required to meet the specified target uncertainty.

According to the GUM viewpoint, *any* measurand Y should be actually evaluated by derived measurement, that is, by firstly identifying its dependence on a set of 'input quantities' X_i, such as influence quantities, calibration parameters, correction factors, and so on. For each X_i, the values $m(X_i)$ should be obtained by statistical or other methods (e.g. as part of instrument specifications), and for each of such $m(X_i)$, the corresponding standard uncertainties $u(m(X_i))$ and covariances $u(m(X_i), m(X_j))$ should be evaluated, again by either type A or type B procedures.

In the case in which the functional relation f is known in its analytical form, the sensitivity coefficients c_i can then be computed; if, on the other hand, the complexity of the measurement system prevents the explicit formalization of (3), each coefficient c_j can be experimentally estimated by a suitable setup of the system in which Y is repeatedly measured while all the quantities X_i but X_j are kept constant.

When at least some $u(m(X_i))$, $u(m(X_i), m(X_j))$, and c_i are available, (6) can be computed to obtain an estimation of the measurand uncertainty $u(m(Y))$, to be compared to the specified target uncertainty.

RELATED ARTICLES

Article 10, Extensions of the Representational Theory of Measurement, Volume 1; **Article 48, Common Sources of Errors in Measurement Systems, Volume 1**; **Article 52, Error Models, Error Budgets and their Calculation, Volume 1**; **Article 55, Uncertainty Determination, Volume 1**; **Article 104, Models of the Measurement Process, Volume 2**.

FURTHER READING

CIPM, BIPM (1981) *Procès-Verbaux des Séances du Comité International des Poids et Mesures*, **49**, 8–9, 26, (in French); Giacomo, P. (1982) News from the BIPM, *Metrologia* **18**, 41–44.

ISO (1993) *Guide to the Expression of Uncertainty in Measurement*, International Organization for Standardization, Geneva.

ISO (1993) *International Vocabulary of Basic and General Terms in Metrology*, 2nd edn, International Organization for Standardization, Geneva.

ISO (1999) *Geometrical Product Specification – Part. 2: Guide to the Estimation of Uncertainty in Measurement in Calibration of Measuring Equipment and Product Verification*, International Organization for Standardization, Geneva.

Mari, L. and Zingales, G. (2000) Uncertainty in Measurement Science, in *Measurement Science – A Discussion* (eds K. Karija and L. Finkelstein), IOS Press, Ohmsha, (pp. 135–147).

Taylor, B.N. and Kuyatt, C.E. (1994) *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*, NIST Technical Note 1297 (available on the NIST website).

55: Uncertainty Determination

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 General Comment on Uncertainty 336
2 Representing Numbers and Uncertainties 336
3 Example of Uncertainty in Determining Temperature 339
4 Rounding Off Measurement Results 339
5 Example of Rounding off of Laboratory Measurements 340
6 Prose Methods of Results Presentation 341
7 Tabular Methods of Results Presentation 341
8 Graphical Methods of Results Presentation 341
9 Use of Equations to Describe Results of Measurement 342
10 Comparison of Methods for Presenting Measurement Results 343
Related Articles 343

1 GENERAL COMMENT ON UNCERTAINTY

Numerical information, like most other kinds of information, can always be presented in ordinary prose or text. However, unless the amount of numerical information is small, this method of presentation is clumsy. Consequently, other methods of presenting technical information are often used.

In many cases, the corresponding values of two or more variables have to be presented. It is, therefore, often more appropriate to use tables, graphs, or equations to show the relationships between variables. Before these methods are discussed in detail, a few points regarding units, uncertainties, and rounding off need to be settled.

Whenever numerical information is presented, it must be made clear what unit is being used; this is obvious but occasionally forgotten. In most technical writing, the units employed belong to the SI system.

2 REPRESENTING NUMBERS AND UNCERTAINTIES

Before continuing it is necessary to introduce the manner in which a number is represented. Suppose the number five thousand three-hundred and forty-two point 00... is to be represented. Following the usual procedure using digits, the number is 5342.00... This is the correct representation. However, if the number is fifty thousand three-hundred and forty-two point 00..., then writing it as 50 342.00... is a contravention of the relevant international standard. The standard recommends that, for numbers equal to or greater than ten thousand, the thousands should be separated from the hundreds by a space and not by a comma. The correct representation is thus 50 342.00....

For numbers greater than this, the digits should be grouped in threes, corresponding to increases of 10^3.

If the long-hand version of 2.35×10^6 is required, then it should be written as 2 350 000. In a similar way, the fractional decimal part should be written by grouping the decimal digits in threes so that 2.354×10^{-1} should be written as 0.2354, while 2.3546×10^{-1} should be written in long-hand as 0.23546 (while the standard calls for these formats, common usage still will place a point or comma as the separator).

When numerical information is quoted, the intention is to achieve a higher degree of precision than would otherwise be possible. However, perfect precision cannot be achieved.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

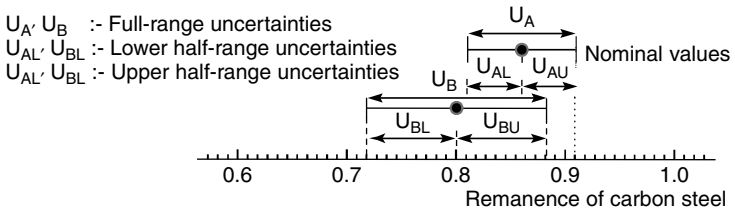


Figure 1. Graph comparing interval estimators and illustrating notation for uncertainties.

Precision is a statement about how exact a particular value is. Almost every quoted numerical value has some degree of uncertainty associated with it.

Quite often there is an interest in knowing just how large the uncertainty is. In particular, it is often impossible to make meaningful comparisons between numerical values unless their uncertainty is known. The combination of the numerical value and the uncertainty is called an *interval estimator*.

Interval estimators may be illustrated as shown in Figure 1. The full-range uncertainty defines the band within which the true value of the measurement is most likely to be located, while the half-range uncertainties, sometimes also called *tolerances*, define the upper bound and lower bound of this range.

Suppose, for example, that the measured remanence of carbon steel is reported as 0.86 ± 0.05 T. Here, 0.86 T is the measured value, and ± 0.05 T is the uncertainty in that value. This style of presentation indicates that the remanence is believed to lie close to 0.86 T and that any value between 0.81 and 0.91 T is also acceptable as being in agreement with the reported result.

If a second report specifies a measured value of $0.80\,\text{T} \pm 10\%$ for the same quantity, this means that values in the range 0.72 to 0.88 T are acceptable. Values in the range 0.81 to 0.88 T are acceptable to both writers. It can, therefore, be seen that the two measured values do not differ significantly. Although such a graphical representation is not of much value where only two values are being compared, it can be useful where three or more values are involved.

Obviously, any comparison of values depends on some knowledge of the uncertainty in both, or all, of the values being compared. In the above example, if only the two measured values of 0.86 and 0.80 T were reported, then it is not possible to judge whether the difference is significant or not. Specification of only these values is widely known as giving point estimators.

Where numerical information is presented in ordinary prose, in text, in a table, or by means of an equation, the statement of uncertainties is usually straightforward. In graphs, the method of representation shown in Figure 2 extended to two dimensions is used. For example, suppose that $X = 10 \pm 1$ and $Y = 3.0 \pm 0.5$. Bars are drawn

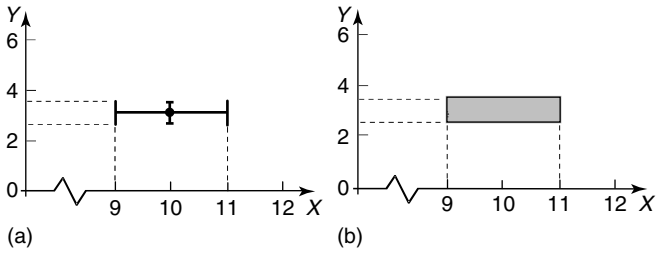


Figure 2. Interval estimators: (a) in the X and Y directions and (b) area of uncertainty of the estimators.

in the X direction from 9 to 11, and in the Y direction from 2.5 to 3.5. Hence, interval estimators are also known as uncertainty bars as the example in Figure 2(a) illustrates. This style of representation leads to drawing the point (10, 3) with the given uncertainty rectangle. However, it is acceptable to say that the 'correct' point is anywhere within the shaded rectangle shown in Figure 2(b).

To appreciate how the interpretation of a graph may be aided by the inclusion of interval estimators, consider firstly the points plotted in Figure 3(a).

Is there a 'kink' in the graph at about $X = 10$, or does the suggestion of a 'kink' result merely from random errors in the measured values?

On the information given in Figure 3(a), it is not possible to answer this question. Figure 3(b) and (c) shows the same plotted points as does Figure 3(a), but with interval estimators added. In Figure 3(b), the uncertainties are small, and the reader accepts that there is a kink in the graph occurring at about the value $X = 10$. This 'kink' probably corresponds to a real physical effect.

In Figure 3(c), where the uncertainties are much larger, it is clear that the information presented does not clearly establish whether any such 'kink' exists.

In practice, it is usually not necessary to show uncertainty bars on every point on a curve. To avoid unnecessary clutter, they can be shown on only two or three points, which are typically one near each end of the curve, with perhaps one about the center as well. If the uncertainties are small, they will be invisible on the graph. In such cases, their values should be stated elsewhere.

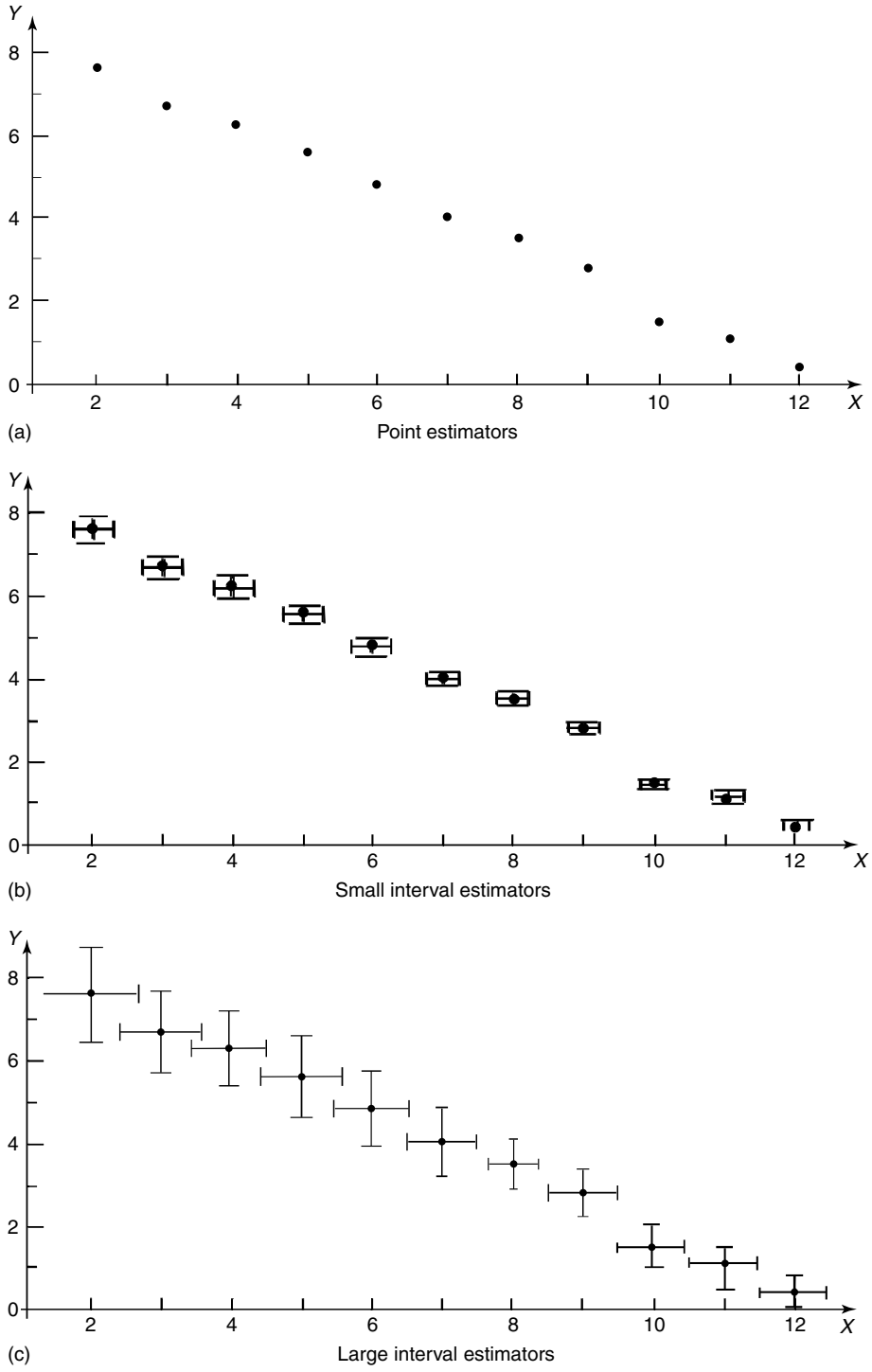


Figure 3. Graphical kink with different types of estimators: (a) point estimators, (b) small interval estimators, and (c) large interval estimators.

The above approach to the subject of uncertainties has been oversimplified. According to it, if a value is quoted as 3.5 ± 0.5, this means that a value of 4.01 will be interpreted as being in total disagreement with the result. This is clearly an unrealistic view.

To develop a more realistic approach to the subject of uncertainties, it is necessary to use statistical concepts.

The ability to estimate the uncertainty that exists in a measured value is an important part of the skill required

by experimenters. It often involves a blend of science and intuition.

The theory of probability and combinations provides the foundation upon which the calculation and treatment of errors is based.

3 EXAMPLE OF UNCERTAINTY IN DETERMINING TEMPERATURE

Four different experimenters measure the same temperature, and report their results as (i) $231.0 \pm 2.5\,^{\circ}\mathrm{C}$ (ii) $245\,^{\circ}\mathrm{C} \pm 5\%$ (iii) $227.2 \pm 5.2\,^{\circ}\mathrm{C}$ (iv) $225.0\,^{\circ}\mathrm{C} \pm 5\%$

Draw a diagram that will allow these results to be compared easily. Comment upon the results. Do they agree?

3.1 Solution

Before sketching the interval estimators it is necessary to convert the percentage tolerances to actual values of the measured temperature. Thus, for (ii), 5% of 245 equals 12.25 and 5% of 225 equals 11.25. The interval estimators may now be drawn as shown in Figure 4.

Examination of Figure 4 shows that the measurements agree due to the overlapping range of uncertainty indicated by the shading. This conclusion is justified for the following reasons.

Measurements (i) and (iii) have nominal values lying within the tolerance band for that of (iv). Hence, measurements (i), (ii), and (iv) exhibit good agreement. The nominal value of measurement (ii) lies outside the tolerance bands of (i), (iii), and (iv). However, as its lower half-range uncertainty totally includes the full-range uncertainty of measurement (i), then measurements (i) and (ii) may be taken to be in agreement. Because of this and in the interests of noncontradictory and consistent analysis, it must be concluded that all four measurements agree within the limits of the tolerances given.

Further justification of this opinion can be obtained by examining the relationship between the interval estimators for measurements (ii), (iii), and (iv). The nominal value and upper half-range uncertainty of measurement (iii) and part of the upper half-range uncertainty of measurement (iv) are also included in the lower half-range uncertainty of (ii).

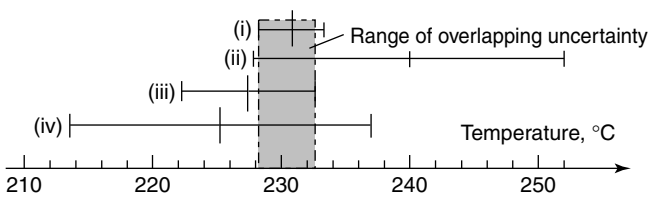


Figure 4. Interval estimators used to compare different measures of the same quantity.

The state of affairs illustrated by this example is far from satisfactory from the measurement point of view. Steps should be taken to see if paying closer attention to the measuring method could reduce the differences between the measured nominal values and the uncertainties, the instrumentation used, and the treatment of the measuring data from which the estimators are obtained.

4 ROUNDING OFF MEASUREMENT RESULTS

When a numerical value is quoted, a choice as to the number of digits to be used will need to be made, that is, it is necessary to decide whether to round-off the number by discarding some of the digits at its right-hand end. Where the first digit to be discarded is 5 or higher, the last digit retained is normally rounded up. Thus, 4.83500 rounded off to three significant figures becomes 4.84.

Suppose a numerical value is quoted as being 5.12. Usually, this does not mean 5.12 precisely; it means any number which, when rounded off to three significant digits, becomes 5.12. Thus, it means any number between 5.1150000... and 5.1249999..., that is, it practically implies 5.12 ± 0.005. Thus, rounding-off to the second figure after the decimal point has introduced an implied uncertainty of ± 5 in the digit immediately following the last digit quoted.

A number that is correctly rounded off fulfills two requirements. Firstly, the implied uncertainty is much smaller than the actual uncertainty, so that no meaningful information is lost in the process of rounding-off. Secondly, the implied uncertainty is not so small that the figures quoted suggest a higher degree of precision than is justified.

In cases in which the uncertainty is not calculated, a rough idea of how large the uncertainty might be is usually available. In rounding off, it is important to ensure that both these requirements are met. Where the uncertainty is calculated, it is possible to be more specific.

Suppose, for example, that it is required to quote the circumference of a circle, having measured its diameter as 20 mm with uncertainty of 1/30 mm in either direction. The circumference is then $\pi(20 \pm 1/30)$ mm and may be evaluated as $(62.831853 \pm 0.10471976)$ mm.

Looking at the uncertainty term, it can be seen that all digits before the decimal point are zero, while there is a 1 in the first place after the point. In the nominal value (62.83...), therefore, the first two digits are not directly subject to any uncertainty, while the third digit, which is the 8, in the first place after the decimal point, could reasonably

go either up or down by 1. With this uncertainty in the third digit, if we wish to ensure that the implied uncertainty is much smaller than the actual uncertainty, all digits beyond the fourth should certainly be discarded. On this basis, it is reasonable to round off the value of the circumference to (62.83 ± 0.10) mm.

This example suggests a useful working rule for rounding off, which can be adopted, as follows:

4.1 Rule

In rounding off, include one digit beyond the first uncertain digit, both in the nominal value and in the uncertainty.

The two exceptions to the rounding-off rule are as follows:

- Where a whole number ends in a zero. For example, 105 000 does not necessarily imply $105\,000 \pm 0.5$. To make the implied uncertainty clear, this number could be written as 105×10^3, $105.00 \times 10^3 \ldots$ as appropriate.
- Where the number quoted has no uncertainty such as in the statement that 1 inch = 2.54 cm, the number 2.54 is absolutely accurate by definition, or when the number quoted can take on only well-defined specific values, such as whole numbers. For example, the valency of oxygen is 2, or the number of cylinders in a certain car engine is 4.

5 EXAMPLE OF ROUNDING OFF OF LABORATORY MEASUREMENTS

A standards laboratory measured the nominal reference value, x_r, of a quantity as 19.5081 and the reference value uncertainty, u_r, as ± 0.8506. Round these values off and sketch the interval estimator representing it.

When the six laboratories, A, B, C, D, E, and F, use different measuring techniques to measure the same quantity, the nominal value and the uncertainty obtained by each method are as given in Table 1. Correct any mistakes in the rounding off, sketch the uncertainty intervals for each method and comment upon the result of each of the measuring methods relative to that of the standards laboratory.

Table 1. Reported measured nominal values and uncertainties of six laboratories.

	A	B	C	D	E	F
Nominal	18.503	20.5	19.0	21.0	19.5	20.0
Uncertainty	0.374	0.205	0.571	0.525	0.585	1.20

5.1 Solution

Applying the rounding-off rule to the reference measured value gives a rounded-up value of 19.51. In a similar way, the rounded-off uncertainty is 0.85. The corrected statement of the reference value is thus 19.51 ± 0.85, which is illustrated in Figure 5.

The correctly rounded-off values for the results quoted in Table 1 are given in Table 2.

The interval estimators may now be drawn as shown in Figure 6.

Although the measurements conducted by laboratories A, B, C, E, and F agree with the reference value, measure D

Table 2. Correctly rounded-off nominal values and uncertainties of six laboratories.

	A	B	C	D	E	F
Nominal	18.50	20.50	19.00	21.00	19.50	20.0
Uncertainty	0.37	0.21	0.57	0.53	0.59	1.2

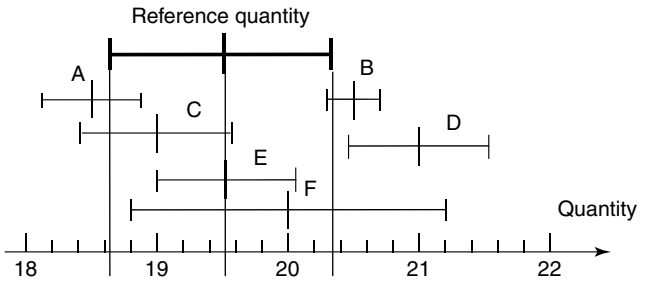


Figure 6. Interval estimators for the nominal values of A to F and their uncertainties.

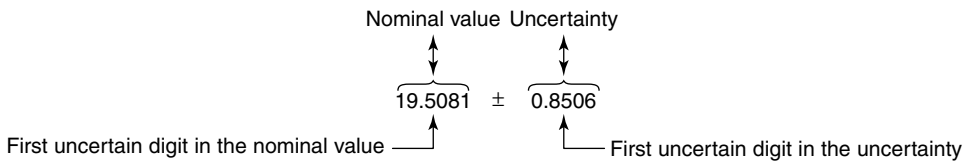


Figure 5. Illustration of rounding-off numerical values.

Table 3. Comments upon comparison of rounded-off values for the measurements of six laboratories.

	Values	Comment
Reference	19.50 ± 0.85	tolerance band is ±5% of nominal value
A	18.50 ± 0.37	nominal value outside tolerance band of reference; equipment requires recalibration
B	20.50 ± 0.21	nominal value outside tolerance band of reference; equipment requires recalibration; check measuring method
C	19.00 ± 0.57	nominal value inside tolerance band of reference; acceptable but check equipment and measuring method
D	21.00 ± 0.53	nominal value outside tolerance band of reference; equipment requires recalibration; check measuring method
E	19.50 ± 0.59	good consistency with the reference; measuring method is acceptable
F	20.0 ± 1.2	nominal value within tolerance band of the reference; check method to see if errors may be reduced

does not. The situation is not at all satisfactory. Table 3 contains appropriate comments.

6 PROSE METHODS OF RESULTS PRESENTATION

As has been mentioned above, the prose method of presentation is unattractive where a large amount of numerical information is involved. It can be useful for small amounts, as for example:

> The sample tested showed a resistivity of 12.42 Ωm ±2% when tested with an applied voltage of 100.0 V ±1% at a temperature of $(20 \pm 0.5)\,^{\circ}$C.

This example is not quite in accord with recommendations given above regarding uncertainties and rounding off. However, even partial adherence to these recommendations can make the writing seem rather pedantic. Consequently, there is a temptation to be less precise. Alternatively, a rearrangement of the information given in the above example is possible:

> The sample tested gave the following result:
> Applied voltage 100.0 V ±1%
> Temperature $(20 \pm 0.5)\,^{\circ}$C
> Measured resistivity 12.42 Ωm ±2%.

Obviously, this kind of list can be extended to include quite a large amount of numerical information without becoming as objectionable as a long prose statement containing a lot of numbers. Such a list in effect constitutes a one-dimensional table.

7 TABULAR METHODS OF RESULTS PRESENTATION

Tables can be useful where corresponding values of two or more variables are to be presented as in the example in Section 5.

A table is preferable to a graph where:

- the amount of information presented is small; there is little to be said for a graph that shows only two or three points;
- the information is to be presented with a higher degree of precision than can be achieved in reading values from a graph.

Compared with a graph, a table has the following disadvantages:

- the form of the relationship between the variables is not immediately apparent;
- interpolation between the values given is usually not easy.

8 GRAPHICAL METHODS OF RESULTS PRESENTATION

A brief examination of almost any manufacturer's data sheet or any scientific or technical specialist journal will show that graphs are more widely used than any other method for presenting a set of numerical information.

A graph allows a fairly quick and transparent impression of the kind of relationship that exists between two variables.

Two or more graphs, plotted on the same axes, allow rapid comparisons to be made. Interpolation between the plotted values is usually fairly easy. Unfortunately, it is usually not possible to read values from a graph with a high degree of precision. Appropriate choice of scale and section of range can make it more precise.

Where there are no limitations in the space allowed, the ideal arrangement is often to present the numerical information in both a table and a graph.

9 USE OF EQUATIONS TO DESCRIBE RESULTS OF MEASUREMENT

When numerical information is obtained as a result of a theoretical analysis, the information naturally makes its first appearance in the form of equations.

In the case of numerical information obtained by measurement, it is also sometimes appropriate to express the information by an equation. This can lead to a statement like the following:

> For the ten cores tested, the measured power loss P (watts) lay within $\pm 5\%$ of the value given by $P = 120B_m^{2.7}$ where B_m is the peak operating flux density in Teslas. This equation is valid over the range $0.1\,\mathrm{T} \leq B_m \leq 1.7\,\mathrm{T}$

This kind of statement can summarize a lot of information in a small space.

Two points in this example are worth emphasizing. Firstly, the statement indicates the range over which the equation is valid. Secondly, it indicates how closely the measured values lie to the values given by the equation. Without these two pieces of information, the communication of the ideas would not be so effective. It is also necessary to exercise care by also giving information on the uncertainties in the measured values of B_m and P.

There are at least three different situations that can give rise to a statement of the kind illustrated in the above example.

In some cases, the equation will have been derived from theoretical considerations. The statement will then constitute a comparison between theory and measurement.

In other cases, when theory may indicate the form of the equation, the parameters will have to be evaluated from the measured results.

In yet another case, the measured values will be used to indicate the form of the equation as well as the values of the parameters. In this last case, the equation is referred to as an *empirical equation* as it has no theoretical basis.

Graphical methods are useful if it is necessary to check or find the form of an equation that measured results fit, and in determining the parameters of the equation. A discussion of this process is called *curve fitting*. Where there are no limitations in the space allowed, the ideal arrangement is often to present the numerical information in both a table and a graph, as also stated above.

Table 4. A summary of methods for information presentation.

Method	Advantages	Disadvantages
Ordinary prose or text	1. Useful for small amounts of information	1. Unattractive where the amount of numerical information is large 2. Writing can seem rather pedantic 3. Temptation to be less precise
Tables	1. Can include a large amount of numerical information 2. Less objectionable than a long prose statement 3. Preferable to a graph if the amount of information is small 4. Higher degree of precision than in reading from a graph	1. Interpolation between given values is not so easy
Graphs	1. Easy interpolation between values 2. Kind of relationship usually transparent 3. Rapid comparisons of two or more graphs possible	1. Usually not possible to read values from a graph with a high degree of precision
Equations	1. Result of a theoretical analysis 2. Appropriate way to express the information 3. Information density is high 4. Possibility to compare theory and experiment	1. Must be supported by experimental measurement

10 COMPARISON OF METHODS FOR PRESENTING MEASUREMENT RESULTS

A summary of the advantages and disadvantages of the above methods of information presentation is given in Table 4.

RELATED ARTICLES

Article 48, Common Sources of Errors in Measurement Systems, Volume 1; **Article 49, General Characterization of Systematic and Stochastic Errors, Volume 1**.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

56: Transfer Characteristics of Instrument Stages

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Representing the Overall Response of a System 349
2 Tachometer Example of System Description 350
3 DC and Dynamic Characteristics Compared 351
4 Use of Simplifications for System Performance Description 352
5 Performance Variation with Signal Parameter Values 352
6 Transfer Function Example for Printed Armature Motor 353
Related Articles 353
References 353

1 REPRESENTING THE OVERALL RESPONSE OF A SYSTEM

To begin explanation of how to specify measuring systems, it is first necessary to understand the role of the various stages of an instrument system and how they change the energy variables from the input form into the required output. This article deals with the transfer characteristic. With that foundation in place, **Article 57, Static Considerations of General Instrumentation, Volume 1** and **Article 58, Description of Accuracy, Linearity, and Drift, Volume 1** describes the static regime with **Article 59, Introduction to the Dynamic Regime of Measurement Systems, Volume 1** then introducing the more complex dynamic regime.

The stages used to form a measurement system can be represented as blocks of hardware that effectively transfer an input to its output either in the same energy form, such as the voltage ratio of an amplifier, or as a converted quantity, such as where a sensor converts relative humidity into a voltage and angular motion equivalents for a visual indicator.

The ratio of output to input quantities is termed the *transfer characteristic*. The magnitude of, and phase between, the two quantities are expressed as numerical quantities in terms of the units of the two variables involved. For example, the sensitivity of a humidity sensor might be designed to be 0.1 mV per 1% RH.

When the system is investigated further, it is seen that this transfer expression (the usual one quoted by manufacturers) is a great simplification of the existing full characteristic; it provides no information about the sensor's response to dynamic changes of relative humidity. What is missing is some form of statement as to how the system will respond to changing input signals. This can be done by adding a frequency-dependent component to the expression already given to represent the static behavior.

As simplification, to cover the static operational region of use only, suffices for many tasks, it is common practice to quote only the static transfer characteristic. The steady state equivalent is given in the case of a carrier system or a purposefully designed system made to operate at a given frequency. The steady state is also sometimes called the *quasi-static condition*.

In practice, the state-variable form of relationship between the input and the output for a stage often cannot be easily arranged in the form of independent input and output signals. For this and other reasons, it is common practice to convert that relationship into its Laplace transform equivalent, which can then be arranged to produce a transformed transfer characteristic called the *transfer function*.

As an example, the relationship between an input variable x and an output variable y existing through a measurement

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

instrument stage might be

$$\frac{d^2y}{dt^2} + \frac{3\,dy}{dt} + 2y = x + \frac{dx}{dt} \tag{1}$$

from which we require the ratio y/x to be formed to obtain the transfer characteristic. It will be noticed that measurement system behavior is generally described in terms of linear differential equations, this being in line with general physical systems description.

As the system is linear in the mathematical sense, the Laplace transform can be taken obtaining

$$s^2Y(s) + 3sY(s) + 2Y(s) = X(s) + sY(s) \tag{2}$$

This can be rearranged to give

$$\frac{Y(s)}{X(s)} = \frac{s+1}{s^2+3s+2} = G(s), \text{ the transfer function} \tag{3}$$

If the process involves a time delay, then terms of $e^{-s\tau}$ form will be included, for example,

$$\frac{Y(s)}{X(s)} = \frac{e^{-s\tau}}{s+1} \tag{4}$$

It is in the Laplace form that transfer characteristics are most usually expressed as they are immediately in a form that can be operated upon or studied directly.

It is generally permissible to use the operator D in place of s, and some authors use p instead of s. When the system is operating with sinusoidal signals (which can usually be the case because complex signals can be broken down into sinusoidal components), it is permissible to replace s by $j\omega$, enabling the response at any chosen angular frequency ω to be calculated.

2 TACHOMETER EXAMPLE OF SYSTEM DESCRIPTION

As an example, the tachogenerator, a device that generates output voltage proportional to angular shaft velocity and, therefore, maps shaft speed into an equivalent voltage representation, has the schematic electrical circuit shown in Figure 1.

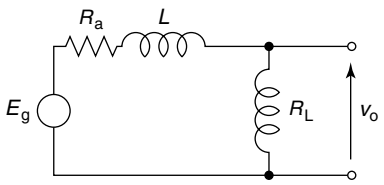


Figure 1. Electrical schematic circuit of tachogenerator.

Being a driven rather than driving mechanical device, its mechanical inertia and friction are not of importance in obtaining the transfer function between angular velocity input θ and the generator voltage output v_o. It is designed such that the voltage generated is proportional to the shaft speed according to $v_o = K_g\, d\theta/dt$; K_g being the generator or transducer constant.

The mechanical parts of the generator are forced (an assumption) to follow input demands and cannot cause lag or lead action because of the generator's compliance or angular mass. The electrical output circuit, however, contains energy storage in the form of inductance of the windings, and this can influence rapidly changing signals. Energy is dissipated via the two resistances, one of which is the load resistance of the next stage that it drives; let $R = R_L + R_a$.

The output voltage v_o, that generated across R_L, is, in Laplace form,

$$v_o = \frac{R_L K_g s\theta}{R + sL_a} \tag{5}$$

which can be rearranged to yield the transfer function:

$$\frac{v_o}{\theta} = \frac{R_L K_g}{L_a}\left(\frac{s}{s + R/L_a}\right) \tag{6}$$

Having developed the expression in this form, for this example, we can move on to explore the three regions of transfer characteristics that could be involved when describing the transfer characteristics of a stage in a measuring instrument or of the whole instrument system.

It can be seen, from the above example, that when s (that is, $j\omega$ for sinusoidal excitation) is zero, the system is operating with zero frequency, that is, at the DC point, in the electrical circuits. The generator is turning to produce an output related to the generator constant K_g, modified by the loading divider effect of R_a and R_L.

In this situation, all frequency-dependent components have no effect on the output. Transients that can arise in the electrical circuit when an input mechanical or an output load transient occur have, in fact, settled down: the generator is producing its steady state output. Thus, the generator constant K_g is, in fact, the response existing after internal transients have settled to be insignificant. It represents the steady state transfer characteristic, and this is generally the sensitivity figure quoted.

The generator will, however, not produce the simple steady state transfer relationship during dynamic changes to its state, that is, during transients. It may be necessary, in application, to make allowances for the transient characteristics, which are more complex. This certainly is the case if rapid-response, closed-loop operation with the

tachogenerator is sought. Phase-shift and gain changes during transients can cause instability.

If the component is to be used in applications where transient response is important, then it will sometimes be found that the manufacturer will quote the necessary additional component values (resistance and inductance in this case) so that the user can construct the full transfer function in order to obtain the transient behavior characteristics needed to condition a tight, stable system.

It is, however, often necessary for the user to conduct tests to obtain the additional values, the only parameter quoted being the steady state characteristic. For example, the data sheets for the previously mentioned humidity sensor contain no information about transient behavior, it being unnecessary for the market intended.

3 DC AND DYNAMIC CHARACTERISTICS COMPARED

There exists another description that can cause confusion for this too is often quoted as though the component only possesses a characteristic mode of behavior: this is the DC response. As the name suggests, this is the response characteristic of the stage when the input, measurement-bearing signal is at zero frequency – direct current or direct-coupled would appear to have formed the symbol DC. Not all systems possess a finite value for DC characteristics, nor is it always relevant.

As an example, consider the DC electronic amplifier. This has a certain gain value when operated at zero frequency. What is often not realized is that it has a transient response during which gain varies and it will exhibit differing gain with signal frequency. When used as an AC amplifier at a given frequency, it then has a certain steady state gain. Thus, use of the apparently simple amplifier block must make allowances for the various modes of transfer characteristic applications. Furthermore, the input signal level can often also alter the transfer characteristic.

It has been shown, by example, that stages used in a measurement system can possess three distinctly different transfer characteristic functions depending upon the time duration and form of the signal. Which are important must be decided for each application. The easiest to measure and quote are DC gain and/or the steady state value. The transient characteristic is much more complex to assess and express and it also varies with signal conditions. Any sensing system usually simply breaks down, by analysis, to yield its DC characteristics, but realization of complete dynamic transfer functions that would enable the transient response to be studied would require expenditure of considerably more effort. The development of transfer functions for very complex systems can often take many years of effort. Systems identification procedures attempt to obtain them from practical testing.

Further explanation about transfer functions, in general, is presented in the classic work of DiStefano, Stubberud and Williams (1976) and Doebelin (1975). Atkinson (1972) explains the difference between steady state and transient characteristics in terms of the solution of linear differential equations having constant coefficients.

Having shown how a stage can be modeled theoretically in terms of its transfer function, and therefore be uniquely described on paper or in a computer model, it is important next to realize that linear systems of interconnected transfer functions, no matter how complicated, can (in theory) be reduced to a single-stage transfer function. This means that a complex measurement instrument system, built from many stages as a serial path including feedback loops with other paths, can be reduced mathematically to just one transfer function that describes its complete overall static and dynamic behavior. This reduction allows the transient, steady state, and DC responses of any system to be calculated or simulated without need to assemble all stages.

There is no need here to elaborate on the techniques used in block-diagram algebra; they are well covered in DiStefano, Stubberud and Williams (1976). It basically stems from the prime fact that transfer functions are commutative, provided it can be assumed that they do not load each other when connected. Interaction of stages destroys the rule, making reduction more complicated. Interaction significantly changes the response, and so it may be necessary to build a slightly more complex system that incorporates buffering between blocks. This being done allows the system performance to be accurately modeled and thus optimized.

When compounding stages together to build up a measurement system, it is, therefore, necessary to consider if the stages interact to a significant degree. In electronic systems, this problem has generally been eliminated by designing integrated circuits to connect on the basis of low output impedance being connected to high input impedance in order to preserve voltage levels. In cases where power must be transferred efficiently, it is necessary to use stage impedances that are equal. For current transfer, another information transfer mode, the opposite to the first given voltage case, applies.

Electronic systems connections generally present few connection problems (provided well-established rules are adhered to), but the same is not true of mechanical, thermal, fluid, and optical stages, for in those disciplines, component design does not enable ideal matching arrangements to be performed so easily.

Transfer characteristics of a stage can be stated in terms of the response in either the time or the frequency domain. Conversion from one domain to the other is achieved via the Fourier transformation. When the signal is of a steady state, time-domain condition, because it is periodic, it can easily be transformed to yield the singular set of frequency components. Signals of transient nature present greater difficulties in transformation for the spectrum changes during the transient.

It is, therefore, quite reasonable for the response of a stage to be stated in terms of the frequency domain. The same information could be expressed in the time domain in terms of the stage's response to a suitable time variant input. In some applications, it is more convenient to obtain or describe a transient in time response form (such as to a step or impulse input) than it is in the frequency domain.

Zero-frequency response is, in the frequency domain, the value of amplitude, or phase, at zero frequency. As frequency response plots are generally expressed using logarithmic scales, zero frequency cannot actually be represented. Amplitudes are generally expressed in decibel form, but for some instrumentation, unconverted ratios may be preferable.

The above remarks apply to stages having at least one each of a clearly identifiable two-terminal input and output port, refer to **Article 108, Transducer Fundamentals, Volume 2**. In use, a two-terminal device, such as a capacitor or a spring, becomes a two-port device, either as a series (through-variable) or as a parallel (across-variable) arrangement, by suitable connection with common lines. The above remarks then apply, provided it is clearly stated what the connection arrangements and the terminating conditions existing are. Alternatively, the two-terminal device can be described in terms of its terminal impedance response, called the *driving-point impedance* (or *admittance*), which can also be fully expressed mathematically in Laplace form.

Simple common components, such as the electrical inductor or mechanical mass, are so well known that they usually do not need to be specified (for response reasons) in greater detail than the numerical constant of the parameter they provide, for example, inductance in henry units or mass in kilogram units. This information is sufficient for users to construct the response characteristics required. When the component is uncommon, a mathematical driving-point impedance expression is needed to provide adequate information about its static and dynamic behavior.

4 USE OF SIMPLIFICATIONS FOR SYSTEM PERFORMANCE DESCRIPTION

In many cases, it is not convenient to express the true response behavior of a terminal port, and approximations are used to map a many-variable parameter situation into a single one. A moving-coil loudspeaker (as an actuator or a sensor) mechanism is generally characterized by stating the impedance it presents at a given audio frequency, this being expressed as a pure resistance of $x\Omega$. In reality, the terminal impedance is a complex electrical quantity having both real and imaginary parts, and it varies greatly with frequency. This applies to the input and output impedances of many stage blocks used. This is, however, not always the case, a notable exception being the transmission line, which provides constant purely resistive impedance at any length, provided it works within certain assumptions of being lossless and properly terminated. Somewhat confusingly, the term impedance is often used synonymously with resistance: correctly, the former is the resistance to flow when both real and imaginary components exist as a complex variable.

5 PERFORMANCE VARIATION WITH SIGNAL PARAMETER VALUES

The tacit assumption often exists that components and subsystem blocks are what they are defined to be for all regions of operation. In practice, a component can behave quite differently for signal frequencies, levels, noise contents, and other parameters from those for which it was purposefully made to handle. For example, a capacitor becomes an inductor with rising frequency; a mechanical mass provides properties of spring compliance under certain conditions.

Ideally, all components and stages used in measurement systems should be defined to provide the user with information about their whole spectrum of response conditions, but this is not practicable. The instrument designer, in particular, must be aware that simplistic definitions, such as those giving no more than steady state or DC information, may not suffice. Lack of observance and understanding of this defect in subsystem description is a prime reason for failure of many designs when first assembled and for incorrect operation when fed with input and influence quantity signals that do not conform to those it is designed for.

It is, therefore in principle, possible to model any complete or part measuring system as a transfer function expression that enables its behavior to be understood without needing to resort to actual hardware. Whether this is worthwhile, however, depends upon the application and the extent of the system. A very practical point is that many of the constants required to provide a boundary to the mathematical model must come from measurement of existing phenomena.

Reduction, in this way, of an arrangement of many subsystems' mathematical models into one, retains all information about the overall behavior but does not allow interstage

signals to be studied. This realization is not to be confused with identification processes and simulations that can also provide overall input–output relationships that apply for some modes of operation and not others and that may be realized by quite different internal processes. If the internal subsystem transfer functions are adequate models of the stage that they represent, so also will be the overall reduction obtained by block-diagram algebra (provided conditions of connection and range of operation are met).

Transfer function representation is only as good as the assumptions on which it is developed. Many are, and need to be, simplifications of the real situation; often, they ignore the higher-frequency terms that can be considered to be of no practical consequence to response, thereby making them only applicable for a given range of frequencies.

6 TRANSFER FUNCTION EXAMPLE FOR PRINTED ARMATURE MOTOR

To illustrate the above ideas, consider the transfer function of a high-performance instrument printed-armature DC motor. This needs only to allow for response lag caused by the mechanical storage aspects of the armature disk, the electrical storage time constants of the armature circuit being much smaller than those of mechanical sources (50 μs compared with 30 ms). A conventional style, field-controlled, DC motor, however, requires a more complex transfer function, one having a second storage term to allow for the delay caused by significant storage effects in the electrical field-coil circuit. The two transfer functions are

Printed motor

$$G(s) = \frac{\omega_{\text{out}}}{\upsilon_{\text{in}}}(\text{s}) = C_1 \frac{1}{\tau_{\text{m}} s + 1} \tag{7}$$

Field controlled motor

$$G(s) = C_2 \frac{1}{(\tau_{\text{m}} s + 1)(\tau_{\text{f}} s + 1)}$$

where

ω_{out} = shaft velocity
υ_{in} = drive voltage
τ_{m} = mechanical time constant
τ_{f} = electrical time constant of field winding
C_1, C_2 = constants related to generator and torque constants and to mechanical friction and field-coil resistance.

RELATED ARTICLES

Article 25, Introduction to Signals in Physical Systems, Volume 1; **Article 26, Signal Classification, Volume 1**; **Article 27, Signals in the Frequency Domain, Volume 1**; **Article 59, Introduction to the Dynamic Regime of Measurement Systems, Volume 1**.

REFERENCES

Atkinson, P. (1972) *Feedback Control Theory for Engineers*, Heineman, London.

DiStefano, J.J., III, Stubberud, A.R. and Williams, I.J. (1976) *Feedback and Control Systems: Schaum's Outline Series*, McGraw-Hill, New York.

Doebelin, E.O. (1975) *Measurement Systems: Application and Design*, McGraw-Hill, New York.

57: Static Considerations of General Instrumentation

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Description of the Static or Steady State Behavior 354
2 Range Factors 354
3 Zero Values of Operation 355
4 Drift 355
5 Hysteresis and Dead-band 356
6 Signal-to-noise Ratio Entities 357
7 Frequency Response Description 358
8 System Constants 358
Related Articles 358

1 DESCRIPTION OF THE STATIC OR STEADY STATE BEHAVIOR

As an example to illustrate the terms introduced in this article, consider the relative humidity sensor in which a humidity-sensitive membrane changes length with moisture take-up. The length change causes a metal spring strip to be strained. On the strip are strain gauges that transduce the strain into a voltage using a simple Wheatstone bridge. The voltage is amplified and adjusted to give a calibration and sensitivity facility. Static performance parameters are first considered to construct a block diagram showing the complete instrument broken down into serially connected paths formed as the stage described above. The spring also drives a mechanical pointer on a scale.

Two measuring paths can be identified from the sensing membrane through the instrument to the two outputs that produce an electrical and a mechanical output signal.

An overall transfer sensitivity (S_1 and S_2) can be developed through each path by multiplication of the static transfer characteristic for each of the stages connected to form a link. This yields

$$S_1 = \frac{V_{\text{out}}}{\%RH} = \left(a\frac{\Delta L}{\%RH}\right)\left(b\frac{\xi}{\Delta L}\right) \times \left(c\frac{\Delta R}{\xi}\right)\left(d\frac{\Delta V}{\Delta R}\right)\left(e\frac{V_{\text{out}}}{\Delta V}\right)$$
$$S_2 = \frac{\theta}{\%RH} = \left(a\frac{\Delta L}{\%RH}\right)\left(f\frac{\Delta L'}{\Delta L}\right)\left(g\frac{\theta}{\Delta L'}\right) \quad (1)$$

where $a, b, \ldots, g$ are the static, or quasi-static, transducer constants of the various stages that relate input to output variables when the stage is in a settled state of response.

Thus, the static sensitivity is found as the product of static transducer constants for all stages in a path. If a different gain is needed, it could be arranged anywhere in the path by altering any one of the constants by design alteration. This expression assumes that the instrument is not operating in a transient state as none of the stages have any frequency terms associated with their transfer characteristic.

2 RANGE FACTORS

Range, the factor that decides the extent of measurement ability from the smallest to the largest signal amplitude that can be accommodated without the system being overloaded or the signals becoming limited (saturated), is decided by the capability of all stages. If the signal level of any one stage moves into a region of improper operation, the whole system will fail to operate correctly.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

The effective measuring or working range is the 'range of values of the measured quantity for which any single measurement, obtained under specific conditions of use of a measuring instrument, should not be in error by more than a specified amount'.

The upper and lower range limits form the limiting boundaries inside which the instrument has been constructed, adjusted, or set to measure the specified quantity. The effective range may not necessarily be all of the instrument's interval between these two limits. Rangeability is the term sometimes used to relate, by some form of expression, the relationship between upper and lower limits of useful range.

In the humidity sensor example, the manufacturer states that the instrument has a range between 20 and 100% *RH*, but on the instrument itself is a warning that use outside of the interval 20 to 85% *RH* may lead to temporary loss of calibration. Furthermore, careful reading shows that the usability is only within given accuracy limits over 30 to 85% *RH*; thus, the range of reliable operation is within 20 to 85% *RH*, rangeability perhaps being given (there is no standard methodology for expressing this) as providing a ratio of maximum to minimum limits of 4.25:1. Rangeability is more useful with instruments having unbounded measurand possibilities, such as occurs in flowmetering where the two limits of practical flowmeters are often rather restricted compared with the interval that might be called for in practice.

A multirange instrument is one in which the range of operation can be selected by manual or automatic means, the latter being referred to as an *auto-ranging instrument*.

When describing the effect on a calibration of influence quantities, it is common practice to state the range over which such quantities can be tolerated. This is defined as the *reference range*. The humidity meter performance is stated for a reference range of 0 to 40 °C temperature variation.

As a guide, ambient environmental conditions may usually be taken as having a mean temperature and reference range of 25 ± 10 °C; relative humidity of 90% or less for general use (but 40–60% for storage of paper materials); and barometric pressure lying within 90 to 110 kPa (900–1100 mbar).

Many measurements can be performed more easily, or with improved accuracy, if the instrument is built to provide a limited effective range. This can be done by designing the lower range limit to be well above zero (the converse holds for measurands having values less than zero). Such systems are called *suppressed-zero systems*. A common example is the voltmeter used to indicate mains voltages in the range 200 to 250 V AC.

For such systems, sensing signal levels will not have a zero level coinciding with the output zero, that is, in the above example, a 200-V input may deliver 0 V to the indicator. Any other system of level shifting might be chosen. The degree of suppression is expressed as the ratio of scale interval omitted, above or below zero, to that presented. A similar and wider concept is when the range is expanded at any point. This is often termed *band* or *scale spreading*.

Range of operation for static and quasi-static use is often different to that in the instrument's dynamic transient mode. The term dynamic range is used to indicate which regime the quoted range relates to. It is generally stated in the same ratio form as is given above for rangeability and is presented in decibel form. At times, its common use would, however, appear to be really defining the static and quasi-static performance, not the time-dependent dynamic behavior.

Related to the range of an instrument is the scale factor that might need to be applied to the actual numerical value obtained in order to arrive at the correct magnitude. Scaling is often practiced to allow a normalized instrument to be applied to many ranges of tasks. Except where the instrument is multiranging, the system should be marked accordingly. Multiranging instruments usually indicate automatically the appropriate scale factor as the range changes.

3 ZERO VALUES OF OPERATION

Many instrument systems require an operation to set the position of the zero. For example, a gas analyzer and its sampling line may need to be purged with a suitable clearing gas before use. Electronic DC systems usually need periodic zero setting to compensate for the drift that has occurred.

4 DRIFT

Drift tends to be regarded as a DC property of a system, but this depends upon the time response needed. In many electronic systems, frequencies of operation less than 10 Hz are regarded as being at DC. In some disciplines, however, examples being earth tides or temperature variations in a cool room, responses may be concerned with frequencies of the order of cycles per year or per day respectively.

Drift, if studied closely enough, will be found to be a slowly moving dynamic phenomenon and may need description to allow for its actual behavior. The simplest form of expression is statement of the linear component in terms of drift occurring per variable of interest – such as time or supply level. In electronic amplifiers, drift with time is often quoted as though it were a linear-time function, but

in reality, it is often formulated on a square-root basis, for drift in this case does not accumulate linearly. Drift is a complex, often poorly understood, parameter that can defy generalization on a causal basis.

Related to the problem of drift specification is how to adequately describe aging effects. These also defy formal generalization. Aging is a factor, like drift, that brings about lack of reproducibility in an instrument because over time the values of components, and hence transducer constants, will change. Aging can also arise because of operational use of the components. There are no general rules to follow to reduce aging, but the problem usually is concerned with appropriate knowledge of material properties.

In the humidity sensor example, aging is to be expected in the membrane, the curved spring strip, in the strain gauges, and in the electronic components. Each of these changes will gradually give rise to a shift in calibration and in the position of the zero. Internal adjustments are provided so that these can be reset. Being an inexpensive design, intended for low-accuracy use, no figures of aging or drift are quoted. Higher-priced instruments usually quote such figures.

The task of deciding when an instrument should be recalibrated, to make up for errors of this kind, is very subjective. The history of an instrument's performance is an important factor in being able to predict when it will have drifted, or aged, out of the calibration tolerance band.

Drift is often allowed for in a measurement by returning to a known stable and fixed measurement point periodically through the test run. It is then possible to extract the drift and adjust the values.

5 HYSTERESIS AND DEAD-BAND

Measuring instruments do not respond in a completely continuous manner, nor do they always give the same value for both directions of closure to a point.

The term threshold is used to signify the smallest level of signal to which the system will respond. All measurements, no matter how well designed, will ultimately be limited by some form of discrete phenomenon. Another example is seen as the threshold of the zero-suppressed voltmeter mentioned above, which occurs at 200 V. Threshold is either caused by a feature of the physical principle involved (e.g. a turbine flowmeter will not respond correctly until there is sufficient fluid in the pipe) or it may be deliberately introduced (such as to suppress the zero). It provides a dead-band of operation.

Many instruments exhibit dead-band effects that are caused by backlash and play in the drives. These will provide different values for each direction of approach to the point of interest. Where it cannot be removed, it is common practice to approach the final value in the same direction each time. Dead-band can become particularly significant as a cause of instability in closed-loop systems; some, however, make use of it to allow the system to settle in the dead-band.

Dead-band produces a kind of hysteresis, but that latter term is more generally used to describe the magnitude of the hysteresis phenomenon arising internally in a material. Figure 1 shows a typical hysteresis loop with its measured quantities. Magnetic material, for instance, does not show the same magnetic field strength for rising and falling induction levels. The magnitude of the strain condition for a cyclically stressed mechanical member will vary depending on the direction of the stress. In each, hysteresis magnitude increases with increasing excursion. Hysteresis magnitude can be quoted as the ratio of maximum difference of the upscale and downscale variations to the full-scale excursion. Other expressions are used. In some applications, the total area is more important than this difference.

In the humidity sensor example, mechanical hysteresis arises in the strained member carrying the strain gauges

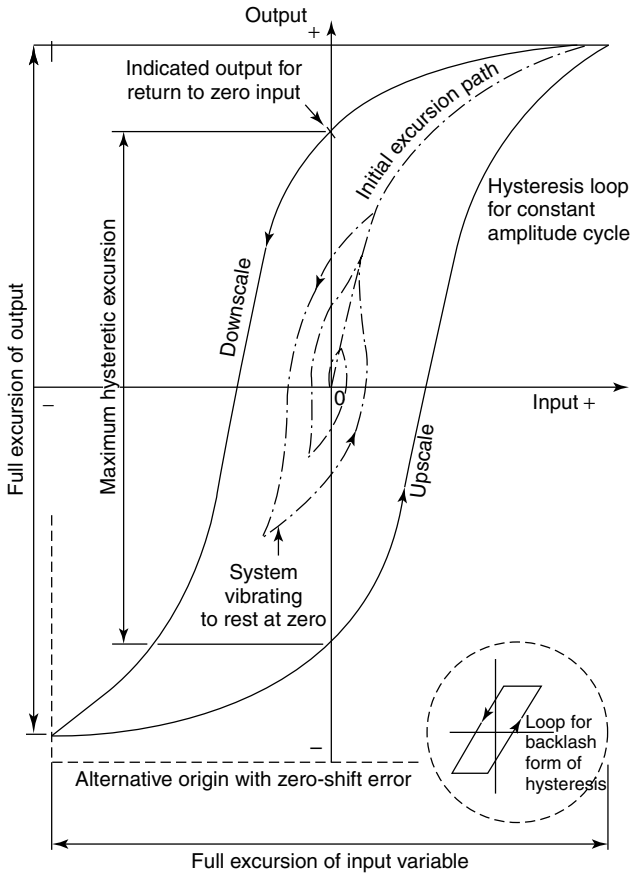


Figure 1. Schematic of a generalized hysteresis loop.

and certainly in the membrane. Backlash arises in the pin joints of the mechanical system (but is kept in control by a biasing tension), in the sector gear driving the pointer pinion, and because of slack in the bearing system. The art of good design is to ensure that these are not significant for the intended purpose and that they do not increase significantly as wear occurs overtime. A highly damped, resonant system and slowly moving signal systems will exhibit greater hysteretic effects than more rapidly moving arrangements.

Larger excursions also increase hysteresis error; systems that can resonate cyclically to rest will usually show less error from this source because the hysteresis loop traversed in each cycle will become progressively smaller, the system finally coming to rest in the center of the loop.

Electronic systems generally do not pose these problems, but some circuits purposefully introduce backlash. Storage effects in semiconductor elements can lead to backlash in fast-acting circuits.

Backlash, hysteresis, drift, creep, and aging are all subclasses of a general noise problem that instruments of measurement must cope with.

Noise is the general name given to unwanted sources of apparent signal that could lead to error in the measurement mapping process. Every energy regime has its own sources of noise. The user of an instrument will often need to know how tolerant an instrument is to common noise sources.

By suitable design, a stage can often be made such that noise quantities (of certain kinds) influence two features of the system in such a manner that their combined effect on the true signal is differenced, thus canceling. This is called *common-mode rejection*. Thus, deliberate introduction, through a second influence point, of correlated noise into the system can be used to dramatically reduce the original, impossible to eliminate, noise effect.

As no process is perfect, these common-mode rejection techniques need some means of stating their effectiveness. This is done through the use of the common-mode rejection ratio (CMRR), which is defined as the ratio of the common-mode noise signal amplitude to the common-mode error signal remaining after rejection. The term is used extensively in the description of electronic systems where it is usually quoted in decibels. It is, however, quite a general concept and can be applied to any common-mode signal-to-noise rejecting system.

As with many parameter statements, the need to keep the statement simple can lead to the assumption that CMRR is a simple fixed ratio. This is not always so; it will often be a function of many system variables. Average figures are frequently quoted. It can, at least in electronic systems, vary with temperature and usually decreases with decreasing frequency of the noise signal. Often, the CMRR quoted is for DC input signals.

6 SIGNAL-TO-NOISE RATIO ENTITIES

In some instances, the need is for definition of how well system signals exceed noise levels: this can be assessed from the signal-to-noise ratio (S/N or SNR). This is the ratio of the signal amplitude to the noise amplitude, again usually expressed in decibels. The amplitude used could be average, root mean square (rms), or peak-to-peak. Steady state AC signals are involved, and they may be of sinusoidal or complex waveform. Rarely are the waveshape or amplitude criteria quoted, leaving considerable room for imprecision, especially if the signal and the noise do not have the identical waveforms that a ratio properly demands.

The introduction of a stage into a measurement system usually degrades the SNR to some degree. The ratio, in decibels, of SNR at the output to SNR at the input is termed the *noise figure*.

Measuring instruments are designed to fulfill a certain task at a specified accuracy level. As increasing the accuracy of a measurement implies an increasingly costlier design, the general rule is to choose an instrument that just copes with the task needed.

It is quite pointless to specify an instrument that has accuracy exceeding the task; its capability will not be utilized and the maintenance costs will be greater than necessary.

To assist the specification and selection of instruments of appropriate accuracy, many disciplines issue standards documents that state accuracy classes. There is no general rule for the development of such classes, a dominant reason being that each kind of measurement presents different ranges of possibility and need.

For example, it is quite feasible to call up time measurements to errors of parts in 10^{12} but photometry can only be best achieved to parts in 10^4.

Where the need has arisen, agreed accuracy classes are defined, within which ‘the instrument will have certain declared metrological properties within specified limits of error’. A class is ‘usually identified by one of a set of symbols, adopted by convention, and usually referred to as the “class index”’. Definition of accuracy often ties in with that of linearity.

The term conformance is sometimes used to relate an instrument to a specified accuracy class. Interchangeability is the term used to describe how well an instrument can be exchanged with another having the same specifications; tolerances of manufacture and differences in the various manufacturers’ designs cause units to be slightly different to each other yet still be suitable for a given task.

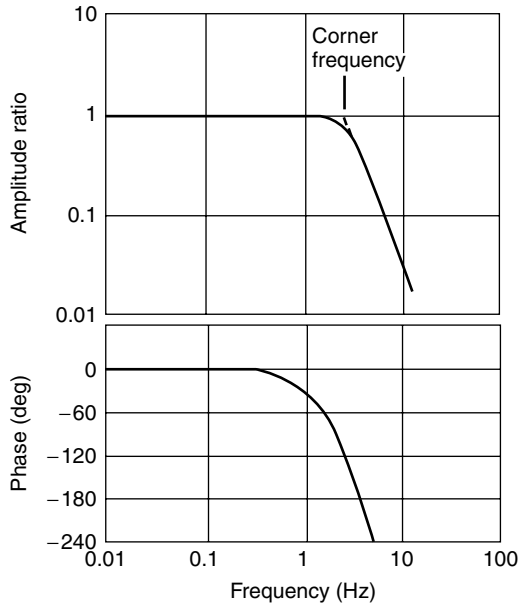


Figure 2. Amplitude and phase responses expressed in Bode diagram form.

7 FREQUENCY RESPONSE DESCRIPTION

Although the response of a stage is dependent upon its dynamic behavior, the frequency response of a stage is usually that of a steady state that is taken after the stage has settled to the final value after excitation has been applied at each frequency.

In the simplest form, the frequency response can be stated in a verbal manner describing it in such terms as 'flat to "*x*" Hz' (or 'between', or 'above'). In such cases, the extremes are denoted, by convention, to be where the response has fallen off (or risen) to the point where the ratio of input to output signal magnitudes has changed by 0.707. In some cases, the 3-dB points are used, both ratios being the half-power points. Some times, the response obtained is a function of signal amplitude. In such cases, it is necessary to state the amplitude of the test.

A more adequate, but harder to publish, description is to provide a magnitude and a phase plot as a Bode diagram. Examples are given in Figure 2. The true Bode diagram uses only straight-line segments to approximate the actual response, but this distinction is not always made. The upper curve is, by convention, the amplitude response.

Systems involving continuous cyclic signals require description of amplitude, phase angle, phase shift, power factor, and true and apparent power. Each of these terms is used in instrumentation in the same way as it is in electrical power systems. Probably, the greatest error that arises in their use in instrumentation is the lack of care in ascertaining that the system is operating with sinusoidal waveforms (which electrical power systems are generally assumed to be), for the parameters are often quoted in terms of sinusoidally based quantities such as r.m.s. and average amplitudes and power levels.

8 SYSTEM CONSTANTS

Constants quoted are often far from being so. Many are functions of several operational parameters. Noise figure for an electronic amplifier, for example, is a function of signal amplitude, frequency of operation, and the actual device used. Constants may vary with time.

There do exist, however, certain constants, called *physical constants*, that are believed to be unchanging. It is therefore often necessary to know more about the formulation of a given constant if more intelligent use is to be made of the number quoted.

The designer or user is usually left to explore such ramifications through reading the primary literature. In essence, consensus obtained from practice provides the published information generally required.

The difficulty is who shall decide what the general demand is, and how does the user know whether the factors given are entirely adequate for the task in hand. Overspecification can be useful but will increase the cost.

Terms used in the description of measuring systems are many. A useful service, but membership is needed, is that provided by the NCLS International http://www.ncsli.org//index.cfm. It is a searchable listing for a glossary of terms and for acronyms and abbreviations.

RELATED ARTICLES

Article 26, Signal Classification, Volume 1; **Article 48, Common Sources of Errors in Measurement Systems, Volume 1**; **Article 56, Transfer Characteristics of Instrument Stages, Volume 1**; **Article 58, Description of Accuracy, Linearity, and Drift, Volume 1**; **Article 59, Introduction to the Dynamic Regime of Measurement Systems, Volume 1**.

58: Description of Accuracy, Linearity, and Drift

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Simplification of Terms for Accuracy 359
2 Linearity Description 359
3 Drift Description 362
Related Articles 362
References 362

1 SIMPLIFICATION OF TERMS FOR ACCURACY

Having introduced the static regime in **Article 56, Transfer Characteristics of Instrument Stages, Volume 1** and **Article 57, Static Considerations of General Instrumentation, Volume 1**, this article concentrates on the description of a measuring system's accuracy, linearity, and drift.

For reasons of simplicity of specification and description, the system's performance features of accuracy and related linearity are often quoted in a simplistic manner by the use of a single numerical value. This practice might hide, from those persons who are not aware of the facts, that these are often gross simplifications of a much more complex reality. To illustrate this point, the following paragraph is taken from the terminology section of a Foxboro-Yoxall statement:

Accuracy includes the combined conformity, hysteresis, and repeatability errors. The units being used are to be stated explicitly. For example, expressed in output units, a typical expression might be $\pm 1\,^{\circ}\mathrm{C}$; expressed in percent of output span, the expression could be $\pm 0.5\%$ of output span; expressed in percent of the upper range value, it might read $\pm 0.5\%$ of upper range value. For indicators, it might read $\pm 0.5\%$ of scale length; or in some cases, it is expressed as a percent of actual output reading, such as $\pm 1.0\%$ of actual reading.

Confrontation with this variety of simplifications leads one to refer to approved standards documents, such as BS 5233 or AS 1514, to find that accuracy is defined only in a general sense and that no suggestion for a uniform methodology is provided. An ISA document, formulated by its SP37 subcommittee, on transducer specification, provided greater depth in the manner given above (issued as S37.1 or ANSI MC6.1:1975.)

In reality, the concept of accuracy is more properly discussed in terms of errors and the uncertainty of a measurement made with the measuring instrument. The term accuracy, one source suggests, should only be used as a general descriptor of system performance.

Implicit in any statement of uncertainty of an instrument is the apparent assumption that the instrument will provide the same absolute or proportionate performance at any part of its usable range. This is rarely a totally valid assumption. Uncertainty is covered in detail in **Article 52, Error Models, Error Budgets and their Calculation, Volume 1**.

2 LINEARITY DESCRIPTION

The manner in which error varies over the effective range of an instrument is expressed as the linearity of the instrument (but which correctly actually relates to the nonlinearity or lack of linearity between desired and actual performance).

A process is said to be linear when the relationship between two quantities is such that a change in one quantity is exactly proportional to the associated change in the other quantity.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

Expression of the nonlinearity of an instrument is achieved in several different ways most of which, again, result in a simple statement that conveys information about how the signal deviates as it ranges over the available span.

Fundamentals of measurement indicate that a single numerical value can only characterize a single attribute of a situation. This is, therefore, the case in the statements of linearity: each method of description conveys different information about the linearity. Seldom is an actual curve presented, the exception being when a calibration chart is provided.

Linearity is expressed as deviation from a specified straight line. Four predominant reference lines might be used.

The first reference line uses a line placed so that it lies in the position satisfying the least squares criterion calculated for the data set available. It is often called the *best straight line* (BSL). It might also be referred to as a *least squares line*. Figure 1(a) is an example of this line.

In many instances, the system provides zero output for zero input, in which case a more appropriate line to use is a BSL that is computed so that it is constrained to make the line pass through the zero point. In suppressed or elevated-zero systems, the point of interest will be the pseudozero used. Figure 1(b) illustrates this line. This method describes the zero-based linearity of the system. One variation of this is to not use a zero-based line but some other suitable reference point. This is termed *BSL* with forced reference linearity.

In terminal-based linearity, a reference line is adopted that links the upper and lower limits of the span, a line drawn between its terminal points. It is also called *terminal linearity* or *end-point linearity*. The line so drawn is sometimes referred to as the *end-point line*. Figure 1(c) shows this line.

Added later in AS 1514 was another line that can be used (Figure 1d), being one that lies in the theoretical linear position of the principle invoked. This is called the *theoretical slope linearity*.

Having first decided which line is the most appropriate to use, the next step is to select a method of description that will express the magnitude of deviations from the line. In general, this will usually be the greatest departure from the line of the output variable. Linearity of instruments is generally determined as the result of a calibration test in which only a limited number of values would be determined. Some calibration processes provide a considerable number of values and if they are of random nature, the deviation might then be expressed in terms of the standard deviation or some other statistical descriptor.

A little thought reveals that the above methods of expressing linearity and accuracy of an instrument are

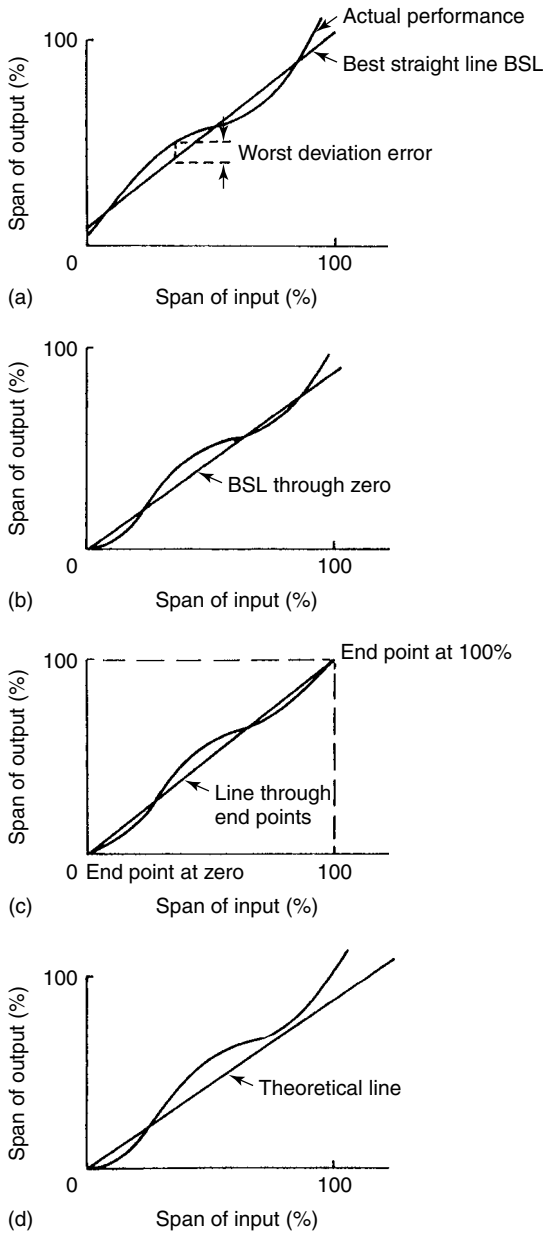


Figure 1. Four methods of specifying linearity of a response.

subject to many causes of error in their determination, expression, and subsequent interpretation. A process with significant hysteresis can exhibit quite different linearity with each direction of signal change and with signal amplitude. The magnitude of influence variables at the time of the determination also can alter the error magnitudes. The method of calculation of the best fit line and the number of data samples available will further influence the value that will be stated as the singular value of linearity of the instrument.

Expressions of the linearity of a system are invariably determined for a system in the steady or static

state. Rarely will the same relationship apply when the stage is in the dynamic state. Storage parameters will alter the instantaneous transfer characteristic according to the signal amplitude and history of previous signal excursions. This point is often overlooked when accuracy and linearity figures are taken from data sheets for application to sensors used for rapidly changing measurands. Pressure gauges, accelerometers, position sensors, thermometers, and the like will give very different instantaneous performance figures to those measured in the steady state and stated in their specification of calibration chart.

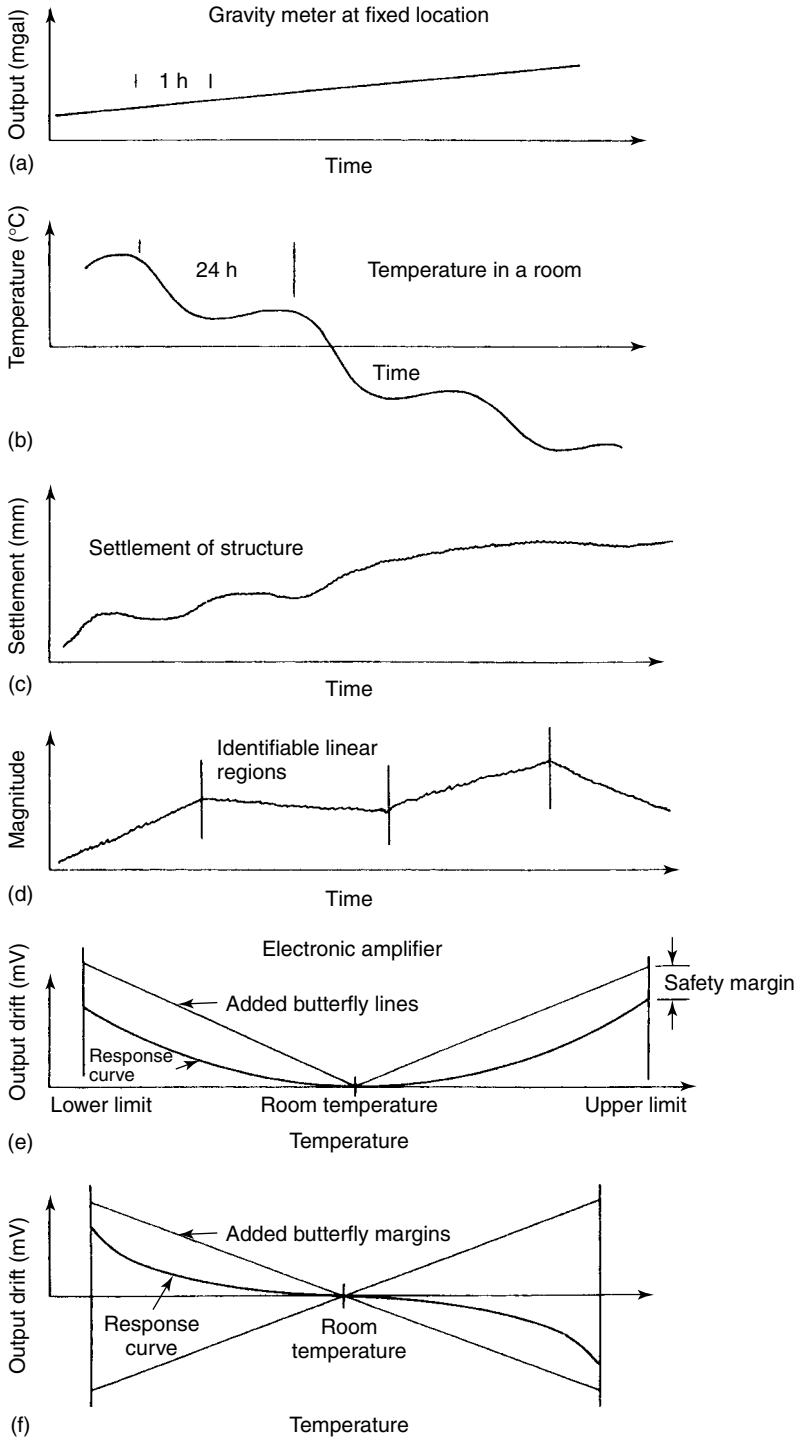


Figure 2. Typical drift curves, for example, instruments and measurands.

Another factor to be considered may be that the stage transfer characteristic may possess adequate linearity of response but that the BSL may slope at the wrong rate because of an inaccurate DC sensitivity coefficient for the stage.

Where possible, it may be preferable to provide a mathematical expression that describes a curve fitted to the data, an example being the relationship used for the output of a thermocouple.

It can be seen, therefore, that usually stated expressions of accuracy and linearity are invariably simplifications. Specifications of these parameters must be generated and interpreted with care and understanding if ambiguity is to be avoided.

Extending these ideas further, to cover other nonlinear responses, again introduces likely confusion as the direct result of seeking simplified statements of performance. Specification of drift is a related area where such problems arise.

3 DRIFT DESCRIPTION

Drift is the feature of a system that characterizes how a system variable, that is intended to be constant, changes with time. The drift of a gravity meter is one example. The term is also used to describe effects of influence quantities on the output, an example being the temperature drift of an electronic amplifier. It is also a term applied to explain how a measurand varies with time or other variables. An example of this use is to characterize settlement of a structure with time. It is sometimes used synonymously with the term stability. Some kind of qualifying statement is needed to uniquely define which context is being used.

Drift conveys very low frequency response information about the measuring instrument or of the measurand. It can, on occasion, be very difficult to separate the two.

Figure 2 shows several often-met drift curves. For each curve, the need is to formulate a simple expression that will convey information about the trend of the output away from a chosen reference value.

When the drift is linear, it is a simple matter to express the slope in terms of the two parameters involved. The gravity meter example of Figure 2(a) might be expressed as drifting at x milligals h^{-1}. Being entirely linear, it would not matter where the slope was determined along the curve.

The situation becomes a little more difficult when it is required to express the drift for the case shown on Figure 2(b). If the characteristic possesses a cyclic component of constant amplitude, then a mean line is easy to determine. If, however, the curve does not have obvious linear features, like that of Figure 2(c), then a best fit line must be fitted from which the drift can then be assessed as for the previous cases.

When the curve clearly possesses two or more linear regions (Figure 2d), each of a different slope, a more extensive statement will be required. No standards of practice appear to exist for such cases, so the description should carry a statement of the additional features rather than simply quote a fixed numerical drift coefficient.

If the drift curve returns to the original value (Figure 2e), quoting the mean drift that has occurred between the two end points would imply that the system is perfectly stable. Furthermore, nonlinear curves, such as that given in Figure 2f, would be grossly misrepresented. One approach, outlined in Stata (1969) for example, is to use a butterfly characteristic curve. Typical boundary curves have been superimposed onto the two amplifier drift curves. Drift is calculated as that occurring from the reference point (which is often room temperature at which the amplifiers were adjusted for the test) to the extremities of workable range, plus a margin of safety. This coefficient then provides a conservative figure within which drift can be expected to lie for a given temperature excursion. Such simplification, however, is at the expense of possible overdesign on the part of the user.

These examples clearly show that the specification of drift can be a complicated matter and that simplistic definition, as with those for nonlinearity, can be misleading.

Manufacturers have been known to use the shortcomings in the understanding of descriptions to their advantage when specifying equipment.

Further detail on accuracy and related concepts is to be found in CORD (1999) and Westgard (1997).

RELATED ARTICLES

Article 50, Errors in Signal Systems, Volume 1; **Article 51, Errors in Digital Signal Systems, Volume 1**; **Article 52, Error Models, Error Budgets and their Calculation, Volume 1**; **Article 56, Transfer Characteristics of Instrument Stages, Volume 1**; **Article 57, Static Considerations of General Instrumentation, Volume 1**; **Article 59, Introduction to the Dynamic Regime of Measurement Systems, Volume 1**.

REFERENCES

CORD. (1999) *Unit 13: Precision, Accuracy & Tolerance Communications*, CORD.

Stata, R. (1969) *A Selection Handbook and Catalog Guide to Operational Amplifiers*, Analog Devices, Inc., Cambridge, MA.

Westgard, J.O. (1997) *Opspecs Manual: Operating Specifications for Precision, Accuracy, and Quality Control*, AACC Press.

59: Introduction to the Dynamic Regime of Measurement Systems

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Introduction to the Dynamic Regime of Measurement Systems Performance 363
2 Forcing Functions 364
3 Application of Forcing Functions 365
Related Articles 366
References 366

1 INTRODUCTION TO THE DYNAMIC REGIME OF MEASUREMENT SYSTEMS PERFORMANCE

The dynamic regime of an instrument stage is generally a more complex aspect to characterize and test than the stage's static performance. To fully understand the behavior during the dynamic state, it is necessary to allow additionally for the transient solution of the transfer characteristic when forced by an input function, a factor that is not important in understanding the steady state and static regime behaviors.

Dynamic behavior of systems has been well explained at the theoretical level if the performance remains linear during the dynamic state. It has been shown that the energy relationships of electrical, mechanical, thermal, and fluid regimes can each be characterized by the same mathematical description and that a system containing a cascaded chain of a mixture of these, as is typical of instrument systems, can be studied in a coherent manner by the use of such mathematical techniques. This body of knowledge and technique lies in the field of the dynamics of systems. (This is not to be confused with a subject area called *Systems Dynamics*, SD, that is somewhat different to that being described here.) It has progressively grown from isolated, unconnected explanations of the analogous situations existing mainly in electrical, mechanical, and acoustical areas, into one cohesive systematic approach to the study of the dynamics of linear physical systems.

The approach has been widened considerably (see Klirs, 1972), to take in philosophical concepts, such as are found in social systems, this broader assembly of knowledge usually being referred to as *general systems theory*. General systems theory certainly has application and potential in some instrument systems, but most designers and users of instruments will find that the more confined and mathematically rigorous assembly of technique and knowledge (systems dynamics) will be that found to be usefully applicable in practical realization and understanding of measuring systems. The use of the word systems is extremely common and carries numerous connotations ranging from a totally general concept to the quite specific use that is based in mathematical explanation.

As indicators of the part that is useful to measuring instrument dynamic studies the reader is referred to some classic texts that sorted out this field well. Karnopp and Rosenberg (1975), MacFarlane (1964), Paynter (1961), Shearer, Murphy and Richardson (1967), and Zahed and Desoer (1963) for general expositions, and to Coughanowr and Koppel (1965), Fitzgerald and Kingsley (1961), Koenig and Blackwell (1961), and Olson (1943) for more specific uses in areas of chemical plant, electrical machines, electromechanical devices, and the so-called analogies,

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

respectively. Other more recent works of the engineering systems kind are Franklin (1997), Hsu and Hse (1996), Nise (2000), Oppenheim *et al*. (1996).

Several authors have extracted, from the total systems knowledge contained in such works, the much smaller part that is needed by measuring systems interests, presenting this as chapters in their works. Such accounts are to be found in Beckwith and Buck (1969), Doebelin (2003), and Neubert (1976).

Notable foundation papers have also been published that further condense the information for measurement systems application (see Bosman, 1978; Finkelstein and Watts, 1978). Finally, many control-theory-based works are of relevance.

There exists, therefore, a considerable quantity of well-organized knowledge about the physical behavior that may be encountered in measurement systems studies. When response is linear it is possible to make use of mathematical models to obtain a very workable understanding of the dynamic behavior. If it is nonlinear, however, then the situation is not so well catered for because no such widespread and generally applicable mathematical foundation has been forthcoming in truly formal description terms. In practice, however, designers and users of instruments can find considerable value in assuming linearity, if only for a limited range of operation. For this reason, it is important to know how to recognize whether a system functions in a linear manner and which techniques apply in such cases. Further, it is often up to the designer to chose methods and, thus, it is recommended that consideration be given to the use of those modules that do respond in a linear manner for their operation can be modeled precisely using long hand math or modeling tools.

The purpose of this article is not to expound the mathematical modeling and design of transducers in a concise and rigorous manner, but to present the commonly seen characteristics of linear dynamic systems that will be encountered so that they can be recognized and operated upon by using a fundamental approach that is based on knowledge of their characteristics, *see* **Article 19, Dynamic Behavior of Closed-loop Systems, Volume 1**. It also provides the basic descriptive terminology needed in specification of response characteristics.

2 FORCING FUNCTIONS

To see how a stage performs to various kinds of changing input signals, it is forced into its transient state by one or more external inputs. These are termed *forcing input* or *excitation functions*. They may be applied to the intentional input of the stage or cause transient behavior by entering, as influence quantities, through numerous other unintentional ports. In reality the true two-port, four-terminal, instrument stage can rarely be realized for virtually all designs are influenced to some extent by unwanted noise perturbations entering through many mechanisms.

When considering the dynamic response that might arise for a stage it is, therefore, necessary to first decide the ports through which a forcing function signal might enter. This decided, the next step is to assess which kind of forcing function is relevant and apply this to the real physical device, or to its correct mathematical model expressed in the state-variable or transfer function form. Alternatively, it might be simulated in a digital computer. When the transfer function model is used, the product of the Laplace transform of the forcing function and the transfer function can be solved to provide the transient dynamic behavior in a reasonably simple manner, *see* **Article 19, Dynamic Behavior of Closed-loop Systems, Volume 1**. This method is dealt with in detail numerous electrical engineering and systems texts. Here are presented some typical forcing functions that might be used to test or study an instrument stage.

Certain types of forcing function lend themselves to analytical solution being easy to apply to the Laplace method of response evaluation. They are also simple to procure as practical test signals that, although not perfect, come close enough to the mathematical ideal. For these reasons, testing and evaluation of systems tends to attempt first to make use of one or more of several basic forcing functions.

These functions (see Figure 1) are the discontinuous unit step and the unit impulse, the ramp plus the continuous sine wave. They are described elsewhere and, therefore they do not need further theoretical expansion here.

These functions are easy to produce and apply in practice, but it must be recognized that they may not provide an adequate simulation of the real forcing signals existing. This point is not always made clear in the general treatment of the dynamics of common linear systems stages.

When such simplifications are not adequate, other more suitable ones must be applied. If they can be transformed with the Laplace transform expression, and if the product with the transfer function can be arranged so that a solution into time-variant terms is obtained, then a theoretical study can be made. In many cases, however, this is not possible within bounds of realistic adequacy, and alternative methods of study must be employed if no way can be seen to simplify satisfactorily the forcing transfer function. Simulation using a computer-based tool often provides the means to a solution.

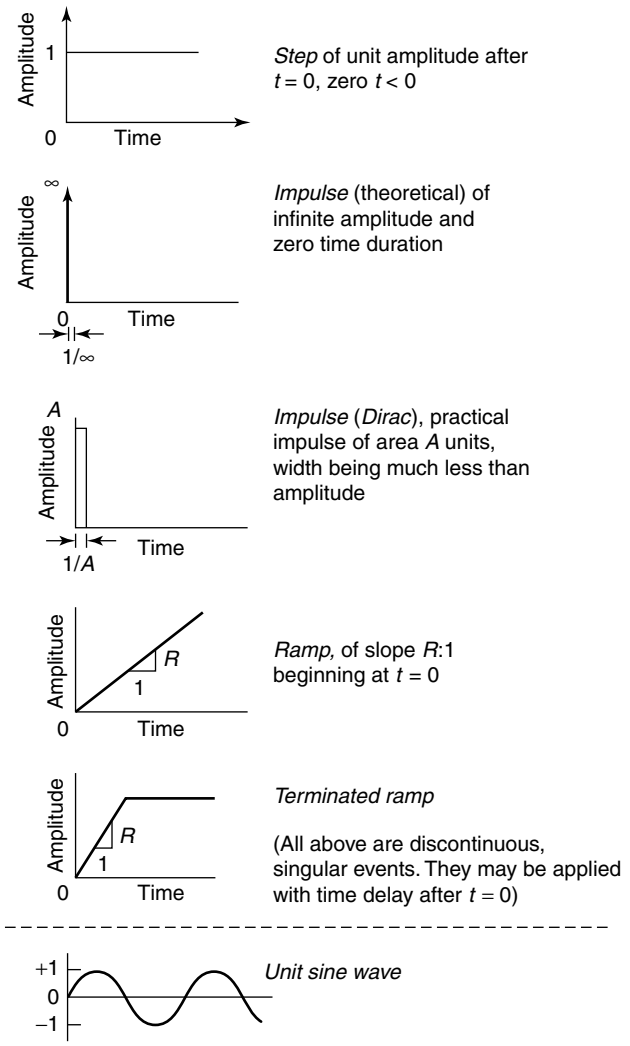


Figure 1. Typical forcing functions used in testing and study of the dynamic response of systems.

3 APPLICATION OF FORCING FUNCTIONS

A step or impulse will excite a system into its dynamic transient response as might a sudden change of demand in a control loop, or a sudden change in a measurand as exists when an indicating voltmeter is first connected to a live circuit. Many physical systems, however, cannot provide such rapidly changing signals. Mechanical and thermal systems often cannot supply a rate of rise that is great enough to be regarded as a step because of the presence of significant storage of energy within the components. For example, attempting to square-wave modulate the dimensions of a piezoelectric crystal at relatively high frequency will produce quasi-sinusoidal output response, not a square wave – the crystal acts as a filter. Similarly, it is not possible to provide a perfect step function to a pressure sensor to obtain its step response. Faster acting systems, such as found in the electronic and optical disciplines, can supply rapid changes.

When the response to a slow-to-rise input signal is needed, the use of the step or impulse function may provide misleading information. A ramp function is more applicable in such cases. Many of the published systematized transient behavior descriptions have not included this particular forcing function, the solutions usually presented being for impulse, step, and sine wave inputs. Sinusoidal excitation can, however, sometimes be used to approximate ramp responses.

The unit impulse function represents the input provided by a sudden shock in a mechanical system – it might be due to slackness in a drive link taking up. In electronic systems, it represents such events as a high-voltage pulse generated by lightning or the switching surges from a nonzero crossing controlled silicon controlled rectifier. In practice, a pseudoimpulsive function, called the *Dirac* or *delta function*, is used instead of the true impulse. As will be seen, the transient solutions of typical linear systems to impulse and step functions are somewhat similar in transient shape.

The third most commonly used input function is the sine wave (or cosine wave, which is the same function, phase-shifted in time). This acts to excite a system in a continuous cyclic manner forcing it to be excited in both the transient and steady state when the sine wave is initially applied, the former dying away to leave the latter as the solution most usually discussed.

In practice, systems are often likely to be disturbed by a continuous complex input waveform. As complex waveforms can be broken down, by Fourier techniques, into a set of sine waves of different frequencies, amplitudes, and phases, the use of sine waves of the correct amplitude and frequency enables the system to be studied one component at a time. Sine wave response is also called *frequency response*, *see* **Article 27, Signals in the Frequency Domain, Volume 1**; **Article 28, Signals in the Time Domain, Volume 1**; and **Article 29, Relationship Between Signals in the Time and Frequency Domain, Volume 1**.

The concept that a complex continuous waveform can be so resolved into separate components rests on the assumption that the system is linear and that superposition applies. Nonlinear systems can behave quite differently to complex signals, creating, for instance, harmonics of lower frequency than those existing in the original signal.

It is often considerably easier, and more reliable to obtain the transient response of a system by practical testing than it would be to develop a mathematical model. This is one of the reasons why the data sheets of instrument products

often include graphical statements of transient response. Further, a detail about forcing functions can be found in chapters of classical control texts such as Atkinson (1972), Coughanowr and Koppel (1965), and Shearer, Murphy and Richardson (1967).

RELATED ARTICLES

Article 18, Nature and Scope of Closed-loop Systems, Volume 1; **Article 19, Dynamic Behavior of Closed-loop Systems, Volume 1**; **Article 27, Signals in the Frequency Domain, Volume 1**; **Article 28, Signals in the Time Domain, Volume 1**; **Article 29, Relationship Between Signals in the Time and Frequency Domain, Volume 1**; **Article 30, Statistical Signal Representations, Volume 1**.

REFERENCES

Atkinson, P. (1972) *Feedback Control Theory for Engineers*, Heineman, London.

Beckwith, T.G. and Buck, N.L. (1969) *Mechanical Measurements*, Addison-Wesley, Reading, MA.

Bosman, D. (1978) Systematic Design of Instrumentation Systems. *Journal of Physics E:Scientific Instruments*, **11**, 97–105.

Coughanowr, D.R. and Koppel, L.B. (1965) *Process Systems Analysis and Control*, McGraw-Hill, New York.

Doebelin, E.O. (2003) *Measurement Systems: Application and Design*, 5th edn, McGraw-Hill, New York.

Finkelstein, L. and Watts, R.D. (1978) Mathematical Models of Instruments – Fundamental Principles. *Journal of Physics E:Scientific Instruments*, **11**, 841–55.

Fitzgerald, A.E. and Kingsley, C. (1961) *Electrical Machinery*, McGraw-Hill, New York.

Franklin, G.F. (1997) *Digital Control of Dynamic Systems*, Addison-Wesley.

Hsu, A. and Hse, W. (1996) *Schaum's Outline of Signals and Systems*, McGraw-Hill.

Karnopp, D. and Rosenberg, R.C. (1975) *System Dynamics: A Unified Approach to Physical Systems Dynamics*, MIT Press, Cambridge, MA.

Klirs, G.J. (1972) *Trends in General Systems Theory*, Wiley, New York.

Koenig, H.E. and Blackwell, W.A. (1961) *Electromechanical Systems Theory*, McGraw-Hill, New York.

MacFarlane, A.G.J. (1964) *Engineering Systems Analysis*, Harrap, London.

Neubert, H.K.P. (1976) *Instrument Transducers*, 2nd edn, Clarendon Press, Oxford.

Nise, N.S. (2000) *Control Systems Engineering*, Wiley, New York.

Olson, H.F. (1943) *Dynamical Analogies*, Van Nostrand, New York.

Oppenheim, A.V., Willsky, S.H., Nawab, A. and Nawad, H. (1996) *Signals and Systems*, Prentice Hall, Englewood Cliffs, NJ.

Paynter, H.M. (1961) *Analysis and Design of Engineering Systems*, MIT Press, Cambridge, MA.

Shearer, J.L., Murphy, A.T. and Richardson, H.H. (1967) *Introduction to Systems Dynamics*, Addison-Wesley, Reading, MA.

Zahed, L.A. and Desoer, C.A. (1963) *Linear System Theory*, McGraw-Hill, New York.

60: Zero-order System Dynamics

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 General Remarks on Linear Systems Responses 367
2 Summary of Systems Orders in Transfer Function Form 367
3 Definition of the Zero-order System Response 368
4 Design Aims for Measurement Systems Behavior 369
5 Examples of Zero-order Measurement Systems 369
6 Specifying Zero-order Systems 369
Related Articles 370
Reference 370

1 GENERAL REMARKS ON LINEAR SYSTEMS RESPONSES

The following discussion applies only if the dynamic behavior of the system of interest can be adequately characterized by a suitable linear ordinary differential equation, this often being found to be the case in instrumentation, but by no means always so. The theoretical foundation given in this article provides the basis for understanding the behavior of linear systems in general. The first-order system is covered in **Article 61, First-order System Dynamics, Volume 1** and the second-order system in **Article 62, Second-order System Dynamics, Volume 1**.

These equations will have the general form (called the *nonhomogeneous linear differential equation*) for input variable x and output variable y and unit amplitude input:

$$a_n \frac{\mathrm{d}^n y}{\mathrm{d}t^n} + a_{n-1} \frac{\mathrm{d}^{n-1} y}{\mathrm{d}t^{n-1}} + \cdots + a_0 y = x(t) \qquad (1)$$

The coefficients $a_n, \ldots, a_0$ are constants. They could be functions of t but not of y. Powers of derivatives do not occur in this class of differential equation. Not all orders of derivatives less than the highest are necessarily present.

The expression on the left-hand side is the characteristic equation of the differential equation. It contains the information that characterizes the physics of the system of interest and how it will provide output when it is excited. The expression on the right-hand side characterizes the forcing function, the input signal, applied to that system. Thus, both sides contain independent information about, on the one hand, the inherent transfer characteristic of the stage and, on the other hand, the forcing function applied to the stage. As both functions are linear expressions, by the above definition of requirement, they can both be transformed into transfer functions by Laplace methods.

System performance can thereby be obtained theoretically if the nonhomogeneous linear differential equation for the specific system can be solved to yield the output versus time relationship.

In providing the explanation of dynamic response, authors can make use of the differential equation format given here but often using other symbols: they may make use of the operator D to replace derivatives; the explanation might be presented directly in transfer function form; differential equations might be presented in complex number form, plus other variations. These tend to confuse the issue somewhat but each produces the same result.

2 SUMMARY OF SYSTEMS ORDERS IN TRANSFER FUNCTION FORM

The above given general differential equation can be written in the specific sense as an ascending order of equations

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

when additional derivative orders are progressively added. For unit amplitude input, these become

$$\begin{aligned}
&\text{zero order} && a_0 y = x(t)\\
&\text{first order} && a_1 \frac{\mathrm{d}y}{\mathrm{d}t} + a_0 y = x(t)\\
&\text{second order} && a_2 \frac{\mathrm{d}^2 y}{\mathrm{d}t} + a_1 \frac{\mathrm{d}y}{\mathrm{d}t} + a_0 y = x(t)\\
&n\text{th order} && a_n \frac{\mathrm{d}^n y}{\mathrm{d}t^n} + a_{n-1} \frac{\mathrm{d}^{n-1} y}{\mathrm{d}t^{n-1}} + \cdots + a_0 y = x(t)
\end{aligned} \tag{2}$$

Zero-, first-, and second-order systems are all that need usually be considered in this discussion of the transient response of instrument stages because higher-order differential equations can be reduced, if need be, to these, through suitable manipulation. Also, establishing analytical solutions, which requires the roots to be found to higher orders can be far more difficult, if at all possible, than for these three cases. This may seem rather short of real requirements but, in general, a very large part of instrument testing and characterization can be done with the understanding of these three types of system response.

Transformed into typical unitized transfer functions, the differential equations become

$$\begin{aligned}
&\text{zero order} && \frac{Y(s)}{X(s)} = 1\\
&\text{first order} && \frac{Y(s)}{X(s)} = \frac{1}{\tau s + 1}\\
&\text{second order} && \frac{Y(s)}{X(s)} = \frac{1}{(\tau_1 s + 1)(\tau_2 s + 1)}\\
&n\text{th order} && \frac{Y(s)}{X(s)} = \frac{1}{(\tau_1 s + 1)(\tau_2 s + 1)\cdots(\tau_n s + 1)}
\end{aligned} \tag{3}$$

where $\tau_1, \ldots, \tau_n$ are termed time constants and $Y(s)$ and $X(s)$ the output and inputs respectively. This groundwork laid now enables the three types of systems to be studied in order. This article covers the simplest, almost trivial one – the zero-order system. *See* **Article 61, First-order System Dynamics, Volume 1** for first order, **Article 62, Second-order System Dynamics, Volume 1** for second order.

3 DEFINITION OF THE ZERO-ORDER SYSTEM RESPONSE

The zero-order system is so trivial mathematically that it is rarely mentioned in texts on dynamic response. In practice, however, it is often (but not always) a most desirable response, one that the designer strives to achieve but rarely obtains in a perfect sense.

The differential equation form, given above, shows that there are no derivatives, meaning that the system does not alter the time features of a time-dependent function introduced to the system by the forcing function. Thus, the zero-order stage cannot introduce phase shifts between the input and output, alter frequencies, or provide changing amplitude with frequency. It can, however, provide constant attenuation or gain and it can transform energy variables. The numerical coefficient a_0 is the static sensitivity of the stage, the constant of transduction. In principle, the zero-order stage provides perfect dynamic response for situations where the input variable must not be processed with respect to its time features as it passes through the stage.

Considering the transfer function form, it can be seen (but this will become clearer when the higher-order systems are studied) that the stage contains no storage mechanism because there is no time constant associated with it.

In many design situations, a zero-order response is highly desirable. For example, consider a resistance strain gauge (Figure 1a) in which extension of the electrical resistance wire, attached to a sample, alters the resistance proportional to the strain so induced. In normal use, the gauge, being apparently purely resistive electrically, will respond completely in phase with strains of the sample. No matter how rapid the input signal is, within practicality, the output follows without delay, without signal amplitude change, and without frequency change (Figure 1b).

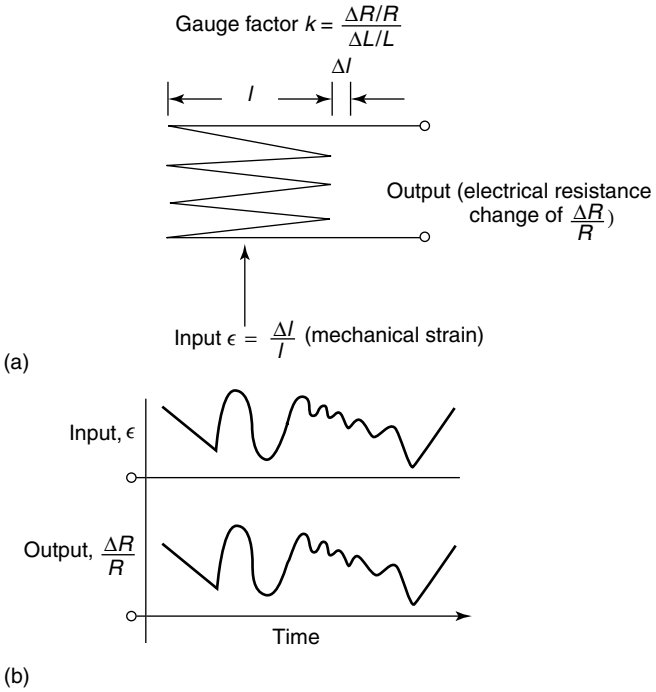


Figure 1. Example zero-order system – a resistive strain gauge: (a) gauge and its mathematical model; (b) responses.

As the gauge is presumably being used to monitor the strain behavior of the sample, zero-order response will yield the observer true information about what is actually happening to the sample with time and in amplitude. In control-loop design, zero-order sensing components often help the designer obtain a very tight, responsive system.

In practice, the designer or user must decide if zero-order characteristics are what are needed (see below). Assuming this to be so, it is then necessary to ascertain that the device does indeed produce them to the dynamic performance level desired, for, in reality, a perfect zero-order system is a mathematical abstraction. All systems include some measure of storage effects; it is these that will degrade the zero-order response into a higher-order system where they become significant.

In the strain gauge example, storage might arise through thermal capacity of the gauge and sample; through after-elastic mechanical effects in the gauge material; through parasitic electrical inductance and capacitance when the frequency of response or excitation increases sufficiently; through mechanical compliance (and its capacity to enable the mass of the gauge to move out of phase with the sample) in the bonding of the gauge to the sample via the thin layer of adhesive; plus other reasons. These, in practice, will each introduce time-dependent features situated over the frequency spectrum of response, from virtually DC to the radio frequencies.

4 DESIGN AIMS FOR MEASUREMENT SYSTEMS BEHAVIOR

The art of good instrument design and application is to select a stage that performs with the response that just suffices. In a relative humidity sensor, it can now be seen that the hygroscopic membrane, the strained metal beam, the strain gauges, and the electrical resistance Wheatstone bridge used can each be regarded as zero-order elements for the purpose of studying and designing the sensor for the application it was built to fulfill.

If a response less than several minutes is needed, then the various elements must be viewed as higher-order components, and a more complex model must be used to study and improve the overall dynamic behavior.

Design of instrument systems does not, however, always require zero-order performance. In many instances, there is need for storage in a system. Storage helps provide such often-desirable effects as integration and differentiation, which, in turn, enable frequency filtering to be devised. Applicability of the relative humidity sensor is enhanced in many applications by the long storage time of the membrane. This helps average (integrate) fluctuations that might occur in a room in which a fan circulates pockets of air differing in %RH that would otherwise cause a closed-loop controller to attempt to smooth out to the detriment of the real need.

5 EXAMPLES OF ZERO-ORDER MEASUREMENT SYSTEMS

Of the texts already referenced concerning dynamic response, only Doebelin (2003) makes specific mention of zero-order systems. His description includes another example – that of a resistance potentiometer used as a position-sensitive sensor.

It will probably now be clear that it is not possible to provide a list of zero-order instrument devices and stages for that attribute is decided by the use to which the stage is applied. As a guide, if the energy storage effects are minimal compared to existing energy levels and if only relatively low-frequency forcing functions are to be applied to the stage, then zero-order response probably occurs. The important factors are the relative values of these parameters and not their absolute magnitudes. A device operating at picowatt energy levels can provide delays of hours and conversely, a megawatt energy system might respond in milliseconds.

It can be very dangerous, in practice, to assume the response order for a given system. Simple testing will easily establish the order for the application in question. This is to be recommended where possible. Many of the problems that arise in measuring system design and application occur as a result of assuming that a sensor or stage provides zero-order response when, in the specific application, it actually operates with higher-order responses present. One example is the use of a classical flat bed plotting table that will filter signals that are too fast for it to respond to. It will still record them, but with diminishing amplitude as the frequency of the component rises. Phase shift will also occur between signals being recorded on multichannel recorders.

6 SPECIFYING ZERO-ORDER SYSTEMS

Each kind of response can be described in simple terms by stating certain simple descriptors. In the case of the zero-order system, these are as follows:

- Systems type is of zero order
- Amplification constant.

There is no need to describe any frequency features for it is known that zero-order systems do not exhibit any frequency varying behaviors. The other orders need additional descriptors.

RELATED ARTICLES

Article 35, Systems in the Time Domain, Volume 1; **Article 36, Systems in the Frequency Domain, Volume 1**; **Article 37, Relationship Between Systems in the Time and Frequency Domain, Volume 1**; **Article 38, Stability Issues, Volume 1**; **Article 56, Transfer Characteristics of Instrument Stages, Volume 1**; **Article 57, Static Considerations of General Instrumentation, Volume 1**; **Article 59, Introduction to the Dynamic Regime of Measurement Systems, Volume 1**.

REFERENCE

Doebelin, E.O. (2003) *Measurement Systems: Application and Design*, 5th edn, McGraw-Hill, New York.

61: First-order System Dynamics

Peter H. Sydenham

GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Descriptive Equations for First-order System and an Example 371
2 Response of First-order System to Step Function Input 373
3 Response of First-order System to Ramp Function Input 374
4 Response of First-order System to Sinusoidal Function Input (Frequency Response) 375
5 Examples of First-order Measurement Systems 376
6 Specifying First-order Systems 377
Related Articles 377
References 377

1 DESCRIPTIVE EQUATIONS FOR FIRST-ORDER SYSTEM AND AN EXAMPLE

The dynamic regime of systems is introduced in **Article 59, Introduction to the Dynamic Regime of Measurement Systems, Volume 1**. For linear signal systems the various responses can be ordered. The zero-order system is covered in **Article 60, Zero-order System Dynamics, Volume 1**. This article deals with the dynamic behavior of first-order systems. Second-order responses are discussed in **Article 62, Second-order System Dynamics, Volume 1**.

The differential equation for a first-order system is, with a unit amplitude forcing function:

$$a_1 \frac{dy}{dt} = a_0 y = x(t) \tag{1}$$

Rearranging the coefficients to yield a unit coefficient for y gives

$$\frac{a_1}{a_0}\frac{dy}{dt} + y = \frac{1}{a_0}x(t) \tag{2}$$

Taking Laplace transforms produces

$$\frac{a_1}{a_0}sY(s) + Y(s) = \frac{1}{a_0}X(s) \tag{3}$$

Rearrangement again gives the transfer function as

$$G(s) = \frac{Y(s)}{X(s)} = \frac{1}{a_0}\frac{1}{(a_1/a_0)s + 1} \tag{4}$$

At zero frequency, and with sinusoidal excitation, the operator 's' validly becomes jω and is equal to zero. Thus:

$$G(s) = \frac{1}{a_0} \tag{5}$$

The coefficient $1/a_0$ represents the static regime transduction constant or amplification factor, often denoted as K. It may involve either attenuation (or gain) of a common input–output variable or relate one form of variable to another.

The ratio a_1/a_0 is termed the *time constant* for reasons that will become clear below. It has dimensions of time and it is generally denoted as τ. All first-order systems have the same characteristic kind of time response to any given forcing function and τ specifies the actual numerical magnitude of that response, this being regardless of the energy regime of the system.

The mercury-in-glass thermometer is what is chosen by many authors to illustrate a first-order system. In use, the

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.

© 2005 John Wiley & Sons, Ltd.

ISBN: 0-470-02143-8.

thermometer is immersed in a fluid bath and is intended to transduce fluid temperature into an equivalent capillary-tube mercury movement without introducing delay or being affected in amplitude by the speed of the input temperature change. In practice, the thermometer can only be regarded as a zero-order system when very slow changes are to be monitored. For changes that take place in a few seconds, the thermal storage effect of the mercury and the heat flow restriction of the thermal film transfer effect that thermally couples the fluid to the mercury, provide first-order action.

Figure 1 is a schematic of the system including all variables that experience shows to be relevant in order that the system can be adequately treated by a first-order model.

The physical operation of the system must first be studied to discover assumptions that can be made in order to obtain an adequately descriptive model that is also reasonably workable, mathematically.

Assumptions in this example relate to such factors as the majority of resistance to heat transfer between the fluid and the mercury being provided by the heat transfer film, the glass being of insignificant thermal resistance. The mercury is presumed to have perfect heat conducting properties and to respond to heat changes producing instantaneous corresponding volume changes. Other factors assumed are that the glass envelope changes volume at the same rate as the mercury, an assumption that is not exactly true when high discrimination is needed since the glass takes considerable time to return to a smaller volume after being expanded. Other factors that must be presumed to hold are that heat transfer is achieved over a constant surface area and that it is not a function of the actual temperature. At least two more important requirements that do not strictly hold are that heat is only stored in the mercury, not in the glass, and that no heat is lost to the environment via the heat conduction path of the thermometer stem.

Despite the obvious presence of these many imperfections, the first-order model that can be developed for the thermometer does provide quite a reasonable estimate of the dynamic performance of the thermometer under changing bath temperature conditions.

Derivation of the differential equations, on the basis of energy balance and knowledge of the physical process by which heat changes cause a rise in the mercury meniscus, leads to the actual development of numerical quantities for the time constant and the transduction constant. This example is worked through in both Coughanowr and Koppel (1965) and Doebelin (2003) each providing a slightly different form of expression of the resulting numerical factors.

In terms of the parameters defined in Figure 1, it can be shown that

$$\left.\frac{\Delta y}{\Delta x}\right|_{\omega=0} = K = \frac{\Delta V \times V}{a} \tag{6}$$

and

$$\tau = \frac{\rho s V}{h A} = \frac{\text{heat stored}}{\text{resistance to heat flow}} \tag{7}$$

Note that the time constant parameter concerns one storage and one energy dissipative parameter.

It is, therefore, possible to determine theoretically not only the transduction constant as a numerical value (as also could a practical test) but it also provides knowledge of

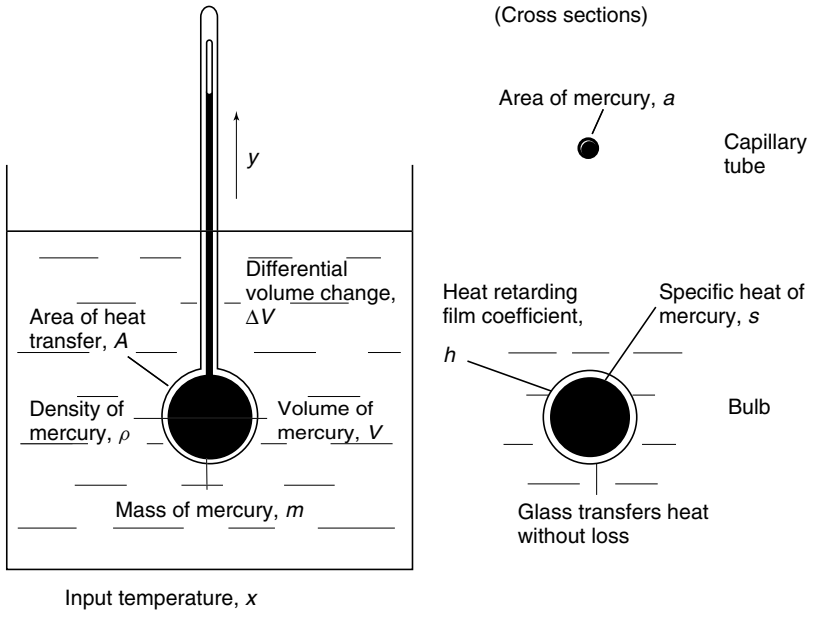


Figure 1. Basic physical system of a mercury-in-glass thermometer modeled at a first-order system level.

the factors of the system upon which it depends; similarly so for the time constant. Knowing these parameters enables the system to be tuned to provide whatever static and dynamic response is needed or to learn that existing materials and other practical constants might not allow a specified performance to be realized. They also enable the system to be easily specified in terms of three statements: the system order, K, and τ. The constants, however, vary greatly depending on the system conditions into which the thermometer is coupled. They cannot be quoted without reference to the model conditions that are applicable.

The importance of τ can now be considered in its capacity as a unique descriptor of the transient behavior of a first-order system when excited by certain given forcing functions.

Of the forcing functions introduced above, the step (the impulse, being similar, need not be covered in the same detail here), the ramp, and the sine wave are the most useful inputs to consider. Other input functions will require a special mathematical study of the transfer function along the lines now outlined.

2 RESPONSE OF FIRST-ORDER SYSTEM TO STEP FUNCTION INPUT

The Laplace expression for a step function of amplitude A is

$$X(s) = \frac{A}{s} \tag{8}$$

Response of a first-order system, which has $G(s) = K/(\tau s + 1)$ (where K is the transduction constant and τ the time constant) is given by

$$G(s)X(s) = \frac{K}{\tau s + 1}\frac{A}{s} \tag{9}$$

Expansion by partial fractions yields

$$Y(s) = \frac{AK}{s} - \frac{AK}{s + 1/r} \tag{10}$$

Using tables of Laplace transformations of common functions to invert this expression gives, for t from zero onward:

$$y(t) = AK(1 - \mathrm{e}^{t/\tau}) \tag{11}$$

This expression describes the dynamic behavior of the first-order system to which a step function of a given amplitude is applied at $t = 0$. For the above to hold when the step is applied, the system must already be in the steady state, meaning that the storage element is completely discharged through the dissipative path; the initial conditions are thus satisfied. If energy is already being discharged, that is, the system is still in the transient state of a previous forcing function application, then the response will be the result of two forcing functions and not the step applied here.

Being generally applicable, this time function can usefully be drawn as a graph in which the input is normalized to a unity maximum and the time scale is expressed in terms of a time-variable ratio. This is shown in Figure 2.

If the step is applied as a decrement, or fall, rather than the rise shown here, the curve is simply reversed, the shape is identical. It will then follow the form

$$y(t) = AK\mathrm{e}^{-t/\tau} \tag{12}$$

It can be seen that τ is an important descriptive constant of the system, deciding the curve's actual magnitude. It also directly describes the initial rate of rise; a tangent to the original line will intercept the final value level at the $1\,\tau$ point. This implies that if the system were to maintain (which it does not) the initial slope, the output would reach the final value in the time of $1\,\tau$. In general, this slope, for any order of system, is termed its *slewing rate*.

It can be seen from Figure 2 that the continued response reaches what might be considered to be close enough to the final value, at 86.5, 95.0, and 98.2%, levels for 2τ, 3τ, and 4τ, respectively. By 5τ, it is certainly close enough to a settled value for most applications. Although the time constant is the key parameter of definition, it must be realized that for final values to be reached, it takes a longer time than that given by the appropriate factors presented above. In 1τ, the response has only reached 63.2% of the final value. Furthermore, it rises to 63.2% of the remainder in the next 1τ period, and so forth.

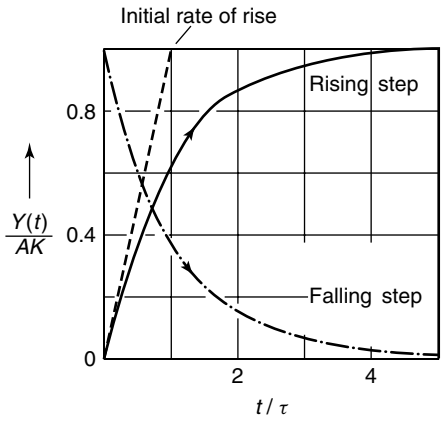


Figure 2. Normalized response of a first-order to a step input function.

Respective values for a falling unit step are that it falls to 36.8% in 1τ, to 13.5% in 2τ, to 5% in 3τ, and 1.8% by the time 4τ has passed.

In order to provide a general descriptive term that applies to any instrument response, linear or nonlinear, instrument users have adopted the term *settling time*. This is usually quoted as the time needed for the system to settle to within given percentage limits of the final value. These limits may be expressed as an error band. For example, it will take a first-order system, excited by a unit step function, a time of 3τ to come within $\pm 5\%$ limits. Note that a first-order system cannot overshoot the final value but approaches it from one side only.

Detailed analysis of impulse input function response is to be found in Coughanowr and Koppel (1965).

3 RESPONSE OF FIRST-ORDER SYSTEM TO RAMP FUNCTION INPUT

Step inputs may give an unrealistically severe rate of rise. Using a step or impulse in such cases will provide an output response that is much greater in transient deviation than what actually occurs. Slower rise rates, as represented by the ramp input function, are often more realistic forcing functions to use. As will be seen, the output response to a ramp follows the input more faithfully but quite an unexpected delay and possibly a progressively increasing deviation error results.

Using the same mathematical technique as was applied in the previous step response explanation, it is first necessary to obtain the Laplace transform for a ramp function of rate of rise, R. Forming the product of this with the transfer function for a first-order system having a transduction constant of K yields

$$Y(s) = G(s)X(s) = \frac{K}{\tau s + 1}\frac{R}{s^2} \tag{13}$$

Rewritten, this becomes

$$Y(s) = \frac{RK}{\tau}\frac{1}{s^2(s + 1/\tau)} = \frac{a}{s^2(s + b)} \tag{14}$$

where

$$a = \frac{RK}{\tau} \quad b = \frac{1}{\tau} \tag{15}$$

Expanding this into partial fractions gives

$$Y(s) = \frac{a/b}{s^2} - \frac{a/b^2}{s} + \frac{a/b^2}{s + b} \tag{16}$$

reverting a, b to physical constants

$$Y(s) = \frac{RK}{s^2} - \frac{RK\tau}{s} + \frac{RK\tau}{s + 1/\tau} \tag{17}$$

These are now in the form from which inverse Laplace transforms can be recognized from standard tables. Three time functions result to make up the whole response function:

$$Y(t) = RKt - RK\tau u(t) + RK\tau e^{-t/\tau} \tag{18}$$

that is, output = ramp + step + exponential transient.

This is interpreted to be a ramp function response of slope rate RK. It will have a different rate to the input when the stage has either attenuation or gain for the common input and output variables (that is, volts in, and volts out, as would occur in an RC stage) and the rate will differ when the rate is not linked to the input because the transduction constant is one of conversion (that is, temperature change yielding a length change output). If the transduction constant is unity, then input and output slopes will be identical.

The second term, a step of $RK\tau$ amplitude, and negative in sign, tells us that the output response begins to occur at $t = 0$ after a time lag. If the rates of the input and output are identical, there will exist a constant lag with time, but if they differ, the lag magnitude will be proportionately related to the time that has elapsed after initiation of the input signal. This lag effect has been interpreted by some authors as meaning that this combination of input to a first-order system will produce a system in which the output measured value at a given time is that of a fixed time before. This observation only holds true if the rates of the input and output ramps are identical. If not, the steady state error changes with time.

Finally, the third component of the total response is a transient signal of exponential form that occurs at the commencement of the ramping action to die away to virtual zero within a short time.

The total response is shown diagrammatically in Figure 3. The form of the transient part of the response, when isolated, follows that given as the falling curve in Figure 2.

Clearly, if the output is to be a faithful replica of the ramp input, then the time constant of the first-order stage must be sufficiently small in magnitude. Furthermore, the transduction constant K needs to be unity. When these conditions are met, a ramp will be followed with negligible droop, with minimum time delay, with minimal transient component, and with a small fixed, rather than changing, magnitude error. (Under such conditions it virtually becomes a zero-order system.)

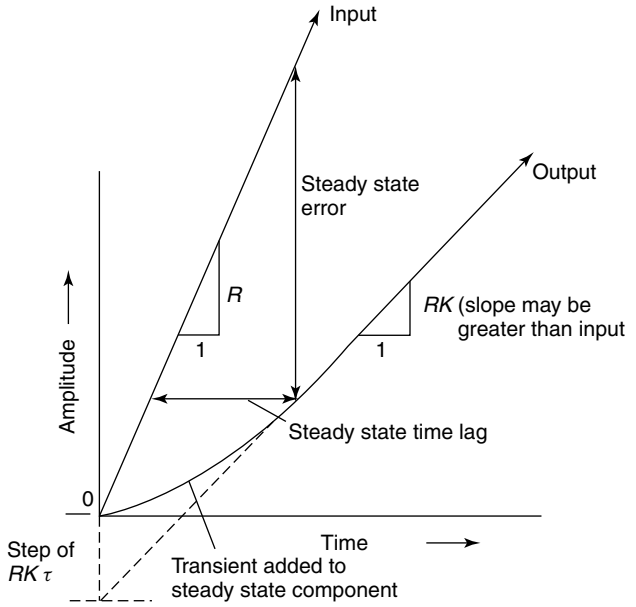


Figure 3. Response of a first-order system to a ramp input function.

4 RESPONSE OF FIRST-ORDER SYSTEM TO SINUSOIDAL FUNCTION INPUT (FREQUENCY RESPONSE)

The requirement for testing, or understanding, may relate to the behavior under conditions of continuous steady state excitation of the system stage. This is fundamentally approached using the sine wave signal train as the forcing function since all other continuous complex signals can be reduced to the sum of such responses. The forcing function in this case in the untransformed time domain is

$$X(t) = A \sin \omega t \tag{19}$$

In Laplace form

$$X(s) = \frac{A\omega}{s^2 + \omega^2} \tag{20}$$

Thus, the response of a first-order network to this will be

$$Y(s) = G(s)X(s) = \frac{K}{\tau s + 1}\frac{A\omega}{s^2 + \omega^2} = \frac{AK}{\tau}\frac{\omega}{(s^2 + \omega^2)(s + 1/\tau)} \tag{21}$$

This can be shown to be as in Coughanowr and Koppel (1965), after rearrangement, taking partial fractions and then inverse Laplace transformation:

$$Y(t) = \frac{AK\tau\omega e^{-t/\tau}}{\tau^2\omega^2 + 1} + \frac{AK}{\sqrt{(\tau^2\omega^2 + 1)}} \sin(\omega t + \phi) \tag{22}$$

where $\phi = \tan^{-1}(-\omega\tau)$; that is,

$$\text{output response} = \text{transient} + \text{steady state components} \tag{23}$$

The transient term is short lived, decaying to leave only the steady state component, that is, the contribution described as the frequency response of a system. The steady state term is a sine wave of the same frequency as the forcing function but it is phase shifted as a lag, by angle ϕ. It will have the so-called DC gain decided by the amplitude of sine wave at that frequency and the transduction constant. The gain factor, however, will progressively decrease (be attenuated) as frequency rises, the magnitude of this attenuation being dependent upon the time constant and the frequency.

It is clear, from the presence of the transient term, that a first-order system will not immediately respond in the steady state form but will undergo initial exponential behavior. Thus, the system will not ring (oscillate) but does need a certain amount of time to establish its intended frequency response action.

The steady state normalized frequency responses of a first-order system with transduction constant K, to input sine wave frequencies of amplitude A are plotted in Figure 4, being given with respect to the amplitude and phase of signals.

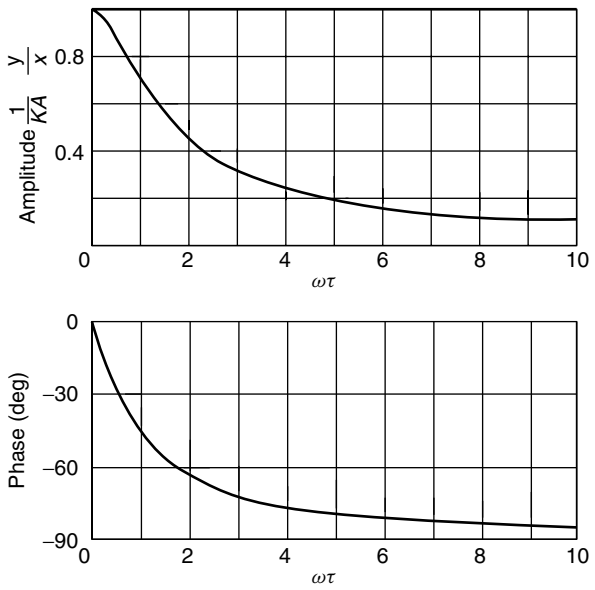


Figure 4. Normalized response of a first-order system to sinusoidal signals of varying frequency.

A first-order system will, therefore, not alter the applied frequency but may attenuate the amplitude and introduce phase shift rising, in a lagging sense, to a maximum of 90°, the degree of these effects depending upon the time constant of the system and the frequency of operation in question.

By way of an example that shows how these dynamic features can provide incorrect measurement action by a first-order sensor, consider a temperature sensor. The example is fully worked through in Coughanowr and Koppel (1965), so only the results need be given here. A thermometer with a time constant of 6 s (that is, for its use in a defined application; it is not a feature of the thermometer alone) is placed in a bath to detect bath temperature variations of $\pm 1\,^{\circ}\mathrm{C}$ amplitude that are sinusoidal at around three cycles per minute. Because of its inherent frequency response, these cycles would be indicated to be half their real magnitude, the sine wave output being of the same frequency but now delayed by some 4 s.

A second example is found when considering the mechanical pen recorder. When adjusted for critical damping, the response of a plotter can be regarded as being essentially of first-order character. If the recorder response is inadequate, it will plot signals at reduced amplitude. If the signal of one channel is of lower frequency than the other but of the same amplitude, the recorder will not record a true indication of the amplitudes and phase existing between the two signals being recorded.

For similar reasons, filters added to smooth out noisy signals (in any energy regime), may also lead to similar inaccuracies when more than one frequency component exists in the forcing function, that is, any complex continuous signal. The response can be found by making use of superposition, which is applicable when the system is linear. This allows each component to be evaluated separately for magnitude and phase values, the set then being added to obtain the total response. A worked example for a forcing function with two components is given in Doebelin (2003).

Complex signals will not, therefore, give exactly the same normalized responses as are given in Figure 4, but they will always produce attenuation and phase shift as τ and ω increase. It can be seen that a square wave, being a set of sine wave components, will produce a response in which each cycle follows a step response profile as the square wave rises and falls, the starting point of each new transition depending upon the level to which it rose, or fell, in the previous cycle.

As has already been mentioned, the time constant used must be the effective value of the combination present when the stage is interfaced. It will not be the same value as that of the stage in isolation, unless the input source and output load added to the stage do not interact through loading mechanisms. For example, a simple electrical low-pass RC stage will have, in isolation, $\tau = RC$. When the input source impedance and the output load impedances are connected, they each modify the effective value of resistance because the load resistance is added in parallel with C and the source resistance is added in series with R.

This situation has been described in Section 1, where the mercury-in-glass thermometer was considered; in that example, the properties of the fluid can greatly vary the effective time constant experienced.

It might be thought that the type of a stage can be recognized uniquely from the shape of the step response. However, this is not necessarily a unique indicator because systems of any order can, if suitably conditioned, exhibit somewhat similar transient responses. This occurs when the time constants of the respective order contributions are such that the components producing them effectively work in phase. As the order rises, however, the shape of the curve changes more noticeably enabling the nonconformity to a first-order response to be more easily recognized.

5 EXAMPLES OF FIRST-ORDER MEASUREMENT SYSTEMS

The suitably thermally lagged thermometer has already been discussed. From that example, it is possible to recognize that a first-order system results when there exists a significant energy storage mechanism and the inability of that stored energy to discharge through a dissipative element in a relatively short time. The direct combination of these two system properties decides the all-important time constant.

As the dissipating property (often called *the loss*) decides how well a response is damped (assuming, in such cases that a fixed level of energy storage capacity exists) this is often called *the damping property*. The simplest form, mathematically, is that in which the rate of damping is fixed independently of the rate of the signal. In mechanical systems, this is called *Coulomb friction*. Almost always, however, it will be found that the damping property in a system is dependent upon influence variables, such as temperature. Friction that is proportional to the velocity of the signal is termed *Newtonian*. Add to these two the fact that the friction coefficient for static systems is generally much higher than for those after breakaway has occurred (nonlinear friction), it will be seen that practical use of linear systems theory is a much-simplified version of reality. Although expressed in now deprecated units, the practical information on damping is contained in a classic statement. Drysdale and Jolley (1924) is very relevant to fine mechanical systems.

In electrical systems, the obvious first-order example is the *RC* combination, such as is met in a low-pass filter stage. Cascading first-order stages raises the order of the system producing a different result from that obtained by the same components arranged to produce the same time constant as a single stage.

In fluid systems, first-order stages occur in tanks of fluid being drained, the resistance to outward flow and the capacity of the tank forming the time constant pertinent to level of the fluid. Chemical systems provide this order action when solutes are mixed into solution. In acoustics, an air enclosure having a baffled (resistance to airflow) output port, as is used in some speaker enclosure designs, is another fluid regime example. The same applies to a similar situation met in pneumatic instrumentation; much of pneumatic equipment, however, is not linear in operation. First-order systems are also referred to as providing, or being, an exponential lag stage.

Further details of practical examples are to be found in Atkinson (1972). Coughanowr and Koppel (1965), Olson (1943), and Shearer, Murphy and Richardson (1967). Other works that may be found to be useful are Oppenheim *et al.* (1996), Hsu and Hse (1996), Nise (2000).

6 SPECIFYING FIRST-ORDER SYSTEMS

First-order systems require the following features to be given when specifying their response:

- system type is of the first order
- amplification factor
- time constant, usually in seconds.

RELATED ARTICLES

Article 37, Relationship Between Systems in the Time and Frequency Domain, Volume 1; **Article 38, Stability Issues, Volume 1**; **Article 59, Introduction to the Dynamic Regime of Measurement Systems, Volume 1**; **Article 60, Zero-order System Dynamics, Volume 1**.

REFERENCES

Atkinson, P. (1972) *Feedback Control Theory for Engineers*, Heineman, London.

Coughanowr, D.R. and Koppel, L.B. (1965) *Process Systems Analysis and Control*, McGraw-Hill, New York.

Doebelin, E.O. (2003) *Measurement Systems: Application and Design*, 5th edn, McGraw-Hill, New York.

Drysdale, C.V. and Jolley, A.C. (1924) *Electrical Measuring Instruments*, Part 1, Ernest Benn, London.

Hsu, A. and Hse, W. (1996) *Schaum's Outline of Signals and Systems*, McGraw-Hill.

Nise, N.S. (2000) *Control Systems Engineering*, Wiley, Chichester.

Olson, H.F. (1943) *Dynamical Analogies*, Van Nostrand, New York.

Oppenheim, A.V., Willsky, S.H., Nawab, A. and Nawad, H. (1996) *Signals and Systems*, Prentice Hall, Upper Saddle River, NJ.

Shearer, J.L., Murphy, A.T. and Richardson, H.H. (1967) *Introduction to Systems Dynamics*, Addison-Wesley, Reading, MA.

62: Second-order System Dynamics

Peter H. Sydenham

GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Descriptive Equations for Second-order System and an Example 378
2 Response of Second-order System to Step Function Input 379
3 Response of Second-order System to Ramp Function Input 380
4 Response of Second-order System to Sinusoidal Function Input (Frequency Response) 381
5 Selectivity Issues for Second-order Systems 381
6 Specifying Second-order Systems 382
Related Articles 383
References 383

1 DESCRIPTIVE EQUATIONS FOR SECOND-ORDER SYSTEM AND AN EXAMPLE

Second-order systems can arise when two energy-storage mechanisms exist in a system and act with some degree of interaction. They must also be able to act with a 180° phase shift between them, for example, a capacitor and an inductor in an electrical system but not two inductors or two capacitors. They need not be of the same energy type, for energy conversion mechanisms make them mathematically and practically compatible. Dissipative qualities are, again, important as this is the factor that decides the damping of responses.

A most noticeable feature of second-order systems is that they can overshoot final values, even providing oscillation (when sufficiently small damping exists).

Characteristics of a second-order stage will first be derived, followed by the study of its response to step, ramp, and continuous sine-wave forcing functions.

The second-order system is described by the nonhomogeneous ordinary linear differential equation:

$$a_2 \frac{\mathrm{d}^2 y}{\mathrm{d}t^2} + a_1 \frac{\mathrm{d}y}{\mathrm{d}t} + a_0 y = x(t) \tag{1}$$

Taking Laplace transforms for $f(0) = 0$ (where the system is initially in the steady state) gives

$$a_2 s^2 Y(s) + a_1 s Y(s) + a_0 Y(s) = X(s) \tag{2}$$

Rearrangement into the transfer function form gives

$$G(s) = \frac{Y(s)}{X(s)} = \frac{1}{a_0} \frac{1}{(a_2/a_0)s^2 + (a_1/a_0)s + 1} \tag{3}$$

For this expression, three parameters of practical implication can be defined. They are

- $K = 1/a_0$ transduction constant, the DC gain or conversion constant.
- $\omega_n = (a_0/a_2)^{1/2}$ the angular natural frequency (with zero damping present); it has time dimensions.
- $\zeta = a_1/2(a_0 a_2)^{1/2}$ damping ratio, a dimensionless number relating, in practice, the magnitude of actual damping to that at $\zeta = 1$, called *critical damping*.

These parameters occur as the result of interpretation of the solution of the above equations that show that the system is oscillatory in nature and that the level of damping present, due to losses, controls the degree of oscillation. Being properties of the stage and not of the stage coupled to any

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.

© 2005 John Wiley & Sons, Ltd.

ISBN: 0-470-02143-8.

particular forcing function, they can be used as descriptors of a second-order block.

The period of the natural angular frequency of the system can be expressed instead of the angular frequency as its 'reciprocal' and can be given the symbol τ. This is done by some authors and is adopted here. It must be made clear that τ is not used in the same context as it was for the first-order system nor is it exactly the period of the oscillating waveform (it is only for a zero damping situation).

Using these alternative symbols enables the transfer function to be expressed either as

$$G(s) = \frac{K}{(1/\omega_n^2)s^2 + (2\varsigma s/\omega_n) + 1} \tag{4}$$

or as

$$G(s) = \frac{K}{\tau^2 s^2 + 2\varsigma\tau s + 1} \tag{5}$$

both forms of which are commonly used.

As the denominator of the transfer function is a quadratic function, it can be factored into a form that shows that the system could be formed by using two suitable first-order systems placed in cascade.

Note that the natural frequency and damping ratio are interdependent, each also being functions of the transducer constant. Thus, gain changes and damping-ratio changes can alter the natural frequency.

Changes of K matched by appropriate changes to the damping ratio can retain a given natural frequency.

The implication of these constants becomes clear when the step response of such a system is studied.

2 RESPONSE OF SECOND-ORDER SYSTEM TO STEP FUNCTION INPUT

Forming the output expression in Laplace form from the Laplace transforms of the transfer function given above and the unit step function yields

$$Y(s) = \frac{1}{a_0}\frac{1}{s}\frac{1}{(a_2/a_0)s^2 + (a_1/a_0)s + 1} \tag{6}$$

This reveals that the response will be one involving roots of a quadratic. As the actual working is not needed to understand the output response and because it is fully presented elsewhere, such as in Coughanowr and Koppel (1965), the method of solution only needs to be considered.

Forming the partial fractions and inverting the Laplace expressions gives the time-domain solution. The solutions can only be obtained mathematically as three separate regions, these depending upon the nature of the roots of the quadratic. The damping ratio ζ is the dominant factor here. The roots are complex when $\zeta < 1$, real and equal when $\zeta = 1$, and real when $\zeta > 1$. These three regions are termed *underdamped* or *oscillatory, critically damped*, and *overdamped* or *nonoscillatory* for the three regions respectively.

It must be clear that the response range itself passes through a gradual gradation from the under to the over-damped response. It is a need of the mathematical method of solution, not a property of the real physical system, that dictates that there are three different regions. The performance characteristic is a smoothly varying effect, not having sudden changes of behavior as the equations might suggest happens.

A family of response curves is presented, in a normalized manner, as in Figure 1; these show the response as the damping ratio is altered from zero to around 5.0. Note that at large damping ratios, the response curve takes on a similar form to that of a first-order system, the two being slightly different in shape at the origin.

It can be seen that second-order system response can overshoot the final value, and if inadequately damped, it can continue to oscillate after a step is applied. It can be seen that the second-order system, with zero damping, acts as a sinusoidal signal generator if the losses are truly zero or are actually supplied to the system.

A step function is the most severe form of input-level change that can be applied. In practical systems, this rate of rise may be unrealistic, and in such cases, a step input would produce a response that is considerably more oscillatory than what reality might yield. The transient is sometimes called *ringing*. A slower rate of rise, as would be simulated by a ramp that reached a terminated level at given amplitude, would often be more reasonable to apply. This

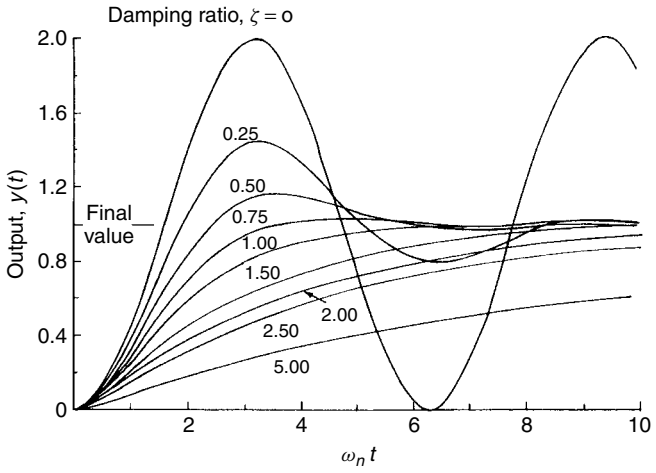


Figure 1. Normalized response of a second-order system to a unit step function.

situation is worked through in Doebelin (2003) to show that with a terminated ramp forcing function, the underdamped ($\zeta < 1$) response can be made to be virtually error free if the natural frequency of the system is large compared with the reciprocal rise-time duration of the ramp applied, this also applies for virtually zero damping. Doebelin makes the point that this means that stages having little damping may be satisfactory if they have high natural frequencies and are not excited in use by rapid rate of rise steps. The example cited in such a case is the piezoelectric accelerometer, when it is used to measure responses of relatively low-pass mechanical functions.

The general relevancy of the step response for many practical systems has given rise to several descriptive terms that relate to the underdamped responses given in Figure 1.

Overshoot is defined as the, less than unity, ratio of the magnitude to which the first overshoot rises over the final-value line to the magnitude of the final value. Numerically, this is given by

$$\text{overshoot} = \exp\left(\frac{-\pi\varsigma}{\sqrt{1-\varsigma^2}}\right) \tag{7}$$

and is, thus, decided only by the damping ratio. Note that the half magnitude of the overshoot cannot exceed the final value.

Decay ratio is defined as the, less than unity, ratio of successive amplitudes of decay cycles. Mathematically, it happens to be the square of the overshoot function given above.

These two expressions enable the damping for a given system to be calculated from a given record or the overshoot to be estimated from the tail end of a record, as occurs when phase delay loses the first part of a record in an oscilloscope.

Their behavior with varying damping ratio is plotted in Coughanowr and Koppel (1965).

Rise-time does not follow the same simple relationship with the defined τ as it did in the first-order case. It is quoted here as the time for the signal to rise from zero to the first crossing of the final value.

Settling time, defined in **Article 61, First-order System Dynamics, Volume 1**, applies with the same meaning to these waveforms. It is also called *response time*.

The period of oscillation, T, the time duration from one peak to the next of the cyclic transient, is not given directly by τ when damping is present, but it is found from

$$T = \frac{2\pi\tau}{\sqrt{1-\varsigma^2}} \tag{8}$$

the term τ being the period of the undamped second-order system. As suggested earlier, the natural frequency of the system is a function of the damping ratio. As damping increases, natural frequency falls according to

$$\frac{\omega}{\omega_n} = \sqrt{1-\varsigma^2} \tag{9}$$

As with first-order systems, although these factors appear to define the system as a stand-alone unit, they are subject to modification by the terminations used. For example, the damping ratio that would be used when studying the response of a high-speed plotter would depend upon the damping added by the plotter to the system of interest.

Impulse response is not covered here, the reader should refer to Coughanowr and Koppel (1965) and Doebelin (2003) for worked derivations and normalized response curves that match those given in Figure 1. A somewhat similar oscillatory response, described in identical terms, results.

Although the treatment has covered the response in a general manner, many texts include worked examples of specific practical systems of mechanical, electrical, electromechanical, acoustical, and pneumatic nature. In the discipline of physics, it is covered most usually under the titles of wave motion and simple harmonic motion. It is also dealt with in many texts on the mechanical-design aspect. Closed-loop control systems often result in second-order open-loop transfer functions and can sometimes be regarded more simply as a stage with this form of response, instead of needing to be seen as a closed-loop stage.

3 RESPONSE OF SECOND-ORDER SYSTEM TO RAMP FUNCTION INPUT

The response of the second-order system to a ramp forcing function is derived by the same procedure as was used for that of a first-order system. It is worked through in Doebelin (2003) and Atkinson (1972).

The response is similar to that of a first-order system fed with a ramp, the exception being that now the system might oscillate about the final ramp line during the initial transient portion of the response curve. Again, there exists a step function component that delays the output ramp by an amount decided by the damping ratio ζ and natural angular frequency ω_n. The resulting steady state time lag and error are both reduced by increasing the natural frequency and reducing the damping ratio. Doebelin (2003) provides a normalized chart of how deviation error varies with varying damping ratio as the transient solution dies away.

4 RESPONSE OF SECOND-ORDER SYSTEM TO SINUSOIDAL FUNCTION INPUT (FREQUENCY RESPONSE)

The forcing function here is that used to establish the second-order frequency response. The methodology used is that given in **Article 61, First-order System Dynamics, Volume 1**. A solution is needed for the Laplace form of the resultant function:

$$Y(s) = \frac{A\omega}{s^2 + \omega^2} \frac{K}{\tau^2 s^2 + 2\varsigma\tau s + 1} \tag{10}$$

This involves solution of the roots of two quadratics. This solution is worked through in several texts (Beckwith and Buck, 1969; Coughanowr and Koppel, 1965; Shearer, Murphy and Richardson, 1967). The resulting time-domain parameters of phase and amplitude are obtained from the output expression:

$$Y(t) = \frac{AK}{\sqrt{[1 - (\omega/\omega_n)^2]^2 + (2\varsigma\omega/\omega_n)^2}} \sin(\omega t + \phi) \tag{11}$$

where

$$\phi = -\tan^{-1}\left(\frac{2\varsigma}{\omega/\omega_n - \omega_n/\omega}\right) \tag{12}$$

or in the form where $\tau = 1/\omega_n$,

$$Y(t) = \frac{AK}{\sqrt{[1 - (\omega\tau)^2]^2 + (2\varsigma\omega\tau)^2}} \sin(\omega t + \phi) \tag{13}$$

where

$$\phi = -\tan^{-1}\left(\frac{2\varsigma\omega\tau}{1 - (\omega\tau)^2}\right) \tag{14}$$

These expressions show that the output is a sine wave of the same frequency as the input signal but that the magnitude can now be greater, or lesser, as the frequency rises. When $\omega = \omega_n$, the amplitude can rise to a theoretical infinity magnitude, the magnitude being decided by the damping ratio.

The phase component of the equations shows that it lags from zero to a maximum while approaching, but never quite reaching, -180° if the damping ratio approaches zero. It is -90° at the natural frequency, ω_n. The expressions are plotted in normalized form as the two plots of Figure 2. In this regard, second-order systems are quite different to first-order stages. These curves are plotted in a particularly accurate manner in Shearer, Murphy and Richardson (1967).

Because the system can produce output signals larger than what is provided to it, it has the ability to provide signal gain or amplification; the output amplitude variable is, therefore, termed the *magnification ratio*.

Whether magnification is to be deliberately adopted or not depends upon the application. In some detection systems, a second-order sensor is purposefully designed to ring (by virtue of absence of damping) at its natural frequency so that magnification is gained. In other systems, it may be desirable to keep the frequency response flat for as wide a bandwidth as possible.

The degree of resonance can be described in terms of the Q-factor, a quality factor or figure of merit. It can be shown (Shearer, Murphy and Richardson, 1967) to be

$$Q = \frac{1}{2\varsigma\sqrt{1 - \xi^2}} \tag{15}$$

or, for low levels of damping, it approximates to

$$Q \simeq \frac{1}{2\varsigma} \tag{16}$$

Alternatively, it is defined in more direct terms of the 'peakiness' of the resonance curve as the bandwidth at the half-power points (where the power level has dropped to half that at the peak), divided into the resonant frequency. Both methods give the same result; direct measurements on the amplitude–frequency response curve can be used to determine the damping ratio.

A word of caution is needed here. In many second-order systems, resonances need time to develop when fed with a swept frequency forcing function. Response curves swept too rapidly may, therefore, yield resonance peaks below their final value if given sufficient time to build.

5 SELECTIVITY ISSUES FOR SECOND-ORDER SYSTEMS

The factor Q is also a measure of the selectivity of the system. This usage applies when the system is used to deliberately detect signals near the resonant frequency, this being achieved by attenuating all others with respect to that small bandwidth.

Quotation of flat-response bandwidth is usually achieved by specifying the range of frequencies over which the response remains within the maximum and the half-power (3 dB) points. From Figure 2, it can be seen that the widest flat response is obtained with a damping factor of around 0.5. By comparison with the frequency response curve

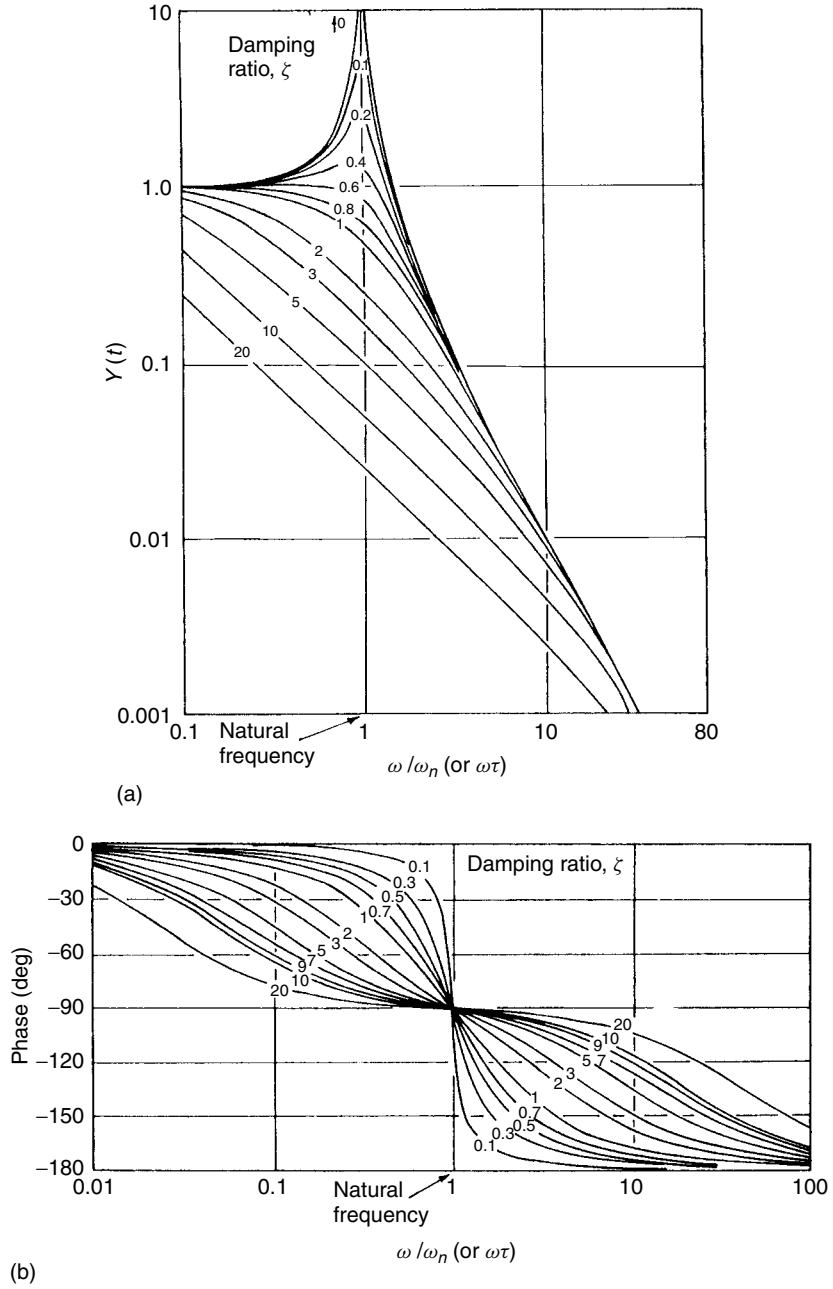


Figure 2. Normalized response of second-order system to sinusoidal excitation: (a) amplitude; (b) phase.

of the first-order system **Article 61, First-order System Dynamics, Volume 1**, it can be seen that a second-order system will provide a flat response of similar bandwidth but the first-order system will not fall off as rapidly. Additionally, the phase shift of each might be important. At the 3-dB point of each, a first-order system introduces approximately -60° shift, while the second-order stage introduces some -110° of lag.

Complex signals fed to a second-order stage can be treated as the sum of sinusoids. If the original character of the signal is to be preserved, then each frequency must be transmitted through the stage without attenuation or phase shift.

6 SPECIFYING SECOND-ORDER SYSTEMS

Second-order systems require the following features to be given when specifying their response:

- system type is of the second order
- magnification ratio at DC
- damping ratio
- resonant frequency.

Other sources of detail on second-order systems are Oppenheim *et al*. (1996), Hsu and Hse (1996), and Nise (2000).

RELATED ARTICLES

Article 37, Relationship Between Systems in the Time and Frequency Domain, Volume 1; **Article 38, Stability Issues, Volume 1**; **Article 59, Introduction to the Dynamic Regime of Measurement Systems, Volume 1**; **Article 60, Zero-order System Dynamics, Volume 1**; **Article 61, First-order System Dynamics, Volume 1**.

REFERENCES

Atkinson, P. (1972) *Feedback Control Theory for Engineers*, Heineman, London.

Beckwith, T.G. and Buck, N.L. (1969) *Mechanical Measurements*, Addison-Wesley, Reading, MA.

Coughanowr, D.R. and Koppel, L.B. (1965) *Process Systems Analysis and Control*, McGraw-Hill, New York.

Doebelin, E.O. (2003) *Measurement Systems: Application and Design*, 5th edn, McGraw-Hill, New York.

Hsu, A. and Hse, W. (1996) *Schaum's Outline of Signals and Systems*, McGraw-Hill.

Nise, N.S. (2000) *Control Systems Engineering*, Wiley, Chichester.

Oppenheim, A.V., Willsky, S.H., Nawab, A. and Nawad, H. (1996) *Signals and Systems*, Prentice Hall, Upper Saddle River, NJ.

Shearer, J.L., Murphy, A.T. and Richardson, H.H. (1967) *Introduction to Systems Dynamics*, Addison-Wesley, Reading, MA.

63: Outline of Systems Thinking

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Definition of Systems Thinking 389
2 Reductionism Versus Postmodernist Thinking 392
3 Soft Systems Thinking 392
4 Hierarchy of Systems 393
5 Introduction to the 5-layer Systems Engineering Model 395
6 Life Cycle Concept 395
7 Emergent Properties 396
8 Systems of Systems 396
9 Application of Systems Thinking to Measurement Systems Design 397
Related Articles 398
References 398

1 DEFINITION OF SYSTEMS THINKING

Too often designs for a system are generated from the 'bottom up'; the core need is hastily matched against an apparently suitable technical solution that appears to fit the need. As the detailed design progresses, it becomes clear that other factors should have been allowed for and that it is by then getting expensive and time consuming to change things.

On the other hand, the more aware and experienced designer will carry out the design from the 'top down' finding appropriate solutions much later when the full requirements have been established. This article and **Article 64, Executing A Measuring System Design, Volume 2** provide the background information to become a 'top down' designer, who rarely needs to make drastic redesign changes.

Systems thinking is a mental attitude to problem solving that intuitively assists one allow for issues of the 'whole' – this is the holistic mode of consideration.

Its implementation is done by application of Systems Engineering (SE) principles that, in turn, leads to the detailed engineering design that is well developed in academic curriculum; systems thinking and systems engineering are seldom developed in an overt didactic manner. The material given here is extracted from Sydenham (2004) where the issues of modern engineering design are covered in detail.

A simple definition of Systems Engineering starts development of this thinking. QinetiQ, UK, staff has developed this conveniently short description of SE:

> 'A set of activities which control the overall design, implementation and integration of a complex set of interacting components or systems to meet the needs of all users.'

This clearly recognizes that engineering design tasks have to include numerous interacting issues (often not properly allowed for in engineering design) in order to obtain a sound solution to the customer's needs.

SE makes use of a set of principles and processes that efficiently apply resources to optimize a development project's progress toward a sound solution of a customer's product or service need. The definition of SE also highlights the fact that it deals with more than the physical energy/mass relationships that are covered by detailed engineering work.

Application of Systems Engineering is not a case of applying a set of rules, but more a way of thinking that extends conventional engineering design practice.

Where the memory requirement needed to track the many issues that arise exceeds one person's 'brain-full'

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

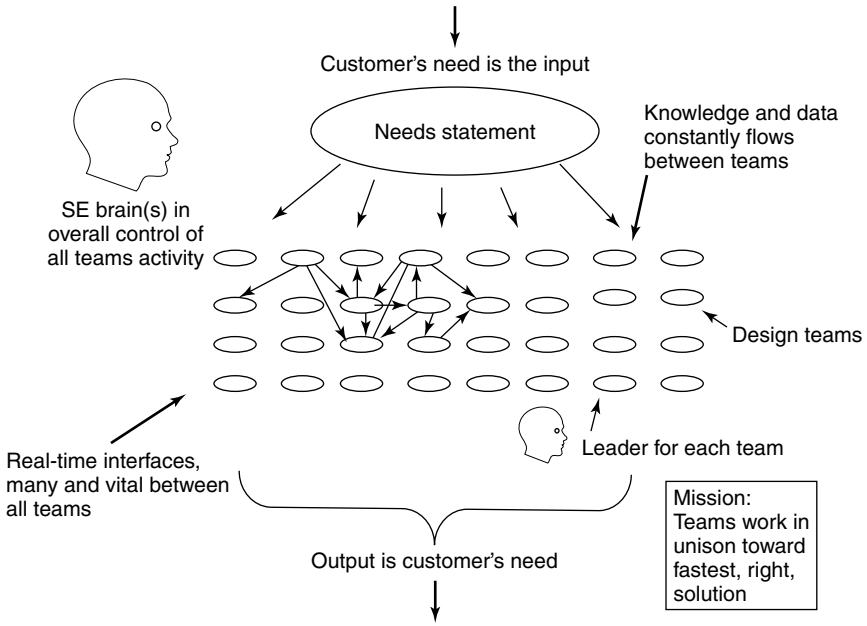


Figure 1. Teaming in an engineering development.

it becomes necessary to employ commonly understood and agreed recorded processes and activities. This helps ensure each person designs the 'right' thing as their part of the whole.

Application of systems engineering is about deciding, for the technical aspects of a project, what should be done by whom, and by what time?

This kind of management task differs from that of office, corporate, or project management; however, the distinctions are not always black and white. Figure 1 provides a simple model of the human teaming aspect of the engineering of a systems task; it represents the overall situation taking place in a development.

All teams must be efficient in their duty in order that the whole set of teams delivers the 'best practice' design needed. Note that numerous interactions will take place between the teams as the project outcomes pass from 'customer need in' to 'satisfactory system out'. An individual must work as a team player; a design team leader is needed to help the process along.

Applying SE to the measuring system design task is a matter of being able to recognize what kind of activity is appropriate to be done at any particular time. This ability is developed by reading, taking courses, working with experienced practitioners, and using every opportunity to bring fresh and better solutions to design situations.

Underpinning the professionalism of SE is an appreciation of Systems Thinking. An appreciation of how engineers and physical scientists go about their thinking is needed at this point.

In the seventeenth century, Descartes suggested that problems are handled by successively breaking down a problem until a level is reached where sufficient understanding exists. This paradigm for problem solving is known as 'reductionism'. It is the basic thinking methodology of science and engineering and has been applied with great effectiveness.

Taking this reduction idea a step further suggests that a system rebuilt from a set of subsystem solutions must be a sound overall solution. This is, however, not necessarily so. Small deviations in subsystem solutions can, all too easily, propagate upward leading to major errors in the final performance.

Difficulty in not meeting requirements is also due to the fact that the traditional limited engineering detail viewpoint often cannot cope with the complexity of real systems. Reductionism needs to be supplemented with other kinds of thinking.

In the reductionism design approach, a closed model of the design situation has to be realized to complete and close the design boundaries. Care is needed in setting up these boundaries – *see* **Article 68, Setting the System Boundaries, Volume 2** and **Article 64, Executing A Measuring System Design, Volume 2**.

Many engineering design situations contain issues that do not lend themselves to reductionism thinking. The ability to recognize the nature and scope of these limiting parameters needs skill in design team operations. Seeking to force reductionist ideas to fit where they are not appropriate can lead to project failure.

Systems Thinking includes attention to

- human activity systems, not just the inanimate physical objects that make up the whole;
- operational readiness and suitability, that is, will it do the job when called upon and will it continue to perform its task for as long as expected?
- systems of all levels and types.

Key statements have been published on Systems Thinking by engineers involved in major projects; Hitchins (1992) is worth visiting for its mind-opening views.

Key tenets of Systems Thinking are as follows:

- It is concerned with the concept of 'whole' and its properties.
- It needs systemic thinking (i.e. including all of the issues) as well as systematic thinking (i.e. being methodical in tackling the problems).
- Systems consist of hierarchies that relate to each other through numerous interfaces, each having their own kind of requirement.
- All parts of the whole are interconnected (*interface* is an alternative term) to a varying degree, some being very dominant and thus having greater influence on the behavior of the whole.

The various methods and parts of Systems Thinking are illustrated in Figure 2.

A model of the layers of system openness starts with the outer total shell that includes everything thought to be of relevance to the problem that the diagram represents.

Inside this layer is placed the study of how the different systems viewpoints are expressed. This has two thinking aspects – philosophical systemic thinking that is often hard to apply, and the various pragmatic working areas that the various kinds of thinking need to advance their problem solving.

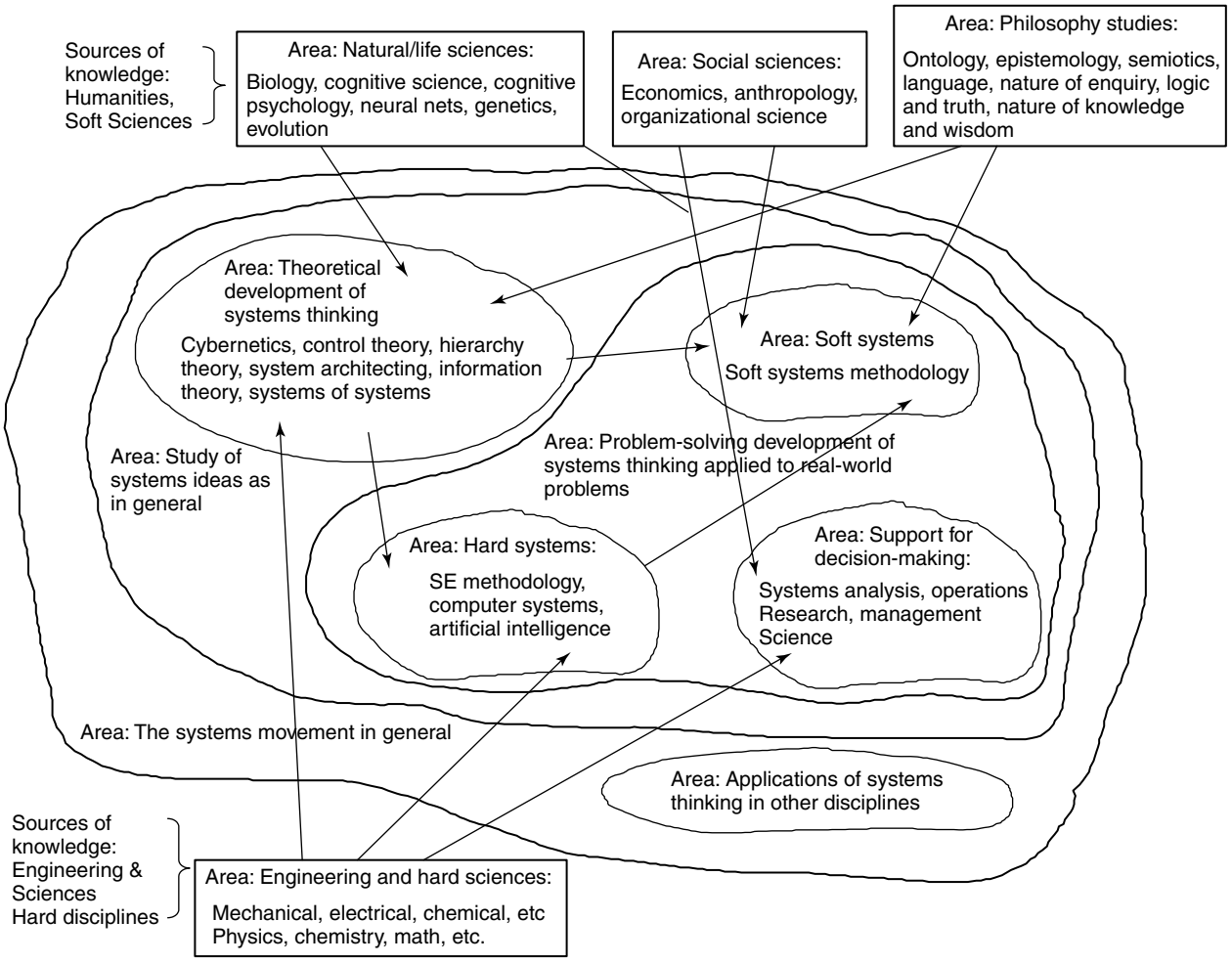


Figure 2. Representation of the relationships of the key disciplines.

Engineering can be seen there to make use of all of the domains shown in Figure 2 with the exception that the application of the soft kind of systems is not well developed – see Section 3.

2 REDUCTIONISM VERSUS POSTMODERNIST THINKING

Reductionism has served engineering and science well, but it cannot provide all of the solutions to systems design – see Section 4. Design engineers overly seek to use the reductionist approach because it is the main method in their design toolbox. They are not as well equipped to find sound solutions for systems involving human activity.

In the 1970s, an engineer in charge of major engineering projects in the United Kingdom was dissatisfied with the fact that the reductionist engineering approaches were not working well enough. He developed a methodology for these situations that is now known as the Soft Systems Methodology (SSM), Checkland (1981), Checkland and Howell (1998). SSM recognizes that not all disciplines make use of reductionist thinking.

In the humanities, it has long been recognized that the human machine has two aspects. The physically existing 'body' part can be explained through the laws of physics. The cognitive 'mind' part, however, cannot – yet they form the same whole. There is currently no meeting of the minds on this dual body problem; it continues to provoke philosophical debate.

This has led to the humanities disciplines solving problems differently to how they are tackled in the hard science disciplines. Their phenomenological approach – often called the *postmodernism approach* – is not as precise as reductionism but it can make progress in soft science situations.

3 SOFT SYSTEMS THINKING

Figure 3 shows the flow of activities for finding and implementing a solution to a system need in a 'soft' situation.

The SSM process begins with the problem being identified as unclear and lacking an obvious reductionist solution. The need is then expressed in writing as that helps clarify the situation.

Purposeful activities are identified and conceptual models generated. These are compared with the starting point of the real situation. Extreme differences are rectified in the model and the implementation adjusted to make it acceptable to the people involved in the change.

This process is repeated until the best available plan results. Considerable iteration may be needed to reach an acceptable situation.

Success cannot be assured with the soft system methodology, but it is one of the few ways known to seek design solutions for problematic system designs. Here the solution solver 'gets inside the problem' making a considered intervention to see what happens.

In sharp contrast, the hard engineering solution path develops a separated model of the whole through metaphorically dissecting that which identifies its subsystem. This set of component subsystems, and their interrelationships, are then simulated using a computer to investigate the

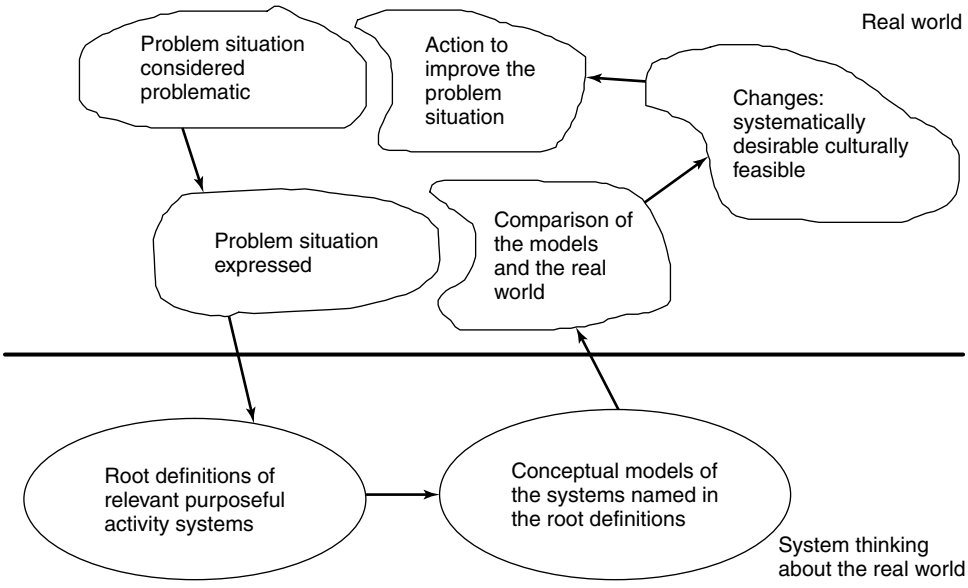


Figure 3. Flow of activities in finding and implementing a solution in a soft system situation.

sensitivities of the various critical issues. Optimization methods are applied to facilitate beneficial changes to the model.

When the right model and its parameters have been realized, the real system is rebuilt to form a new system; altered parts are put back into place to see if the whole works as intended.

Humans cannot be adequately understood as technical machines alone. People within systems can be slow and reluctant to respond to the direct process of change. Changing organizational cultures can take a long time to achieve. It has been summed up as trying to 'herd cats'.

It is often, however, the engineers trained in reductionism who are expected to develop solutions for systems with which they are not well versed, and for which there is often low chance of clear-cut success, as measured in hard science ways.

The design team leader needs to recognize the kind of system class in which the design team is involved and set up appropriate team membership.

When designing measuring systems, any human element present needs careful treatment to ensure it will operate properly. This arises when generating the needs statement in order to establish what should be measured, and for what purpose. This stage – *see* **Article 64, Executing A Measuring System Design, Volume 2** – usually involves thinking in the soft regime as customers develop their ideas. The operational regime of measuring systems – the knobs, dials, panel layouts, handling properties, colors, and emitted sounds – all involve soft thinking as the human factors – *see* **Article 94, Human Factors Engineering, Volume 2**; **Article 95, Human–Machine Interface, Volume 2**; **Article 96, The Domains of Human Factors Integration, Volume 2**; and **Article 97, Design Methodology, Volume 2**.

4 HIERARCHY OF SYSTEMS

Representing the whole of all systems activities and relationships would be a massive task; there are too many issues and variations to cover. Instead, a collection of thought provoking models is used that develop insight into aspects of the whole.

Three different models are now given, each revealing different characteristics of the same generalized whole.

The first relates the groups of people involved, the sciences, and their thinking, Figure 4.

Three key kinds of interrelated activities are shown – the natural world, the sociopolitical system, and engineers and scientists at work. The needs of all three must be met for a system design to be successful. Engineering, until recently, largely neglected the other two; today sociopolitical and natural world aspects must be taken into account. Installing a measuring/detection system can sometimes seriously impact people's perceptions producing disastrous outcomes, as did the 'Star Wars' missile protection system program.

Each of the three regimes is represented by a triangle. At the base of each triangle sits the scientific, formal, quantitative thinking workers. Moving up each triangle, the thinking style used changes from essentially quantitative to almost entirely qualitative, for at that level it is taking in people's feelings and emotions.

Engineering and science disciplines operate best in the lower areas.

The middle ground is where the use of Systems Engineering finds effective application. At the top, all manner of

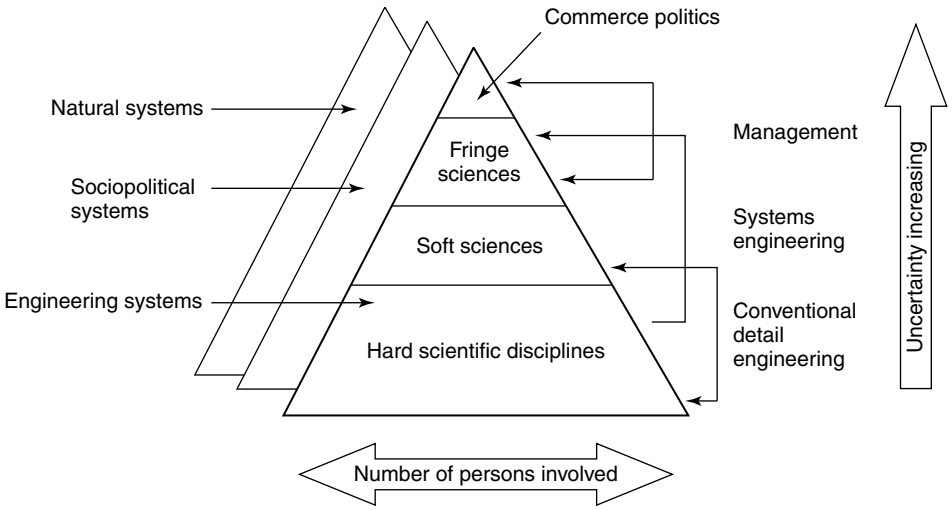


Figure 4. 'People and Science' model of a systems activity.

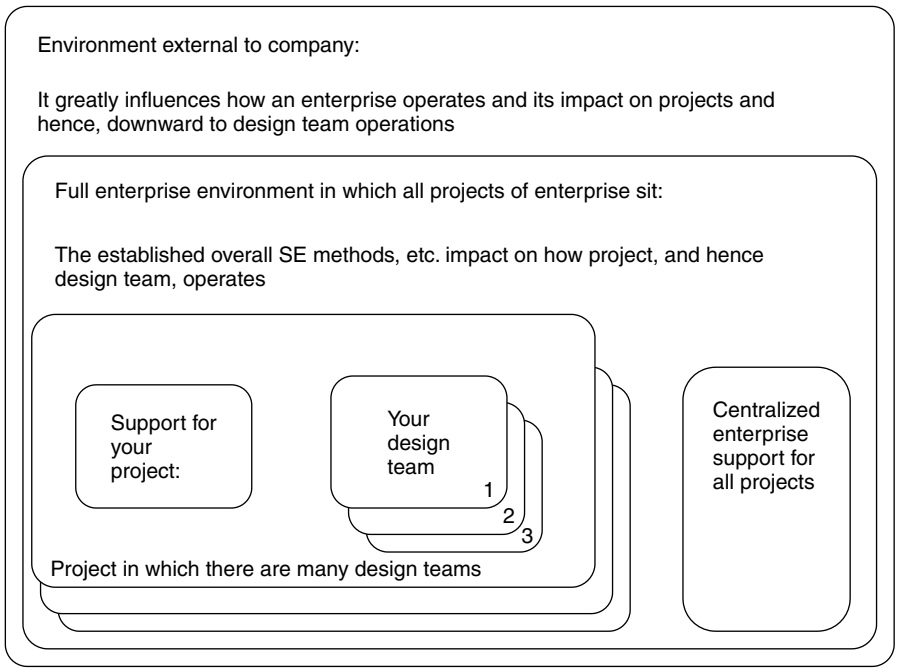


Figure 5. Environment layers in which a design team works.

often inexplicable decision-making takes place, not because of lack of skills but for lack of any formalized way to do it better.

The width of the triangle at any given level crudely indicates the number of people involved. It is interesting to note that as little as one person at the top can decide how the many people below use their resources and skills; large groups of designers are involved in taking the ideas of a few to fruition. A designer generally has little influence over the top-level needs and has to work within given requirements.

The second model, Figure 5, shows how the design team works within a multilayered set of quite different environments. For overall success, a project must make allowances for the nature of the limitations and controlling factors that exist for the type of enterprise in which the design team works. These issues vary greatly. For example, a private organization does not have to disclose as much information about its processes to the public as does a government institution. To get on top of many problematic issues, it pays to appreciate the higher layer affairs that are impacting on a design's situation.

The third model given in Figure 6, after Boulding (1956), assists appreciation of the classes of types of system design that can arise. It helps a designer identify what kind of difficulties the design team might expect to encounter.

This model is based on mapping the various kinds of systems that exist onto a modeling space represented by two variables – the degree of disagreement on systems

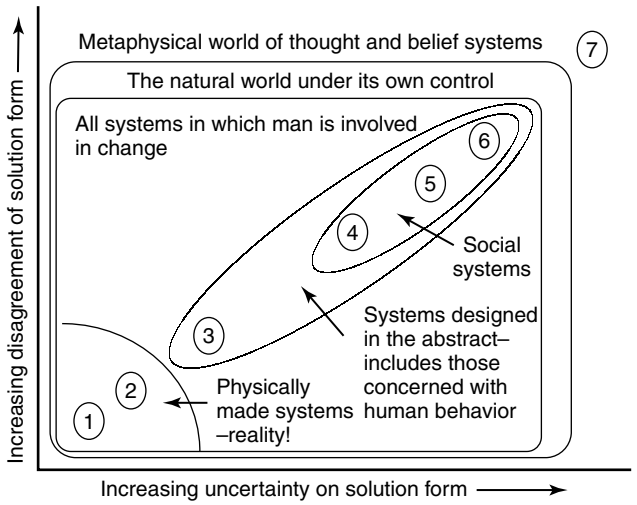


Figure 6. Diagrammatic representation of a hierarchical classification of systems.

issues that exist versus the degree of uncertainty of their characteristics.

The types of systems shown in the diagram are

1. straightforward technical design tasks that inherit considerable know-how and have low risk in execution if done by experienced experts;
   (Examples: simple road bridge; electronic amplifier board; weighing platform)
2. technical tasks with a modest degree of design change, thus including a clear degree of risk;

(Examples: original network for 3G mobile phones; major automobile model change with advancing functionality such as moving to all-wheel drive from dual wheel drive, measuring instrument using novel principle of detection)

3. engineering systems involving considerable human control and intervention in their operation, but not so much in the overall organization;
   (Examples: production line manufacturing systems instrumentation; transport systems)
4. systems where their major subsystems components are associated heavily with human organization; here engineering risk issues are low compared with the risks of understanding the human behavioral aspects;
   (Examples: building detection and evacuation systems; flight control systems; educational support systems)
5. systems where human attitude is dominant and largely unpredictable;
   (Examples: change management taking place in a factory; speed control on roads; engendering professionalism. Improving the quality of the design of measuring systems)
6. systems that are as complex as man thinks he can build and so often attempts to make;
   (Examples: fighting wars, peace making, societal policy systems. Government's economic measurement models.)
7. systems that can only yet be represented by abstraction in the thinking world of science fiction and theology. (Examples: Utopian worlds; Godlike abilities of design.)

As the risks rise and the systems nature becomes more problematic, it becomes increasingly impossible to be certain about numerous critical systems issues. Those involved are increasingly unable to agree on what kind of solution to use. Without clear definition of the critical issues it is problematic for a sound engineering solution to be achieved – *see* **Article 3, Measures and Metrics; Their Application, Volume 1**.

Most commercial projects sit in the lower two classes of 1 and 2 because they tend to exploit proven technologies and because they need to work in relatively low-risk areas.

The measuring system engineering detail design team is usually working, by necessity of delivering a reliable and low-risk outcome, in the high-certainty and low-disagreement area with respect to their design solutions.

However, they can sometimes get involved in the execution of tasks in many of the classes shown. As the magnitude of the position number increases, the detail engineering design component becomes of lesser importance compared with execution of the whole as it, in itself, is unable to provide solutions to the problems.

It is important to be able to recognize the class of system in which a design task is working; the surrounding climate of thinking can make a large impact on progress and on the type of solutions that will be accepted.

5 INTRODUCTION TO THE 5-LAYER SYSTEMS ENGINEERING MODEL

The various holistic levels of a project task have been identified by Hitchins and are reproduced here from his Web site www.hitchins.net, where it is free to download.

Overall human system endeavors fit into levels as follows:-

Level 5: Socioeconomic Engineering, what constitutes regulation and government control at national and international operation.

Level 4: Industrial Systems Engineering, or engineering of complete supply chains/circles. Many industries make a socioeconomic system. A global wealth creation philosophy.

Level 3: Business Systems Engineering – many businesses make an industry. At this level, systems engineering seeks to optimize performance somewhat independent of other businesses.

Level 2: Project or System Level – many projects make a business. Western engineer managers operate at this level, principally making complex artifacts.

Level 1: Product Level – many products make a system. The tangible artifact level. Many engineers and their institutions consider this to be the only 'real' systems engineering

The design of measuring systems is most likely to fit into Levels 1 and 2. Detail of the factors of each of the layers, provided on the Hitchins Web site, provide appreciation of the kind of problems, and thus the boundary limits that a particular development needs to consider.

6 LIFE CYCLE CONCEPT

The popular way to provide a representational foundation for systems development uses the system life cycle model. Development and use of all product or service systems follow the same generic sequence of life cycle activity stages.

While various specific expressions of this life cycle exist they follow the generic one illustrated in Figure 7. The phases (stages) are the following:

- conceptual formation of ways in which the task might be solved;

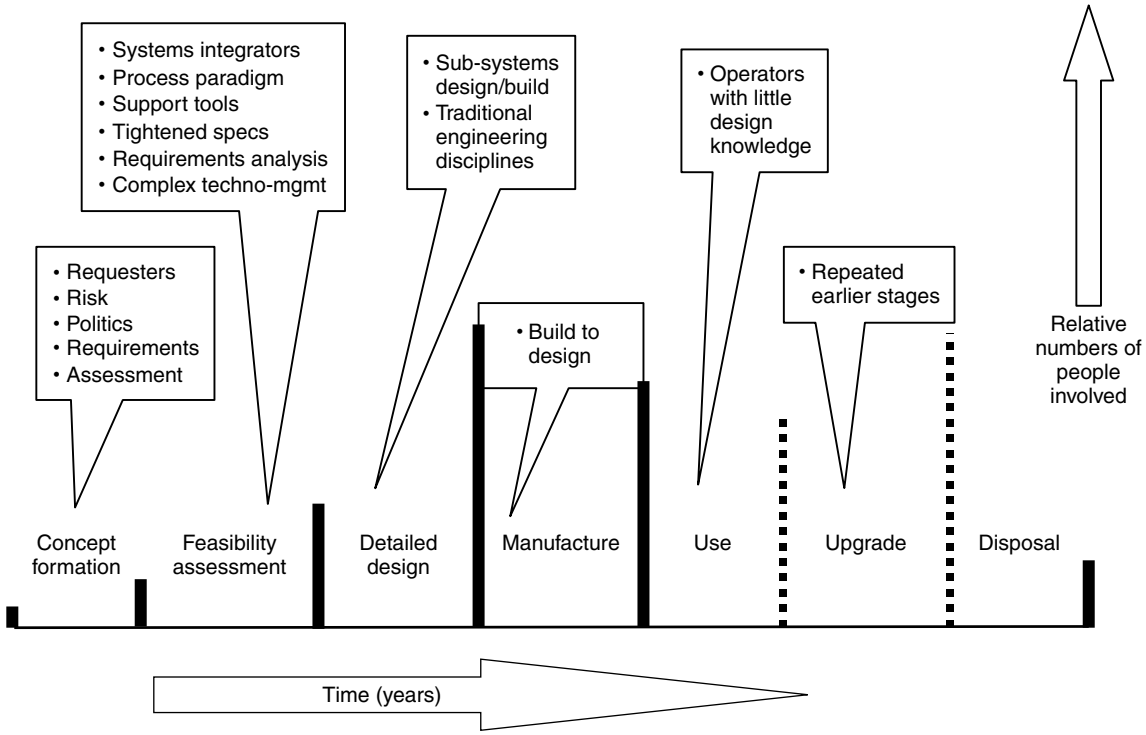


Figure 7. Life cycle stages of a product or service development.

- feasibility assessment of selected, apparently sound, design ideas;
- detailed design of the component parts of the chosen design;
- manufacture of the real system based on detailed engineering design, using physical or digital model prototypes to prove the design;
- use in the application for which the system was designed;
- upgraded for further use with improved or different requirements;
- disposal after use, to follow regulations, and so on.

Good design and effective application of systems thinking are key issues within the tasks of all of the stages. Each is not a stand-alone stage and all need some level of consideration from commencement of a project. To appreciate the differences consult **Article 66, Phases of System Life Cycle, Volume 2** and **Article 65, Life Cycle Concept, Volume 2**.

7 EMERGENT PROPERTIES

Parts of the whole will have their own important 'emergent properties'. These are key performance parameters that may, or often may not, have been expected as the result of implementation.

They can exert a great influence on the other systems with which they interface. Unexpected, nonbeneficial, emergent properties can become very apparent once the system is nearing, or after being put into service.

For example, it might have been decided to use a microminiature wireless telemetry system to communicate temperature data from inside the flying suit of a pilot, only to find after commissioning that it causes the flight navigation system to be inaccurate. Today, this is an obvious design factor to expect, but that was not always the case or we would not have to turn off mobile phones in hospitals and in aircraft.

8 SYSTEMS OF SYSTEMS

As man has learned how better to organize and design technical systems they have grown in size to an enormous extent. Many of today's major man-made systems have evolved by progressively combining systems to form very large conglomerates.

For example, the first railway system in Britain created emergent requirements for engine water and fuel supply depots, spare parts provision, hotels en route, signaling, track rules, special tools for making parts, training schemes, and so on. What might have been overlooked as it was developed was the effect it would have on the existing water

canal transport system that it replaced, and the impact it would have on the nature of the industrial revolution taking place at that time.

In recent years, such complex and extensive systems were seen to be too large to be considered as suited to the usual methods of management and design. The name Systems of Systems (SoS) has been coined for such systems.

A main driver for SoS developments has been in defense systems. These evolved as the following:

- First there existed the personal weapon system.
- Then came team use of weapons combining the various forms of firepower with behind the lines support logistics and intelligence inputs.
- This was followed by many kinds of platforms combined with the necessary command and control needed in a campaign structure.

The sophistication and number of cooperating systems has continuously increased in defense, civilian commercial, and government systems, and in the search for solutions to societal and humanitarian problems.

It became obvious that the former paradigm of first building general utility platforms (the ship, the airframe, the armored vehicle, etc.) on which are then mounted control and command systems, weapons, and other systems, all as separate entities, was inadequate. It needed the SoS approach. Similar thinking is essential in such systems as civil aircraft control systems that now span countries, and in integrated power grid operations.

So, what are the differences between Systems Thinking and SoS thinking? This is a matter of degree. SoS is an extension of general holistic considerations and has the following characteristics:

- High complexity comprising relatively independent systems that can each be regarded as a sophisticated system in its own right
- Continuously evolving as the emergent properties of each system interact
- No obvious start or end-point goals for their existence
- Parts are often geographically distributed
- Viewed as a set of interacting, separate, systems
- Component systems retain much of their independence, pursue their own goals, and have independent management
- Systems Engineering activity is dispersed and loosely controlled
- Understanding the behavior of constituent systems needs transdisciplinary (each is learning from the other) approaches, not just multidisciplinary approaches (each does its own thing usually with an insufficient number of disciplines)
- Measuring systems are unlikely to form into SoS units in their own right but their common nature and universal methodologies will be used across those systems comprising an SoS assembly.

9 APPLICATION OF SYSTEMS THINKING TO MEASUREMENT SYSTEMS DESIGN

Applying the systems thinking culture is a case of 'being a specialist at being a generalist'.

When a design team is part of a large systems provider house, a person will usually be appointed who is responsible for organizing the systems support for all teams. In such situations, teams usually will be provided with the following:

- a company specific SE Process Manual;
- a computer-based support tool system and support tutors;
- special development facilities as are needed according to the project;
- sound archiving and communications for design records and the configuration management system;
- safety control process;
- design controls;
- sources of advice and mentoring for junior staff;
- in-house training and more.

As a guide, design team leaders need to be familiar with the basics of systems thinking and the culture of systems engineering. They need to have on hand copies of foundational books on SE practice such as Blanchard and Fabrycky (1998), Sage and Rouse (1999), Sydenham (2004) and they need to refer to these for concepts and methods to employ as project problems arise. A sound source of SE knowledge is found in the general pages, and also in those of the technical committees, of the International Council on Systems Engineering, INCOSE (2004).

The US DOD Military Handbook on Systems Engineering MIL-STD-499B, USDoD (2004), although being comparatively old as SE standards go, is still an excellent source of ideas for use at the various stages of the life cycle. Other works giving various views of Systems Engineering are Stevens *et al*. (1998), Buede (2000), Westerman (2001), Faulconbridge and Ryan (2003).

An issue that will arise in the design of measuring systems is how much of the SE process should be carried out when executing the systems engineering for a project. Some design tasks are so small as not to need any ordered process. Some, like that for flight-testing an aircraft where

over 100 000 test points might be set up, will need to adopt the complete best practices of SE.

Overhead costs of an SE support process are not always easy to justify because the benefits of its use are abstract in nature, cover long-term issues, and appear not to produce easily measured value-adding components to a project. For this reason, SE activities are often seen as costly luxuries by those who control the project budget.

They are, however, as important to project success as accounting and management activities in that they assist in

- controlling the final performance obtained;
- providing early detection and control of design error.

Who in their right mind would make a development journey without sound assurances that all design work is on track and that the design is moving forward in the right direction?

The design team leader has to use personal judgment in setting the scale of use when applying the techniques or support mechanisms of SE. A single team comprising a dozen or so staff working in a start-up company will probably not be able to devote the time to writing a dedicated SE manual. They might instead simply align with an SE standard. The team leader in that case, however, still needs to apply SE principles as part of routine technical management. Article **Article 64, Executing A Measuring System Design, Volume 2** moves this discussion on to show how a project should progress.

RELATED ARTICLES

Article 65, Life Cycle Concept, Volume 2; **Article 66, Phases of System Life Cycle, Volume 2**; **Article 68, Setting the System Boundaries, Volume 2**; **Article 69, Requirements Allocation, Volume 2**.

REFERENCES

Blanchard, S.B. and Fabrycky, W.J. (1998) *Systems Engineering and Analysis*, 3rd edn, Prentice Hall, Upper Saddle River, NJ.

Boulding, K.E. (1956) General Systems Theory – The Skeleton of Science. *Management Science*, **2**(3).

Buede, D.M. (2000) *The Engineering Design of Systems: Models and Methods*, John Wiley, New York.

Checkland, P. (1981) *Systems Thinking, Systems Practice*, Wiley, Chichester.

Checkland, P. and Howell, S. (1998) *Information, Systems and Information Systems*, Wiley, Chichester.

Faulconbridge, I. and Ryan, M. (2003) *Managing Complex Technical Projects: A Systems Engineering Approach*, Artech House, Norwood, MA.

Hitchins, D.K. (1992) *Putting Systems to Work*, Wiley, Chichester, Free download version is available from www.hitchins.org/prof, April, 2002.

INCOSE. (2004) International Council on Systems Engineering, www.incose.org.

Sage, A.G. and Rouse, W.B. (eds) (1999) *Handbook of Systems Engineering and Management*, (ed. A. Sage), Wiley, New York.

Stevens, R., Brook, P., Jackson, K. and Arnold, S. (1998) *Systems Engineering: Coping with Complexity*, Prentice Hall PTR.

Sydenham, P.H. (2004) *Systems Approach to Engineering Design*, Artech House, Boston, MA.

USDoD MIL documents, (2004) http://astimage.daps.dla.mil/quicksearch/.

Westerman, H.R. (2001) *Systems Engineering Principles and Practice*, Artech House, Norwood, MA.

64: Executing A Measuring System Design

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Understanding the Whole 399
2 Closed-systems Design 400
3 The Open-system Challenge 401
4 Specifying the Need 402
5 Requirements Generation 404
6 Conversion into a System Design 406
7 Partitioning 406
8 Interface Design 407
9 Test and Evaluation 408
10 Applying SE to a Project 409
Related Articles 409
References 410

1 UNDERSTANDING THE WHOLE

Having introduced system thinking in **Article 63, Outline of Systems Thinking, Volume 2**, it will be clear that a sound design of a measuring system begins with a general appreciation of many system issues before a commitment is made to firm design parameters.

Obviously, the main aim of a design exercise is to produce a system that satisfies the need. Getting to that state needs systematic activities that drill down from the top of the situation.

The whole process will pass through several life-cycle stages such as

- formation of some suitable solution concepts
- assessment of two or three likely candidate concepts
- detailed design of the chosen concept that has become more defined by now
- manufacture of the measuring system
- installation and commissioning of the system
- operation by obtaining the desired measurements
- removal from service when useful life has expired.

The first step is to obtain a sound briefing of the need from the customer or their representative. Sounds easy! Customers are not always experienced enough in stating a need in terms that clearly lead to design parameters. Do not expect a swift resolution of this issue. Usually it requires a lot of patience as both the designer and the customer undergo a joint learning process. Development of the requirement is a journey of discovery; more on that later.

It is very important to glean as much as possible about the need and its influencing factors as the starting point to success.

Issues to be explored are

- the overall statement of customer need, in customer's terms;
- purposeful requirement – what capability is to be fulfilled;
- operational requirements – how well, and in what circumstances it is to be operated;
- the extent of the limits of influencing issues in the technical, social, and legal arenas;
- the key critical issues that were introduced in **Article 3, Measures and Metrics; Their Application, Volume 1**;
- personnel that will operate and maintain the system;
- decide when it is needed.

It is necessary to appreciate the role of the various types of players in the development. Figure 1 shows their relationships as the project progresses.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

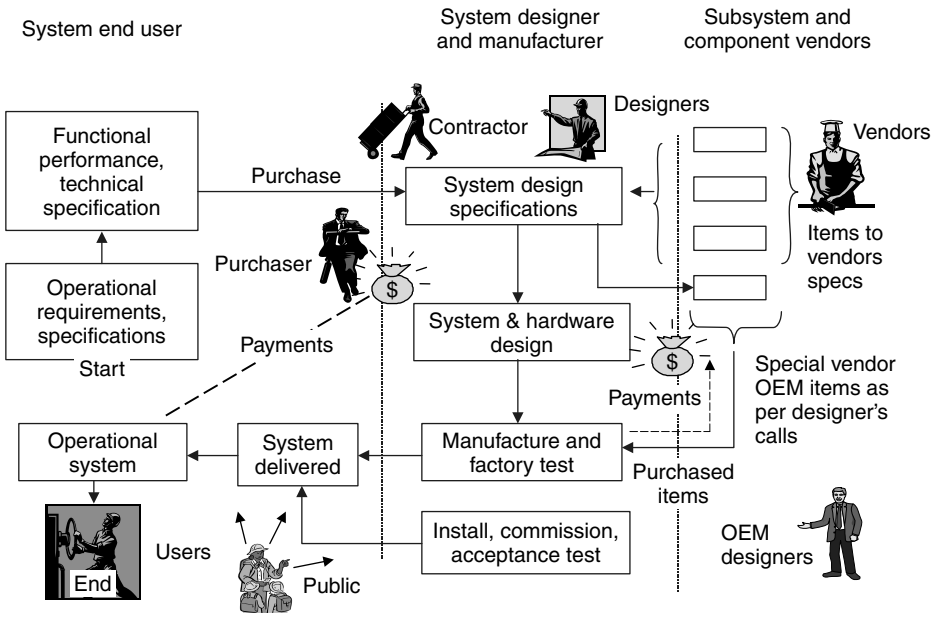


Figure 1. Role of the various players in a system development.

Each player – customers, users, designers, contractors, vendors and the public – has different reasons for participating. Their incentives are different, and the designer needs to appreciate these differences, Sydenham (2004).

2 CLOSED-SYSTEMS DESIGN

Begin with a general assessment of the type of design situation, using the tenets of Systems Thinking given in **Article 63, Outline of Systems Thinking, Volume 2**. Ask the following questions and reflect on the findings.

- Is it going to easily lead to a closed-system design?
- What is the level of system in the layered Boulding model given in **Article 63, Outline of Systems Thinking, Volume 2**?
- Does the degree of human activity suggest a reductionist approach that will need the use of a Soft System Methodology (SSM), as described in **Article 63, Outline of Systems Thinking, Volume 2**?
- What is the impact of the characteristics of each lifecycle stage? For example, is there a serious disposal problem?
- What is the useful lifetime likely to be and will it need upgrading?

After obtaining an appreciation of these issues, start building an overall contextual systems model of the situation. This will help clarify matters and provide material for briefing others who are associated with the project.

This does not need to be that complex to start with – include just enough detail to portray the overall situation. Figure 2 shows such a model for a measuring system being set up for monitoring an oil refinery system.

Just how the system will finally look is still to be developed. For example, the sensors may well be connected on a circular digital bus instead of the star configuration shown. At this stage, do not rush to start the detailed design; the specifications need to be well developed before that commences.

A good way to develop a concept is to develop a Concept of Operations (ConOps) – *see* **Article 67, Principle of Concept of Operations (ConOps), Volume 2**. This is a case of working through the things that need to be done to achieve the right outcomes for the system in use. Figure 3 shows one developed by Ring (2000) for a program to train systems thinking design engineers.

While this particular ConOps is developed for a system in which humans are the subject, the same rules of development apply for purely physical or mixed systems; apply the main rule – what is it that happens to the flowing entity as it migrates from an input to an output?

At some stage, the overall system architecture becomes clear enough to define the boundary limits of the whole. This can depict the extent to which influencing effects will impact the system, for good or bad. Figure 4 shows such a diagram set up for a temperature measuring system operating in a high hazard situation. The influence parameters that will heavily impact the actual sensor system are defined.

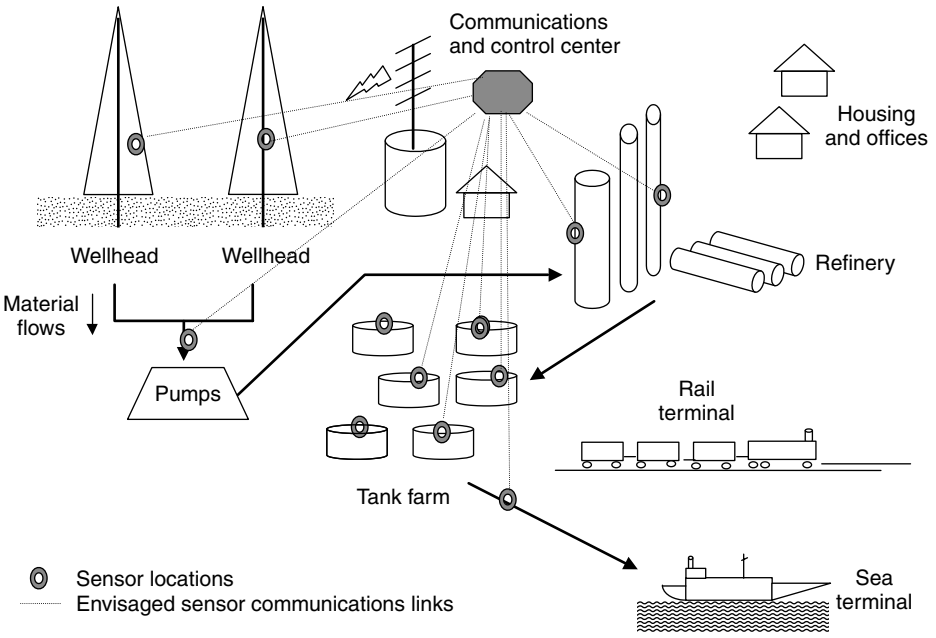


Figure 2. First cut measuring system layout.

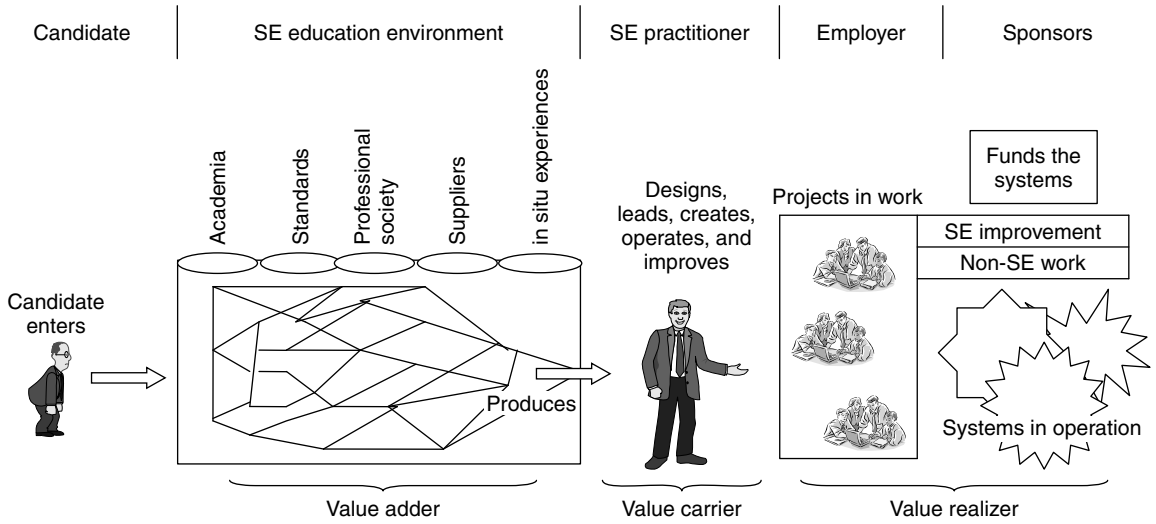


Figure 3. ConOps developed to train engineers.

It is useful to form a matrix of the likely influence parameters as the design proceeds. For each parameter, list the

- parameter name
- effect by which it arises
- likely magnitude, frequency, and phase of the influence signal
- likely effect on the output signal in terms of error
- suggested means for its sufficient reduction.

It will not be possible to fill in all of the details of these various mechanisms until after the requirements have been well established. They, however, need to be understood as the requirements are developed; they settle down as investigation proceeds.

3 THE OPEN-SYSTEM CHALLENGE

If humans are involved, the task can be considerably complex to design. After gaining oversight of the whole system, and its working environment, prepare a list recording where human activity is likely to be needed in the construction, installation, operation, and the maintenance of the measuring system.

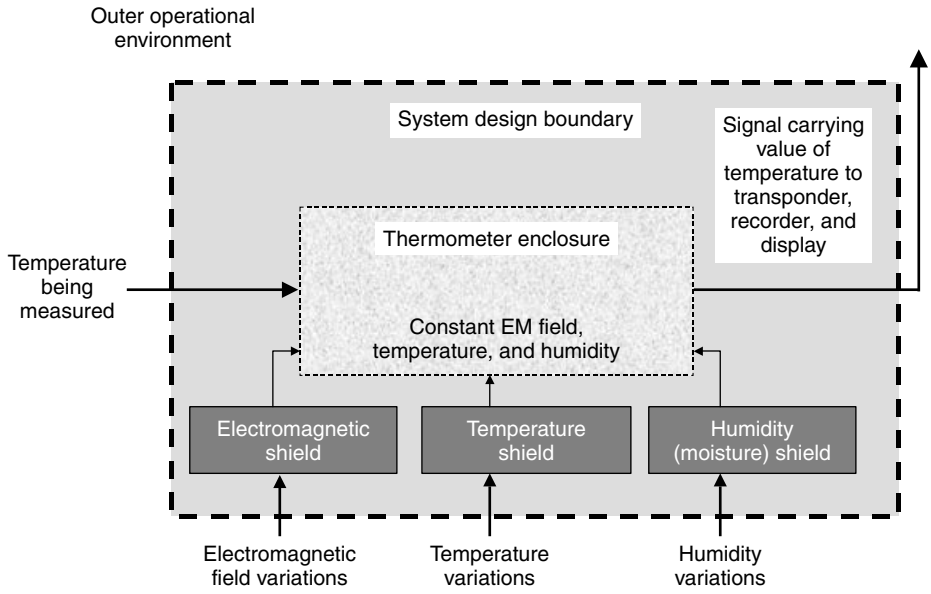


Figure 4. Simple closed-system diagram for a small sensor system.

Each of these human situations will need human factors (HF) consideration – *see* **Article 94, Human Factors Engineering, Volume 2**; **Article 95, Human–Machine Interface, Volume 2**; **Article 96, The Domains of Human Factors Integration, Volume 2**; and **Article 97, Design Methodology, Volume 2**.

Assess whether the situation is suitable for reductionist methods by testing it for the following:

- Is it only concerned with energy or mass flows and not people activities?
- Can a truly closed system be modeled?
- Will *all* elements be able to be modeled with formal models?

If the system boundaries are not clear, then it is of little value to advise the client that the system has no clear-cut solution. The design has to proceed with the fuzzy boundaries that are modeled as best as one can as things progress. Where they are fuzzy, application of the soft systems methodology is suggested – *see* **Article 63, Outline of Systems Thinking, Volume 2**.

Explore how the boundary might be limited to a closed system by employing risk reduction measures, such as reducing human involvement, or by circumventing a difficult area with bypass type of solution.

Areas of design uncertainty must not be ignored as they have a habit of showing up at the wrong time. Also, make sure they are recorded so that they are not overlooked.

To this point, little detailed engineering design has been undertaken, but what to do is becoming clearer.

It will also become clear as to what type of life-cycle development model should be used – *see* **Article 66, Phases of System Life Cycle, Volume 2** for a description of the alternatives. This has to be decided early as it dictates what process is needed to develop the early SE (systems engineering) items and specifications.

4 SPECIFYING THE NEED

If not already familiar with design, this is the time to consider the engineering design process. It has the same elements as the SE life-cycle process but it is expressed as a flowchart of tasks with key feedback paths incorporated – shown in Figure 5 as links back to (1). Detail designers usually view the design process as shown in Figure 5.

There may seem to be too much process management needed, but it is imperative to follow the general thrust of the flowchart or serious design errors can appear late in the process. Design errors are expensive to fix if not picked up early. For example, a resistor in an electronic circuit that does not have a sufficient power rating will be hard to replace once the circuit boards have been manufactured, for the following reasons:

- too little space is then available for the larger packaging needed;
- there may be overheating of adjacent components;
- a board change will be needed that entails removal from service, freight, replacement, and testing cost.

To change the design at the early design stage, the cost will be as little as a dollar, but once in service, it shoots up to hundreds/thousands of dollars if it gets to the recall situation.

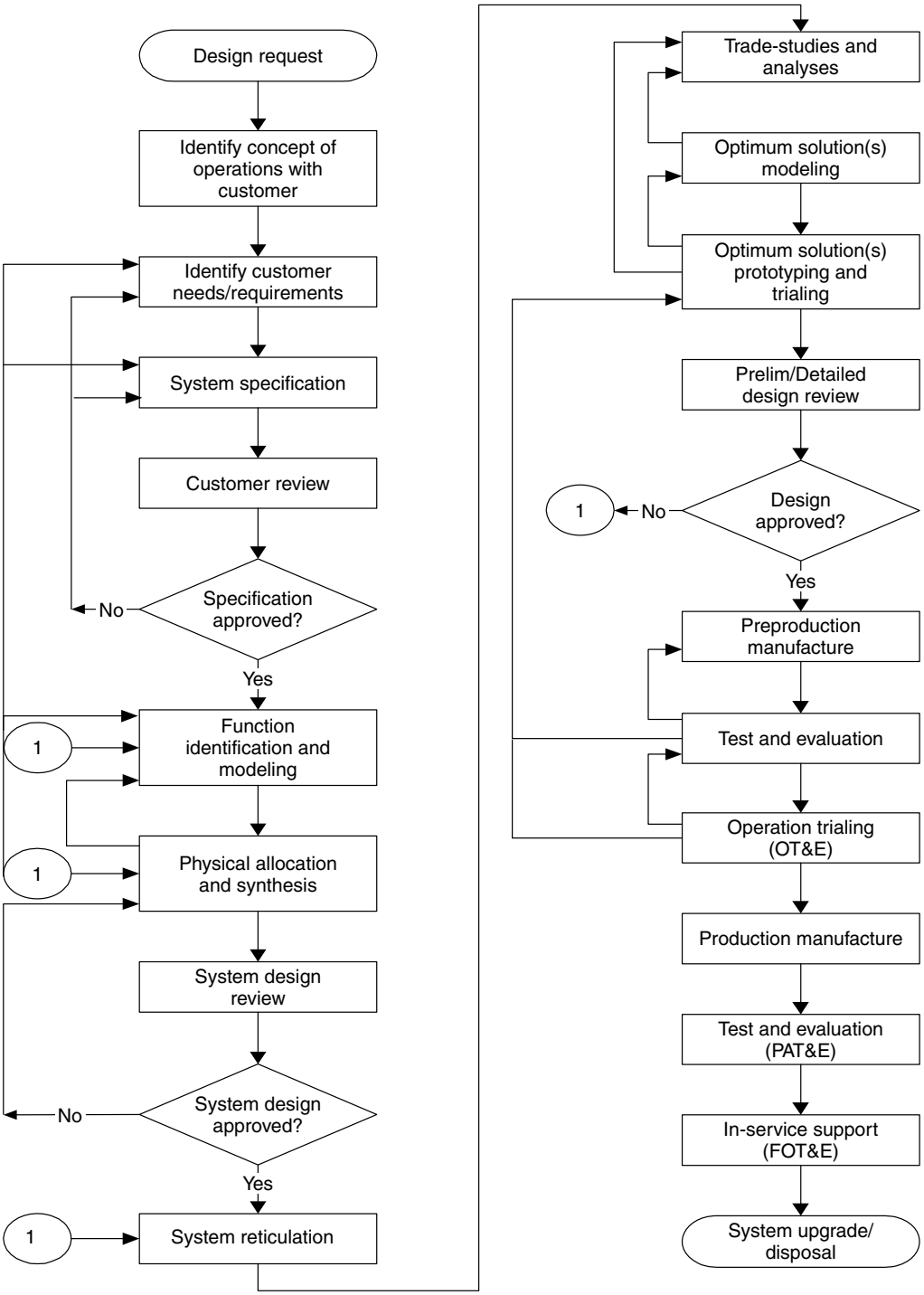


Figure 5. Flowchart of the design process (Courtesy: Tim Welburn).

Getting a design 'right' requires considerable effort in the early phases of the life cycle. As a guide, considerable design resource should be placed into this front-end loading (FEL). Figure 6 indicates the differences in FEL.

The area under the curve is an indication of total cost. It will be seen that while the large FEL at the start seems to have more resources than needed – at least that is too often the perception of senior managers – the total is less overall.

There is little doubt that allocating the right level of FEL expenditure is a wise move; but there are many forces at work against proper allocation.

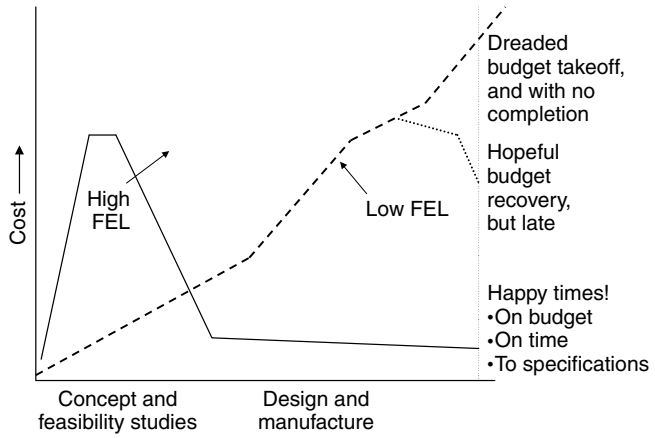


Figure 6. Effects of different FEL in project outcomes.

The first difficulty is that most situations are operating in a fixed financial supply climate. This means, overall, managers will tend to opt for more project starts by cutting the FEL budget on each in order to allocate funds to more new starts. When the project overruns the forecast time and budget expectations, the lack of the earlier FEL budget is not seen as the culprit.

High FEL can be interpreted by some as a luxury situation wherein designers will waste the funding by doing too many unnecessary things, at too slow a pace. If the senior managers have little faith in use of high FEL, they will probably also feel that they are not keeping back sufficient funds to fix the system errors that always seem to arise in projects around delivery time – a circular argument that leads to failure all too often. Early planning is needed to avoid the risk of late errors.

5 REQUIREMENTS GENERATION

Determining a sound set of specifications is one early task that needs appropriate resources. The requirements extraction activity generates the inputs for System Specifications, which in turn, establishes design requirements.

Establishing the set of requirements for a project needs patience, understanding, and experience. It can take a considerable time to assemble a sound set of issues that characterize the customer's need. They must be set up carefully and completely.

This section can only give an introduction; more detail is found in several texts and standards on the topic and in numerous published papers and standards documents such as IEEE (1993), Sage and Rouse (1999), Bray (2002), Hull, Jackson and Dick (2002), INCOSE (2004), and Sydenham (2004).

Requirements need to cover much more than the operational task that the measuring system has to meet.

Need should be based on real deficiencies, that is,

- the current system is inadequate, is
- unavailable or unsupportable, or is
- too costly to operate.

The statement of need should be presented in specific qualitative and quantitative terms, as sufficient detail is provided to allow the systems engineering process to proceed in functional terms.

Requirements extraction is just that – extraction of hopefully 'hard' objective detail from what is initially often quite abstract thinking. It is usually necessary to provide assistance to the customer to transform the initial expressions of need into the format and knowledge content needed to later flow down tasks to the detailed engineering activity stage.

Customers are not often experts at this task and would not have usually approached it by establishing their need on well-structured lines that suit those who follow. To get a project approved, they would have been working in a different direction to the designer's need, perhaps coming up, instead, with a business plan or budget bid that makes its case differently to the needs of requirement extraction.

Requirements extraction starts with the formulation of the key factors – often called the *Critical Issues* (CIs) – via a process of discovery, Sage and Rouse (1999). A representation of the steps of the complete development process is given in Figure 7.

Although the customer will usually provide a written needs statement, it is the contractor that prepares the first requirements list. This developing document needs to be regularly checked with the customer to ensure whether this is what is needed. In this activity, it is important to ascertain that the customer is doing a sound job of verifying suggestions; they may well be distracted by other more urgent internal matters.

Always keep in mind that this is a most critical step in the life-cycle process; this is where directions of effort of many people are decided.

A client needs a system that

- provides operational capability, that is, does the right task;
- possesses adequate operational suitability, that is, does the above tasks when needed, for as long as needed;
- meets certain cost conditions for ownership and ongoing operation;
- fits within a host of legally required environmental conditions;
- meets appropriate health and safety regulations.

To provide this large list of 'must haves', it is necessary to tease out a long set of requirements. A sound way to

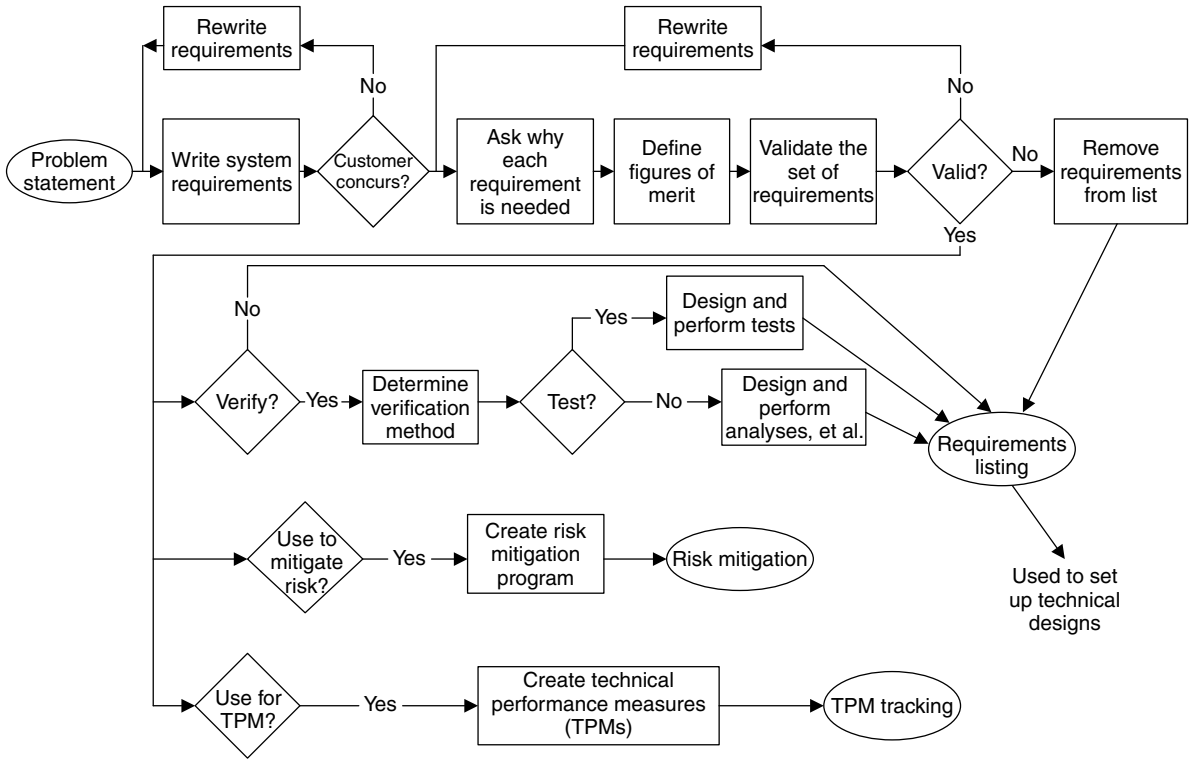


Figure 7. Requirements development process. (Reproduced from *Handbook of Systems Engineering and Management*, Sage and Rouse, Copyright © (1999), Reprinted with permission of John Wiley & Sons, Inc.)

get started is to generate an operational concept or ConOps document for the project, as is discussed in Section 2.

After becoming familiar with every statement about the project that is then available, the next step is to explore the nature of the problem to be serviced. Key issues to address are as follows:

- Why is this measuring system needed?
- What role does it fulfill?
- Is it a new system with few precedents or is it an updated copy needed to give the customer a place in a market or to suit an enhanced operation?
- How does it provide for that need?
- Is it that a major activity will be needed to push forward the technical edge?
- Is it a new innovative use of existing largely proven technology, such as a new-generation digital measurement and control system?

Differences realized here will highlight many requirements issues. This is also the time to make a short study of the type of system needed in terms of its hierarchical and organizational natures as discussed in **Article 63, Outline of Systems Thinking, Volume 2**. The kind of development it will be will then emerge, revealing the generic kind of engineering solutions that might be applicable.

This activity is looking for directions as well as for specific engineering needs. Engineering design must start into its detail only when the needs and constraints are adequately appreciated. It should not leap off with a specific solution at this early stage, but be top-down driven from a sound requirements extraction.

A situation assessment should be developed. General approaches to apply here are as follows:

- What should it be?
- What are the factors that matter over time?
- What to do and how to do it?

An important aim of requirements development is to lead to the recommendation of some candidate solutions. The most likely approaches need their performance, effectiveness, maintenance, and logistic support to be made clear.

This critical identification step is all too often jumped into using ‘intuition’ alone to start an investigation of requirements. This may well start to motivate the mind, but its outcomes must be followed up with analytical assessment of the findings.

It will not be feasible to address every issue that is discovered. Concentrate on the Critical Issues (CI); large systems will identify hundreds of CIs. Learn from similar past projects from which many CI definitions and metric units may be reused.

CIs group into four kinds; these have to do with the following:

- Operational task (What is to be done?)
- Suitability (Will it perform when needed?)
- Political, social, legal, and environmental (What are the key constraints?)
- Programmatic ability (How well will the contractor perform in the development?).

Being able to provide adequate operational service when needed is as important for a system design as is making sure that it does the right job. Ability of a system to successfully fulfill its mission objective is highly dependent on the effectiveness of the support infrastructure provided by the design. This is variously called the R&M (*Reliability and Maintenance*), 'ilities' or 'special functions' aspect, Sydenham (2004).

Life-cycle costing could be a major consideration at the requirements extraction stage.

A Maintenance Concept should therefore be developed that leads to a maintenance plan.

Requirements are best prepared as descriptions of the operation of the system under development and do not state any design solutions – they emerge later. Avoid writing any requirement that dictates a technical solution.

Small measuring system projects can maintain a simple hard-copy record of the requirements. However, when the number of individual requirements exceeds a few hundred, it becomes essential to use a computer-based requirements management tool. All these are based on the creation of a tree structure by intelligently generated functional decomposition activities, starting out with the CIs. Caution is needed. As tools allow rapid generation of statements, it is easy for poorly stated requirements to be entered. What happens later depends on these statements being well thought through.

Detailed design should only start after a careful study of the requirements has developed a sound set of requirement statements that have been fully validated with the customer. Pointers on the features of a good requirement have been published, Sage and Rouse (1999) or Sydenham (2004).

6 CONVERSION INTO A SYSTEM DESIGN

With the requirements in place, it is now the right time to put effort into detailed design. Each requirement statement will outline something to be done and what it is to. It will not describe how to achieve it; that is the technical designer's task.

From the requirements statements, the technical specifications (also called *technical requirements*) are identified that say how the design will be approached. Specifications of many kinds are prepared from these.

Many types of specifications are used in a project. Table 1 gives a summary of the essential types that are usually involved in the larger projects. USDoD MIL documents are given as examples as many of them are available for free download. There exist numerous others. The ones to use are those mandated for a project in its SE Plan or elsewhere.

Other kinds of specification are for

- documentation
- installation
- packaging and transport
- standard items
- modified items
- special subcontracts
- incoming inspection
- use of vendor specifications in projects
- service and maintenance
- safety
- security.

The specifications that are to be used in a project are usually designated in the Systems Engineering manual for the project. Clearly, small projects will roll these together, or often not have them! The list does give an idea of the aspects that a design project has to address. Standard specifications are also useful reference works, for they often embody practical methods and workable ways of proceeding.

7 PARTITIONING

As the initial ideas develop for a concept design solution, there arises the need to break down the whole into parts, each of which is then broken down again. This is called *system partitioning*, *box cutting* or *system reticulation*. This activity concerns the physical association of component parts and how their interfacing is done in the real world and in the design meta world.

A single sensor measuring instrument can be simple, its breakdown being into a serial chain of blocks – *see* **Article 71, Modeling Methodology, Volume 2**. In these cases, the partitioning is reasonably obvious.

As the number of subsystems that interface grow, just how they are grouped is crucial.

For example, development of a multiple sensor monitoring system for an agriculture project will involve sensors, processors, communication channels, central processing, maintenance, calibration, training, and spares replacement.

Table 1. Essential types of specifications, Sydenham (2004).

Specification type	Description	Example format
Operational requirements specification	A specification describing the requirements of the proposed system in terms of user needs, its relationship to existing systems or procedures, and the way in which it will be used and maintained.	MIL-STD-498 DI-IPSC-81430 'operational concept description'.
System/subsystem specification	A specification describing the requirements to be satisfied by the system and the methods to be used to verify that the requirements have been satisfied. The SSS is used as the basis for design and verification of the system.	MIL-STD-498 DI-IPSC-81431 'system/subsystem specification'.
Interface requirements specification	A specification describing the requirements associated with a system that defines its interface(s) with other systems. Interface requirements may be defined as part of the SSS or as a stand-alone IRS.	MIL-STD-498 DI-IPSC-81434 'interface requirements specification'.
Subcontract manufacture/quality specification	A specification identifying requirements and standards of workmanship for items of a system supplied by a subcontract agency. Such requirements may be identified in a subcontract Statement of Work (SOW) as part of a contract or the SOW/Contract may refer to the suggested stand-alone specification.	MIL-STD-498 DI-IPSC-81431 'system/subsystem specification', tailored to identify requirements associated with safety, system quality factors, and design and construction as well as other SOW-style clauses associated with the manufacture and quality of items.
Acceptance test specification (subcontractor, factory, field)	A specification identifying the criteria for acceptance of system requirements described in the SSS. The specification may be tailored to address acceptance criteria for requirements verified by a subcontractor, at the facilities of the contractor, prior to delivery to the end customer or in the field (i.e. when the system is delivered to the end customer).	N/A
Installation specification	A specification identifying requirements associated with the installation of the system in its target (end-use) environment.	MIL-STD-498 DI-IPSC-81431 'system/subsystem specification', tailored to identify requirements associated with installation of the system in its end-use environment only.

If the system development is divided to have all sensors being developed as one kind of subsystem and maintenance as another set of activities, there could well be the need to call all sensors in to carry out the maintenance. Alternatively, they could be designed with the sensor and preprocessing, with maintenance being done for each sensor unit. An example of this going wrong was seen in a fighter plane for remote airfield use that needed 20 plus different maintenance kits, each with different training needs. The project folded when the airfield maintenance staff made it clear that such a division could not be supported in the field.

8 INTERFACE DESIGN

A physical interface is where signals, components, or subsystems connect together. The same interface concept applies to software modules. It also applies to design activities wherein the design of an item needs the data of another item's design to be completed and also the data on where the activities of operation link up.

This means a design exercise always progresses with numerous information interfaces. Each one is a high-risk design-error situation.

For example, the design of an aviation electronic amplifier will need to interface with the packaging, weight, power supply, cooling, and shielding designs. The design requirements of all must be satisfied simultaneously.

Practice has shown that in design activity, there will be almost as many interface design links as there are available interfaces, this can be a huge number.

If the many interfaces involved are not well thought through, prohibitive mixes of activities can arise. There was

once a space telescope that was designed in its testing and calibration functions such that it could only be designed and built when the complete system was assembled and the unit placed into a flight simulation chamber. Any thing found to be wrong at that late stage could only be fixed by iteration back to the very roots of the design.

By partitioning the system in different ways, the number of interfaces can be kept to a minimum. Another way to regard this is by seeking a high degree of isolation of design items. There exists a formal way to investigate this using the n^2 matrix method.

For example, in the design of a sensor, it is sound practice to combine the preamp, supply voltage regulation, conversion, and shielding packaging in one unit sending a high-level signal over the communication system.

A poor design situation was one in which a television-set maker used resistors as a support to wind inductor windings with the insulation set up to deliberately incorporate capacitance – very ingenious but hard to troubleshoot as components could not be isolated for testing or even be seen.

9 TEST AND EVALUATION

The three drivers for all systems developments are cost, time, and performance – the CTP factors. Improvements in all three are always necessary. At first sight, optimizing all three together seems to be impossible. Process improvements and improved design support systems have, however, actually improved all of these in many projects.

As projects are always set up using normal business practices, it is reasonable to assume that the control of cost (accounting) and time (project management, PM) will be in place for a project.

Management of performance, however, is often left to the design engineers to implement on an ad hoc informal basis – *see* **Article 3, Measures and Metrics; Their Application, Volume 1** for an overview of the use of measures and metrics.

This general area has been called *Test and Evaluation* (T&E). Too many people think of this only as testing, that is, measuring a Technical Performance Parameter (TPP) without appreciating T&E is really about evaluating the performance of a system as it develops to ensure it delivers the expected performance throughout its life.

Implementation of sound through-life T&E practices can provide ongoing data on the maturation of the system's CIs. These can be used to tell managers, designers, clients, and financiers that the development is moving toward completion on time, within budget, and with the performance required. T&E should be regarded as a whole-of-life process, not just as a set of tests made at strategic times.

The need for T&E is summed up by asking three key questions of a system development to see if they are in place:

- What are the system's development teams trying to achieve?
- How will those concerned know when the performance objectives are reached?
- Who has the responsibility for a satisfactory performance outcome?

A well-run project will not be using testing as an experiment to find out what has been developed but to verify that the performance of the system is where it is expected to be. A 'no-surprises' project situation should be the aim – T&E is a key mechanism to achieve that condition.

As such, the first text on T&E seems to be the one by Stevens (1989). The case for T&E to be given more status in systems development and operation has been well made in Crouch (1992) and Reynolds (1996). There is, however, little other published material on this topic. It is quite surprising to see project after project being done without sound T&E.

Investing more resources in T&E for a project has the potential to prevent cost overruns and failed systems. Unfortunately, all too often, the T&E budget component is the first to be pruned when overruns arise.

T&E is often practiced in an ad hoc informal manner as a band-aid activity to find out things when a project is not going well. In this form, it has the following deleterious features:

- No adequate traceable or recorded control process exists.
- Success relies on the various designers' abilities to know when and what to test, after which they often have no adequate records addressing the three T&E questions given above.
- There is a real chance that the system elements that the various design teams are developing will not integrate without considerable rework. This situation can arise, not because of lack of competency in performing a good detail design but by simply designing the wrong thing.
- Omitting the overhead of planned T&E activity can indeed save short-term cost. Doing this, however, can increase the risk of not obtaining final success. T&E expenditure can save wasteful rework later by detecting early design errors.
- Last-minute decisions are often made on what to test. This can lead to poor testing as the materials and equipments, not being planned ahead, are often not available.

- There exists too much flexibility in setting up tests and in how to process the data for evaluation purposes. This leaves things very open for biased tests to be implemented to obtain an apparently satisfactory result.

An important thing to recognize is the need to plan T&E activity from the commencement of a project, not at the end of a development stage. A suggested methodology for implementing T&E is given in Sydenham (2004). A T&E Master Plan is essential. This outlines the what, how, and when of the T&E activities to be done.

Three main planning needs of a successful T&E program are as follows:

- Ensuring that scarce test resources efficiently address the project's CIs. Every possible system parameter cannot be tested, there are far too many of them.
- Ensuring that those that need to know can observe how well the CIs are maturing according to well thought through expectations.
- Providing information on the plan and its maturity to all who need to know about these issues. For example, those who plan the use of test resources need considerable lead time for organizing availability of test objects, observers, test equipment, test suite programming, and so on.

As development moves through its stages, the CI determinations should, by tight design control, gradually trend toward the target value, with their uncertainties reducing to acceptable values by the required time.

With such a system available, the project managers and design team leaders can see how the CIs are maturing toward their goals. Time trend graphs can show the current state of performance maturity compared with the target value. Severe deviations can then be given more design attention to bring a particular CI back on track. An openly observable system such as this also exposes errors early, where it can be less expensively corrected with less impact on the careers of the staff.

10 APPLYING SE TO A PROJECT

When a measuring system design team is working within a large systems provider house, there will be a person responsible for organizing systems engineering support for all teams. In such situations, teams usually will be provided with a

- company-specific SE process manual
- computer-based support tool system and support mentors
- special development facilities as are needed according to the project
- sound communications
- effective design records archiving
- configuration management system
- safety control process
- design controls
- sources of advice and mentoring for junior staff
- in-house training
- key documents library.

As a guide, design team leaders need to be familiar with the basics of systems thinking and the culture of systems engineering. They need to have at hand copies of foundational books and standards on SE practice and refer to these for concepts and methods to employ as project problems arise.

Finally, when executing SE and T&E practices for a project, there is the question of 'how much is enough'.

Overhead costs of a support process are hard to justify when the results of that process are of an abstract nature, cover long-term issues, and do not appear to produce easily measured value-adding components to a project. SE and T&E activities are often seen like this – as costly luxuries. They are, however, as important to success as are accounting and project management in that they also assist early detection and control of design error. Who in their right mind would make a major development journey without sound assurances that all design work is on track and moving the design forward is in the right direction at the right speed and within budget?

The design team leader has to use personal judgment in setting the scale of use when applying these techniques or support mechanisms. A single team comprising a dozen or so staff working in a start-up company will probably not be able to devote the time to writing an SE manual. They might instead simply align with an SE standard. The team leader in that case, however, still needs to apply SE principles as part of routine technical management. It is a matter of judgment. First, you need to know what is at stake, then, what tools are available to assist, and then to have the experience to reflect on what to do.

RELATED ARTICLES

Article 63, Outline of Systems Thinking, Volume 2; **Article 66, Phases of System Life Cycle, Volume 2**; **Article 67, Principle of Concept of Operations (ConOps), Volume 2**; **Article 68, Setting the System Boundaries, Volume 2**; **Article 69, Requirements Allocation, Volume 2**.

REFERENCES

Bray, I. (2002) *An Introduction to Requirements Engineering*, Addison-Wesley, New York.

Crouch, V.H. (1992) Test and Evaluation as an Important Emerging Discipline, in *Proceedings Australasian Instrumentation and Measurement Conference*, Adelaide, (pp. 7–17).

Hull, E., Jackson, K. and Dick, J. (2002) *Requirements Engineering*, Springer Verlag, Berlin.

INCOSE. (2004) International Council on Systems Engineering, SE Management Technical Committee, Requirements Working Group, www.incose.org/rwg.

Institute of Electrical and Electronic Engineers. (1993) *Guide for Developing Systems Requirement Specifications*, (IEEE P12330), New York; IEEE Standards Department., and later, Institute of Electrical and Electronic Engineers, *Guidance for the development of the set of requirements*, IEEE Std 1233, 1998 Ed. (R2002) (Includes IEEE Std 1233–1996 and 1233a-1998), IEEE Standards Department, New York, 1998.

Reynolds, M. (1996) *Test and Evaluation of Complex Systems*, Wiley, Chichester.

Ring, J. (2000) *Development of Educational Methodology*, Work in progress for Education and Research Technical Committee of INCOSE, www.incose.com,.

Sage, A.G. and Rouse, B.W. (eds) (1999) *Handbook of Systems Engineering and Management*, Wiley, New York.

Stevens, R.T. (1989) *Operational Test and Evaluation*, Krieger, FL.

Sydenham, P.H. (2004) *Systems Approach to Engineering Design*, Artech House, Boston, MA.

65: Life Cycle Concept

Floyd Guyton Patterson, Jr
George Mason University, Fairfax, VA, USA

1 Concept of the Life Cycle of a System 411
2 The Serial Nature of the Life Cycle 411
3 Life Cycle Versus Team Working Model Comparison 412
4 Typical Life Cycle for a Measurement System 414
Reference 415
Further Reading 415

1 CONCEPT OF THE LIFE CYCLE OF A SYSTEM

The life of a system is modeled by its *life cycle*. In the case of engineered systems, systems that are built using the principles of engineering, the word life cycle is synonymous with *process model*, and it refers to the processes by which the system is conceived, specified, designed, built, deployed, and retired. We will see that life cycles may be quite generic, or quite detailed. A detailed life cycle is shown in Figure 1 for an arbitrary large, complex system. Use is also made of other life cycle models, *see* **Article 66, Phases of System Life Cycle, Volume 2**.

Everything that is divisible into parts can be considered to be a system. It is often said of systems that *'the whole is greater than the sum of its parts'*. This means that there are attributes or properties of the whole that did not exist or that were not manifest in the collection of parts from which the system was assembled. These attributes may be properties of the structure, function, or purpose of the system. Such attributes are called *emergent properties* of the system. It is the study of emergent properties of systems that is the principal subject of *systems engineering*.

Emergent properties of systems, when they are sufficiently well understood and properly managed, provide both solutions and constraints that directly influence designs. On the other hand, unanticipated emergent properties can cause products to fail to meet structural, functional, or purposeful requirements of the system. It is well known that system defects, which require rework of the product specifications, are much less costly to correct when the problems are anticipated – detected and corrected early – rather than later in the design after the product system, and perhaps the production and delivery systems, systems developed to enable the production and delivery of the product as well, have already been built.

The development of a system has both a positive and a negative aspect. In the negative sense, it is in the nature of engineering a system that every design decision, beginning with the development of the initial concept, forecloses options for the structure, function, and purpose of the product. A poorly considered decision early in the design could, in theory, eliminate excellent design alternatives before they are ever considered. Engineering a system is *subtractive – each decision subtracts the options not selected from the total option space* – when viewed in this way, since every decision reduces the size of the option space. The positive aspect will be discussed in the next section.

2 THE SERIAL NATURE OF THE LIFE CYCLE

In the previous section, it was noted that the development of a system has both a positive and a negative aspect;

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

DEFINITION

1. Elicitation of requirements
2. Classification of requirements
3. Analysis of requirements
4. Prototyping of requirements
5. Requirements documentation and specification
6. Preparation of transition documents: request for proposal and a proposal, containing a statements of work, a systems engineering management plan, a work breakdown structure, and negotiated list of other deliverables

DEVELOPMENT

1. Creation of a system architecture and an organizational structure consistent with the system engineering management plan
2. Establishment of configuration control
3. Identification of subsystems
4. Production of subsystem specifications and subsystem test requirements
5. Specification of subsystem components
6. Design of subsystem components
7. Fabrication and testing of hardware components
8. Coding and unit testing of software and firmware components
9. The development of environments and test equipment for hardware testing and scaffolding for software unit testing
10. Integration of components into subsystems
11. Integration testing of each subsystem
12. Validation of the built and tested subsystems
13. Integration of subsystems into a system
14. Integration testing of the system
15. System validation against the system specification
16. Integration test procedures
17. Provide product information to planning & marketing
18. Preparation of transition documents and other aspects of preparing to deliver the system, such as installation materials and procedures, training manuals and special training equipment, maintenance equipment and manuals, and user's guides

DEPLOYMENT

1. Delivery and installation of the system
2. Acceptance testing of the system
3. Operational testing and evaluation of the system
4. System acceptance by the customer
5. Formal maintenance agreement with a contractor that includes a maintenance plan, maintenance test plans, and maintenance test procedures
6. Trade-off study to compare options for continuing to support the business process
7. Decision to decommission the system
8. Provide current system description to requirements team (linkage to new cycle)

Figure 1. A thirty-two step life cycle. (Reproduced from *Handbook of Systems Engineering and Management*, Sage and Rouse, Copyright © (1999), reprinted with permission of John Wiley & Sons, Inc.)

and the negative aspect was explained. In the positive sense, it is also true that each decision opens up a new set of viable options, prompting yet another decision. The sequence of options and decision from concept to product can be represented as a decision tree, which can provide a useful record to support subsequent developmental efforts, provided enough information is captured. Engineering a system is additive when this viewpoint is adopted, since each decision leads to more options, which lead to more decisions, ad infinitum. Neither point of view is altogether correct or incorrect. However, in both cases, the importance of early decision in constraining or creating later decisions is self-evident. For this reason, systems engineering places strong emphasis on the earlier design decisions, those upon which later design decisions are based.

In all this, there is the concept of growth. A given system begins as an embryonic entity: a concept, idea, vision, or the like. Through the repeated application of the process of deciding among options as described above, the system grows in size and maturity, perhaps changing state several times, until the system is realized as a product. The different states may include the mental image, a written concept, a drawing, a prototypephase, subsets of the total development process that are relatively complete, or self-contained, with respect to their structure, function, or purpose.

In broad terms, the simplest life cycle of any system can be described in terms of

- a definition phase,
- a development phase, and
- a deployment phase.

This is referred to as the *DDD* model, shown in Figure 2.

The *DDD* framework facilitates understanding of the resource requirements needed by the system during the various stages of its existence. The *DDD* model is recursive, in that any of the *D*s may be further broken down into steps by application of the *DDD* model to it (Patterson, 1999).

The utility of this approach to engineering a system is found in the ability of the management to treat each phase or subphase as a self-contained project, with discrete inputs and outputs, and with its own schedule and cost profile. As we shall see, the phases are arranged serially, such that the output from any phase represents the starting point for the successor phase; and the input for any phase is provided by its predecessor.

3 LIFE CYCLE VERSUS TEAM WORKING MODEL COMPARISON

It has been pointed out that the systems engineering life cycle provides a simple, useful management tool for estimating, allocating, and accounting for resources during the life of a system. It is an oversimplification to say that phases of the life cycle are self-contained from the standpoint of engineering the total system.

The life cycle concept is a growth model that embodies the concept that early design decisions determine options for later design decisions throughout the life of the system.

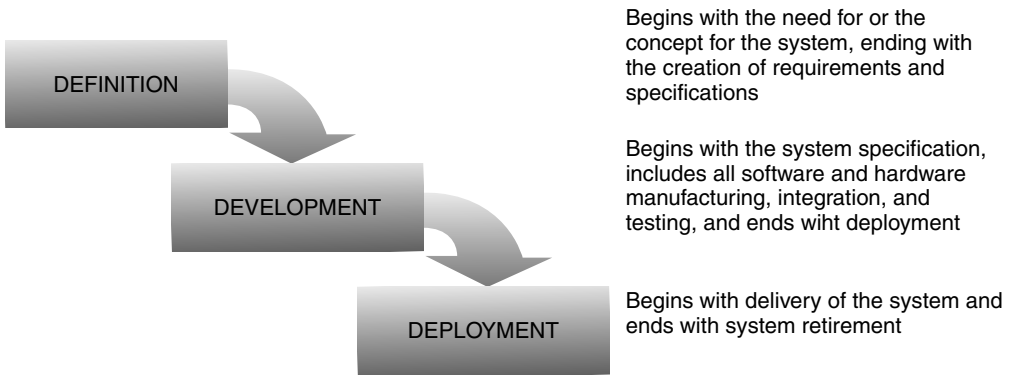


Figure 2. The DDD life cycle. (Reproduced from *Handbook of Systems Engineering and Management*, Sage and Rouse, Copyright © (1999), reprinted with permission of John Wiley & Sons, Inc.)

It follows that no subset, including a life cycle phase, can be entirely self-contained. While it is true that certain very well-defined and well-understood parts of the work effort can be considered to be miniprojects, these parts are always subject not only to verification against a specification, but also to validation in the context of the system of which it is a part.

A common error is to define subsets of the system with nontrivial or poorly defined interfaces to other parts of the system or with insufficient acceptance criteria to ensure the successful validation – testing of the system with the goal of determining whether the customer's expectations have been met – of the miniprojects. It is very difficult, for example, to imagine that one group of engineers can write a set of requirements for a system and deliver the requirements to a different group of engineers for development of a design with the expectation of success. While such an effort may pass a verification test (for example, a design may meet each and every one of a set of written requirements), it is much more difficult to imagine that such a design could be validated against the expected result.

From the standpoint of the engineer, it is much more important to create, communicate, and develop a common vision of the system than it is to attempt to create or to follow a perfect specification. Because documents, drawings, and other design representations have a very low information content relative to the amount of knowledge carried by an engineer who has participated in the design process, various strategies have been proposed to avoid the loss of experience and information through the introduction of new people. Such strategies typically involve the formation of a *system team*, charged with the responsibility of communicating the common vision to the group. The system team remains with the project throughout its lifetime.

Because difficulty of communication varies with the size of a project team and also with the amount of data that one needs to communicate, large teams with large projects

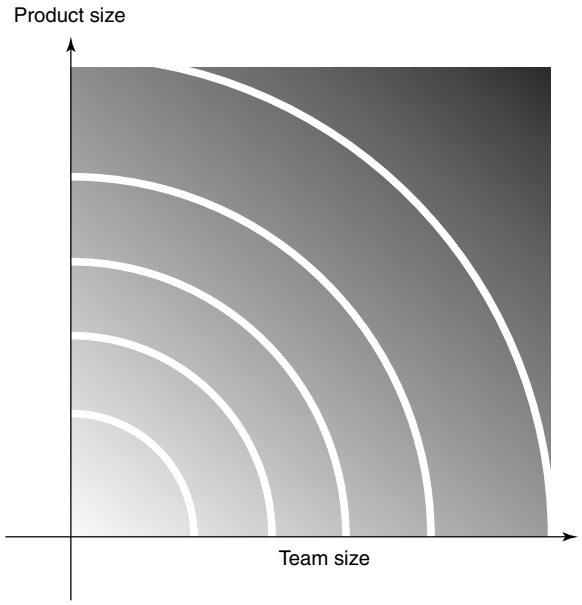


Figure 3. Relationship of project risk to product and team size. (Reproduced from *Handbook of Systems Engineering and Management*, Sage and Rouse, Copyright © (1999), reprinted with permission of John Wiley & Sons, Inc.)

or products experience much more risk – the probability of loss of time, money, or quality – than small teams on small projects. This is depicted in Figure 3, where darker shading indicates more risk.

Organizations organize themselves around processes according to some partition of the enterprise. Their strategic plan is much enhanced by introducing a life cycle model with proven success for the domain. An ad hoc organization will occur – this is the message of Stogdill – if not templated – provided with a structural pattern – especially if the organization is involved in an enterprise with which it is unfamiliar. Introducing the systems engineering life cycle template saves much time and helps to assure success.

4 TYPICAL LIFE CYCLE FOR A MEASUREMENT SYSTEM

The life cycle of a system begins when there is sufficient need to motivate the formation of a concept. The life cycle ends when the system is decommissioned, the *DDD* model can be applied to a measurement system in a very straightforward way. Recursive application of *DDD* yields increasingly detailed descriptions of the activities that compose the effort that is to be managed. It is worth noting here, before discussing the Figure, that the Figure is only a partially complete illustration. The effort required to develop a life cycle that is custom built for the development of system is repaid in many ways. It is a necessary operational framework that needs to be reinvented for each new system development project.

Figure 4 depicts a partially completed life cycle for a generic measurement system, dividing the life cycle into three phases, definition, development, and deployment. At this point, it is appropriate to articulate the expectations for each of these three phases, especially the inputs and outputs.

The input to the definition phase is a need for a particular measurement system that has a well-defined purpose and, perhaps, other success criteria in terms of the structure or function of the system, and any known constraints on the design. The output of this phase is a well-researched and analyzed list of attributes or properties to be measured by some means.

The list of attributes is the input to the development phase, where metrics and measures, which may be qualitative or quantitative, are proposed, validated, and packaged as a metrics framework for use. Thus, the deployment phase can be concerned with using the metrics framework as a tool for actually measuring attribute levels in the data – *see* **Article 3, Measures and Metrics; Their Application, Volume 1**.

In Figure 4, the *DDD* model has been reapplied to each of the three phases to yield subphases as illustrated by the color coding. To further illustrate, the definition subphase of the definition phase has also been refined. This recursive descent into greater levels of detail may be continued until further detail yields no further useful information.

Further information on systems engineering and the life cycle model application is found in the part on '**Part 5, Measuring System Design, Volume 2**'.

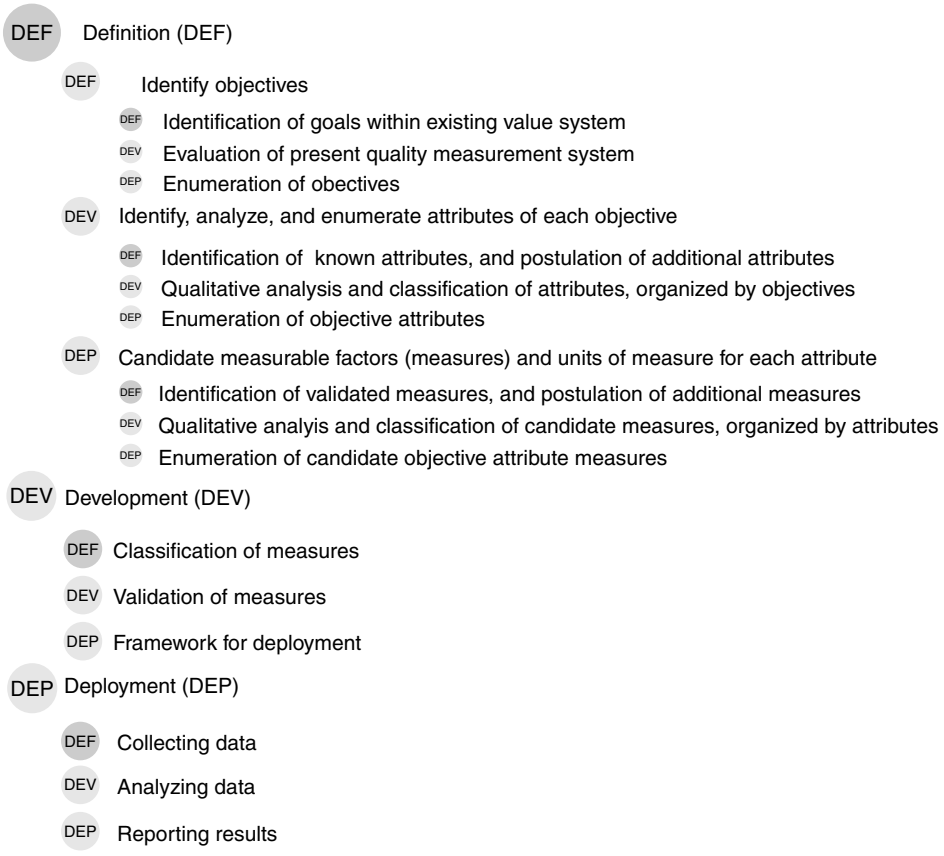


Figure 4. Measurement example using the DDD model.

REFERENCE

Patterson, F.G. Jr. (1999) Chapter 1: Systems Engineering Life Cycles: Life Cycles for Research, Development, Test, and Evaluation; Acquisition; and Planning and Marketing, in *Handbook of Systems Engineering and Management*, (eds A.P. Sage and W.B. Rouse), John Wiley, New York (pp. 59–112).

FURTHER READING

Bass, B.M. and Stogdill, R.M. (1990) *Handbook of Leadership: Theory, Research, and Managerial Applications*, 3rd edn, The Free Press Division of MacMillan, New York.

Sage, A.P. and Rouse, B. (1999) *Handbook of Systems Engineering and Management*, John Wiley, New York.

66: Phases of System Life Cycle

Kim R. Fowler
Technical Consultant, Cool Stream, Windsor Mill, MD, USA

1 Life Cycle Approach to the Engineering of Systems 416
2 Concept Phase 418
3 Detailed Engineering Design 419
4 Manufacture 420
5 Operation 420
6 Upgrade 420
7 Disposal 421
8 Concurrent Development 421
9 Evolution of the Life Cycle 421
References 421

1 LIFE CYCLE APPROACH TO THE ENGINEERING OF SYSTEMS

Systems engineering necessarily takes a life cycle approach. It views the end from the beginning and considers carefully everything in between.

1.1 Standards and life cycle variations

The life cycle for each product or system varies according to its function and market. You tailor the life cycle approach to fit the product; a medical device could emphasize the US Food and Drug Administration (FDA) Design Control Guidance, whereas an electric food mixer will focus on manufacturability to reduce final cost.

Standards for specific markets drive the life cycle processes: software for avionics, for instance, requires very closely defined procedures and methods and, in the United States, follow the stringent requirements of the Department of Defense (DoD) DOD-178B standard; military devices might follow, instead, DoD STD 2167; finally, commercial companies often adhere to the International Standards Organisation (ISO) ISO 9000 standard for producing sytems. Each one of these standards has different specific requirements but they all have similarity, if not the exact same terminology, in their phases. These phases are described in detail below.

1.2 Development processes

Development can move through these phases in several different ways. You can use the top-down or waterfall method to layout and build your system (Figure 1); its premise is that the requirements completely specify a design and the development follows precept upon precept. Alternatively, you could build from the bottom up using the prototyping method (Figure 2); you take a basic idea, add some features to it, and try it out. Finally, you can combine the waterfall and prototyping methods into a spiral development (Figure 3) where requirements and development occur incrementally (Fowler, 1996; Sydenham, 2004).

Each stage of the waterfall process, Figure 1, aligns closely with the phases described below. The waterfall process is the original model for life cycle development. It may be the most well understood but it fails to accurately portray most development. It does not account for major changes in requirements without starting over. It assumes that iteration can only happen between adjacent phases of development, which is not typically the case. Problems in one stage, such as manufacturing, may force you to go right back to revisit the requirements and change some of them.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

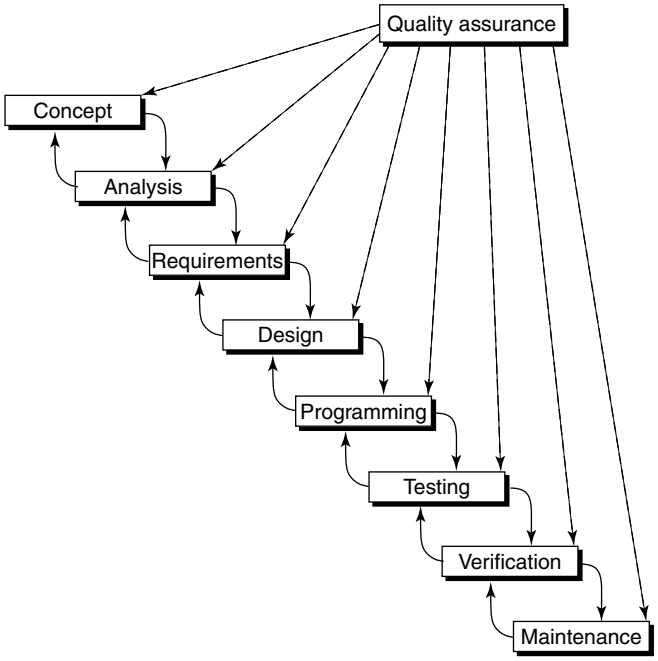


Figure 1. Waterfall method of development. (Reproduced from "Electronic Instrument Design: Architecting for the Life Cycle", Fowler K.R. Copyright (1996) © Oxford University Press.)

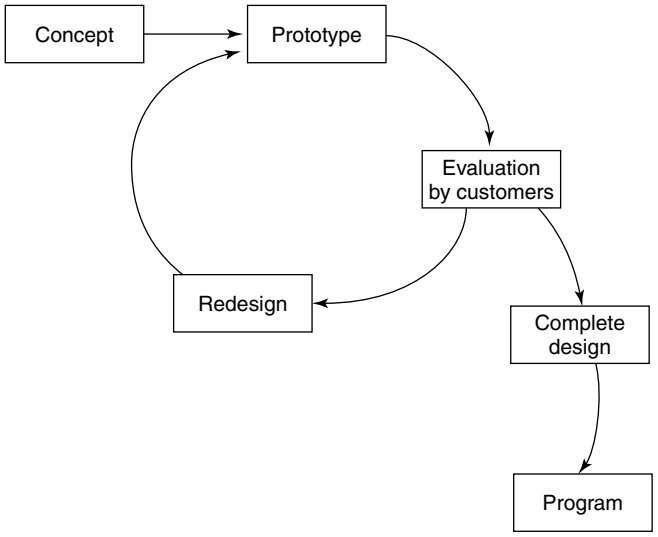


Figure 2. Prototyping model of development. (Reproduced from "Electronic Instrument Design: Architecting for the Life Cycle", Fowler K.R. Copyright (1996) © Oxford University Press.)

The prototyping method, Figure 2, often suffices for proving a concept, but it does not have the rigor to transfer sufficient information and knowledge for manufacturing and deployment. Its utility resides in one-of-a-kind systems and custom development.

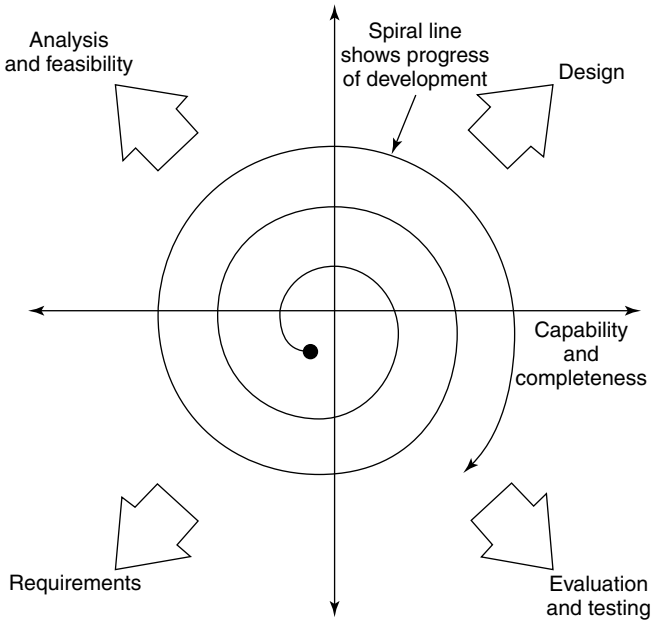


Figure 3. Spiral model of development. (Reproduced from "Electronic Instrument Design: Architecting for the Life Cycle", Fowler K.R. Copyright (1996) © Oxford University Press.)

Spiral development, Figure 3, evolved from the perceived strengths of the other two methods to supersede their shortcomings. It takes a basic set of requirements and develops a workable system with a reduced set of desired features. This allows you to understand more fully the interactions between critical components and subsystems early in development. You then add more requirements that lead to additional features and proceed around the spiral to integrate, test, and prove greater functionality. It is often favored in software development.

1.3 Judgment

Understanding the principles of systems engineering and applying the appropriate life cycle development are not enough to ensure an optimal solution. These are necessary but not sufficient conditions for success. Most final products are a suboptimal set of compromises and trade-offs; a fighter aircraft, for example, is never a perfect solution – it trades off many concerns, such as threat, mission, speed, maneuverability, range, weapons loads, and reliability. Most products and services of a mature and new kind will be of this nature.

Ultimately, good judgment, based on knowledge, experience, and wisdom – some call it refection – will do the most toward guaranteeing a good result. We can only get there by learning and applying, which means that we need

to risk some elements of failure. Learning should perfect your engineering judgment, and the only way to get there is through experimentation, iteration, and failure. 'No one wants to learn by mistakes, but we cannot learn enough from successes to go beyond the state of the art' (Petroski, 1985).

2 CONCEPT PHASE

The life cycle begins by defining the problem. You have a number of sources for understanding the problem better; they include the following:

- customer objectives (which are not necessarily the users),
- user needs (users are not necessarily the paying customer),
- mission (defines the regions of operation),
- constraints from physical processes and environment,
- government regulations,
- commercial standards.

The real brilliance in setting requirements is to see what people *do*, not what they *say*. Only by doing this can you really understand the true need (Christensen, 1997). People do not necessarily understand the capabilities of technology and innovation, so they do not often express their true needs. They tend to express their desires, clouded as they are with misconceptions and misperceptions. Setting up the need is usually a journey of discovery for both the client and the designer.

2.1 Requirements

Requirements generate the '*what*' to which designers then prepare the detailed descriptions (schematics, source listings) that give the '*how*'. A sizable portion of time is devoted to generating requirements in the Concept Phase. Requirements may include some or all of the following: performance, reliability, maintainability, availability, human factors, safety, fault tolerance, operation regimes, environment, and logistics. *See also* **Article 69, Requirements Allocation, Volume 2**.

2.2 Analysis

After defining the problem, you enter an iterative cyle of analyses to refine the possibilities toward a potential solution. Functional analysis defines basic operations and then breaks them down into successively more detailed functional components. These functional blocks or components eventually specify *who* does *what*, and *when* and *where* they do it. This kind of analysis is much like tracing the roots of a tree down to finer and finer rootlets. It can be an exhaustive but worthwhile exercise to understand all the possibilities and interactions.

Once you define the functional components, you then model the problems and their solutions through mathematical analysis, simulation, and prototyping. Mathematical analysis can be tedious and limited in its efficacy, but it often can give bounds or constraints to the problem and its solution. Simulation provides insight into the interactions between components and allows 'what if' type of studies. Prototyping produces a subset or subscale model that allows indepth studies of functionality without committing to a complete production cycle – *see also* **Article 72, Mathematical Methods of Optimization, Volume 2**. These methods and techniques allow you to characterize the potential solutions, which then enter the feasibility analysis.

Typically, a CoDR or Conceptual Design Review completes this phase of development – *see also* Concept of Operations in **Article 67, Principle of Concept of Operations (ConOps), Volume 2**.

2.3 Feasibility

Feasibility is the evaluation portion that refines the potential solutions. Typically, you will iterate several times through definition and refining requirements, functional analysis, modeling, and feasibility analysis (Figure 4). As a problem grows more complex, the analyses become more detailed and time consuming.

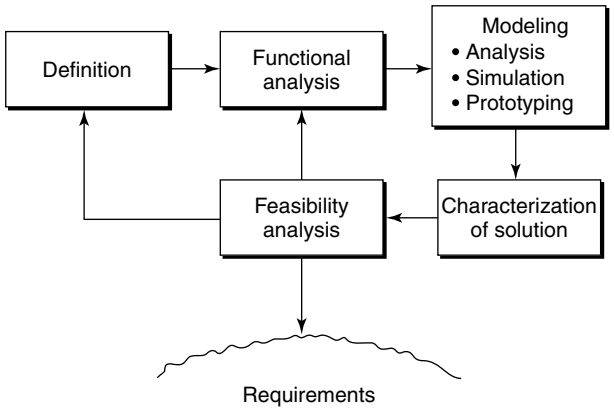


Figure 4. The iterative nature of concept development and feasibility analysis. (Reproduced from "Electronic Instrument Design: Architecting for the Life Cycle", Fowler K.R. Copyright (1996) © Oxford University Press.)

Feasibility has several categories, business, physical, and risk, which can overlap in their outcomes. The business case looks for the best outcome, most likely to maximize revenues or profit. The physical case evaluates whether the proposed solution violates any physical principles and other needs such as safety and legal issues. Finally, risk analysis evaluates the potential problems and margins of tolerance; these can include Fault-Tree Analysis (FTA) and Failure Modes Effects Analysis (FMEA) – *see also* **Article 101, Assessing and Demonstrating Safety, Volume 2**; **Article 100, Safety Analysis Methods, Volume 2**; and **Article 99, Safety Organization, Volume 2**. These two types of risk analysis are complementary. FTA takes each identified potential fault in the system and traces the cause(s) down to the possible originating components or interactions. FMEA hypothesizes a failed component or interaction and then traces the effects of its failure up through the subsystems to its ultimate manifestation at the system level.

3 DETAILED ENGINEERING DESIGN

Detailed engineering design takes the *'what'* from requirements and molds it into an integrated whole. In other words, design synthesizes the *'how'* of the system. Detailed engineering design follows after the Concept Phase and feasibility.

For many complex systems, detailed engineering design splits into two design phases: preliminary design and critical design. Typically, you complete each of these phases after a formal review that satisfies all open action items or concerns.

Design, and indeed all development, should follow a simple, iterative cycle: plan, execute, review, and report. Every activity should follow this process. Even the overall phases should follow this general pattern. It is a useful habit to inculcate.

3.1 Documentation

The primary output from detailed engineering design is documentation. Documentation includes source listings from software, schematics of hardware, plans, and procedures, checklists, memos, messages, manuals, reports, and reviews. Documentation records the plans, the execution, and the reviews and tests. These components of documentation become the directions and instructions for manufacturing, producing, and using the product or system.

Documents need version control to record an accurate historical picture and to maintain a contiguous string of updates. Version and configuration control has been manual until the recent past, software and systems now exist that assist control of versions and archive documents.

3.2 Execution

Execution is the actual design work but it is usually a very small portion of the total effort. This is the creative part of development where coding happens or schematics originate. You should maintain a disciplined habit of keeping a detailed engineering notebook, or its equivalent in software; it becomes indispensable when problems arise or modifications are needed in the future. The notebook needs to provide clues and reasons for your work.

3.3 Review

Review has a variety of forms. It may be formal, such as a design review. It may be informal, where you consult with a colleague about a problem. Formal review can include some or all of the following specific instances:

- Peer review – a process of examining a module or subsystem, particularly software that evaluates the implementation for soundness. Usually four or five people take part – a moderator (best being a person not too close to the project) to facilitate the review, the author/designer, two colleagues to do the review, and sometimes a note taker to record comments and action items.
- CoDR – conceptual design review, which typically completes the Concept Phase. It evaluates the basic approach and the feasibility.
- PDR – preliminary design review, which completes the Preliminary Design phase. It evaluates the basic approach and the initial designs.
- CDR – critical design review, which completes the Critical Design phase. It evaluates the final design and approves the production go-ahead.
- Manufacturing transfer or production hand off (PHO) – this is an additional step that follows a CDR in a volume production scheme. It evaluates the manufacturing issues and gives final approval for production.

3.4 Test

Tests usually begin during the detailed design phase and evaluate the individual components, modules, and subsystems. Tests can cover one or a variety of concerns, such as

performance, failure modes, and environmental tolerance. Examples of tests include the following:

- Software white box testing – where knowledge of the code leads the tester to exercise all reasonable paths of operation (white box for where only internals are known).
- Software black box (for where only externals are available) testing – where the tester exercises the module in all the anticipated operations without knowledge of the code operations.
- Hardware thermal testing – cycling the circuitry through cold and hot temperatures to find potential areas of operational noncompliance.
- Vibration and shock – testing the circuitry and mechanical hardware for potential problems and stress points.
- Integration – exercising the system as subsystems are attached for functionality and correct operation.
- Verification – testing for conformance to the metrics in the requirements. These metrics may originate in performance, electromagnetic compatibility (EMC), fault tolerance, reliability, and maintainability.
- Validation – testing of the system for apparent fulfillment of the customer's or users' desires. Usually, you exercise the user interface or graphical user interface (GUI) for operational responsiveness, utility, and ease that leads to user satisfaction.
- Field-testing – is a combination of integration, verification, and validation. It can accurately portray the environment and functionality of a system in actual use. These types of tests are extremely valuable but tend to be the most expensive; they also often require a completed system, but today the trend is to make use of a sophisticated model of the system.

4 MANUFACTURE

Manufacturing is an entire field of study unto itself. Concerns for manufacturing should be addressed during the previous phases of concept and detailed engineering design. You should consider design for manufacture, where fewer, simpler parts are almost always better than greater complexity. Another area for consideration is design for test; this will help you and those who follow you to also diagnose problems more quickly and efficiently as well as show compliance.

One particular manufacturing concern that you should consider is building the system in-house or outsourcing the work. If you have the assets and resources in the company, which are available and which can be efficiently and wisely used (meaning keep them busy full time for most situations), then in-house production is reasonable. Otherwise, outsourcing the production effort is a worthy consideration, particularly if your company does not have the expertise to manufacture the product. It does, however, pass control out and thus the supplier has to be reliable.

5 OPERATION

Just designing and building something is not the end of the matter. You need to see it into operation and ensure that it functions reliably. This means that you need to consider the effort to install the equipment, train the operating personnel, and provide technical support.

One area of this that is continually overlooked is the user manual. Successful operation requires a good manual, training, and access to appropriate information. For more complicated systems, a troubleshooting and repair manual is essential.

Once operational, you need to consider the logistics and inventory. What are the consumable resources that your system needs – fuel, lubrication, batteries, or auxiliary supplies, for example, paper for printers? Then there are maintenance and repair, which are two completely separate activities. Maintenance involves periodic care for the equipment – labor, spares, inventory, and equipment downtime are all factors. Repair is the fixing or replacement of a failed component; here the same concerns as maintenance, such as diagnostics, prognostics, and test equipment accompany additional issues.

6 UPGRADE

If your system operates with some degree of success, then invariably someone will want to upgrade its capabilities in the future. You need to consider this when you do the initial design. Prepare design hooks in the original design. These are openings or attachment points for potential upgrades. Careful coding of software can make the addition of new modules and capabilities easier in the future. Providing a margin in memory, performance, size, weight, and power will ease the addition of future hardware.

Designing a system for expandability does add complexity to the system. It generates new possibilities for unforeseen circumstances and interactions. Consequently, complete and correct documentation is absolutely essential. Necessary documentation includes

- software source listings,
- hardware schematics, and
- design documentation that explains the theory of operation.

The engineering notebook can be indispensable for upgrades. It provides the reasons for why and how things were done. It can also help with explaining the theory of operation.

7 DISPOSAL

Environmental concerns for disposing off equipment that is no longer in use are rising. You need to consider, during the early phases of concept and detailed engineering, how to dispose of your system. Plan for recycling of the materials, particularly of the enclosure and packaging aspects. Reuse of components may be an avenue for disposal as well. Toxic compounds, such as heavy metals used in solder, circuit boards, and components, are becoming an environmental crisis; think about how you might reduce their use in your products.

8 CONCURRENT DEVELOPMENT

The life cycle of development necessarily involves parallel streams of effort. Someone must evolve the hardware designs as someone else designs and codes the software. A different group must prepare the mechanical mechanisms and packaging alongside the circuits and software. This is the business of systems engineering oversight.

One principal concern is the definition of interfaces within the system. These interfaces will control the architecture and the eventual operation of the system; they also affect the concurrent development of the system. A general philosophy of interface coupling, whether loose or tight, will help define the architecture.

Loose coupling means that the interface is straightforward with the minimum in communications between modules. It tolerates changes and tends to be more robust. Loose coupling can reduce rework and redesign and can increase reuse. It aids concurrent and independent development, can adapt to the market, and ease the test burden. The disadvantage with loose coupling is that it tends to produce a bigger, heavier system.

Tight coupling optimizes performance or size by spreading function across interfaces. Its advantage is that it can reduce the initial cost of manufacture with more optimized structure or smaller form factor. Concurrent development, however, is much more difficult and testing is more extensive at the system level. Redesign is more difficult because it affects the entire system; changing one component more readily affects others within the system.

9 EVOLUTION OF THE LIFE CYCLE

The concept of the life cycle has evolved into various forms from its origins in the 1950s and 60s. The waterfall model gave way to several different models including the prototyping and spiral models that better fit actual development. Complex systems, for example, a fighter aircraft might use a combination of approaches – first it is built to a set of requirements that provides a major subset of functions (basically a waterfall development), then it goes through extensive flight tests that determine other capabilities. These flight tests are iterative and can cause changes to requirements, now development looks more like the spiral method.

The facility provided by the Internet is forcing further evolution of systems design and life cycle development. Embedded systems are gaining Internet ports so that software upgrades may occur quickly and easily in the field. These kinds of upgrades significantly reduce downtime and allow more frequent tuning of operation within embedded systems. The capability of the Internet can now combine field-testing with operation; successive versions of the software can be far more numerous. This type of life cycle development strongly resembles prototyping but it is much faster. The iteration time is much shorter. The problem with fast field upgrades is that people will be tempted more than ever to release 'buggy' code that does not work quite right to make market deadlines and hope to 'fix' the problems later through field upgrades.

Engineering knowledge and judgment is the one area over which you have some control during your career. You should be continually learning, monitoring, optimizing, and improving what you do and how you do it. Your experience can lead to changes and improvements in the life cycle of development. There is much more to engineering than the design itself!

REFERENCES

Christensen, C.M. (1997) *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*, Harvard Business School Press, Boston, MA (p. 188).

Fowler, K.R. (1996) *Electronic Instrument Design: Architecting for The Life Cycle*, Oxford University Press, New York (pp. 407, 415, 416).

Petroski, H. (1985) *To Engineer is Human: The Role of Failure in Successful Design*, St. Martin's, New York (p. 62).

Sydenham, P.H. (2004) *Systems Approach to Engineering Design*, Artech House: Boston, MA.

67: Principle of Concept of Operations (ConOps)

Jack Ring
Innovation Management, Scottsdale, AZ, USA

1 Definition of a Concept of Operations (or Operational Concept) 422
2 The Concept of Operations Applied to System Development 423
3 Allocations of Measures of Effectiveness to the Conops Model 423
4 Documentation 424
Further Reading 425

1 DEFINITION OF A CONCEPT OF OPERATIONS (OR OPERATIONAL CONCEPT)

A Concept of Operations (ConOps) describes how an observer intends to use a contemplated measurement system in a specific actual or anticipated situation. It is the starting point for engineering a measurement system – *see* **Article 63, Outline of Systems Thinking, Volume 2** – and other articles on that topic.

In the vein that 'understanding the problem is 90% of the solution', a ConOps purposefully treats the intended measurement system as a black box. ConOps authors describe the operational situation – what the users need to accomplish and how a measurement system can help make sure that the right issues are reached. Authors must studiously avoid making any assumptions about the internal content and structure of the eventual measurement system.

The ConOps focuses on both the stimuli to which the measurement system is expected to accommodate and the user-visible responses to such stimuli. Further, the ConOps describes the effect the responses are intended to have on the operations situation. This latter facet sets the stage for the Measures of Effectiveness (MOEs) by which the eventual measurement system will be judged.

A ConOps describes why a measurement system is needed, how the measurement system will be used, and the consequent value of such a system. The user's mission objectives and goals regarding the phenomenon to be measured are mapped to measurement system capability scenarios. It tells a story about the expected behavior of the phenomenon to be measured and about a day, week, month, or other episode in the life of the users of such measurements and thus of the desired behavior of the contemplated measurement system.

A ConOps is not the same as an Operations Concept. The ConOps describes the usage of the intended system by articulating 'we need a system that will enable us to do the following: "The authors are standing in problem space looking out"'.

The other view, the Operations Concept or Concept of Operation, describes the operators view of a system as designed or produced. The authors are standing in solution space looking toward the problem space.

This subtle distinction is very important because it emphasizes an understanding of the operational need without the clutter, confusion, and complexity of implementation ideas.

For example, a ConOps says, 'here's what we need to measure about the weather and how we intend to use those measurements'. In contrast, an Operations Concept document says 'given that we have designed this ocean temperature, salinity, and current flow system, here is the way it should be deployed, operated, and maintained'. Accordingly, the Operations Concept is a response to the ConOps, not an alternate name for the same thing.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

A good ConOps may reflect 'out-of-the-box' thinking and is not concerned with immediate perceptions of implementation feasibility. Because a ConOps describes a measurement system yet to be designed, it is necessarily speculative – a vision, though hopefully not a hallucination.

To many readers and especially to system designers and developers, a ConOps can be irritatingly vague, presumptive, or unrealistic. However, a ConOps should be judged solely on how well it describes the situation in which the measurement system is to be used.

A ConOps is not the same as a Use Case. The ConOps is a big picture scenario of the total usage envisioned, whereas a Use Case presumes to describe what the system will do internally to serve each user. Use Case Scenarios are similar to Operations Concepts rather than to ConOps.

2 THE CONCEPT OF OPERATIONS APPLIED TO SYSTEM DEVELOPMENT

A ConOps often serves to converge multiple stakeholders toward a common understanding regarding the usage and value of the desired measurement system. This should be accomplished before getting into details about what the measurement system will look like, what technologies it will employ and what its cost and development schedule are expected to be. In other words, it is not about implementation detail.

The ConOps description included all of the stimuli and responses to be monitored by the measurement system. These include all interfaces between the intended measurement system and the User, Phenomenon System, and all collateral systems. Interface types must be specified. Examples are as follows:

- Coexist = two or more entities in contact.
- Compete = two or more entities in tension regarding contact.
- Connect = two or more entities structurally involved.
- Conduct = two or more entities passively exchange energy.
- Communicate = two or more entities passively exchange data.
- Cooperate = two or more entities change behavior in order to exchange energy or information (usually as function of energy or information).
- Collaborate = two or more entities change structure or content in order to exchange energy or information (usually as a function of result being pursued).

This level of information enables an analysis of the effect of all the responses relative to user objectives and goals. These effects are the basic Measures of Effectiveness. Next, the acceptable performance levels for each MOE are decided. These are the basis for the measurement system acceptance criteria.

The design contract for the measurement system should be based on the ConOps and the MOEs. Of course, this is what Customer Oriented, Voice of the Customer, and similar ministrations in a variety of business publications have meant for the last 10 years.

Products can fail in the marketplace because the contract document was based on the Specification for the System rather than on its value to the system users.

Time and again, marketing and prospective customers collaborate to describe the problem that exists and how they would like their world to be, and even the tolerable therapies they will endure to get there only to watch as engineering people proceed to reexpress the intent in the form of requirements, then 'manage requirements'. In response to this, they can fend off corrections with threats of cost and delivery time explosion.

After consensus is reached on the ConOps, MOEs, and acceptance criteria, the system development process can proceed to draft Use Cases about each of the stimuli (and their sources) on the system. These begin to reveal the system requirements.

3 ALLOCATIONS OF MEASURES OF EFFECTIVENESS TO THE CONOPS MODEL

Although a measurement system may have to handle hundreds of stimuli and generate responses for each, it is necessary that the MOE is the minimum set that adequately characterizes the effect of the measurement system on the user situation. Some call these the system's critical issues (CI) that then can be analyzed to reveal its MOEs – *see also* **Article 14, Hierarchical Aspects of Measurement Systems, Volume 1**.

The generic MOEs for a measurement system each address and include the following:

Adequate: The system must measure and report enough attributes of the phenomenon of interest to satisfy the user needs. For example, if the phenomenon is a single event, then the measurement of that event is adequate. However, if the phenomenon is the popularity rating of a politician, then dozens of attributes may have to be measured as well as the demographics and psychographics of those interviewed.

Accurate: The inherent precision and uncorrectable disturbance error in each measurement are important as is the frequency of measurement and duration of observation.

Timeliness: This factor concerns the time lapse from (a) occurrence of the phenomenon to (b) the availability

of the measurement in useful form. A highly accurate measurement too late is not valuable to the user.

Interoperability: This factor concerns the behavior of the measurement system with respect to other systems in the operational situation. This goes beyond mere existence of interfaces to consider the interoperation scenario (e.g. control vs. slave and multiple modes of interoperation).

Resilience to operator error(s): Often called the *ease of use*, which is too ambiguous; the better term is the *effectiveness factor* of the measurement system's resiliency to disturbances caused by operator error, and especially to a scenario of errors.

Operational Availability index: This factor describes the likelihood that the measurement system will be operational at all times it is required to be operational. The operational availability (OA) index is a combination of reliability, maintainability, and operations calendar. An important subfactor is the calendar date for the first intended use of the measurement system. This, of course, is an attribute of the measurement system development project rather than of the measurement system itself, but the OA index goes to zero if the measurement system is not delivered on time when it is needed.

Total cost of ownership: Total cost of ownership (TCO) is important even if the cost is not the direct responsibility of the user.

The use of Standards of Acceptance is illustrated in Table 1. For each MOE (generic ones are used here), the acceptable values for each are specified. The 'X(i)' values indicate the capability of the system if it is to be of maximum value to the users. The 'Y(i)' values indicate the effectiveness level that the users will find acceptable. Because the TCO of the system typically increases with the level of system effectiveness, a measurement system can be judged acceptable while significantly below the Max Value Threshold.

Table 1. Relationship of MOE factors to quantified level of each MOE.

MOE	Max Value Threshold	Acceptance Threshold
Adequate	X1	Y1
Accurate	X2	Y2
Timely	X3	Y3
Interoperability	X4	Y4
Resilience to operator error(s)	X5	Y5
Operational availability index	X6	Y6
Total cost of ownership	X7	Y7

4 DOCUMENTATION

The ConOps is often produced as a file document but the better ConOps is formed as an executable model (or a model and simulation) such that the dynamic interaction of the ConOps factors can be observed.

Regardless of form, the following checklist of ConOps content should be used:

1. Mission of the Measuring System (MS).
2. Scope/context of the MS.
3. Anticipated phenomenon system (content, structure, behavior, pertinent exogenous attributes, pertinent endogenous attributes, persistence, variability, glint, scintillation, position, etc.).
4. Anticipated User objectives and resulting needs for a measurement system, such as

   - the existence of something, both the exogenous attributes and the endogenous attributes;
   - the location of something, including the related factors;
   - the persistence of something, including the related factors;
   - the variability exhibited by that something (such as range of values, the rep rate of pertinent changes, the least count of observations).

5. Desired observation and interpretation scenario (regardless of implementation feasibility). Includes usage of the measurements, such as

   - surveillance,
   - metrication,
   - alarm,
   - control.

   Also describes the roles, accountabilities, authorities interfaces and responsibilities of the operations personnel.
6. The boundary of the MS (and what is not MS such as the transport system that positions the MS or the reference standards that enable calibration of the MS.
7. Any technology constraints or opportunities consistent with operational considerations.
8. The permissible support and maintenance scenario for ensuring OA of the MS.
9. Test capability the MS must provide or endure.
10. Operator readiness capability the MS or a supporting subsystem must provide.
11. The preferred or permissible Cutover scenario.
12. The MS evolution that must be provided for within the OA timeline.

13. The interoperability with related systems (communicating, cooperating, collaborating, coevolving).
14. Permissible environmental impact considerations.
15. Permissible societal impact considerations.

Effort expended in developing a sound ConOps can save the existence of late errors that are costly to fix. Errors are far less costly when fixed early.

FURTHER READING

Buede, D.M. (2000) *The Engineering Design of Systems*, Wiley, NY, (p. 42).

Wray, R. (ed.) (2000) *Systems Engineering Handbook*, version 2.0, INCOSE, Seattle, WA, URL: http://www.incose.org.

68: Setting the System Boundaries

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Boundary Perspective of Measurement Systems 426
2 Extrinsic and Intrinsic Boundaries 427
3 Application of Boundary Concept to Systems Identification 428
4 Influences on a Real System 428
Related Articles 429
References 429

1 BOUNDARY PERSPECTIVE OF MEASUREMENT SYSTEMS

The wider system, within which a human–machine system is embedded, is sometimes referred to as *the supersystem* (McGhee, Henderson and Sankowski, 1986; McGhee *et al*., 1996).

Since each of the subsystems, described in **Article 6, Substructure of Human–Machine Systems, Volume 1**, may also be considered as systems in their own right, this means that the human–machine system is the supersystem in which each is embedded. Hence, it can be seen that measurement systems, which may be considered as subsidiary blocks of the information system, cannot exist in isolation.

The human–machine system environment, or supersystem, within which they function, regards them as subsystems of itself. Hence, the bilateral impacts of this linking must be considered.

Instruments perform a purposeful function. They also possess an opaqueness and obscurity of structure, characteristic of complexity. This functional complexity of the combination of the associated components justifies treating them as systems. An exemplary aspect of the systems approach is the totality of the considerations that are taken into account.

A basis for highlighting all of the effects impacting on a measurement system may be adapted from the boundary view of human–machine supersystems, given in McGhee, Henderson and Sankowski (1986), in McGhee *et al*. (1996), and in **Article 6, Substructure of Human–Machine Systems, Volume 1** where it has been shown that this boundary perspective indicates the importance of a detailed consideration of all impacts in measurement systems. It also allows a distinction to be drawn between the three groups of inputs and outputs as shown in Figure 1, where the systems engineering approach, described in **Article 5, Humans in the Real World, Volume 1**; **Article 2, Organization of Instrument Science, Volume 1**; and **Article 6, Substructure of Human–Machine Systems, Volume 1**, can be applied fruitfully by visualizing a boundary view of an instrumentation system.

In the same way as for human–machine systems, as described in **Article 6, Substructure of Human–Machine Systems, Volume 1**, the three types of input and output may be grouped as Resources, Information, or Contamination.

These inputs and outputs in measurement systems may appear in any of the COMETMAN energy forms described in **Article 109, Structure and Energy in Sensor Systems, Volume 2**. The principal resources in measurement systems are the energy forms involved in supporting the flow of information.

In the case of a measurement system, there is no need to emphasize the presence of a resource output since the information output is also the resource output.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

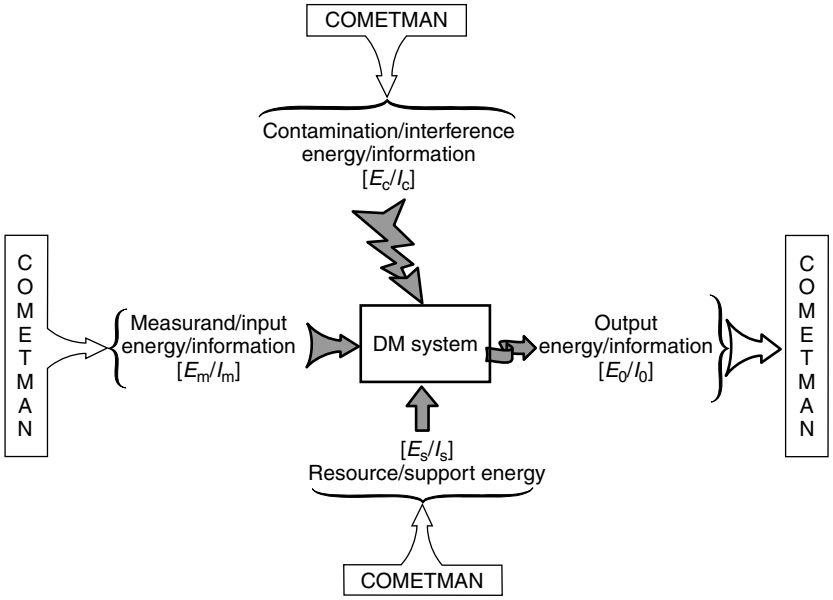


Figure 1. Boundary impacts in an instrumentation system.

Contamination, which is also known as an *influencing input*, or simply as interference, consists of disruptions, impurities, and disturbance power flows. Although not shown in Figure 1, it is generated at the output of the, say, digital measurement (DM) system represented in this figure. It may also appear as interference in other information carrying channels. Disturbance power flows are the primary source of deviations and errors in measurement systems.

2 EXTRINSIC AND INTRINSIC BOUNDARIES

It is also important to distinguish whether the sources of contamination occur from inside or outside the measurement system. The systems approach uses the principles of recursion and reticulation (or subdivision) of systems. Consistent with this view is the possibility of distinguishing between extrinsic and intrinsic system boundaries. The reticulation, or breaking down, is based upon the methods of taxonomy described in **Article 2, Organization of Instrument Science, Volume 1**.

Consequently, a boundary may be visualized between the measurement system and the human–machine supersystem. This boundary may be referred to as an extrinsic, or external, boundary (McGhee, Henderson and Sankowski, 1986; McGhee *et al.*, 1996) as represented in Figure 2. Impacts from outside the instrument penetrate the extrinsic boundary and cause effects within the system. These external interference sources distinguish the extrinsic boundary from the intrinsic, or internal, boundary.

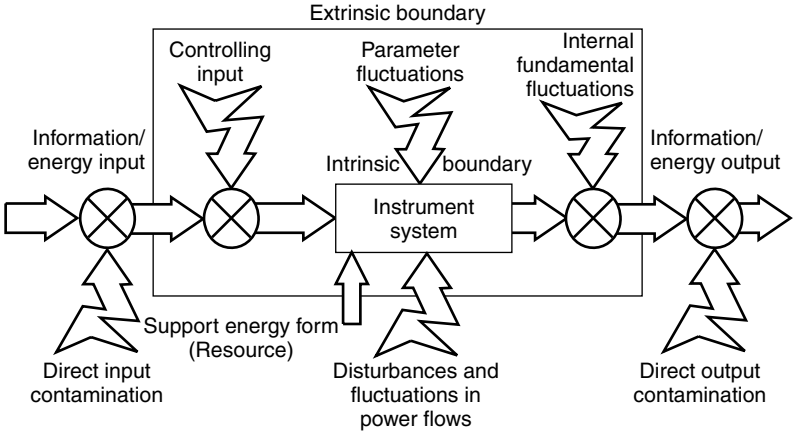


Figure 2. Extrinsic and intrinsic boundaries in instrument systems.

Linking the extrinsic boundary with another boundary called an *intrinsic boundary*, as shown in Figure 2, provides a way of classifying different extrinsic or intrinsic interference sources. Taking an intrinsic boundary into account allows a separation and distinction between the various impacts within the system from those outside the system.

Thus, modeling instruments using the extrinsic and intrinsic boundaries illustrated in Figure 2 is essentially the same as modeling, for example, an operational amplifier. The real operational amplifier is enclosed within an imaginary extrinsic boundary. An idealized noise and drift-free operational amplifier is contained within its intrinsic boundary. The internal noise and drift in the real amplifier are accounted for by including appropriate models for these effects within the space between the intrinsic, or internal, boundary, and the extrinsic, or external, boundary. Proposing extrinsic and intrinsic boundaries allows a separation and distinction between effects in the system due to external or internal causes. It should be noted that inputs and outputs in Figure 2 could be in any of the COMETMAN energy forms. They have not been included for the sake of clarity. The COMETMAN acronym is explained in **Article 109, Structure and Energy in Sensor Systems, Volume 2**.

3 APPLICATION OF BOUNDARY CONCEPT TO SYSTEMS IDENTIFICATION

Systems Identification (Eykhoff, 1984; Ljung, 1985; Söderström and Stoica, 1988), which has special importance in model building for control engineering purposes, is essentially a problem in high precision inferential measurement (Finkelstein, 1994). It is shown in **Article 5, Humans in the Real World, Volume 1** that inferential measurement is one of the two classical problems in measurement.

It is concerned with inferring the physical properties of a system, which characterize its physical ability to store, dissipate, transmit, or transform energy. The close symbiosis between Systems Identification and Physical Measurement, which is manifested in their many and diverse similarities, gives the identification problem a predominant relevance in measurement.

This example provides a convenient means of demonstrating the importance of a boundary perspective for systems. This view is especially appropriate, as the systems' context of identification highlights the importance of the holistic aspects of the Systems Engineering approach when applied to Instrumentation Systems (McGhee, Henderson and Sankowski, 1986; McGhee *et al*., 1996). From the vantage point provided by this view, it is possible to propose that Figure 2 is a general representative model of a process system. It is also valid to look upon this as a representation of an identification channel.

In the case of identification, the extrinsic boundary of Figure 2 may be appropriately called the *identification boundary*. Impacting upon this boundary are the demand input and the contaminating influences. These interferences can affect the system in a direct manner at input and output. Alternatively, they may be distributed throughout the system. An important class of extrinsically arising interference is that caused by disturbances and fluctuations in the various types of power flow.

4 INFLUENCES ON A REAL SYSTEM

As systems operate in a real environment, they are subjected to the full range of influences, which this environment can exert. The boundary perspective provides an important aid to the visualization of these impacts, which may be from energy or information sources. Energy can be classified in a number of different ways, as described more fully in **Article 109, Structure and Energy in Sensor Systems, Volume 2**. The classification of energy using the COMETMAN acronym, which may be regarded as complementing other energy classification schemes (*see* **Article 109, Structure and Energy in Sensor Systems, Volume 2**), is useful because it is a good memory aid. It may also be conveniently extended to the measurement system shown in Figure 2.

The inputs then become the energy and signal form of the input, the resource energy form supporting the flow of information, and the energy and signal form of the contamination. Thus, the resource and contamination classes have a possible total of eight energy form inputs each. The signal at the output of the identification channel has captured information about the ability of the system to store, transmit, transform, or dissipate energy. This information is the resource output of the channel.

The intrinsic boundary is subjected to two kinds of internally occurring disturbance or interference.

The first of these gives rise to inevitable fundamental limits, which are important in all information handling systems. If the information handling operation is measurement, then a fundamental measurement limit exists. Fundamental fluctuations occur as a result of the equipartition of energy between the thermodynamic state and the states of the other energy forms.

In mechanical systems, this is manifested in Brownian motion, while Nyquist or Johnson noise occurs in electrical systems, and temperature fluctuations occur in thermal systems – the latter being also the result of mechanical molecular motion. Although the fundamental fluctuations

have been shown at the output of the system in Figure 2, in practice they are more likely to be distributed throughout the system.

Components used to build systems are not everlasting. Aging of the elements gives rise to a change in their properties, which can affect performance. If burnout failure occurs, then a disruption of the operation of the system occurs. Parameters of the sensor system, which are instrumental in determining the performance of the system, give the system some level of parameter sensitivity. Parameter sensitivity is caused by the inexorable aging of all of the components that make up the system structure.

All of the above sources of interference, in whatever energy form they manifest themselves, must be considered in data measurement systems. Any synthesis or analysis of measurement systems that does not take them into account will almost certainly have very serious shortcomings. Interfering influences, which arise from inside the measurement system, are due to either parameter sensitivity or internal noise.

External interference is caused by disturbance power flows or by other forms of direct contamination. Highlighting the various sources of contaminating inputs exposes the limits of the ultimate resolution of a measurement system in executing its primary function.

RELATED ARTICLES

Article 5, Humans in the Real World, Volume 1; **Article 6, Substructure of Human–Machine Systems, Volume 1**; **Article 63, Outline of Systems Thinking, Volume 2**; **Article 95, Human–Machine Interface, Volume 2**; **Article 109, Structure and Energy in Sensor Systems, Volume 2**.

REFERENCES

Eykhoff, P. (1984) Identification Theory: Practical Implications and Limitations. *Measurement*, **2**, 75–84.

Finkelstein, L. (1994) Measurement and Instrumentation Science – An Analytical Review. *Measurement*, **14**(1), 3–14.

Ljung, L. (1985) *System Identification – Theory for the User*, Prentice Hall, Englewood Cliffs, NJ.

McGhee, J., Henderson, I.A. and Sankowski, D. (1986) Functions and Structures in Measurement Systems: A Systems Engineering Context for Instrumentation. *Measurement*, **4**(3), 11–119.

McGhee, J., Henderson, I.A., Kulesza, W. and Korczynski, M.J. (1996) *Scientific Metrology*, ISBN 83-904299-9-3, printed by A.C.G.M. LODART, Lodz.

Söderström, T. and Stoica, P. (1988) *System Identification*, Prentice Hall, Englewood Cliffs, NJ.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

69: Requirements Allocation

Andrew Kusiak

The University of Iowa, Iowa City, IA, USA

Fang Qin

The University of Iowa, Iowa City, IA, USA

1 Introduction 430
2 Requirements Forming 430
3 Decomposition 431
4 Solution Approach 431
5 Structural Decomposition 432
6 Module Design 433
7 System Design 435
8 Vending Machine Configuration 435
References 436

1 INTRODUCTION

Design is information-processing activity resulting in the creation of an object. An early design stage is referred to as *conceptual design* (Pahl and Beitz, 1988) – *see* **Article 64, Executing A Measuring System Design, Volume 2** and **Article 66, Phases of System Life Cycle, Volume 2**. In conceptual design, design requirements are transformed into a functional and then physical description. At an early design stage, requirements are formed and then allocated to functions. This article covers the methodology.

To date, relatively little research has been done in the area of design specifications. Kota (1990) developed a function-decomposition hierarchy to identify a set of basic design building blocks. Qualitative functional specifications and synthesis for conceptual design of microelectromechanical systems are presented in Crary and Kota (1990). A methodology for hydraulic circuit synthesis based on functional and structural design building blocks is presented in Kota and Lee (1990). Kannapan and Marshek (1990) presented procedures for specification and synthesis using algebraic and logical transformation rules. Rinderle and Finger (1990) used a graph-based language to describe the behavioral specifications of a design as well as the behavior of the available physical components. Morrell (1988) described the concept of quality function deployment (QFD) applied to improve product quality based on customer requirements.

2 REQUIREMENTS FORMING

Design *requirements are* 'demands' and 'wishes' that clarify the design task in the space of needs (Pahl and Beitz, 1988). They provide an abstraction of the design task from the most general demand (the overall requirement) to more specific demands (subrequirements). A design *function* indicates the behavior that is required for the device to satisfy a given requirement (Kota and Ward, 1990). Requirements and functions are domain specific and represent part of the knowledge base of the design system. A designer, as the needs or his/her design experience changes, can also add them.

A design task is given overall requirements that are domain dependent. They are further decomposed into a set of subrequirements. The number of levels of requirements depends on the complexity of the design task.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.

© 2005 John Wiley & Sons, Ltd.

ISBN: 0-470-02143-8.

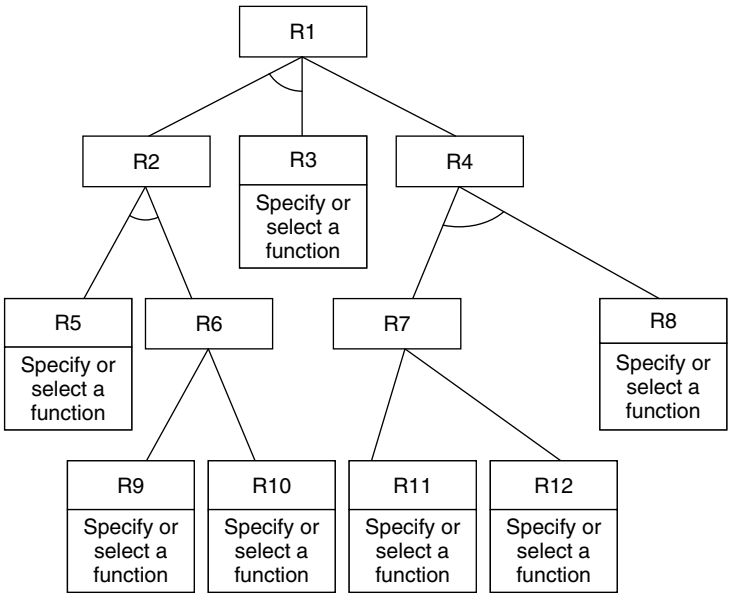


Figure 1. Decomposition of requirements.

Figure 1 shows the overall requirement R1, its decomposition into subrequirements, and the corresponding functions (Kusiak and Szczerbicki, 1992).

An arc between nodes of the tree in Figure 1 represents a conjunction. A node without an arc represents a disjunction. In the example presented in Figure 1, the overall requirement R1 is satisfied by each of the following four sets of subrequirements:

1. {R11, R8}
2. {R12, R8}
3. {R10, R5, R3}
4. {R9, R5, R3}

Each of the above sets may lead to a different conceptual design. Some such designs are discussed in Kusiak (1999).

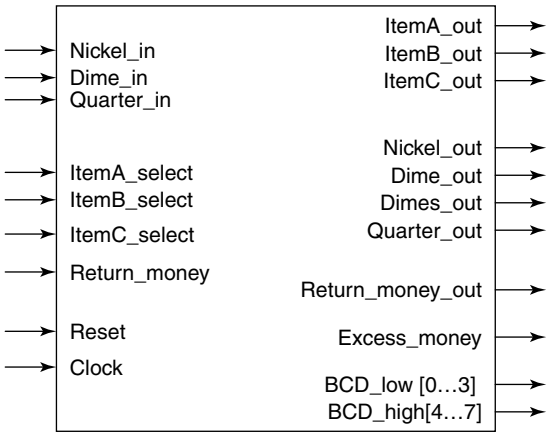


Figure 2. Model of a vending machine controller.

3 DECOMPOSITION

Decomposition is useful in analysis of requirements and functions (Kusiak 1999). It reduces the complexity of the problem at hand. The application of decomposition will be demonstrated here using the design of a vending machine controller. The controller of a typical vending machine has 9 inputs and 11 outputs as shown in Figure 2.

The vending machine needs to be designed for purchasing items, for example, A, B, C, by a providing a change and then selecting an appropriate button to get the item. If the change exceeds the price of the item selected, the vending machine returns the difference. If the Return_money button is selected, the controller will return all the coins deposited since the last transaction (Return_money_out will be activated). The machine accepts the money only up to 95 cents. If more change is inserted, then the Excess_money output variable activates a bypass chute and extra change inserted is routed directly to the change return. Two 4-bit binary coded decimal code combinations represent the current amount of money (in cents) the user has deposited in the vending machine.

4 SOLUTION APPROACH

Having formed the design requirements for the vending machine, the next step is to analyze the various Product Functions. The main functions are now discussed.

1. **Coin counting**
   The function 'count coins' of the vending machine is illustrated in Figure 3.
2. **Money return**
   The vending machine returns change if the button of Return_money is selected (see Figure 4).
3. **Excess money**
   The excess of 95 cents is returned (see Figure 5).
4. **Item selection**
   When the customer selects an item and the amount deposited in the vending machine is equal to or greater than the price of the selected item, the item is delivered (see Figure 6).
   Note: Item_select includes itemA_select, itemB_select, itemC_select.
   Item_out includes itemA_out, itemB_out, itemC_out.
5. **Change calculation**
   If the customer drops in more money than the price of the item selected, the change needed is calculated. The change equals the total amount deposited in the vending machine minus the price of the selected item (see Figure 7).
6. **Change out**
   After the vending machine has calculated the amount of change, the proper change amount is returned to the customer (see Figure 8).
7. **BCD indicator**
   The vending machine needs to indicate the current amount of money (in cents) the user of the vending machine has deposited with two 4-bit BCD code combinations (see Figure 9).

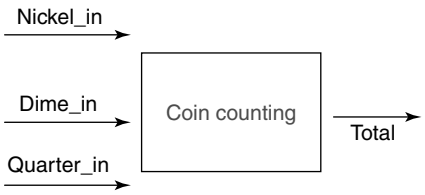


Figure 3. Coin counting.

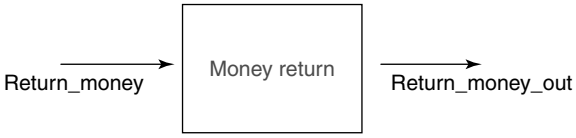


Figure 4. Return money.

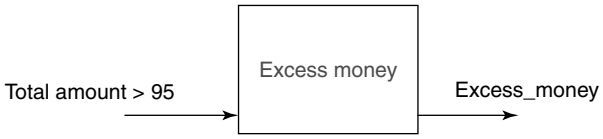


Figure 5. Excess money.

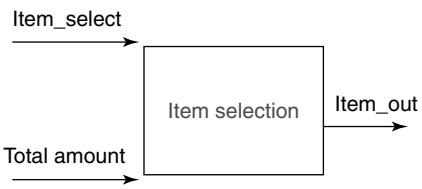


Figure 6. Item selection.

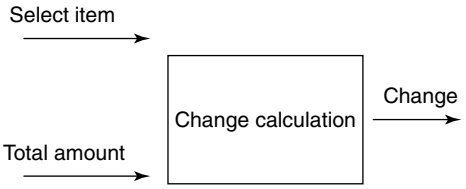


Figure 7. Change calculation.

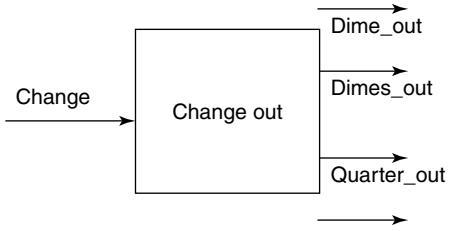


Figure 8. Change out.

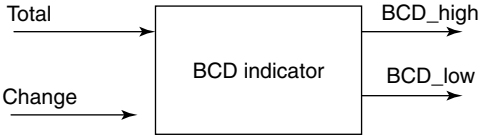


Figure 9. BCD indicator.

5 STRUCTURAL DECOMPOSITION

On the basis of the functions defined in Figure 2 through to Figure 9, the function-input/output matrix is formed (see Figure 10).

Each "*" in Figure 10 indicates a relationship between the corresponding function i, $i = 1, \ldots, 7$ and input/output j, $j = 1, \ldots, 16$.

A list of functions is obtained from the rows in Figure 10:

1. Coin counting
2. Return money
3. Excess money
4. Item selection

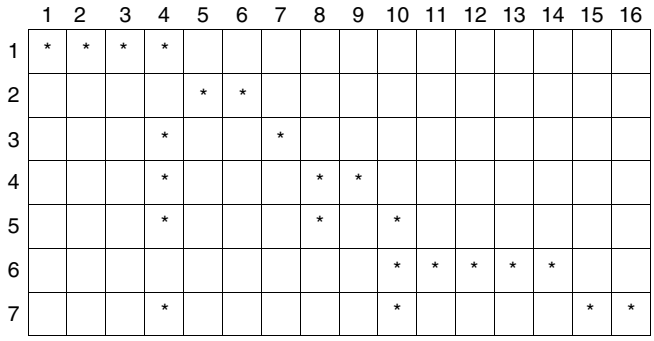


Figure 10. Function-input/output incidence matrix.

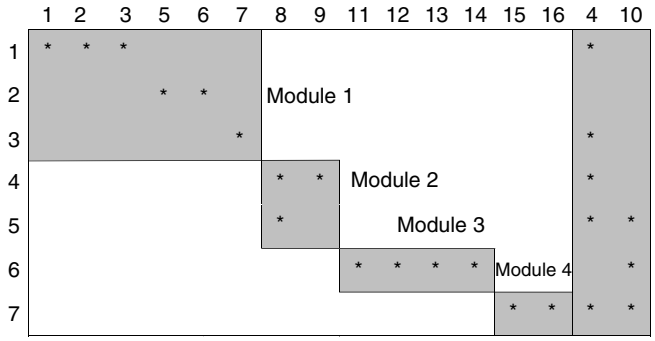


Figure 11. Function-input/output matrix of vending machine clustered with the branch-and-bound algorithm.

5. Change calculation
6. Change out
7. BCD indicator.

List of inputs/outputs (columns in Figure 10):

1. Nickel_in
2. Dime_in
3. Quarter_in
4. Total_amount
5. Return_money
6. Return_money_out
7. Excess_money
8. Item_select (includes ItemA_select, ItemB_select, ItemC_select)
9. Item_out (includes ItemA_out, ItemB_out, ItemC_out)
10. Change
11. Nickel_out
12. Dime_out
13. Dimes_out
14. Quarter_out
15. BCD_high
16. BCD_low.

Clustering the matrix with the branch-and-bound algorithm (Kusiak, 2000) results in a new matrix (Figure 11) that is more suited to design allocation. Four groups of functions – {1, 2, 3}, {4, 5}, {6} and {7} – and four groups of inputs/outputs – {1, 2, 3, 4, 5, 6, 7}, {8, 9}, {11, 12, 13, 14} and {15, 16} – are visible in Figure 11. Inputs/outputs 4 and 10 interact with more than one function.

Introducing redundant inputs/outputs for performing the functions allows decomposition of Figure 11 into the four mutually separable submatrices shown in Figure 12 (for details see Kusiak (1999)).

6 MODULE DESIGN

The four function modules are introduced to fulfill the requirements from Figure 12. VHDL logic is now used to describe the behavior of each module (Perry, 1994).

Module 1: Coin handler

A coin handler counts coins, calculates the total amount, returns money, and handles excess money according to the VHDL logic presented next and the function illustrated in Figure 13.

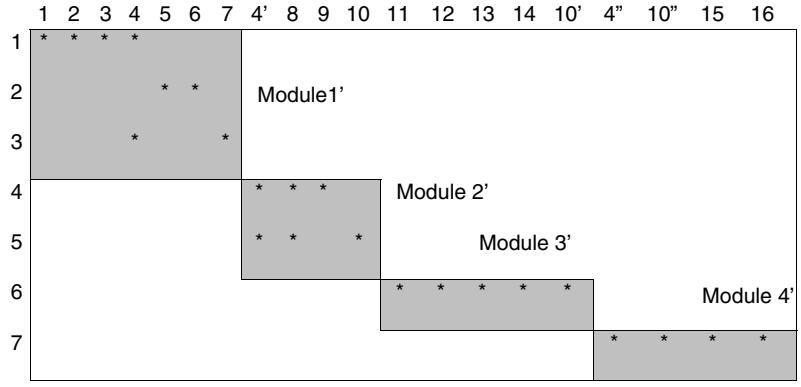


Figure 12. Structured function-input/output matrix of vending machine after introduction of redundant inputs/outputs.

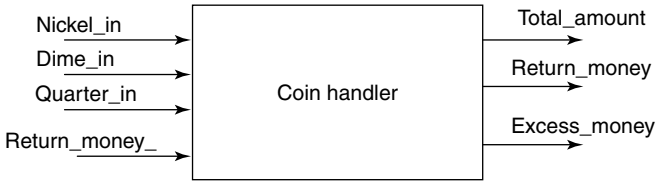


Figure 13. Coin handler.

1. **Count coins**

```
IF nickel_in = '1' THEN
  total_amount := total_amount + 05;
END IF;
----------
IF dime_in = '1' THEN
  total_amount := total_amount + 10;
END IF;
----------
IF quarter_in = '1' THEN
  total_amount := total_amount + 25;
END IF;
----------
```

2. **Return money**

```
IF return_money = '1' THEN
  return_money_out = '1';
END IF;
```

3. **Handle excess money**

```
IF total amount > 95 THEN
  excess_money = '1';
END IF;
```

Module 2: Item processor

The Item processor accepts purchase requests and determines if the money entered is enough to purchase an item (see Figure 14).

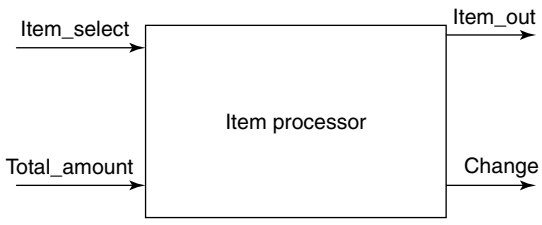


Figure 14. Item processor.

Item	Price(cents)
A	55
B	65
C	75

Figure 15. Item price list.

The price of each of the three items A, item B, and item C is listed in Figure 15.

1. **Select Item A**

```
IF ItemA_select = '1' THEN
 IF total_amount >= 55 THEN
  ItemA_out = '1';
  Change := total_amount −55;
 END IF;
END IF;
```

2. **Select Item B**

```
IF ItemB_select = '1' THEN
 IF total_amount >= 65 THEN
  ItemB_out = '1';
  Change := total_amount −65;
 END IF;
END IF;
```

3. **Select Item C**

```
IF ItemC_select = '1' THEN
 IF total_amount >= 75 THEN
  ItemC_out = '1';
  Change := total_amount −75;
 END IF;
END IF;
```

Module 3: Change maker

Change maker outputs provide change to the customer (see Figure 16).

The input/output of the change maker is shown in Figure 17:

The following VHDL statements illustrate the process of handling a change of 35 cents.

```
IF change = 35 THEN
  Nickel_out = '1';
  Dime_out = '1';
  Dimes_out = '1';
  Quarter_out = '0';
  Total_amount := 0;
  Change := 0;
END IF;
```

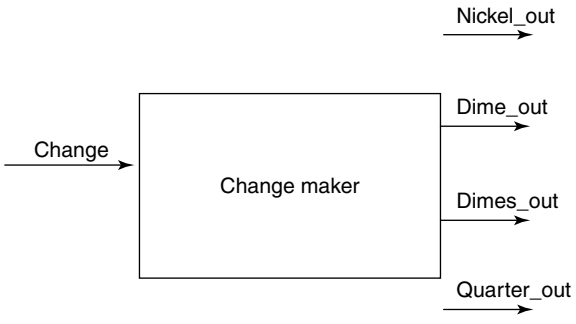


Figure 16. Change maker.

Change (cents)	Nickel_out (5 cents)	Dime_out (10 cents)	Dimes_out (20 cents)	Quarter_out (25 cents)
0	0	0	0	0
5	1	0	0	0
10	0	1	0	0
15	1	1	0	0
20	0	0	1	0
25	1	0	1	0
30	0	1	1	0
35	1	1	1	0
40	1	1	0	1

Figure 17. Input/output table of the change maker.

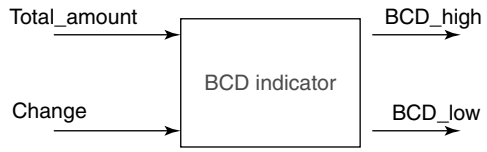


Figure 18. BCD indicator.

Module 4: BCD indicator

Two 4-bit BCD code combinations represent the current amount of money (in cents) the user of the vending machine has deposited. Before purchase, the amount equals the total amount, after purchase, the amount equals the change (see Figure 18).

The input/output of change maker is presented in Figure 19.

Suppose the amount is 35, the following VHDL sequence is used:

```
IF amount = 35 THEN
  BCD_LOW = ''0101'';
  BCD_HI = ''0011'';
END IF;
```

7 SYSTEM DESIGN

7.1 Module interaction

From Figure 11, it was determined that input/output 4 (total amount) and input/out 10 (change) interact with multiple modules. The relationship of the modules is obtained from Figure 11 and shown below. Thus, the inner connections of modules are established (see Figure 20).

Amount (cents)	BCD_high	BCD_low
0	0000	0000
5	0000	0101
10	0001	0000
15	0001	0101
20	0010	0000
25	0010	0101
30	0011	0000
35	0011	0101
40	0100	0000
45	0100	0101
50	0101	0000
55	0101	0101
60	0110	0000
65	0110	0101
70	0111	0000
75	0111	0101
80	1000	0000
85	1000	0101
90	1001	0000
95	1001	0101

Figure 19. Input/output of BCD indicator.

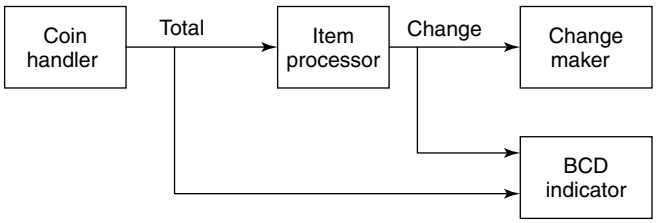


Figure 20. Inner connections of modules of vending machine.

8 VENDING MACHINE CONFIGURATION

After the module design is finished and the inner connections established of modules, the top-level vending machine configuration is obtained. Clock and Reset are added to synchronize the operation of each module (see Figure 21).

This process has systematically worked its way through the many options. By reorganizing the relationship matrix, several units have been identified that have less interfaces to other units than was first developed. With less interfaces, the chance of system failure is reduced and design made simpler. Measuring systems can similarly be broken down in this manner. More information on the nature and scope of requirements is found in Sage and Rouse (1999) and Sydenham (2004).

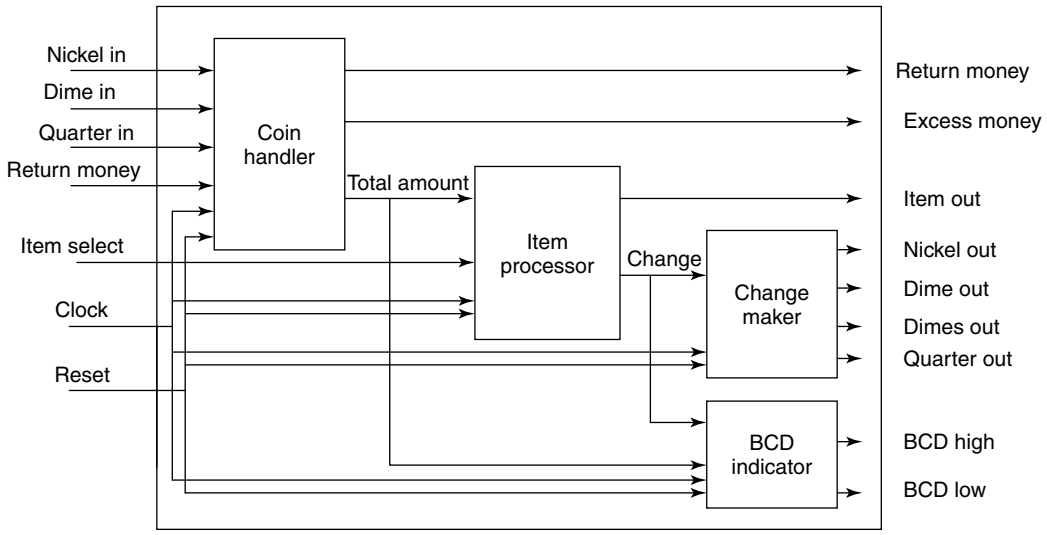


Figure 21. Vending machine.

REFERENCES

Crary, S. and Kota, S. (1990) Conceptual Design of Micro-Electro-Mechanical Systems, in *Proceedings of the Micro Systems Conference*, Berlin.

Kannapan, S.M. and Marshek, K.M. (1990) An Algebraic and Predicate Logic Approach to Representation and Reasoning in Mechanical Design. *Mechanism and Machine Theory*, **25**, 335–353.

Kota, S. (1990) Qualitative Motion Synthesis: Towards Automating Mechanical Systems Configuration, *Proceedings of the NSF Design and Manufacturing Systems Conference*, Society of Manufacturing Engineers, Dearborn, MI, (pp. 77–91).

Kota, S. and Lee, C.-L. (1990) A Computational Model for Conceptual Design: Configuration of Hydraulic Systems, *Proceedings of the NSF Design and Manufacturing Systems Conference*, Society of Manufacturing Engineers, Dearborn, MI, (pp. 93–104).

Kota, S. and Ward, A.C. (1990) Functions, Structures, and Constraints in Conceptual Design, in *Proceedings of The 2nd International Conference on Design Theory and Methodology*, (ed. J. Rinderle), Chicago, IL, (pp. 239–250).

Kusiak, A. (1999) *Engineering Design: Products, Processes, and Systems*, Academic Press, San Diego, CA.

Kusiak, A. (2000) *Computational Intelligence in Design and Manufacturing*, Wiley, New York.

Kusiak, A. and Szczerbicki, E. (1992) A Formal Approach to Specifications in Conceptual Design. *ASME Transactions: Journal of Mechanical Design*, **114**, 659–666.

Morrell, N.E. (1988) Quality Function Deployment; Disciplined Quality Control. *Automotive Engineering*, **9**, 122–128.

Pahl, G. and Beitz, W. (1988) *Engineering Design*, Springer-Verlag, New York.

Perry, D. (1994) *VHDL*, McGraw-Hill, New York.

Rinderle, J.R. and Finger, S. (1990) A Transformational Approach to Mechanical Design Synthesis, *Proceedings of the NSF Design and Manufacturing Systems Conference*, Society of Manufacturing Engineers, Dearborn, MI, (pp. 67–75).

Sage, A.G. and Rouse, W.B. (eds) (1999) in *Handbook of Systems Engineering and Management*, (ed. A. Sage), Wiley, New York.

Sydenham, P.H. (2004) *Systems Approach to Engineering Design*, Artech House, Boston, MA.

70: Measuring System Design Methodologies

Ludwik Finkelstein
City University, London, UK

1 Measuring System Design Task 439
2 The Design Process 440
3 Design Concept Generation 441
References 442

1 MEASURING SYSTEM DESIGN TASK

The design of a measuring system is the transformation of knowledge about the measurement task to be performed into a specification of a system that can perform the required task in such a way that the system can be realized.

The methodology of measuring system design is determined by the fact that measurement is an information process, measuring instruments are information machines, and that measuring instrumentation can be treated for the purposes of analysis and design as a system.

To discuss the principles of design of a measurement system, it is convenient to do so on the basis of a simple model of such a system – see Figure 1.

Measuring systems are invariably a subsystem of a large system. The measuring system consists of a number of subsystems. There is firstly the system under measurement (also called the *system under test* SUT). This is connected to a sensor system and acts on the sensor by a flow of matter or energy.

The sensor converts this flow into a signal, maintaining a functional relation between the input flow and the information carrying characteristics of the signal. There is usually a signal-conditioning block that converts this signal into a symbol that may be conveniently handled by the following block, which performs any required functions of information transformation and communication.

This system passes the information to the effector block, to further processing, or to the human operator. The measuring instrument system operates under the control of a control block. An important part of the system is the operator or human–machine interface. Through this interface, the operator effects supervisory control of the measurement process. The interface embodies also any displays.

Measuring system analysis and design uses description by mathematical or like models – *see* **Article 71, Modeling Methodology, Volume 2**; and **Article 104, Models of the Measurement Process, Volume 2**.

Models involve abstraction, that is, the omission of irrelevant detail. This is required to reduce complexity, ensure tractability, and give access to analytical tools. In the case of instruments, we may distinguish a number of levels of abstraction in increasing levels of abstraction:

1. Embodiment models in terms of the physical variables and the shape, dimensions, and material properties of the system components and their physical interconnection.
2. Models of flow of power and power transformation; models of signals and their transformation.
3. Models of flow of symbols and of symbol transformation.
4. Models of information and knowledge flow, that is, representation and transformation of information and knowledge by the system.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

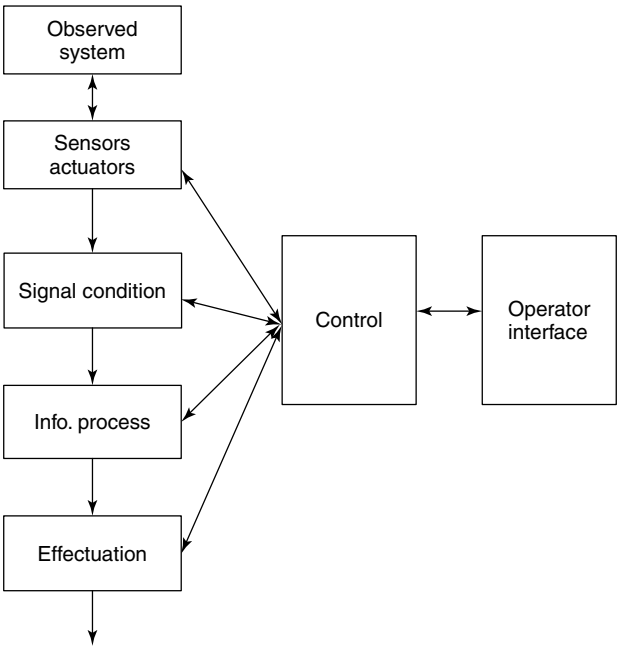


Figure 1. Measuring system.

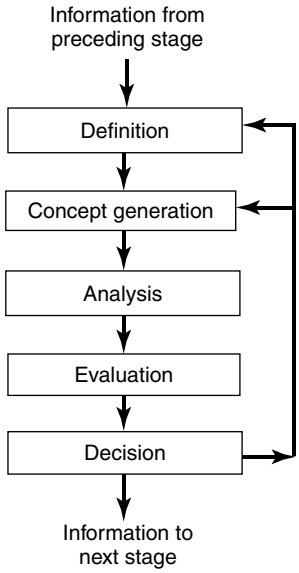


Figure 2. The design process.

We may also distinguish between functional models that represent only the transformation of input into output and embodiment models that represent some details of the physical realization of the system.

In modern measuring systems, the information processing block, the control block, and to some extent the effectuation and signal-conditioning blocks are designed and implemented by the standard methods and equipments of information technology.

In general, functional blocks for performing the required operations are available. Special problems arise in the sensor block and in the interaction between the measuring system and the observed system. They need to be designed at the embodiment level.

2 THE DESIGN PROCESS

The design of measuring systems follows the classical model of the design process.

Classically, design activity is represented as a process consisting of a sequence of stages, starting at the perception of need and terminating at the communication of the final firm knowledge necessary to implement or realize the artifact, or system. Each stage is itself an elementary design process, which starts with an initial concept and refines that concept. The stage is a sequence of steps, subprocesses, or operations.

Consider the elementary design process depicted in Figure 2.

The initial step of the process is that of definition of an initial design concept to be refined. The model and related knowledge is, in general, provided by the previous stage of design. Associated with this model is a set of evaluation criteria, or a value model, which express the degree to which the concept meets the requirement.

The next step in the process is the generation of a more refined candidate design concept. The generation of candidate design concepts is central to the design process. The candidate design concept is represented as a configuration, that is, in terms of attributes of its construction.

It is next analyzed to determine its performance attributes and evaluated using the value model previously established. A number of such design concepts are normally generated and stored.

The evaluation step passes information to a decision step in which the candidate design concept judged to be the most satisfactory is accepted for implementation, or as the initial concept on which the next stage of design will be based.

If none of the candidate design concepts is acceptable, it may be necessary to return to an earlier stage in the process, for example, to alter the value criteria, or to generate more candidate design concepts.

Viewed overall, design proceeds from a global view of the artifact or system to progressively more localized considerations, and from abstract and fluid descriptions to concrete and firm ones.

The design stages typically start with a definition of the overall system that arises from the original requirement. This establishes the main features of the system as a set of interconnected subsystems and also fixes the nature and

specifications of the subsystems. In turn, the subsystem design stage uses this information to establish the main features of the subsystem and the nature and the specifications of its simpler components.

This sequence of stages proceeds to the design of elementary components. If at any stage it appears that it is not possible to meet the specification generated by the preceding stage, it is necessary to return to an appropriate point of that latter design stage.

3 DESIGN CONCEPT GENERATION

The core of the design process is the generation of design concepts; its basis is decomposition and abstraction.

The components of a measuring system that handle knowledge and information are generally realized by the selection of available functional blocks and the use of essentially algorithmic methods of design. The design methodology of those parts of the system is not specific to instrumentation and is basically that of information technology.

In the case of sensors and the interaction between the sensors and the object under observation, it is commonly necessary to find an appropriate working principle and a physical embodiment of the principle.

There are a number of basic methods of determining such working principles and finding their embodiments. They are useful heuristics rather than algorithmic procedures.

3.1 Use of established equipment and design concepts

Established design concepts may be used in the design of a system. They may be retrieved from systematically organized knowledge bases and tools for searching them.

Formal knowledge bases of sensor principles may be organized on the basis of their mathematical models or in terms of their function.

Catalogs of commercially available sensors and directories of such catalogs are increasingly available on the Internet.

Searching through literature, also increasingly facilitated by the Internet, is an important method.

3.2 Design concepts deduced from the relevant physical effects

Instrument components that maintain a functional relation between physical variables may be deduced from the relevant physical effects.

One approach is the systematic listing of physical laws that involve a physical variable of interest, or that relate two physical variables that are to act as the input and output of a device. For example, if we require to generate concepts for sensing electric current, we may list the physical laws involving electric current, each of which may suggest a design concept. One law is that of force between current-carrying conductors, which suggests that we can sense current by sensing the force between two coils, as in a dynamometer instrument. The law of force on a current-carrying conductor in a magnetic field suggests the principle of the moving-coil ammeter. Ohm's law relating voltage and current suggests sensing current by sensing the voltage across a standard resistor. Joule's law of heat generated by a current in a resistor suggests the principle of the hot-wire ammeter. This list is illustrative but not exhaustive.

The second approach is the examination of a single law involving the variables in question that then suggests a design concept. Consider the generation of design concepts for instruments to measure viscosity. The relevant law is $F = -A\ (\mathrm{d}v/\mathrm{d}x)$, where F is the force acting on an area A normal to the velocity gradient $\mathrm{d}v/\mathrm{d}x$ of a fluid. This suggests two viscometer principles:

- generating a velocity gradient in a fluid and sensing the resultant shearing force;
- generating a shearing force and sensing the resultant velocity gradient.

From this we can proceed sequentially to elaborate variants of the principles. Taking two solid bodies with fluid between them, we may have the two bodies fixed with fluid moving, or one body fixed with the other body moving, or all may move. The mode of motion may be translation or rotation, and the motion may be steady, harmonic, or transient. We may then derive the idea of two coaxial cylinders with fluid between them. We may drive one cylinder (inner or outer) at constant angular velocity and measure the torque on either moving cylinder (rotor) or stationary cylinder (stator). Alternatively, we may drive the rotor at constant torque and measure the resultant velocity of stator or rotor. This systematic process then suggests oscillatory motion of the rotor. A complete range of variant design concepts is thus generated.

3.3 Analogies

A method of generating design concepts is the consideration of analogies. When considered in terms of an abstract description, a system may seem to be essentially similar to another and aspects of the second may suggest design solutions for the first.

Physical analogies such as those between electrical, mechanical, fluid flow, and thermal systems are often useful.

Another important source of analogies for instrumentation and information processing are living organisms, especially the animal system of receptors, nerves, brain, and effectors.

3.4 Divergent or lateral generation of concepts

Design concepts can be generated by creative thinking in which ideas diverge and move laterally. Design concepts are considered in abstract form, neglecting irrelevant details until the latest possible stage. Every attempt is made in the first instance not to consider an established practice, authority, or apparent obstacles so as not to allow promising concepts to be abandoned prematurely. Finally, for the same reason, idea generation and evaluation are separated.

3.5 Transforming concepts

Some existing design concept usually forms the first step in the process of generating a design. This may then act as a starting point for generating new concepts by transformation. Starting with a particular design, one may proceed to derive new concepts.

The basic steps are the functional description and decomposition. The instrument is considered as a system of components and each component is described in terms of the functions it performs. Variations are then derived by systematic analysis of the concept and its systematic variation to generate new options.

Foundational material on engineering design is available in Finkelstein and Finkelstein (1983, 1994) and Sydenham (1994, 2004).

REFERENCES

Finkelstein, L. and Finkelstein, A.C.W. (1983) Review of Design Methodology. *Proceedings of the IEE*, **130**, Part A, 213–221.

Finkelstein, L. and Finkelstein, A.C.W. (1994) Instruments and Instrument Systems: Design Principles, in *Systems and Control Encyclopedia*, Vol. 4 (ed. M.G. Singh), Pergamon Press, Oxford (pp. 2533–2536).

Sydenham, P.H. (1994) Disciplined Design of Sensor Systems. *Measurement*, Pt 1 Types of Design, **14**, 73–80; Pt 2 Knowledge Based Systems, **14**, 81–87.

Sydenham, P.H. (2004) *Systems Approach to Engineering Design*, Artech House, Norwood, MA.

71: Modeling Methodology

Peter H. Sydenham

GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Models Relevant to Measurement Systems 443
2 Types of Modeling Representation 445
3 Modeling Inside Blocks 446
4 Process for Creating Computer-based Models 446
5 Setting up the Architecture 447
6 Identifying Equations 448
7 Assignment of Model Coefficients 449
8 Exercising the Model 449
9 Limitations of Models 450
10 Model Validation 450
Related Articles 451
References 451

1 MODELS RELEVANT TO MEASUREMENT SYSTEMS

Designers will invariably need to develop some form of model of the system or service being developed or operated (Neelamkavil, 1987; Holland, 1996). Models enable the system to be studied in the computer-based virtual domain. They allow the designer to study the behavior of a system before it is physically built. The place of measurement in the life cycle of a system is shown in Figure 1.

No one modeling tool suits all needs. There exist many factors of the system that will decide which particular tool has to be used. The first to consider is what basic kind of model is appropriate.

Three types of models are used. They will be based on one or more of the following:

- laws of physics – formal description
- process involved
- hardware artifacts – called *icons*.

1.1 Physics-based formal models

In these, a real-world activity, existing or not yet existing, is represented by the formal laws of physics that are expressed in terms of mathematical equations. The ideally desirable level of representation of the complete enough real world open system requires extensive mathematical development that is not always available to the degree of completeness needed. Corners are often cut by making unproven assumptions that can be overlooked later to the peril of the user.

Much of engineering practice has been modeled by this physical means, but it has been long recognized that not all systems aspects can be handled by use of 'hard' formal descriptions. Adding in empirical data is sometimes needed. Examples of this are found in the flight dynamics of aircraft, temperature distribution on an electronic circuit board, wing lift, and other fluid-flow designs.

1.2 Process-based models

Here, the modeled domain is represented by a set of heuristic and formal rules embedded into an algorithm or flow diagram. This 'softer' thinking approach can sometimes be superior to the hard-science model in terms of applicability and can often yield a model where the nature of the problem totally defeats formal description.

Examples of these types of models are the queuing of traffic in a road system, a logistics supply system in

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.

© 2005 John Wiley & Sons, Ltd.

ISBN: 0-470-02143-8.

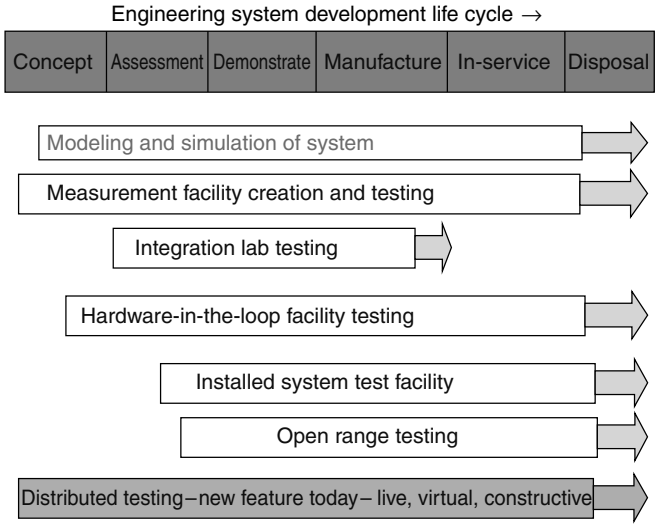


Figure 1. Place of measurements in the systems development life cycle.

operation such as loading a cargo ship in best time, flows in a manufacturing facility, and an Expert System control room support system.

This type can also accommodate the softer human issues – but not that well – using the class of problems called *Systems Dynamics* (SD).

1.3 Iconic models

An icon is an object representing something. It can be an image or a 3-D object. It is used to represent the physically existing model of something.

Examples abound: a scale model of an automobile used in a wind tunnel; a full-size aircraft cabin mock-up; and a scaled-down 3-D model of a large rainwater catchment area.

Being largely hand-crafted, these

- are costly to make;
- take a long time to build;
- are easily destroyed or damaged in testing;
- cannot be altered with ease;
- need special crafts skills to build;
- can only represent a small range of conditions;
- require costly testing instrumentation installations.

Once commonplace, they are rapidly becoming a thing of the past as the economic means for supporting design. Computer simulation is overtaking the need for these, for that alternative is often superior to an iconic model.

Furthermore, the computer model can also be placed in a virtual world environment to put the system under virtual test. In this domain, little is physically built.

For this reason, there is now a diminishing demand for physical model-testing facilities. Reduced demand pushes up costs, which can render them unaffordable. Some large wind tunnels, for example, have been closed down for lack of sufficient business. Physical test facilities are now used more for periodically testing a model under development and for application in design optimization.

Setting up a mathematical model to represent the simple first-level principle of the issue under investigation, as represented by known laws of physics, is rarely sufficient for engineering design use. Useful models usually must also allow for the numerous second-level effects caused by the imperfections of real materials, design inadequacies, and external influence effects. These greatly extend the model's sophistication. Models rapidly expand to be complex, needing ever-increasing effort to establish the particular equation-coefficient values that the model needs to run.

Over time, these have, however, been developed for some industries to the point where they can adequately represent the need – but at a cost that is not always evident, for much of the development has come from past projects and before that from university style research. Sophisticated models shown as selling demonstrator applications do not reveal how much effort has been expended in their development!

Some regimes have been particularly good for modeling realistic systems. One example is that of electronic circuitry where sophisticated models and tools can deliver an almost complete virtual development environment. Here it is often possible to design the systems in the model domain – few changes will be needed after it is made. Usually no prototype build is needed to tune the system.

Now, computational fluid dynamics (CFD) and mechanical and manufacturing design also have tool support for the design and development that needs minimal final physical testing.

Thermal modeling of instrument circuit boards and cases is also highly developed today.

With use of the tools, the new knowledge gained is fed back to continuously improve the model. The successful development of many quite remarkable tool packages may give the impression that all engineering design can, and should be, model based. That is a fine aspiration, but there are many areas where such support is coming slower for reasons of complexity not allowing adequacy of modeling or, perhaps, because the industry sector is too small to support the costs involved in developing the special tools it would like to use. This is the situation for much of the measuring systems design. For example, it is affordable, indeed essential, to develop models of a spacecraft instrument system but it is far less so to fully model a custom-designed single, weighing platform.

In the measuring system field, the problem in tool use is that the particular class of design need often cannot support the development of the sophisticated tools needed. Derivate use of tools built for other uses is usually necessary.

Modeling of the formal part of the information technology (IT) aspect of a measuring system is, however, well supported with virtual instrument modeling tools like LabView – *see* **Article 105, Modeling with LabVIEW™, Volume 2**.

Most design regimes – mathematical, electrical, mechanical, optical, chemical, thermal, and vibrational – have excellent tools available. However, the mixed regime use needed for many measuring system investigations is often complicated by the need to integrate the findings of several tools that are not directly compatible. It will be necessary to repeatedly transfer datasets between tools in instrument design. The time taken to become truly skilled in the use of a tool, and its cost, often limits its use to the higher-volume runs – that are not characteristic of measuring systems.

1.4 Aim for simplicity in model building

Models should be kept as simple as is just necessary to extract the information needed.

Models are NEVER a complete representation of a topic. They are built to exhibit chosen aspects of behavior and will be set to certain limits of dynamic performance.

Static, or steady state, models have their uses, but the most used are those that allow the dynamic behavior of the modeled system to be investigated – they are much more complicated.

If the external environment conditions are changed from those for which the model was designed, it is not likely to give accurate outcomes.

2 TYPES OF MODELING REPRESENTATION

Formal models of real-world systems can be formed by virtue of the fluxing entities of a design situation. The entity of interest will either flow as a

- distributed continuous flow (mechanical stresses in the skin of an aircraft; weather-flow patterns; pollutants entering a water channel; thermals in an electronic circuit board) or
- channeled flows, wherein the 'substance' flows in distinct channels and/or with varying discrete arrival times (manufacturing materials flow; public transport movements; digital control system; alarm instrumentation system).

Complex situations will often be formed as hybrid combinations of the above.

Flows of both types in a formal model can be formally represented by different descriptions of behavior:

- Deterministic
  - Future behavior is formally predicted from knowledge of the model on a single-point basis.

For example, given the laws of heat flow, the temperature rise of a thermometer allows upcoming values to be calculated at that point in the system space.

- Stochastic
  - Future behavior can also be predicted, but only for the behavior of a group of values having a known statistical relationship.

For example, the height of the individual vehicle driver who will next use a truck driver's seat cannot be predicted, but the range of heights for the set of likely drivers can be predicted.

- Chaotic
  - Behavior appears to be random but does not follow well-known stochastic laws; it appears to be haphazard, that is, chaotic. Mathematical expressions can handle some forms of chaos using different means from the above two behavior descriptions.

For example, work activity in a system engineering process often follows process model descriptions similar to that of a pile of sand grains that is sliding down from the top – that has been formalized as sand-pile chaos.

To make a model perform, they need a stimulator (driver, forcing function, excitation input) to make them represent a particular situation. The resulting system behavior depends both on the characteristics of the stimulator and the model.

Stimulators take several forms:

- Time driven
  These are the equations given as a function of time, such as heat transfer as an item is heated. A suitable time interval and discrimination is required to use the model. The time variable can be speeded up or slowed down in the formal mathematical models, which is not always possible in physical prototypes. This is useful for prediction of future behavior and for understanding events that are seemingly stationary or are arising very fast.
- Event driven
  These lead to a process-orientated model, such as operators conducting an assembly operation in which their task has one kind of statistical variability and the arrival of the parts they need has another.

3 MODELING INSIDE BLOCKS

Models of a system, or part of it, can be in a black or white box form.

3.1 Black box models

Here, the internal working of the model is not necessarily operating in the same way as the internals of the modeled system do. The output/input relationship is, however, a faithful representation of the characteristics of the block's ports as it would be in the real thing.

An example of this is the use of a simple low-pass filter model to represent the torque-speed characteristics of an engine in an overall automobile performance model. Here, the output/input does represent the torque behavior of the engine as a whole, but with this model, it is not possible to investigate such things as thermal behavior and fuel flow of the engine.

3.2 White box models

In this form, the internal workings are transparent and are modeled as a replica of the internal functions. Each subsystem, down to an appropriate level, is formed of white boxes that simulate true internal operation. These can be much harder to build than the black box form and are often not needed. For instance, to model an automobile engine as a fully white box model would be the most complex task, for the overall model would have to cover numerous energy regimes in a fully integrated manner.

Care is needed to appreciate which type of box is being used in a model. When constructing a model, it is necessary to look well ahead to ascertain whether white box modeling will be needed at a later stage of development. Starting with a black box model realized by, say, system identification methods cannot be delved into to give reticulated behavior of internal behavior.

Several methods exist for building the formal mathematical description of models. These are then converted into a suitable form for digital computation. The methods are

- equation – differential equations (DEs)
- finite element – (FE)
- discrete – nodal.

Engineering system or component models using DEs are based on there being available an in-depth understanding of the dynamic behavior, on the basis of the laws of physics that can be expressed in DEs.

Such models in engineering often have to use 'fiddle' factors to facilitate the modeling when the sufficient understanding is not available due to lack of time to develop more knowledge or where no explanation can be realized. For example, in fluid-flow modeling, various 'relationship' numbers are used, such as Reynolds Numbers.

It is highly useful that the formal descriptive equations for electrical circuits also are the same as that for the mechanical and the acoustic regimes. This gives rise to what is known as the method of the analogies. Basically, the DE set for each describes the same simple harmonic motion (SHM) behavior.

Real systems often involve a mix of energy regimes. For example, an audio loud speaker has a mechanical cone suspended by an elastic mount being driven in and out of the magnet assembly by an electromagnetic moving coil actuator. The air column acting on the speaker cone dampens motion. Thus, provided all movements are operating in the linear mode, the system exhibits SHM. Similarly, an automobile suspension system is formed with a coil spring and thus has energy storage in the spring and in the mass of the suspended structure, and there is damping of the energy effects with the shock absorber. This is also a SHM system. Thus, linear SHM systems can all be modeled with the same set of equations, provided the appropriate coefficients are used. It also allows one regime to be modeled with a set up in another. Equivalent electrical Resistor–Inductor–Capacitor (RLC) models are usually used as they suit the regime of operation of the electronic computer.

Given certain system assumptions (or some trickery), it is possible to write formal equations that relate the behavior of most energy regimes into one set of coherent integrated differential equations. This is the basis of many modeling tools. Details of SHM system behavior are covered in **Article 59, Introduction to the Dynamic Regime of Measurement Systems, Volume 1**; **Article 60, Zero-order System Dynamics, Volume 1**; **Article 61, First-order System Dynamics, Volume 1**; and **Article 62, Second-order System Dynamics, Volume 1**.

4 PROCESS FOR CREATING COMPUTER-BASED MODELS

There is a clear-cut process for developing a formal, and to some extent other forms, model. The following key steps need to be undertaken with care and quality consideration (refer to Figure 2).

1. Study the situation and all known requirements in order to ascertain the particular features that decide the characteristics of the type of model needed, as are described above.

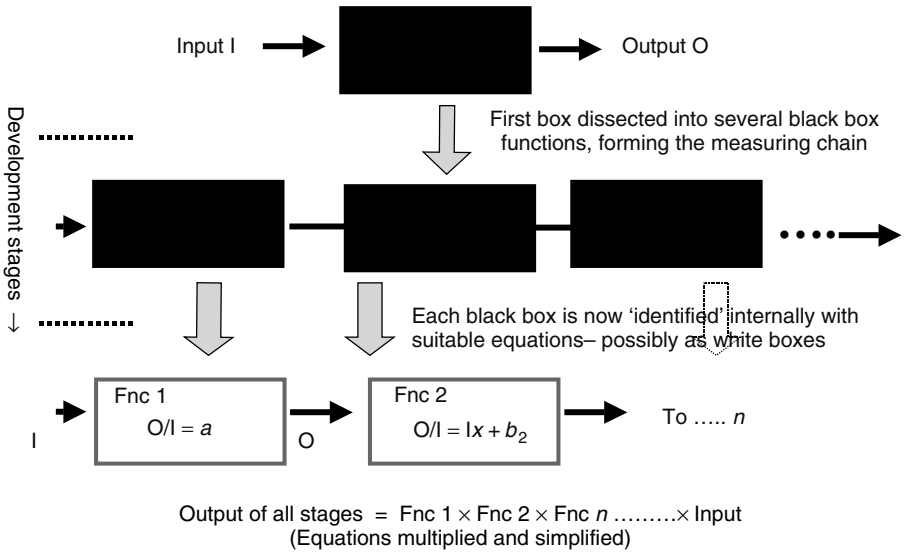


Figure 2. Simplified development of an identified model of a measuring instrument.

2. Decide the limits of the situation for modeling the subject as an adequately closed system – *see* **Article 64, Executing A Measuring System Design, Volume 2** and **Article 68, Setting the System Boundaries, Volume 2**.
3. Set up the architecture of the system in terms of the fluxes of energy or mass, process steps, or discrete activities.
   Consideration about which modeling tool to use commences at this stage. The final choice of model and tool may well take place as the result of many trial starts in which the features of a chosen tool become clearer to the user and are seen to be not suitable after all.
4. Assign blocks and their connectivity to system functions that can be described in terms of black or white boxes.
5. Assign a suitable equation for the output/input relationship of each box.
6. Assign suitable values to the coefficients to characterize the equations of the boxes for the task in question.
7. Integrating the whole equation set into one total equation may be useful here. This may be done automatically for the user in a modeling tool.
8. Set in the boundary conditions to limit the operation to that of interest.
9. Set up the appropriate stimulating functions, their coefficients, and initial conditions.
10. Run calibrating tests to exercise the model, using stimulations that have well-known and proven outcomes.

The model is ready to use in an investigative manner once all of these issues have been decided.

Many general-purpose formal mathematical tools support the above process – examples are MatLab™, LabView™, Ithink™, and so on.

Specialized tools are often available for particular applications.

5 SETTING UP THE ARCHITECTURE

Generally, measuring system architecture is developed mainly by the intuition driven by one's past experience of how things were done before and from the knowledge of new technologies and marketed 'block' products.

It is important to also decide where in the hierarchy of models does the one being developed sit (Davies, 1996). The integration of models in a hierarchical architecture needs addressing early in development as this can determine key characteristics needed.

Although measuring systems are usually thought of as small subsystems sitting inside a much larger whole system, some are very large indeed. For example, the model of a measuring and simulation system for testing aircraft warfare engagement can demand the use of numerous sensors and major computing. Figure 3 shows one such system.

A defined and methodical way to generate the architecture of a measuring system model is to apply the energy matrix method – *see* **Article 112, Systematic Description of Sensors, Volume 2** and **Article 110, Signal/Energy Matrix Modeling, Volume 2**. This is based on the fact that most measuring systems are formed as a chain of energy conversion stages that progressively convert the measurand

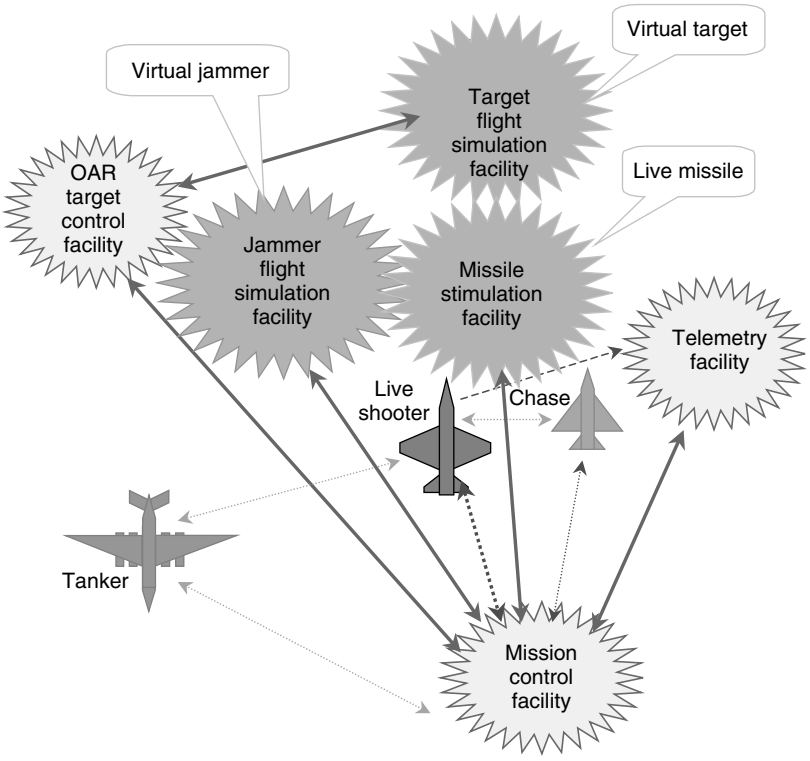


Figure 3. Example of a major measurement and simulation system.

of interest into the usually needed electrical entity required for the following communication and processing system.

Each conversion in the serial chain can usually be treated as a black box. White box models of each will be used where appropriate, but for much of measuring system modeling in support of design, black box representation suffices.

For each box, a suitable dynamic (or static in the simpler cases) describing equation can be assigned to obtain, by overall consolidation, the whole measuring system dynamic performance model.

6 IDENTIFYING EQUATIONS

Where the box behavior is to be selected by design – as is often the case in measuring system design – it is wise to keep them linear in operation because that allows easier and more accurate modeling. If nonlinearity exists, the SHM models cannot be used, except in piecewise modeling operations. Modern modeling tools can handle nonlinear components, but that is not 'keeping it simple'.

For each box, examine the known performance and structural and physical characteristics of the thing it represents. From this study, it is usually clear which type of mathematical description will adequately represent the operation of the box. An exact equation match is not usually necessary and not feasible.

Assign to each box a suitable equation type. Table 1 shows the simplest SHM equations used for modeling measuring systems stages.

For example, consider the design situation wherein force is being converted by use of a load cell on which a strain gauge is mounted to monitor the change in strain of the calibrated mechanical ring. This system clearly possesses

Table 1. Equation forms to select for a linear model stage.

Equation type	Example equation	Measuring example	Comments
Zero-order linear SHM system	a	R-only circuit; amplifier, or stage, gain or attenuation	Fixed value in any situation
First-order linear SHM system	$bx + a$	R–C or R–L circuit: temperature change on heating or cooling	Rise, or decline, without any oscillation, in exponential manner
Second-order linear SHM system	$cx^2 + bx + a$	Weighing scale spring balance; R–L–C circuit	Some level of oscillatory motion, depending on degree of damping

mass, springiness, and a low level of damping so that it can to be modeled with the second-order SHM equation. At the early stage of development, the higher second-order level is selected, ruling it into a lower order when the coefficients of the real situation show that simplification is sufficient.

In some cases, it is immediately clear that a constant (e.g. an amplifier working only in its constant gain frequency range), or first-order (e.g. a temperature sensor immersed in a fluid) description will suffice.

If the describing equation is already known, then use it after reduction to the simplest form that suffices. An example of this availability is the dynamic describing equation often quoted on data sheets for operational amplifiers and instrument motors.

In many modeling tools, a library of linear and nonlinear functions and statistical distributions is provided for selection in mathematical, tabular, or graphical form. The need to appreciate the mathematics has been lessened by these tool features, but one still needs to appreciate what is being used from a foundational viewpoint lest the wrong ones be selected. A safeguard is that good testing of the eventual model will reveal faulty versions – but at that stage, rework can be tedious and expensive.

7 ASSIGNMENT OF MODEL COEFFICIENTS

When all blocks have equation forms assigned, it is time to characterize the model. This process identifies the value of several coefficients of the equations. For example, a spring-mass-damper system of the weighing platform unit will possess gain, damping factor, and resonant frequency, each of which will need to be identified in numerical terms.

When these are assigned, the system should be evaluated by inspection to verify if this higher-order description is still needed. The degree of damping, or the resonant frequency, could well be so high that the system can be modeled with a first-order model, or even zero-order model, thus simplifying the model.

When all blocks have appropriate equation types and coefficients assigned, the serial chain of equations is multiplied together to arrive at the overall model of the specific measuring system being modeled.

Again, inspection of relative coefficients may well show that some stages are insignificant compared to the others and that they can be approximated with a lower-order model.

With modern electronic components and systems, the following communication and processing will often be of zero order (i.e. no phase shift) and, thus, not deteriorate the overall dynamic performance. This always needs to be verified as such, for latency and cable-filtering effects may be significant.

It is best to reduce the forms in the above sequence. Seeking reduction after all equations have been integrated may make it less clear as to which blocks are less significant.

As an example, in the design of the serial set of blocks for an electromechanical seismic recorder system (an inherently low-pass system), the electronic amplifier response will usually be well below its cutoff frequency, meaning that it can be modeled as a constant (zero-order system). In contrast, in the case of an acceleration sensor (which is a high pass system), the amplifier response may be critical and it must be modeled as a second-order system,

7.1 Initial conditions for models

All models need to have their initial conditions set correctly when they are started up. This usually means setting conditions to zero, but it is not always so. System components that need setting are storage elements, positions, voltage levels, fluid levels, and the like.

It is not that well appreciated that some kinds of systems will, given the same initial condition, start up differently giving different solution paths. For example, there exists a class of vibration situations where Eigen values are involved that give differing trajectories.

8 EXERCISING THE MODEL

It is possible for a person with good knowledge of the model and situation to set up tests that seemingly work to show that a model is sound – when it is not! A well-developed test must exercise the system well – not just at obvious points.

The person accepting a model must be familiar with the range of modeling needed. For example, an electronic warfare (EW) instrumentation detection system may well be able to respond as needed for an aircraft flying through a zone containing just one EW threat. This is, however, not a realistic situation, for most real cases will have many more than one hazard present.

Testing should commence from a simple situation for which there is a well-established output. For instance, a new signal spectrum analyzer model can be fed a sine wave. The output should then be a single spike, as shown in Figure 4. A square wave should give a series of reducing spikes.

As confidence grows, the input is expanded to be more complex.

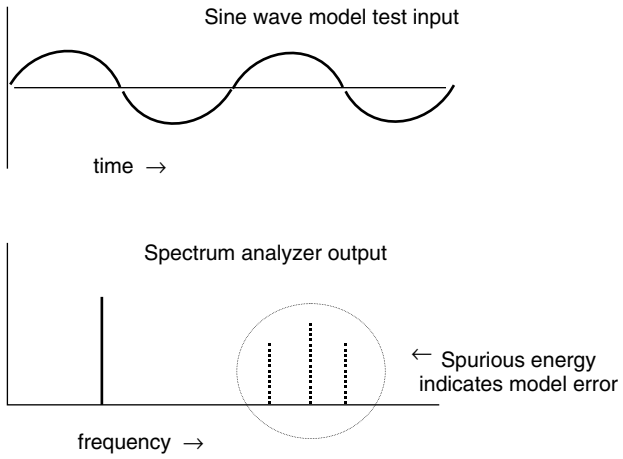


Figure 4. Use of predictable tests to verify models.

9 LIMITATIONS OF MODELS

Models are not perfect representations. The following list of commonly met model performance parameters can be applied to ascertain limitations:

- Shared data consistency – how well does it facilitate data sharing with other models?
- Real-time interaction – does it work well enough in real time or is it really working in 'near real time' that is not 'near enough'?
- Scalability – can the model parameters be scaled up or down and with what loss of precision?
- Extensibility – can its boundaries of operation be extended as the modeling exercise grows?
- Bandwidth – how fast can it model behavior? Lack of adequate response can attenuate the results without any obvious indication that it is now in error.
- Reliability – obviously, one seeks a tool that does not need constant repair and resetting.
- Latency – this refers to the lag time for data to arrive. This is not the same as lack of dynamic response, for lag produces a different kind of error in a response.
- Heterogeneity – does it work uniformly over its working range or do strange and unexpected things happen in certain situations?

10 MODEL VALIDATION

Verification, Validation, and Accreditation or VV&A are the cornerstone activities for ensuring that any model is a faithful representation and can be used with the confidence that whatever is built from it, or whatever uses it, will have a sound basis. Each term is covered in turn in the order in which they are applied in practice. The terms tend, quite incorrectly, to be used synonymously; each of them is quite different in purpose.

10.1 Verification

In general, verification is the process applied first to determine that a model implementation accurately represents, to a sufficient level, the developer's conceptual description and specification. In short, is it what was intended?

Verification is done largely by study of the model needs, as can be found from documentation.

10.2 Validation

Validation is the process of determining the manner and degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model, and also the level of confidence that can be placed on the assessment.

This is performed largely by testing and evaluating the model against the earlier verified requirements.

10.3 Accreditation

Accreditation is the official certification given that a model is acceptable for use for a specific purpose. Obtaining a clear understanding of the 'specific purpose' is a problem at times – models are used in other contexts all too easily.

This is performed after reviewing the verification and validation results. If all is adequate, accreditation is given once a clear statement of its purpose and scope of use is recorded to ensure it is not pressed into incorrect use. This should be carried out by an independent authority who has not been too deeply involved in the model's development.

These three sequential steps seem to be clear-cut, but they can rarely be done on a totally crisp and formalized basis in large systems. Conversely, in smaller systems, they are all too often ignored to save time and money!

VV&A can easily be done against the inappropriate level of documentation and with too narrow and selective a viewpoint.

There is no such thing as absolute truth – not even a real test will necessarily duplicate all of the important real parameters – they are just not known until the actual uses take place. Models must always be used with care, and regular real-world testing must be used to ensure they are adequate.

More detail on modeling in design and its role in prototyping is available in Sydenham (2004). Several texts are devoted to modeling, but they do not address measuring systems in particular – Severance (2002), Zeigler, Kim and Praehofer (2000), Fjeldly, Ytterdal and Shur (1997).

RELATED ARTICLES

Article 57, Static Considerations of General Instrumentation, Volume 1; **Article 59, Introduction to the Dynamic Regime of Measurement Systems, Volume 1**; **Article 61, First-order System Dynamics, Volume 1**; **Article 62, Second-order System Dynamics, Volume 1**; **Article 70, Measuring System Design Methodologies, Volume 2**; **Article 72, Mathematical Methods of Optimization, Volume 2**.

REFERENCES

Davies, D.R. (1996) Using Structured Modelling Strategy to Build Validated Computer Model, in *Proceeding of Second T&E Institute of Aeronautics Forum*, RAes, 25–27 June, American Institute of Aeronautics and Astronautics, London AIAA-96-3335-CP, (pp. 25–33).

Fjeldly, T.A., Ytterdal, T. and Shur, M.S. (1997) *Introduction to Device Modeling and Circuit Simulation*, Wiley, New York.

Holland, R. (1996) *The Virtual Reality Homebrewer's Handbook*, Wiley, New York.

Neelamkavil, F. (1987) *Computer Simulation and Modelling*, Wiley, New York.

Severance, F.L. (2002) *System Modeling and Simulation: An Introduction*, John Wiley, New York.

Sydenham, P.H. (2004) *Systems Approach to Engineering Design*, Artech House, Boston, MA.

Zeigler, B., Kim, T. and Praehofer, H. (2000) *Theory of Modeling and Simulation*, Academic Press, New York, NY.

72: Mathematical Methods of Optimization

Halit Eren
Curtin University of Technology, Perth, Western Australia, Australia

1 Use of Mathematical Analysis in Optimization 452
2 Setting Up the Generalized Mathematical Description 453
3 Exercising the Mathematical Description 454
4 Comparison of Mathematical Description and Modeling 454
Further Reading 455

1 USE OF MATHEMATICAL ANALYSIS IN OPTIMIZATION

Optimization is applied in many disciplines, such as engineering, science, control systems, cybernetics, information theory, computer systems and networks, economics and other social sciences, and so on.

In many engineering, science, and technology applications, optimization is used to maximize or minimize the value of a function chosen as the performance index or cost function, J.

The performance is optimized for numerous reasons such as improving the quality, increasing production, maximizing profits, decreasing waste, obtaining greater efficiency, maximizing safety, saving time and energy, and so on. However, in many problems, boundary conditions exist because of limited power, storage capacity, capability of machinery, various temperature constraints, limits in speed, force, acceleration, and so on.

The cost function (see Figure 1) can be continuous and/or discrete. Appropriate selection of the performance index for a given system is important since it determines the nature and complexity of the problem. The system may be linear or nonlinear, stationary or time-varying, continuous or discrete, static or dynamic, constrained or unconstrained. Therefore, appropriate formulation of performance index involves a compromise between a meaningful evaluation of the system performance, availability of feasible mathematical descriptions, and application of meaningful algorithms for their solutions.

It is important to understand that an optimal system under a selected performance index may not be optimal for another performance index.

Figure 1 illustrates a typical optimization problem. In this figure, the sign of the first derivative $\mathrm{d}J/\mathrm{d}U$ of the cost function changes as the function passes through the maximum point. This method commonly uses an initial estimated point (1) where the first derivative is considered (guessed) to be zero. This point (2) is then used, in the direction of steepest ascent (3) or descent as the case may be, for the next point until the absolute maxima or minima is found without getting trapped in local extreme points.

An optimization problem involves two basic stages; first stage is the appropriate mathematical description of the system and the second stage is the solution of the problem using suitable techniques and algorithms. For solving an optimization problem, the following steps may be necessary:

1. A mathematical description has to be formulated.
2. The desired form of the objective function has to be specified.
3. The variables have to be specified.
4. The constraints have to be determined.
5. A suitable optimization method needs to be selected and implemented.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

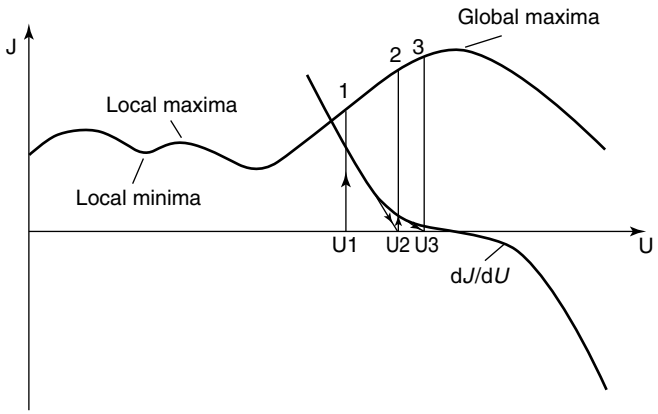


Figure 1. A typical cost function with local and global extremes.

6. Once the problem is solved, solutions must be related to the physical system.

Physical systems are defined (modeled) mathematically usually by identifying the related parameters and using the known properties. The application of optimization allows the selection of those parameters that give the best properties for a chosen preference function.

Once described mathematically, the optimization problem may be solved using computers backed up by powerful software tools, such as linear and nonlinear programming methods, stochastic methods, tabu search algorithms, genetic algorithms, artificial intelligence methods, and so forth.

For the solution of an optimization problem, depending on the nature of the problem, many different methods may be used. For example:

1. If the performance index and constraints can be formulated as linear algebraic functions of the variables, then selection of an appropriate linear programming method, such as simplex programming, is sufficient.
2. If the equations describing the system are nonlinear, then the solution may involve nonlinear techniques or linearization of the problem in some subregions.
3. If the problem involves determination of the largest value of a function with two or more variables, then the steepest ascent (or descent) method, sometimes termed as *hill climbing*, or gradient method may be a good choice, see Figure 2. In this figure, the steps taken to find the optimum point are shown by $S_{T_0}, S_{T_1}, \ldots, S_{T_5}$.
4. If the problems are complex (termed *NP hard problems*), traditional optimization methods based on local information are not adequate to solve them; global optimization methods need to be applied. Some of these methods are Lagrange duality in partly convex programming, robust minimizers, adaptive optimization, decomposition techniques, iterative topographical methods, and so on.

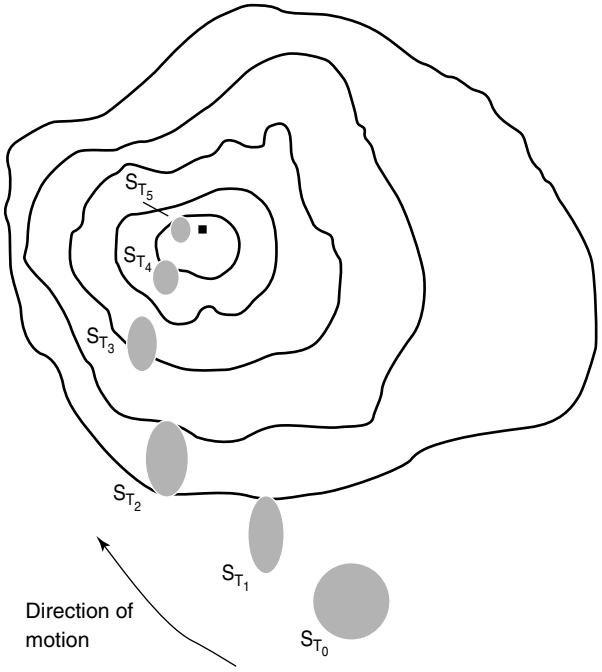


Figure 2. Illustration of hill-climbing with gradient method.

2 SETTING UP THE GENERALIZED MATHEMATICAL DESCRIPTION

Optimization has two main objectives. The first is to identify a set of conditions for a function, f, where min or $\max\{f(x, \theta)|x \in D(\theta)\}$ for constraints, D, will exist as a solution. The second is characterization of the set of optimal points and the application of algorithms.

The mathematical problem involves the determination of a set of values of variables x_i such that the cost function $\varphi|x_i|$ is as small (or as large) as possible. x_i represents the set of all possible variables for $i = 1, 2, \ldots, n$.

However, each variable may be subject to equality or inequality (or both), linear or nonlinear constraints in the form of

$$f_j(|x_i|) = 0, \quad j = 1, 2, \ldots, p \tag{1}$$

or

$$g_k(|x_i|) \geq 0, \quad k = 1, 2, \ldots, q \tag{2}$$

2.1 Linear systems

Many practical problems lead to linear systems of equations, which can involve hundreds of variables thus requiring the solutions of numerous simultaneous linear equations.

The simplest case is the optimization of a single variable, $\varphi(x)$. In simple cases in which there are not many variables,

linear programming can be employed yielding to solutions in n-dimensional space.

If a linear programming problem has a unique optimal solution, then it is an extreme point; if not unique, there is an optimal solution at an extreme point, thus the solution obtained is within a finite set of extreme points.

2.2 Nonlinear systems

Many problems lead to nonlinear (transcendental) systems of equations, which can only be solved by computational methods.

A problem is nonlinear if the objective function $f(x)$ in which x is a vector or any of the constraint functions is nonlinear. Nonlinear functions are classified according to whether their defining functions have the appropriate convexity.

Solutions of nonlinear functions may often be dependent on the differentiability of the constituent functions. If all constituent functions are differentiable, there is an analogous result on the necessary conditions of optimality that can be closely related to the Kuhn–Tucker conditions.

3 EXERCISING THE MATHEMATICAL DESCRIPTION

Standard computational methods for solving nonlinear optimization problems may be divided into three main groups:

1. Zero-order methods (coordinate comparison, simplex methods, and stochastic methods);
2. First-order methods (gradient and quasi-gradient, methods of associated directions, variable-metric methods);
3. Second-order method (Newton method).

Nonlinear optimization aims at mathematical formulation of problems and implementation of algorithms. In the implementation stages, many different algorithms exist, such as numerical optimization, random walk (Figure 3), convex analysis, calculus variations, stochastic methods, game theory, statistical and probability techniques, and more recently, genetic algorithms, and neural networks.

Stochastic methods consist of the calculation of the values of the cost function at a large number of selected points. The selection is made such that each point has an equal probability of being selected. The best points are determined by comparing the function values at a large number of points. Although requiring intensive computational time, stochastic methods are useful for finding the global optimal points, that is, unlike the conventional methods, their likelihood of getting trapped in local minimums is small.

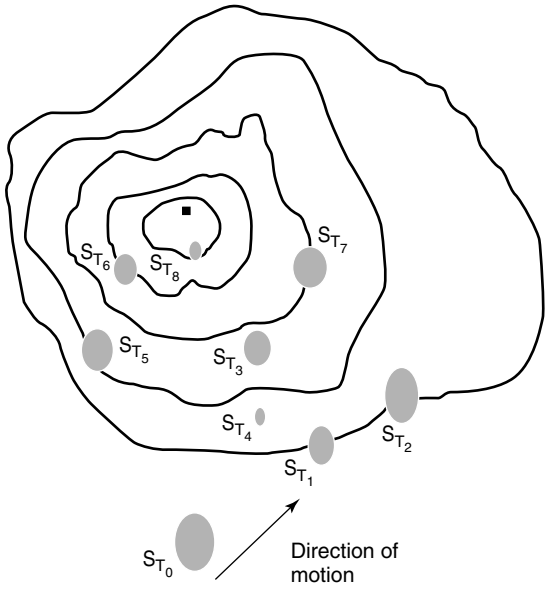


Figure 3. Illustration of random walk method.

Therefore, they find the nearest minimum point as quickly as possible and search for all potential solutions in an optimum manner.

Recently developed, genetic algorithms are used for optimizing the parameters of an unknown system whose mathematical description is either too complicated or unknown. The genetic algorithms are usually applied to neural networks in two levels. The first level determines the suitable weights in the neurons and in the second level, the number of hidden layers as well as the number of neurons is determined.

4 COMPARISON OF MATHEMATICAL DESCRIPTION AND MODELING

The nature of the mathematical description and computational method indicate if the type of problem is to be linear or nonlinear. For example, a linear program is defined as the optimization of a linear cost function whose variables satisfy a system of linear equalities or inequalities.

The term mathematical programming usually refers to linear programs. Mathematical programming has a long history and has been influenced by operations research, game theory, mathematical economics, control theory, dynamic programming, calculus variations, and combinational theory.

The standard form of a linear program is expressed as an instruction to maximize or minimize a linear expression in n nonnegative variables.

Any system of linear inequalities can be reduced by simple substitutions. Associated with each linear programming

problem is a duality as a second linear programming problem. In this association, the original problem is referred to as the primal.

However, the general study of nonlinear programs usually involves such items as integer programs, stochastic programs, and network flow theory.

One broad class of nonlinear programming problems is that concerning minimizing the cost function subject to no constraints; another is the linearly constrained nonlinear programs that include subclass quadratic programs concerning minimization of quadratic function subject to linear constraints.

There are many different ways of solving nonlinear programs depending on whether they display convexity properties and/or differentiability.

Some optimization problems are difficult to solve by conventional numeric optimization methods, even if they converge to an acceptable optimum point in a reasonable time. To solve these problems, other methods have been developed. For example, despite the large size of information to be handled, the stochastic methods reach the global optimum point easily.

In the hill-climbing algorithm, it is possible that an extreme is reached without finding the global minimum. Here, the starting solution is generated at random. Once a local solution is obtained, it is used as the center of a new neighborhood in which the optimization is repeated to determine the other extremes.

Even this way, this method might revert to the local minimum that has already been passed in a previous step, called the *problem of looping*. The program may need to be repeated several times with different randomly generated initial values.

The tabu search algorithm is based on a hill-climbing method that tries to avoid the problem of looping. The hill-climbing algorithm has a short-time memory in which it remembers the optimal solution transformations to generate new centers for iterations. These optimal solution transformations are prohibited (hence the name tabu) when the new neighborhood is created for a given current solution. A hill-climbing algorithm modified in this way systematically searches the entire area in which the global minimum of a function is to be found.

The genetic algorithm is used to optimize systems with unknown parameters. It is useful when other algorithms fail in such problems as where computational time is an exponential or factorial function of a number of variables. In the applications, it is sufficient to know the price of each individual in the population. Therefore, the pricing selection for each individual bears special importance as it is directly related to the outcome of the optimized problem. By application of artificial neural networks in two stages, genetic algorithms are used to determine population, pricing, and optimum points. There are improved versions of genetic algorithms such as scatter search, path relinking, star-paths, and so on.

FURTHER READING

Agrawal, K.S. and Fabien, B. (1999) *Optimization of Dynamic Systems*, Kluwer Academic Publishers, Dordrecht.

Bishop, R.H. (2002) *The Mechatronics Handbook*, CRC Press, Boca Raton, FL.

Floudas, C.A. and Pardas, P.M. (1996) *State of the Art in Global Optimization-Computational Methods and Applications*, Kluwer Academic Publishers, Dordrecht.

Kuo, B.C. (1992) *Digital Control Systems*, 2nd edn, Harcourt Brace Jovanovich, New York.

Lewis, L.L and Syrmos, V.L. (1995) *Optimal Control*, 2nd edn, Wiley, New York.

Rapcsak, T. (1997) *Smooth Nonlinear Optimization in R^n*, Kluwer Academic Publishers, Dordrecht.

Rubinov, A. and Glover, B. (2001) *Optimization and Related Topics*, Kluwer Academic Publishers, Dordrecht.

Webster, J.G. (1999) *Encyclopedia of Electrical and Electronic Engineering*, Wiley, New York.

Whittle, P. (1996) *Optimal Control-Basics and Beyond*, Wiley, UK.

73: Overview of Electrical and Electronic Technique

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Relevance of Electrical and Electronic Technique 459
2 Historical Development of Electrical Regime 460
3 Systematic Nature of Electrical and Electronic Regime Engineering 461
4 Literature of Electrical and Electronic Regime Engineering 462
5 Signals in the Electrical Regime 465
Related Articles 466
References 466

1 RELEVANCE OF ELECTRICAL AND ELECTRONIC TECHNIQUE

Although electricity was the most recent technological discipline to develop, it was very rapidly found to be the most appropriate for many tasks required in a measuring system. Certain classes of instrument can be constructed without recourse to electrical principles, examples being the microscope, the micrometer, the direct recording water level gauge, and many more. However, if an application needs extensive information processing, signal transmission over long distances, uniformity of manufacture at low cost, ability to be extended, and use systematic design procedures, then electrical technique will usually (but not always) provide a better design solution than other alternatives.

In the context of the thousands of years of measuring systems development, this situation has only arisen in the last half century, wherein solid-state integratable electronic circuitry has provided unprecedented ability to produce low cost, very powerful processing and data handling ability.

Here is presented an introduction to the capability and methodology of electrical and electronic techniques. It must be made clear that although electrical and electronic technique is so extensively used and is very attractive it is not always the optimum means of solving a measurement systems design need. Indeed, there are many applications where the use of electrical methods are either not applicable or they are not preferred, even if usable, for a variety of reasons.

Despite the fact that electrical methods appear to be able to supply every measurement system data handling needs, it must be remembered that the natural world is usually the system that must be interfaced to man's inventions; that world does not possess many signal sources having the same form of physical manifestation as occurs in man-made electrical systems. Thus it is that the other energy regimes will always be important in measurement system design to assist formation of the input and output transudation interfaces. By itself, an electronic system has no practical use – it must be interfaced to an application with some kind of transformation device – the sensor.

Electrical technique includes both the traditional electrical and the more recent electronic disciplines. Initially, the two aspects could be reasonably well divided for purposes of exposition and administration. Prior to 1910, electronics was virtually unknown and electricity was, therefore, clearly defined.

Electrical methods were partially displaced by electronic methods (that grew out of the application of thermionic devices) since electronics provided the circuit theory considered necessary, and initially much of the practice. As time progressed, electronic methods were devised that could also handle most power needs, previously only possible through the use of electrical devices. Today the

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

distinction between electrical and electronic methodology is very blurred, although many institutional mechanisms and groupings continue that might imply otherwise.

Until very recent times, electronic technique was generally only suitable for stages following the sensor. In the 1970s decade, however, silicon integratable devices were first developed to yield a new generation of electronically based sensors that are now in prolific use. As time progresses, electronic capability is steadily being extended; the sensor principles are still as current as when first applied.

It is certainly true that in any new application requiring a measuring instrument the designer must now consider the appropriateness of electrical methods for the specific task in question.

2 HISTORICAL DEVELOPMENT OF ELECTRICAL REGIME

The origins of electricity and electronics can be traced to ancient Greece. The distinction between the two is usually based on the premise that both are concerned with electron flow, the former being at a macrolevel, the latter at the discrete electron level; this is somewhat artificial in modern terms of application. Their contribution was, however, minimal compared with advances that began in the seventeenth century A.D. Furthermore, their knowledge was rarely put to use and there was certainly no ordered use of principles of design.

Interest in, and application of, electricity started in earnest with work on electrostatic devices of the form of charge generators and detectors, this being in the century preceding 1800 A.D. Knowledge of charge, its quantity and storage, its transfer from place to place by the use of insulated wires, and its physical effects gave eighteenth century experimenters a tool for further experimentation. Out of their work, gradually, there grew useful applications. Electrical telegraphy first began in the late eighteenth century in a static electricity form. Static electricity led to knowledge about the quantity and potential of electricity and to the understanding that there exist two forms of electric charge: positive and negative. What did not develop so well at first was an appreciation of charge flowing as a current in wires.

Experiments with static electricity, coupled with the improvement in vacuum technique, gradually provided knowledge of discharge effects in gases. These provided a place for the understanding of the behavior of electrons, a maturity that took some 200 years to reach the point where the electron was finally proven to exist as a discrete entity in J. J. Thomson's experiments of 1899.

Although there was much activity in seeking knowledge about the nature of electricity in the eighteenth century, significant progress came only after the invention of the simple primary cell by Volta in 1800. Prior to that time experimenters had only very high-voltage sources possessing high output impedance. It was not exactly the preferred apparatus to use in an investigation of the practical applications of electrical knowledge!

In 1800, the electric primary cell became generally available and, being simple to make, was rapidly adopted by experimenters. Its low output impedance, reasonably long duration of operation, low voltage, and ease of manufacture was admirably suited to the needs of the gentleman scientists of those times. Progress was steadily made in gaining understanding of the fundamental nature of practically useful electrical circuits and devices and in their everyday use.

During the nineteenth century, the laws of circuits, both DC and AC, were formulated through careful experimental and theoretical work. As new laws were published, they were taken up to form technological devices. For example, the relationship between electricity and magnetism was established first as the motor effect in the 1820s. This enabled electrical indicators and electromagnetic actuators to be made. Better indicators of electrical quantities (meters) enabled more laws to be discovered and weaker effects to be applied.

The laws of electromagnetic radiation were also enunciated in the middle of the nineteenth century. By 1888, Hertz had experimentally proven that the theory was correct and that practical apparatus could be constructed that enabled it to be put to use. Marconi, in the last decade of that century, assembled the first practical radio system using simple, purely electrical, components and understanding.

By 1900, many of the basic circuitry laws and physical principles, which are today used to devise instruments, were known, if only as scientific entities. The Victorian era produced much of the electrical systems groundwork needed in twentieth century instrumentation.

A chance encounter occurred in the latter part of the nineteenth century, this being discovery of the Edison effect. It proved, experimentally, that a thermionic device could rectify alternating current. Edison did not, however, make use of this at the time. By 1900, designers of telegraph, telephone, and radio communications needed three important improvements in technological capability. Telecommunication needs provided strong commercial reasons to seek them. These needs were: how to amplify a weak electrical signal, how to rectify radio frequency, alternating currents, and how to generate radio frequency currents. Each of these needs was being catered to at the time by very unsatisfactory methods.

The first decade of the twentieth century saw these three basic needs being met by the invention of one

device, the thermionic valve. It first emerged as the purposefully built diode in which the Edison effect was utilized. The diode was capable of rectification and signal generation. Soon after, the triode valve was devised; it could perform all functions of amplification, rectification, and generation. By the 1920s, the thermionic device had evolved into many forms satisfying many needs, not only in telecommunications but also in the emerging, derivative, instrumentation areas.

Electrical and electronic techniques began their partnership from the onset of the electronic discipline in the 1910s. Tuned circuits, special purpose devices built with valves as their basis, developed; the bistable flip–flop and the DC amplifier being examples that initially fulfilled some special need.

The First World War proved that electronic devices were practical. Training in radio, given to military servicemen, helped the general public to subsequently take an interest in electronics for wireless applications.

Gradually electronic technique spread into industrial usage. By the 1930s, industrial electronics was clearly established. It first began to have an impact through the use of the electric eye, a photocell coupled to a suitable electronic amplifier.

Thermionic components, along with purely electrical devices such as the magnetic and amplidyne amplifiers, were able to cope with most of the heavy current industrial needs that arose; but, in general, electronic method was, by modern standards, slow to be accepted. Basically valve-based electronic circuitry was then too expensive, too unreliable, and too sophisticated for the times to find the widespread use that we have come to accept as the norm of the electronics of today.

It is the field of computing that helped generate incentive to find better technological ways of carrying out the functions developed for valve-operated apparatus. Digital computing, especially, provided impetus for activity, for, by the 1940s, it was clear that very powerful computers could be built that far exceeded those economically possible with valves. A superior replacement for the amplifying valve was needed. Valves used too much power, they ran at too high a temperature, they were too large, they cost too much, and they were not reliable enough to be used in huge numbers.

In the 1940 decade, the transistor was devised at the Bell Laboratories. It did not emerge quite as suddenly as it might seem but came about from gradual evolution of earlier work on the solid-state diode of the 1900 era and after. Valves, in fact, being able to satisfy many of the designers' early needs and expectations, tended to slow down development of solid-state devices.

From the time of the invention of the practical transistor, there has been great increase in the pace of electronic development. Integration made devices smaller, cheaper, and far less power consuming. They enabled the basic level of systems to be gradually, at a quickening pace, extended in sophistication. As an example, in the 1960s a digital counter would normally have been assembled from discrete components using individually picked transistors. Today, the whole multidecade counter, with its display, would normally be purchased as a single basic commercial chip unit that is vastly smaller, far more reliable, and much less costly.

The discipline of electrical and electronic engineering has, for use in the design of instrumentation systems, passed from a stage where originally only relatively small-extent systems could be handled by the average person to one in which very complex and powerful systems are assembled with few technological constraints being placed upon the designer. It is now very much a procedure of being able to more fully express a designer's innovative powers and ability with extensive systems built from basic, very powerful, marketed units, doing this to the limit of his or her imaginative ability.

Space only allows a very short history to be given here. A detailed account is to be found in Sydenham (1979a), where reference to other historical studies is given in the general history of instrumentation presented there.

3 SYSTEMATIC NATURE OF ELECTRICAL AND ELECTRONIC REGIME ENGINEERING

Electrical engineering, as has been described above, developed with a well-ordered basis of theoretical understanding. Indeed, much of what resulted would not have emerged so rapidly without the theory being known. Unlike mechanical engineering, where a large part of its output was able to emerge through the technological path of experimental experience and intuition, most electrical topics can only be understood and applied if the user has a good theoretical grounding in the principles involved. The physical nature of the subject material has allowed its constituent knowledge to be ordered systematically.

Electronic technique rests upon well-understood, theoretically enunciated principles. When used at the level that most users assemble systems (from commercially marketed modules), it also requires a considerable degree of inventive skill. It can, however, be efficiently applied using reasonably straightforward and simplified theoretical aids, such as designing procedures and rule of thumb relationships. The nature of the range that electronic parameter values cover implies that the components, and their assembly, do not usually require very tight tolerances of design.

People with very little training in the basics of electronics can build worthwhile systems. Naturally, specially trained persons can handle the topic with greater ease and better performance.

This fact should not be interpreted as meaning that electronics does not somewhere require in-depth scientific understanding and application of sophisticated methodology. That has become today more the role of the component and subsystem designer, those people have provided a buffer for the end user and general systems designer.

Marketing forces have assisted simplification of electronic and electrical instrument design. The quest for sales, in large numbers, has enticed manufacturers to provide a truly useful and extensive application service. They have sought ways by which sophisticated devices can be used effectively by semitrained persons.

This account is aimed at the semitrained person. It is not intended for the specialist in some aspect of electronic or electrical engineering. It begins by presenting information about the most basic components used to form electronic systems. Whereas many of these may not be used as discrete devices very much in these times, it is, nevertheless, important to have an understanding of the concepts and hardware entities that go into a marketed system module. These basics presented then allows discussion to rise to a higher level in the systems hierarchy wherein description can be given of the marketed building blocks that are commonly used to assemble the specific system needed in a given application. It is then feasible to rise yet further to the final design level in order to discuss how extensive systems are generated.

Other conceptual aspects of system design that need consideration are sources of noise in these systems and the power supplies that supply energy for the building blocks.

As an example of the systematic nature of electronic systems consider the diode-based electronic thermometer shown in Figure 1. It illustrates how a system is progressively broken down into subsystem blocks, each of which in turn can be further broken down, the process continuing until adequately basic levels of description are reached. In many systems, there is no need to progress to the individual component level, as the necessary modules can often be assembled from inexpensive high-performance units that only need wiring into the system in the appropriate manner.

The above procedure allows any system to be portrayed on paper at the level needed to adequately specify its functional architecture. Often the reverse occurs and the user may be initially confronted with the final subsystem details, not the primary stage block diagram. At first this may well overwhelm the beholder but the use of the systematic realization procedure to form the constituent, higher-level blocks will reveal how the system operates as each circuit group is identified as a known and familiar building block.

4 LITERATURE OF ELECTRICAL AND ELECTRONIC REGIME ENGINEERING

It is of value to provide an overview of the extensive printed literature before continuing. There exist numerous texts on electronic and electrical engineering. They range, in marketing approach, from trivial expositions to esoteric academic theoretical works. Each has its own usefulness. Although popular texts are often denigrated by experts, they have, over the years, risen in quality, now providing excellent introductions for persons who wish to use and understand electrical and electronic technique at the level of practical use yet not become formally trained in this way. Academic works may often be found to be too advanced for middle-levels users. The reader is cautioned to study a text carefully to establish if it adequately matches the need.

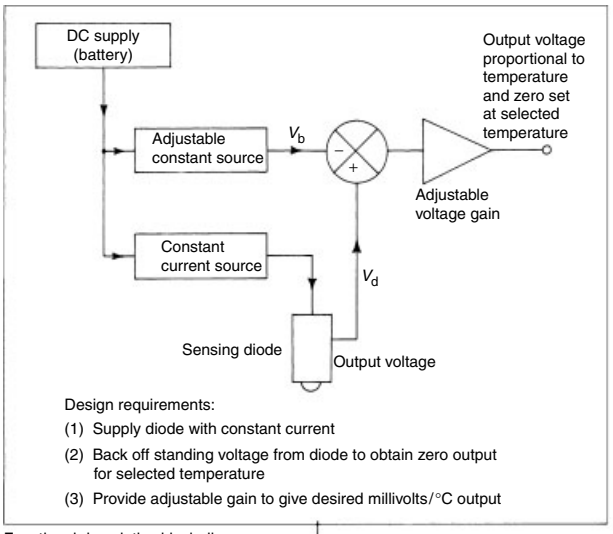
The following selection of published works is presented to assist selection. It must, however, be stressed that many others are available. Electrical engineering principles are presented in Baitch (1972), Bureau of Navy (1970), and Smith (1976) as introductory courses. Similar material is also available in the many handbooks of electrical engineering, but where it is not usually presented in a didactic manner.

Electrical engineering overlaps electronic technique in the area of circuit networks. Linear networks are the easiest to design and theorize (see Leach, 1976; Van Valkenburg, 1974). Nonlinear systems are covered in Fox (1957), Van der Ziel (1977), and Willson (1975). Older texts are still useful for they lay down enduring principles.

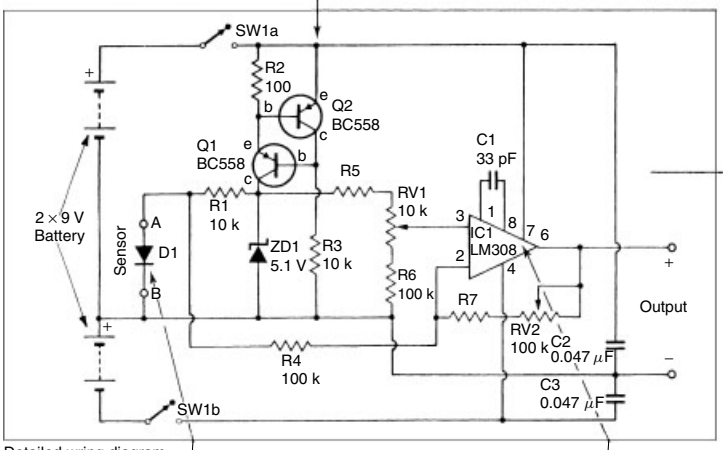
Systems of circuitry originally were developed by analyzing given arrangements, but a more positive design approach was developed through which the system can be synthesized to provide a prestated performance: Fox (1952) is but one text following this approach.

One particular class of network is the so-called filter. Modern design procedures for these use previously developed mathematical expressions to arrive at the desired performance using feedback placed around an active amplifier element. Several texts exist on these forms of filter (refer to Burr-Brown, 1966; Hilburn and Johnson, 1973; Huelsman, 1977; Spence, 1970).

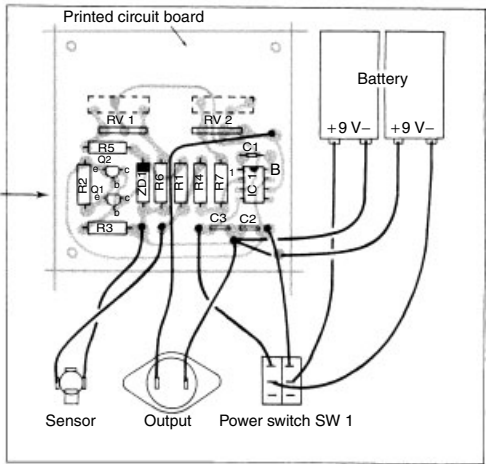
Figure 1. Breakdown of a solid-state, diode-based, temperature meter system showing how an electronic system is formed from basic components arranged in a systematic order.



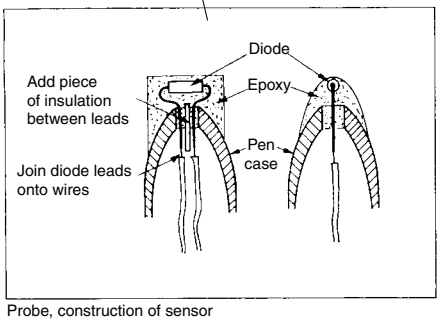
Functional descriptive block diagram



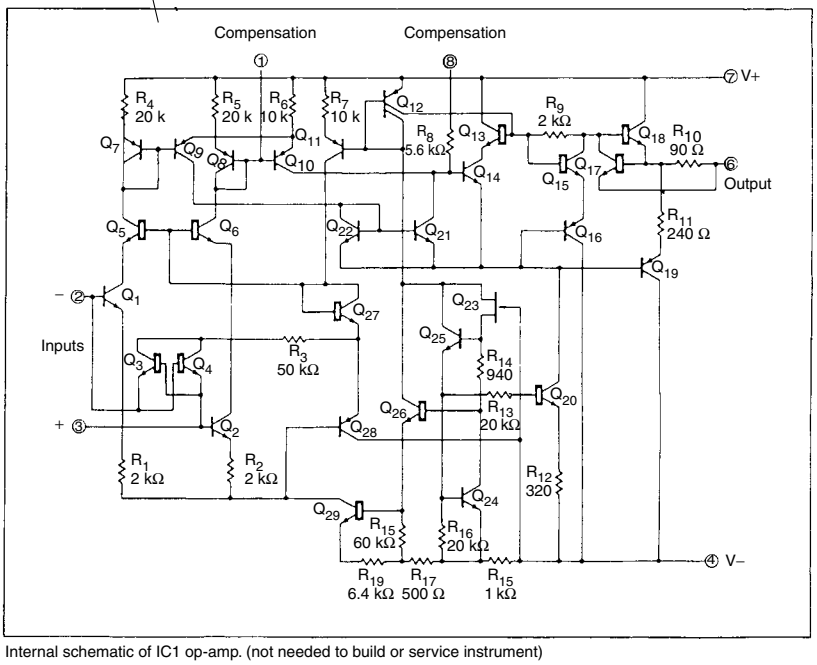
Detailed wring diagram



Constructional layout



Probe, construction of sensor



Internal schematic of IC1 op-amp. (not needed to build or service instrument)

Computers have made simple circuitry design almost trivial and more complex systems possible. Director (1974) and Szentirmai (1974) describe computer-aided design of circuits. This aspect has matured greatly and tool suites are available to support efficient design – once the user is familiar with the tool and can afford to purchase the tool system.

As in electrical engineering, but probably more so, there exist many works covering general electronic principles and method. A selection spanning the ranges of depth and of age (early texts contain useful conceptual information and theoretical derivations) is Brophy (1977), Carson (1961), Churchman (1971), Delaney (1969), Langford-Smith (1955), Lowe (1972, 1977), Millman and Halkias (1976), Mitchell (1951), Pender and McIlwain (1950), Ruiter and Murphy (1962), Squires (1967), Starr (1959), Sydenham (1979b), and Waters (1978). Several of the electronic equipment and component suppliers offer packaged courses, which often include aural recorded tapes, experimental apparatus. An example is Heathkit (1978).

Electronic circuits (these involve electrical circuits plus specialized components) are the subject of Jones (1978), Millman and Halkias (1967), Sands and Mackewroth (1975), and Senturia and Wedlock (1975). Lowe (1974) deals with drawing practice for electrical and electronic engineering: caution is, however, needed in using works on this aspect due to the existence of several different standards of practice.

Electronics, as did electrical engineering earlier, spawned numerous new words in its technical vocabulary. Dictionaries are often needed to understand terms used in written and spoken descriptions. Such definitions are given in Handel (1971), IEEE (1972), and Radio Shack (1975). There is a surprisingly close, but not identical, agreement on terminology across the world in this discipline but very few people appear to make general use of the standard terminologies that do exist, so minor localized differences often occur.

It is difficult to separate electronic instrumentation from the total field of electrical and electronic engineering. For this reason, many of the above mentioned general texts may well provide the information sought. There are, additional to these, many books available that indicate from their title that they relate to electronic instrumentation and testing, the main theme of this article. As with all works, they must be consulted to establish their usefulness in a given situation. The following list refers to those specifically on instrumentation and testing: Banner (1958), Booth (1961), Diefenderfer (1979), Edwards (1971), Gregory (1973), Herrick (1972) IEE (1979), Kidwell (1969), Kinnard (1956), Lion (1975), Maeison (1974), Malvino (1967), Mansfield (1973), Norton (1969), Oliver (1971), Partridge (1958), Piller (1970), Prensky (1963), Regtien (1978), Soisson (1961), Studer (1963), Sydenham (1980), Terman and Pettit (1952), Thomas (1967), Turner (1959, 1963), Waller (1972a, 1972b), Wedlock (1969), Frenzel (1997), Gibilisco (1997), Grob (1996), Horowitz and Hill (1989), Mano (2001).

Texts continuously appear that describe new devices as they arrive in the market. To the titles given here must be added, in any search for information, the enormous amount of application notes and internal publications provided by the component suppliers and marketing agencies. Electronic engineering is one of the areas of technology where tradition has given rise to sales backed by extensive published assistance about product use. Some of these publications are included in the following list on devices and components: Cleary (1969) and Thornton *et al*. (1966) on transistors; Lancaster (1976) on the transistor–transistor-logic assemblies, called *TTL*; and Lancaster (1977) on the text advance on TTL, but not its replacement, the alternative CMOS circuitry. Integrated circuits are covered in many books (see Millman, 1972; Rosenthal, 1975; Turner, 1977). Other components are dealt with in Sheingold (1972, 1974). Turner (1978) is on field-effect transistors (FETs).

Thermionic valves are still used in a general way in some countries and will be found in older equipment; their operation is explained in the numerous electronic subject works compiled before 1960.

Often poorly considered when designing electronic aspects of a measurement systems is the need to supply adequately conditioned power. Texts on power supplies and on handling power include Csáki *et al*. (1975), Dewan and Straughen (1975), H-P (1973), Kepco (1966), Marston (1974), and Motorola (1961, 1966, 1967).

As will be explained in more detail below, one regime of electrical and electronic systems is those systems using continuously varying signals that can take any level of voltage or current within defined limits. These are called *analog signals* and the components for use with such signals have become largely known as *linear devices*. The operational amplifier is a salient linear circuit unit. A selection of works on the linear regime includes Burr-Brown (1964), Clayton (1971), Connelly (1975), Faulkenberry (1977), Graeme and Tobey (1971), Gupta and Singh (1975), Melen and Garland (1978), NS Corporation (1976), and Signetics (1977).

The other kind of signal regime uses a signal form that can only exist at two, or sometimes three levels, the former being the more usual. This is known as a *digital signal*. As will be explained below, several basic units exist for building digital systems. They are variously described as units or as combinations in Blakeslee (1975), Bouwens (1974), Breeze (1972), Helms, Kaiser and Rabiner (1976), Kostopoulos (1975), Middleton (1977), Millman and Taub

(1965), Namgoster (1977), Peled and Liu (1976), Rabiner and Roder (1973), and Wakerly (1976). General works on electronic method, mentioned earlier, also discuss the two regimes of signal use.

Manufacturing aspects of electronic systems using computer aids are covered in Cassell (1972). The works Dummer and Griffin (1962), Simpson (1976), and Waller (1972c) deal with testing and reliability of components and materials. Servicing electronic systems has become an important aspect due to the large size extent of equipment used and the responsible position held by it in the general integrity of a plant. Information about this topic is given in Garland and Stainer (1970), Sloot (1972), and Waner (1979).

Finally, in this introduction to the literature of electrical and electronic systems used for measurement, mention is needed of the many books and booklets available from the commercial press and the marketing companies on the use of electronic instrumentation in specific fields. These often repeat some of the basic information in their introductions before moving onto the details of their specific topics. Two examples are Van Santen (1967) on weighing and Piller (1970) on electromedical instrumentation.

5 SIGNALS IN THE ELECTRICAL REGIME

Measurement is the process of obtaining meaningful information about a topic. In the sense of relationship with the real physical world, this is achieved by interpreting received modulations of energy or of mass- transfer. The entity that conveys the information is known as a signal. Electrical and electronic systems make use of electrical energy for information transfer, so their signals occur in some form of variation in the amplitude or phase relation of either voltage, current, or impedance. These three entities are related according to Ohm's law in linearly operating circuits and through other relationships in nonlinear systems. For example, the system may use a constant voltage-source impedance, variations in the, say, sensor causing current modulations to occur at the receiver. Alternatively, the current may be held constant, the impedance variations giving rise to voltage fluctuations. By suitable conversion, through the use of an impedance varying stage, a constant-voltage supply system can be used to supply varying voltage signals.

If the current flows only in one direction, it is called *a direct current (DC) system*. If the current alternates backward and forward in the connecting link, it is known as an *alternating current (AC) system*. Where the system has DC bias added to the AC signal (Figure 2) the current

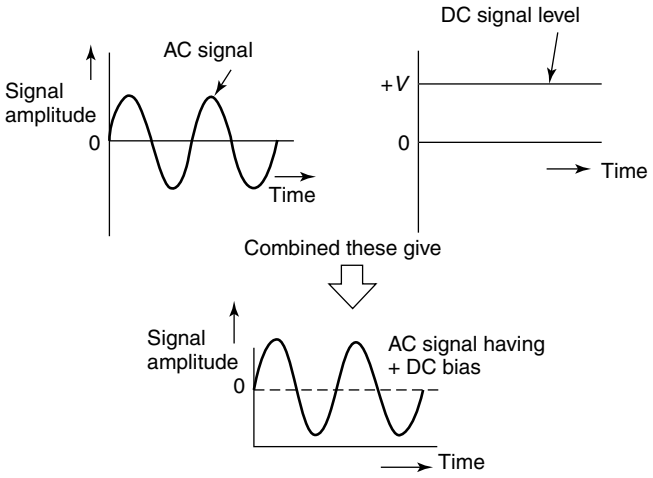


Figure 2. Amplitude-time graph of an electric AC signal having DC bias.

may, in fact, always move only in one direction. This is, however, still regarded as an AC system (possessing a DC level) because the AC signal can be recovered, as can the DC level. A DC level cannot carry information, unless it is modulated with an AC signal, but it is essential in many forms of electronic arrangement.

Signals varying in time (the same concept applies to those varying in space) can be broken down into their Fourier frequency components. Sine wave signals are, therefore, the most basic signal form. Many of the concepts and explanations arising in electrical and electronic engineering rest upon the assumption that the AC signal in question is sinusoidal. If it is not, then considerable error can occur in the inappropriate use of presented information.

Where no qualification is given as to the waveshape or to the averaging criteria used, it is usually safe to assume that the signal is indeed sinusoidal and that the root mean square (rms) value is that which is given.

As amplitudes of the signals and the frequencies that are used in instrumentation range over numerous decades, the practice of compressing the scale from a linear expression is used extensively. The most usual form is the decibel method, which provides compression of magnitude on a logarithm basis.

The range of frequencies involved in signals and systems explanation and operation begins at the cycle period of thousands of years (as arises say, in considerations of the movements of the orbiting planets), passing through the most commonly encountered region of around 10 Hz to 20 kHz (due to widespread interest in audio systems), rising to much higher frequencies for communications and then into the lesser known and used areas above X-ray frequencies. Cosmic rays occur in the 10^{25} Hz region. The most dominant energy frequency spectra used are those

of the electromagnetic spectrum and the audio spectrum. Spatial frequency systems, as arise in optical systems, closely follow many of the fundamentals expressed within the material contained in electrical and electronic method.

This account will now begin at the most fundamental system level at which most users of electronic and electrical systems will need to go, that is, at the components level. Systems will then be considered at progressively higher levels of extent and sophistication.

RELATED ARTICLES

Article 74, Basic Electronic Components, Volume 2; **Article 76, Electronic Systems Design, Volume 2**; **Article 78, Embedded Systems, Volume 2**; **Article 79, Testing Electronic Systems, Volume 2**.

REFERENCES

Baitch, T. (1972) *Electrical Technology*, Wiley, Sydney.

Banner, H.W. (1958) *Electronic Measuring Instruments*, Chapman and Hall, London.

Blakeslee, T.R. (1975) *Digital Design with Standard MSI and LSI*, Wiley, New York.

Booth, S.F. (1961) *Precision Measurement and Calibration; Selected Papers on Electricity and Electronics (3 Vols)*, National Bureau of Standards, Washington, DC.

Bouwens, A.J. (1974) *Digital Instruments Course* (several parts issued), N. V. Philips Gloeilampenfabrieken, Eindhoven.

Breeze, E.G. (1972) *Digital Display Systems: Fairchild Application Note 212/1*, Fairchild Camera and Instrument Corporation, Mountain View, CA.

Brophy, J.J. (1977) *Basic Electronics for Scientists*, McGraw-Hill-Kogakusha, Tokyo.

Bureau of Navy (1970) *Basic Electricity*, Dover, New York.

Burr-Brown (1964) *Handbook of Operational Amplifier Applications*, Burr-Brown Research Corporation, Tucson, AZ.

Burr-Brown (1966) *Handbook of Operational Amplifier Active R.C. Networks*, Burr-Brown Research Corporation, Tucson, AZ.

Carson, R.S. (1961) *Principles of Applied Electronics*, McGraw-Hill, New York.

Cassell, D.A. (1972) *Introduction to Computer-aided Manufacturing in Electronics*, Wiley, New York.

Churchman, L.W. (1971) *Survey of Electronics*, Rinehart Press, San Francisco, CA.

Clayton, G.B. (1971) *Operational Amplifiers*, Butterworths, Sevenoaks.

Cleary, J.F. (ed.) (1969) *Transistor Manual*, General Electric Company, Chicago, IL.

Connelly, J.A. (1975) *Analog Integrated Circuits*, Wiley, New York.

Csáki, F., Ganszky, K., Ipsits, I. and Marti, S. (1975) *Power Electronics*, Akadémiai Kiadó, Budapest.

Delaney, C.F.G. (1969) *Electronics for the Physicist*, Penguin Books, Harmondsworth.

Dewan, S.B. and Straughen, A. (1975) *Power Semiconductor Circuits*, Wiley, New York.

Diefenderfer, A.J. (1979) *Principles of Electronic Instrumentation*, Saunders, Philadelphia, PA.

Director, S.W. (1974) *Computer-aided Circuit Design: Simulation and Optimization*, Dowden, Hutchinson & Ross, Stroudsburg, PA.

Dummer, G.W.A. and Griffin, N.B. (1962) *Environmental Testing Techniques for Electronics and Materials: Series on Electronics and Testing*, Vol. 15, Pergamon, Oxford.

Edwards, D.F.A. (1971) *Electronic Measurement Techniques*, Butterworths, Sevenoaks.

Faulkenberry, L.M. (1977) *An Introduction to Operational Amplifiers*, Wiley, New York.

Frenzel, L.E. (1997) *Crash Course in Electronics Technology*, Butterworth-Heinemann.

Fox, J. (1952) *Modern Network Synthesis*, Wiley, New York.

Fox, J. (1957) *Nonlinear Circuit Analysis*, Wiley, New York.

Garland, D.J. and Stainer, F.W. (1970) *Modern Electronic Maintenance Principles*, Pergamon, London.

Graeme, J.G. and Tobey, G.E. (1971) *Operational Amplifiers – Design and Application*, McGraw-Hill, New York.

Gregory, B.A. (1973) *Electrical Instrumentation; An Introduction*, Macmillan, London.

Grob, B. (1996) *Grob Basic Electronics*, Glencoe McGraw Hill.

Gibilisco, S. (1997) *Teach Yourself Electricity and Electronics*, McGraw-Hill.

Gupta, K.C. and Singh, A. (1975) *Microwave Integrated Circuits*, Wiley, New York.

Handel, S. (1971) *A Dictionary of Electronics*, Penguin Books, Harmondsworth.

Heathkit (1978) *Electronic-Courses for Learn-at-Home: Course 1 DC Electronics; Course 2 AC Electronics; Course 3 Semiconductor Devices; Course 4 Electronic Circuits; Advanced Digital Technique Course; Microprocessor Course*. Available with experimental back-up and optional cassettes, Heath-Schlumberger, London.

Helms, H.D., Kaiser, J.F. and Rabiner, L.R. (1976) *Literature in Digital Signal Processing*, Wiley, New York.

Herrick, C.N. (1972) *Instruments and Measurements for Electronics*, McGraw-Hill, New York.

Hilburn, J.L. and Johnson, D.E. (1973) *Manual of Active Filter Design*, McGraw-Hill, New York.

Horowitz, P. and Hill, W. (1989) *The Art of Electronics*, Cambridge University Press.

H-P (1973) *DC Power Supply Handbook*, Hewlett-Packard, Palo Alto, CA.

Huelsman, L.P. (1977) *Active R.C. Filters*, Wiley, New York.

IEE (1979) *Electronic Test and Measuring Instrumentation: Testmex 79*, IEE, London.

IEEE (1972) *IEEE Standard Dictionary of Electrical and Electronic Terms*, Wiley, New York.

Jones, M.H. (1978) *A Practical Introduction to Electronic Circuits*, Cambridge University Press, Cambridge.

Kepco (1966) *Power Supply Handbook*, Kepco Inc., Flushing, New York.

Kidwell, W.M. (1969) *Electrical Instruments and Measurements*, McGraw-Hill, New York.

Kinnard, I.F. (1956) *Applied Electrical Measurements*, Wiley, New York.

Kostopoulos, G.K. (1975) *Digital Engineering*, Wiley, New York.

Lancaster, D. (1976) *TTL Cookbook*, Howard W. Sams, Indianapolis, IN.

Lancaster, D. (1977) *CMOS Cookbook*, Howard W. Sams, Indianapolis, IN.

Langford-Smith, F. (1955) *Radiotron Designers Handbook*, AWV, Sydney (numerous editions).

Leach, D.P. (1976) *Basic Electric Circuits*, Wiley, New York.

Lion, K.S. (1975) *Elements of Electrical and Electronic Instrumentation*, McGraw-Hill, New York.

Lowe, J.F. (1972) *Experiments in Electronics*, McGraw-Hill, Sydney.

Lowe, J.F. (1974) *Electrical and Electronic Drawing*, McGraw-Hill, Sydney.

Lowe, J.F. (1977) *Electronics for Electrical Trades*, McGraw-Hill, Sydney.

Maeison, E.C. (1974) *Electrical Instruments in Hazardous Locations*, Instrument Society of America, Pittsburgh, PA.

Malvino, A.P. (1967) *Electronic Instrumentation Fundamentals*, McGraw-Hill, New York.

Mano, M.M. (2001) *Digital Design*, Prentice Hall.

Mansfield, P.H. (1973) *Electrical Transducers for Industrial Measurement*, Butterworths, Sevenoaks.

Marston, R.M. (1974) *Thyristor Projects using SCRs and Triacs*, Butterworths, Sevenoaks.

Melen, R. and Garland, H. (1978) *Understanding IC Operational Amplifiers*, H. W. Sams, Indianapolis, IN.

Middleton, R. (1977) *Digital Equipment Servicing Guide*, H. W. Sams, Indianapolis, IN.

Millman, J. (1972) *Integrated Circuits: Analogue and Digital and Systems*, McGraw-Hill, New York.

Millman, J. and Halkias, C.C. (1967) *Electronic Devices and Circuits*, McGraw-Hill, New York.

Millman, J. and Halkias, C.C. (1976) *Electronic Fundamentals and Applications*, McGraw-Hill, New York.

Millman, J. and Taub, H. (1965) *Pulse Digital and Switching Waveforms*, McGraw-Hill, New York.

Mitchell, F.H. (1951) *Fundamentals of Electronics*, Addison-Wesley, Cambridge, MA.

Motorola (1961) *Power Transistor Handbook*, Motorola Inc., Phoenix, AZ.

Motorola (1966) *Silicon Rectifier Handbook*, Motorola Inc., Phoenix, AZ.

Motorola (1967) *Zener Diode Handbook*, Motorola Inc., Phoenix, AZ.

Namgoster, M. (1977) *Digital Equipment Trouble Shooting*, Reston, Reston, VA.

Norton, H.N. (1969) *Transducers for Electronic Measuring Systems*, Prentice-Hall, Englewood Cliffs, NJ.

NS Corporation (1976) *Linear Applications Handbook*, (Vol. I, 1973; Vol. 2, 1976), National Semiconductor Corporation, Santa Clara, CA.

Oliver, B.M. (1971) *Electronic Measurements and Instrumentation*, McGraw-Hill, New York.

Partridge, G.R. (1958) *Principles of Electronic Measurements*, Prentice-Hall, Englewood Cliffs, NJ.

Peled, A. and Liu, B. (1976) *Digital Signal Processing*, Wiley, New York.

Pender, H. and McIlwain, K. (1950) *Electrical Engineers Handbook: Electric Communication and Electronics*, Wiley, New York.

Piller, L.W. (1970) *Electronic Instrumentation Theory of Cardiac Technology*, Staples Press, London.

Prensky, S.D. (1963) *Electronic Instrumentation*, Prentice-Hall, Englewood Cliffs, NJ.

Rabiner, L.B. and Roder, C.M. (1973) *Digital Signal Processing*, IEEE, New York.

Radio Shack (1975) *Dictionary of Electronics*, Radioshack (Tandy Electronics), USA.

Regtien, P.P.L. (1978) *Modern Electronic Measuring Systems*, Delft University Press, Netherlands.

Rosenthal, M.P. (1975) *Understanding Integrated Circuits*, Hayden, Rochelle Park, NJ.

Ruiter, J.H. and Murphy, R.G. (1962) *Basic Industrial Electronic Controls, Holt*, Rinehart and Winston, New York.

Sands, L.G. and Mackewroth, D. (1975) *Encyclopaedia of Electronic Circuits*, Prentice-Hall, New Jersey.

Senturia, S.O. and Wedlock, B.D. (1975) *Electronic Circuits and Applications*, Wiley, New York.

Sheingold, D.H. (1972) *Analog–Digital Conversion Handbook*, Analog Devices, Norwood, MA.

Sheingold, D.H. (1974) *Non-linear Circuits Handbook*, Analog Devices, Norwood, MA.

Signetics (1977) *Analog Data Manual*, available from N. V. Philips Gloeilampenfabrieken, Eindhoven.

Simpson, A. (1976) *Testing Methods and Reliability – Electronics*, Macmillan, London.

Sloot, W. (1972) *Solid-state Servicing*, H. W. Sams, Indianapolis, IN.

Smith, R.J. (1976) *Circuits, Devices and Systems: A First Course in Electrical Engineering*, Wiley, New York.

Soisson, H.E. (1961) *Electronic Measuring Instruments*, McGraw-Hill, New York.

Spence, R. (1970) *Linear Active Networks*, Wiley, Chichester.

Squires, T.L. (1967) *Beginner's Guide to Electronics*, Newnes–Butterworths, Sevenoaks.

Starr, A.T. (1959) *Electronics*, Pitman, London.

Studer, J.J. (1963) *Electronic Circuits and Instrumentation System*, Wiley, New York.

Sydenham, P.H. (1979a) *Measuring Instruments: Tools of Knowledge and Control*, Peter Peregrinus, London.

Sydenham, P.H. (1979b) *Electronics – It's Easy (3 Vols)*, Modern Magazines, Sydney.

Sydenham, P.H. (1980) *Transducers in Measurement and Control*, Adam Hilger, Bristol.

Szentirmai, G. (1974) *Computer-aided Filter Design*, IEEE, New York.

Terman, F.E. and Pettit, J.M. (1952) *Electronic Measurements*, McGraw-Hill, New York.

Thomas, H.E. (1967) *Handbook of Electronic Instruments and Measurement Techniques*, Prentice-Hall, Englewood Cliffs, NJ.

Thornton, R.D., Linvill, J.G., Chenette, E.R., Boothroyd, A.R., Willis, J., Searle, C.L., Albin, H.L. and Harris, J.N. (1966) *Handbook of Basic Transistor Circuits and Measurements*, Wiley, New York.

Turner, R.P. (1959) *Basic Electronic Test Procedures*, Holt, Rinehart and Winston, New York.

Turner, R.P. (1963) *Basic Electronic Test Instruments*, Holt, Rinehart and Winston, New York.

Turner, R.P. (1977) *ABC's of Integrated Circuits*, H. W. Sams, Indianapolis, IN.

Turner, R.P. (1978) *ABC's of FET's*, H. W. Sams, Indianapolis, IN.

Van der Ziel, A. (1977) *Nonlinear Electronic Circuits*, Wiley, New York.

Van Santen, G.W. (1967) *Electronic Weighing and Process Control, Philips Technical Library*, N. V. Philips, Gloeilampenfabrieken, Eindhoven.

Van Valkenburg, M.E. (1974) *Circuit Theory: Foundations and Classical Contributions*, Dowden, Hutchinson and Ross, Stroudsburg, PA.

Wakerly, J.F. (1976) *Logic Design Projects Using Standard Integrated Circuits*, Wiley, New York.

Waller, W.F. (1972a) *Electronic Measurements*, Macmillan, London.

Waller, W.F. (1972b) *Electronic Testing and Measurement*, Macmillan, London.

Waller, W.F. (1972c) *Electronic Component Testing*, Macmillan, London.

Waner, W. (1979) *Trouble Shooting Solid-state Circuits and Systems*, Reston (Prentice-Hall), Englewood Cliffs, NJ.

Waters, F.J. (1978) *ABC's of Electronics*, H. W. Sams, Indianapolis, IN.

Wedlock, B.D. (1969) *Electronic Components and Measurements*, Prentice-Hall, Englewood Cliffs, NJ.

Willson, A.N. (1975) *Nonlinear Networks: Theory and Analysis*, Wiley, New York.

74: Basic Electronic Components

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Materials Used in Electrical and Electronic Regime 469
2 Conductors and Insulators 469
3 Semiconductors 470
4 Passive Elements 471
5 Active Devices 472
6 Supplementary Devices 474
Related Articles 475
References 475

1 MATERIALS USED IN ELECTRICAL AND ELECTRONIC REGIME

The features of materials and their use that places them in the electrical and electronic regime are their electrical characteristics. Suitable components and assemblies, however, must be made of some form of substance, and design will always rest on possession of knowledge about other properties of both electrical and nonelectrical materials. For example, a transistor is formed of silicon, shaped and arranged with various electrical properties, arranged in various positions to which conducting leads are attached to make electrical connection. The whole is then mounted in a suitable container. Thermal and mechanical requirements of the transistor are as important as electrical characteristics.

Materials are formed of combinations of the basic elements. Atoms form the smallest level of discreteness that generally concerns the electronic and electrical component designer. Atoms comprise electrons, of relatively small mass, that are in orbit around a central, much larger mass, nucleus formed of protons and, perhaps, neutrons. The electron carries, by convention, the negative electrical charge, the nucleus possessing positive charge that retains one or more electrons attached to it.

Normally, charges balance to neutrality, and they usually attempt to establish this position in a material system. Removal of an electron leaves an atom positively charged; this is the positive ion. Addition of an electron forms a surplus of negative charge; this is the negative ion. Atoms are combined to form larger molecules, and these can form myriads of compounds or mixtures that make up all solids, liquids, and gases.

When material is formed of charge-stable arrangements of similar, or different, elements, it becomes unclear as to which atom is attached to which atom of the matrix. This (simplified explanation) results in a matrix formed of immovable atom nuclei between which there exist electrons moving around in what is known as the *sea of electrons*.

This primitive depiction of the make up of materials is far from the complete, real situation but does provide for a superficial understanding of the classes of materials that find different use in electrical and electronic technique.

2 CONDUCTORS AND INSULATORS

A material in which the sea of electrons is very mobile, in which any particular electron can easily be swept away and another attracted into the space resulting, is known as a *conducting material*. Typical electrical conducting materials are the metals including silver, copper, aluminum, gold, and tin. It only requires a small energy level to cause a flow of electrons, the current, in these materials, hence the name conductor. As well as solids can, certain liquids and gases can be good conductors under certain conditions.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

Some materials require very high energy potential to obtain electron flow. These are the insulators or resistors. Rubber, wood, glass, most plastics, and ceramics are typical insulating materials. Liquid and gaseous insulators also exist.

Conductors and insulators are at the two ends of the range of resistivities of materials. In reality, there are different materials available with specific resistivities covering the range from one extreme to the other.

Electrical engineering in the main is (or was so in the past) concerned with insulators and conductors and their application. The former are used in suitable arrangements to contain charge (Figure 1) and its flow so that electric current flows in space where it is required. In a wire transmission system, for example, conducting wires are insulated to ensure that the electrical energy arrives at the desired destination with as little loss as possible. In an electrical generator, insulators force current to flow in conducting wires that are revolving in suitably placed magnetic fields.

Materials that form insulators and conductors generally obey the linear Ohm's law given by

$$V = RI \tag{1}$$

where V is voltage potential applied to exert forces upon the electrons, I is the electric current that flows, and R is the resistance to flow, a constant of proportionality for the material.

The electrical resistance between two points on a piece of material is decided by its specific resistance, the cross-sectional area of the material, and the distance between end connections.

3 SEMICONDUCTORS

These form a class of material in which current flow is not necessarily proportional to impressed electrical potential. They have resistivity values just above those of metals. The physics of semiconductors can be most complex. Fortunately, electronic systems development is more likely to need only understanding of what semiconductors can do than of why and how they actually provide their characteristics. For this reason, it is not necessary to provide an in-depth study here. As time passes, the electronic systems designer moves further away from such fundamental considerations, the only people concerned being device

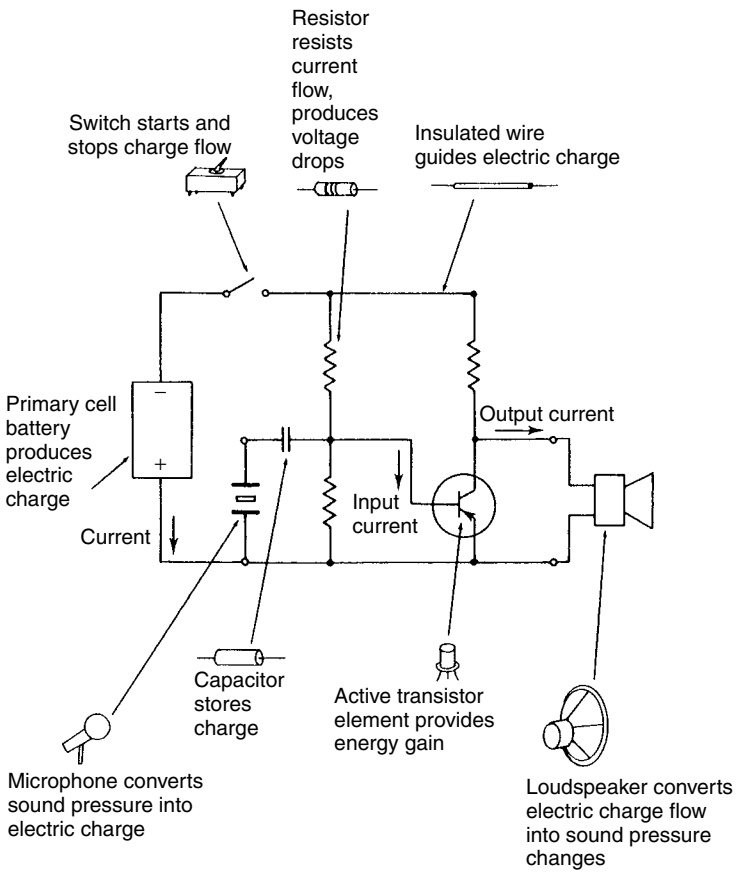


Figure 1. In a simple circuit (a sound amplifier is shown), selected materials provide electrical charge, controlling current flow into the speaker where it is converted back into acoustic energy at greater than input level.

developers and research workers seeking new devices and principles.

Semiconducting materials can usually be distinguished from conductors by their property that thermal energy will cause electron movement, which is not the case for conductors. This is usually (again simplistically) discussed in terms of the energy gap existing in the semiconducting material. If additional energy causes electrons to vacate their positions, causing current flow, the material is denoted as *n-type*. The spaces left are termed *holes*. A second form of current flow can occur in which adjacent valency electrons move to fill the holes. The materials in which this happens are termed *p-type*. As electrons move, they leave holes, thus giving the appearance that the holes migrate.

Suitable additions of impurity into extremely pure semiconducting materials (germanium was first used but silicon is now more generally the basic material in semiconducting devices and with gallium-arsenide for high-temperature uses) can enhance the n-type and p-type behavior.

When a conductor is interfaced physically with a semiconductor, or two of the latter materials are so connected, a flow of electrons occurs across the boundary, reaching some form of equilibrium situation. These connection regions are called *junctions*.

4 PASSIVE ELEMENTS

Subsystem components used to build energy conversion systems can be divided into passive and active kinds, these two being combined into a system to provide the required response.

Passive components are those devices that operate upon signals to provide conversion between voltages and current without being able to increase the output energy level beyond that of the input signal. The output signal energy is always less, because of losses, than the input signal energy. A combination of resistors, for example, can be used to attenuate a signal's energy but not to increase it. It is, however, possible to transform the energy product of voltage and current so that one is made larger than the original form at the expense of the other that is then made smaller. Examples of such transformations are the voltage step-up, or step-down, transformer, and the resonant circuit.

The mainly used passive components are the resistor, the capacitor, and the inductor. The resistor acts to dissipate energy and to resist energy flow by wasting what it does not allow to flow through it. The capacitor and the inductor, on the other hand, can store energy; they do possess losses but these are normally small compared with the storage energy capability. It is the energy storage property that enables signal frequency processing to be obtained. Purely resistive systems exhibit zero-order dynamic behavior; capacitive and inductive systems can be used to realize higher-order dynamic systems as their number is increased in suitable arrangements. The behavior of systems of passive components forms much of the material of introductory electrical engineering.

Calculation of circuit behavior for resistive systems is relatively simple compared to those circuits having storage elements. The simpler resistive circuit theory can, however, often be applied to storage element systems, provided the storage element is considered as a frequency-varying impedance.

The capacitive reactance (its apparent resistance) of a capacitor is given by

$$X_{\mathrm{C}} = \frac{1}{2\pi fC} = \frac{1}{\omega C} \tag{2}$$

and the inductive reactance of an inductor by

$$X_{\mathrm{L}} = 2\pi fL = \omega L \tag{3}$$

where f is the frequency of interest, in cycles per second (Hz), ω the angular frequency in radians per second (rad s^{-1}), C the capacitance in farad (F), and L the inductance in henry (H).

The time-transient behavior of either a capacitor or an inductor combined with a resistor is a first-order response, that is, the rise or fall of signal level to a step input follows an exponential change given by the time constant τ of the combination in question; ($\tau = RC$ or R/L). Resistance acts to alter the time constant and it gives rise to losses in RC and LC combinations. Such losses are not necessarily of importance. Much of electronic engineering makes deliberate use of practical arrangements in which loss occurs, this being quite unlike power electrical engineering where reduction of losses is generally a prime design target. In recent times, engineering of low energy level electronic information signal systems has also concentrated on less lossy arrangements as a means to increase the density of components and to reduce the size and cost of the necessary power supply.

Networks comprising only resistances can be converted into equivalent networks or be reduced to a single equivalent resistance value between two terminals. Inductors and capacitors can only be reduced to equivalent resistive values at a given frequency of operation and when one kind does not interact with others. Generally, this means only series or parallel combinations of a like kind (of inductor or capacitor) are reducible. When the two forms interact to form a second-order or higher-order system, reduction is more complex as it must take into account the vectorial, amplitude, and phase nature of the two components.

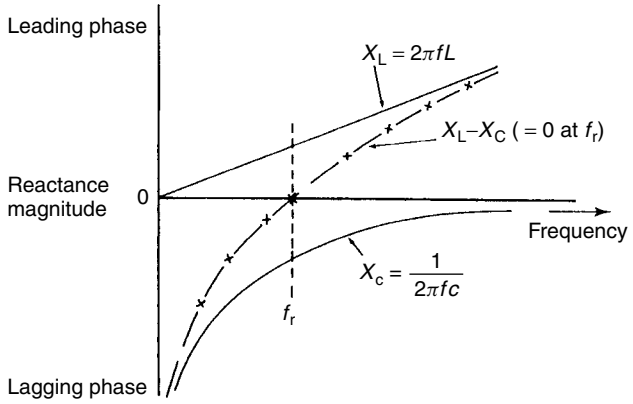


Figure 2. Behavior of inductive and capacitive reactance with frequency f_r is the resonant frequency.

When the two different forms of storage elements are combined, second- or higher-order systems are formed. Losses arising from resistive elements can still be significant but they are, there, usually of secondary importance in basic theoretical considerations.

Figure 2 shows how inductive and capacitive reactance vary with frequency. At a certain frequency, called the *resonant frequency*, where the two reactances X_L, X_C are equal in magnitude, their vector addition results in a combined impedance of zero. This is because they each possess an opposite phase angle. In terms of the complex algebra description, the imaginary components cancel, leaving only the zero, real component to resist current flow.

Resonance, therefore, occurs at

$$\omega L = \omega C \qquad f_r = \frac{1}{2\pi\sqrt{(LC)}} \tag{4}$$

where ω is the angular frequency, f_r the resonant frequency, L the inductance in henry, and C the capacitance in farad.

In practice, no resonant system is free of loss; the real, resistive part has a finite value. This acts to limit the current in the resonant circuit at resonance: the less the resistive component, the higher the series resonance current.

Resonance using an inductor and a capacitor can be established as a series or as a parallel connection of the two, and each of these can be used in a series or a shunt connection in a system. The availability of four options can lead to confusion about how a resonant stage influences the system. It is, therefore, necessary to verify which connection is used. In electronic method, where power loss is often permissible, shunting is commonly used to selectively remove signal frequencies that are not needed. This apparently wasteful but useful practice in electronic systems could not be tolerated in power electrical and electronic engineering.

As the windings of an electrical transformer form an inductor, it is possible to resonate each winding with a capacitor. This forms the tuned transformer, each or one side only being used to increase the sharpness of the frequency selection process provided by resonance. Selectivity is quantified in terms of the Q-factor. The higher the Q-factor, the sharper the resonance; its magnitude depends very much upon the resistance losses in the resonant system. Such losses can be made up by the use of active amplifying elements.

While the resistor, the capacitor, and the inductor are very common in electrical and electronic systems, there do exist other passive elements. Some of these are the transformer, rectifier, temperature sensitive resistor (thermistor), light-generating device, the panel indicator, and other transducers. These, however, can each be reduced, in theory, to combinations – called *the equivalent circuit* – of resistance, inductance, and capacitance.

They are consequently generally regarded in theoretical circuit considerations in an equivalent form. A thorough understanding, therefore, of the characteristics of resistors, capacitors, and inductors, and their combination forms a most basic and important part of electrical and electronic method.

5 ACTIVE DEVICES

Networks formed only of passive devices can be devised that will provide much of the signal processing needed in electronic systems. However, their inability to maintain the signal energy level at, at least, that driving the system stage is a serious disadvantage for the signal progressively degrades down to the noise energy level and becomes lost. Furthermore, the signal input energy may be insufficient to drive the required output transducer. As an example, a signal can be mathematically integrated reasonably well with a simple RC, low-pass filter stage. To obtain accurate integration, however, it must also provide high attenuation. Some method must usually be added to the output to restore the signal level – this is called an *amplifier*. (Early circuit designs did just this but it is often more powerful to use the amplifier in an alternative less-obvious arrangement to obtain an improved overall performance.)

If the network has the ability to make up the circuit losses or increase the energy at the output to more than occurs at the input, the arrangement is known as an *active system*.

Any arrangement that has active properties is capable of providing amplification. Amplifiers are able to provide larger energy signals at their output than occur at their input. At first sight, it appears that the network provides the extra energy needed, but in reality the amplifier acts as

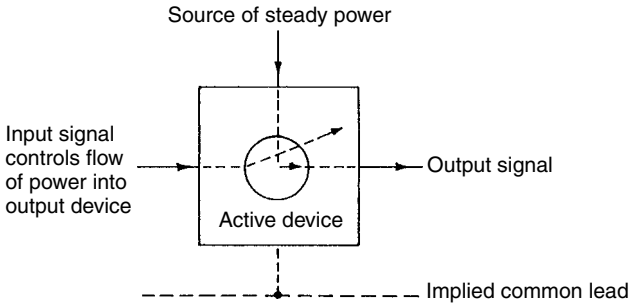


Figure 3. Generalization of a three-terminal amplifier element.

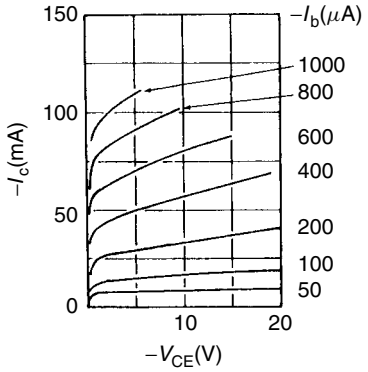


Figure 4. Characteristic curves for a typical, small-signal p–n–p transistor. I_c is the collector current, I_b is the base current, and V_{cc} is the collector to emitter voltage.

a control that allows flow of energy from a supply into the output in accordance with some relationship with the input signal. The output signal is often not electrically connected to the input signal. Figure 3 shows the generalization of an amplifier.

In electronic systems, amplification, also called *gain*, is used in many ways for many purposes. Each of the uses has a different terminology. For example, a preamplifier conditions a low-level signal ready for connection to a following stage. A stage inserted to obtain a more optimum stage match is called a *buffer amplifier*; as the name implies, it buffers one stage from the other so that they do not interact. At the output end of the cascaded system, there is often a need for a power amplifier; this raises the signal power providing the specific matching needed to suit, say, the actuator coupled to the output.

Another is the instrumentation amplifier. These are vaguely defined units that may carry out any, or all, of these functions: the term generally implies that they are able to preserve signal purity and thus retain the meaning of measurement signals passing through the system chain. The operational amplifier is an instrumentation amplifier designed to use feedback in an analog computational mode: they are popularly called *op-amps*.

Each of the above active amplifiers needs to be designed to suit the application in question. Although there are a certain number of basic designs in existence, the specification of an amplifier can only be achieved properly and satisfactorily when the task and the amplifier are well understood.

It is the availability and use of active devices for processing information that largely distinguishes electronics from electrics. Prior to the thermionic valve era, that began around 1910, it was very difficult indeed to obtain gain in a system. The thermionic valve overcame that prime need, allowing numerous active networks to be devised and implemented. It enabled electronic technique to develop. Following the valve came the solid-state semiconductor transistor element. This performs the same role as did valves, operating it in quite similar ways, although not at the electron level of understanding. Electrical method also has some active elements available. The magnetic amplifier, the rotating electromechanical amplidyne amplifier, and the Ward–Leonard motor-generator arrangement are methods for obtaining gain. They are, however, not practical at very small power levels, the devices being too large compared to even valves for them to have been adopted in signal processing in the widespread manner that transistors have over the past years.

The most basic amplifier, in the general sense, is a three-terminal device having characteristics between the three terminals that can be expressed along the lines shown in Figure 4 for a representative electrical component. Transistors, valves, and any other amplifier can be described in this way. Gain occurs when signal input energy can be used to reproduce the same signal at a larger power level at the output. Several other forms of characteristic describing curve can be drawn.

The transistor is the most-used active electronic element today. It is capable, in different forms, of handling very small (picowatt) signals through to very large (kilowatt) signals. It is not practical to use exactly the same actual transistor assembly for all cases; the designer chooses that unit which is most suitable for the purpose in question. Since the development of the transistor, there have appeared many other active semiconductor devices that can provide gain. Each has its special virtues; devices include the silicon controlled rectifier (SCR), the field-effect transistor (FET), the thyristor, and many more. Sophisticated junctions have been devised using a variety of manufacturing methods, these including multiple junction devices. For most practical purposes, it is only necessary to learn about the actual physics of operation within the device on rare occasions. Generally, use of makers' published characteristic curves and numerical values suffices.

As it stands, the transistor, as were valves, is not capable of amplifying AC signals centered at a zero DC voltage

level. This is because the basic transistor p-n-p, or n-p-n, junctions can only pass current in one direction. Only one polarity of the alternating current signal waveform fed into the base connection will be amplified, the other being severely attenuated. It is, therefore, necessary to add extra passive components to the transistor to obtain practical operation. The design of transistor amplifiers used to be very important to electronic system builders but the introduction of low-cost, high-quality integrated-circuit amplifier modules enabled highly specialized amplifier designers to apply considerable effort to the design and to use many more active elements in the design than would have been used in former times. Thus, it is common practice now to assemble systems using commercially available units chosen to suit the need. The user needs to know little about internal amplifier design; only the external characteristics.

The active element can be used to produce gain for use with continuous linear signal analog systems, or it can be used to provide gain through a switching action for use in digital systems. In early times (1840s onward), the electromechanical relay was able to provide gain; it was a prime component in enabling telegraphy to be developed so successfully.

The design of amplifiers is covered in many texts. Of those already listed, refer to Brophy (1977), Carson (1961), Clayton (1971), Cleary (1969), Csáki *et al*., (1975), Delaney (1969), Faulkenberry (1977), Graeme and Tobey (1971), Heathkit (1978), Lowe (1977), Malvino (1967), Millman and Halkias (1976), Senturia and Wedlock (1975), Sydenham (1979), and Thornton *et al*. (1966).

6 SUPPLEMENTARY DEVICES

To the range of passive and active devices mentioned so far must be added a very wide range of supplementary components and devices. These are combined with the previously described active and passive elements to form electronic instrument (and other) systems. It is not practical, or necessary, to cover all of these as a description of a few will give the understanding needed here.

The variety available is most easily seen by reference to electronic component suppliers' catalogs. Electronics always has had a popular following. It began with radio in the 1920s, being added to by audio equipment and television interests. Popular interest then moved somewhat toward all manner of instrumentation for the domestic person's entertainment and use. A typical suppliers' catalog has around a hundred pages of listings.

Systems are generally built assembling the commonly available parts but design needs, where a very large market is assured, will allow new products to be considered. An example are new modules that are appearing as the result of electronic technique being used to a greater extent in mass-produced motor vehicles.

Thermionic devices, although no longer used much for basic amplification, still find some places where they have not yet been replaced by superior solid-state equivalents. Examples are the cathode-ray tube used in television monitors and in the cathode-ray oscilloscope and the older forms of video camera. Thermionic devices are also used to obtain spectral emission in gas analysis equipment. The electron microscope also requires electron emission from a cathode.

Such supplementary devices are not normally built by the less specialized user but are obtained ready to use. If they fail, they are replaced rather than being repaired.

Another example of a supplementary device is the number and letter alphanumeric display device. Such devices again are purchased ready to wire into the circuit. Originally, these were manufactured by the user and were a significant part of the cost of a system. Today, they are often integrated into the complete, ready-to-use, subsystem module complete with any decoding and power drivers required.

Electronic systems need mechanical frames to support the various circuit boards, switches, variable controls, and connections to other stages. These are called *chassis* in older terminology and more recently *frames*. Most network systems are today formed in their final version on printed-circuit boards. These are boards in which all wiring connections between the commercially made components are made by selective etching of a copper coating to form conducting tracks. Holes are then drilled in the copper tracks, so formed, to take the leads of the components, which are then soldered into the copper by hand, or solder-flow methods. Wire-wrapped connections are also used, especially where greater reliability is called for. Printed wiring boards, called *p.w.b. boards*, are also used with wiring on both sides, selected holes being plated through to connect both sides where needed. The actual layout of components on a board more often is to suit manufacturing convenience and does not then follow the schematic block diagram layout.

Printed wiring boards are often connected into the complete system using plug and socket connectors. This makes for easy repair and faultfinding. It also modularizes the system design.

A smaller circuit assembly can be made by printing suitable materials onto a usually ceramic substrate – called *printed circuitry*. These plus, vacuum-deposited conducting metal connecting tracks, make up the passive part of hybrid circuits. To these are added unencapsulated semiconductor active elements formed by the monolithic process. A complete hybrid circuit may then be totally encapsulated to provide environmental protection.

Still smaller assemblies are made by the monolithic methods using masks to lay down progressively conductors, junctions, and components onto the usually silicon chip base. A very large capital investment is needed to create integrated circuits (ICs). The number of circuit elements possible on a single IC now ranges from the original 15 or so to greater than one million.

A myriad of major and minor supplementary components are also used (see Figure 5). Some have already been mentioned: the switch, plug and socket, printed wiring board. It is the ready, and relatively inexpensive, availability of electronic components that allows people with relatively little training to form quite sophisticated electronic instrumentation systems. In many instances, the module required can be purchased in kit form ready for insertion of the components onto a ready etched and plated, printed wiring board. Very little skill is needed to assemble these. The need for skill and understanding arises if they do not function properly!

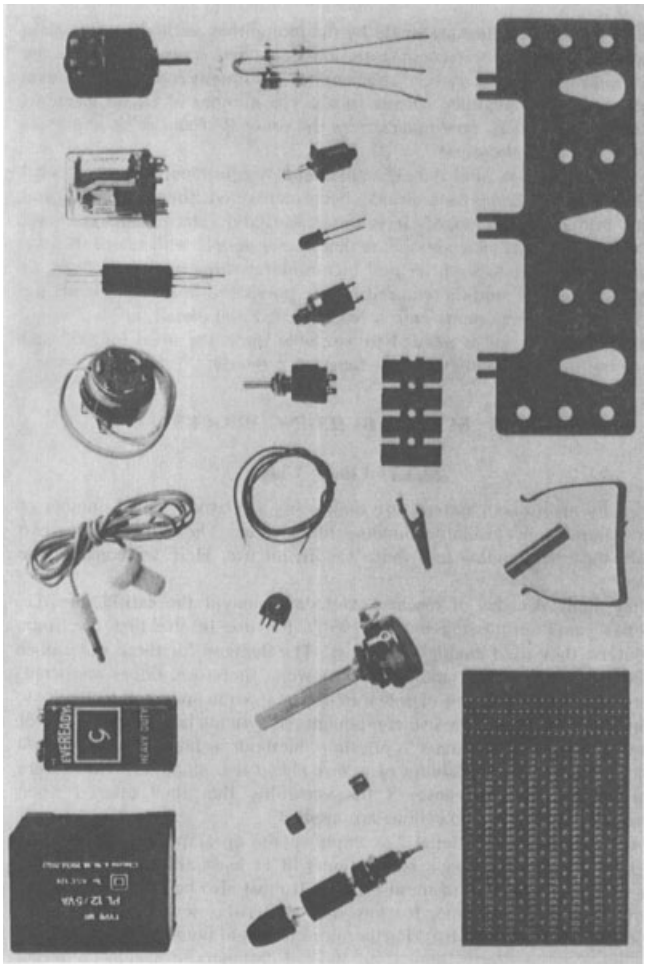


Figure 5. A range of supplementary devices used to form electronic systems.

RELATED ARTICLES

Article 73, Overview of Electrical and Electronic Technique, Volume 2; **Article 75, Electronic System Building Blocks, Volume 2**; **Article 76, Electronic Systems Design, Volume 2**; **Article 122, Operational Amplifiers, Volume 2**; **Article 123, Instrumentation Amplifiers, Volume 2**; **Article 124, Specialized Amplifiers for Measurement Systems, Volume 2**.

REFERENCES

Brophy, J.J. (1977) *Basic Electronics for Scientists*, McGraw-Hill-Kogakusha, Tokyo.

Carson, R.S. (1961) *Principles of Applied Electronics*, McGraw-Hill, New York.

Clayton, G.B. (1971) *Operational Amplifiers*, Butterworths, Sevenoaks.

Cleary, J.F. (ed.) (1969) *Transistor Manual*, General Electric Company, Chicago, IL.

Csáki, F., Ganszky, K., Ipsits, I. and Marti, S. (1975) *Power Electronics*, Akadémiai Kiadó, Budapest.

Delaney, C.F.G. (1969) *Electronics for the Physicist*, Penguin Books, Harmondsworth.

Faulkenberry, L.M. (1977) *An Introduction to Operational Amplifiers*, Wiley, New York.

Graeme, J.G. and Tobey, G.E. (1971) *Operational Amplifiers – Design and Application*, McGraw-Hill, New York.

Heathkit (1978) *Electronic-Courses for Learn-at-Home: Course 1 DC Electronics; Course 2 AC Electronics; Course 3 Semiconductor Devices; Course 4 Electronic Circuits; Advanced Digital Technique Course; Microprocessor Course*. Available with experimental back-up and optional cassettes, Heath-Schlumberger, London.

Lowe, J.F. (1977) *Electronics for Electrical Trades*, McGraw-Hill, Sydney.

Malvino, A.P. (1967) *Electronic Instrumentation Fundamentals*, McGraw-Hill, New York.

Millman, J. and Halkias, C.C. (1976) *Electronic Fundamentals and Applications*, McGraw-Hill, New York.

Senturia, S.O. and Wedlock, B.D. (1975) *Electronic Circuits and Applications*, Wiley, New York.

Sydenham, P.H. (1979) *Electronics – It's Easy*, (3 Vols.), Modern Magazines, Sydney.

Thornton, R.D., Linvill, J.G., Chenette, E.R., Boothroyd, A.R., Willis, J., Searle, C.L., Albin, H.L. and Harris, J.N. (1966) *Handbook of Basic Transistor Circuits and Measurements*, Wiley, New York.

75: Electronic System Building Blocks

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Linear Units 476
2 Amplifier Terms 476
3 Digital Units 479
4 Gates 479
5 Astable Device 480
6 Bistable Device 480
7 Monostable Device 480
8 Inverter Device 480
9 Schmidt Trigger and Comparator 480
10 Overall Features of Digital Systems 481
Related Articles 482
References 482

1 LINEAR UNITS

Electronic systems are built using a relatively small number of various kinds of building block units. They are divided into two main groups: those using analog signals and those using digital operations. Here the former is considered.

After many decades of research and development, a satisfactory general purpose amplifier came into being in the 1950s for use in the first electronic computers: they used analog techniques. The demand for these was rather specialized and their cost and availability were, therefore, rather restricted. However, the introduction of integrated circuit semiconductor technology in which the active and passive components of systems that have many tens to hundreds of elements in their construction and are made by mass replication methods using photomechanical methods, led to the availability of a very cheap and stable DC amplifier. These were taken up en masse because of the versatility that they offered, when appropriate feedback connections are applied.

2 AMPLIFIER TERMS

The satisfactory operational amplifier, the op-amp, must have open-loop gain (without feedback connections) of at least around 10 000, and draw a relatively insignificant input current. It must also be reasonably stable with temperature and time, for any drift of its DC levels internally will provide a related output drift. Furthermore, it should be able to handle both polarities of input signal. Simple integration allowed designers to package a design (Figure 1) with around 20 active junctions, plus some 12 passive resistors, and a few capacitors. Inductors are generally avoided in integrated circuits due to their relatively large size in the integrated form at low frequencies of operation. The cost of the IC op-amp is so low as to now be almost insignificant when compared with assembly costs to place it into use.

Naturally, a general purpose op-amp may not be suitable for more exacting uses and more expensive designs are marketed for such cases. The selection of the op-amp must be made to suit the application. Guidance is generally given in texts on op-amp use and in makers' application notes.

To understand why the op-amp is so versatile, consider the general case of an amplifier with a single feedback path and multiple inputs, as shown in Figure 2. To simplify this explanation, only resistors are considered but in practice the same derivation can be made for complex impedances in place of the resistors.

It can be shown that the transfer characteristic between the inputs $V_1, \ldots, V_n$ and the output V_o, if the amplifier

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

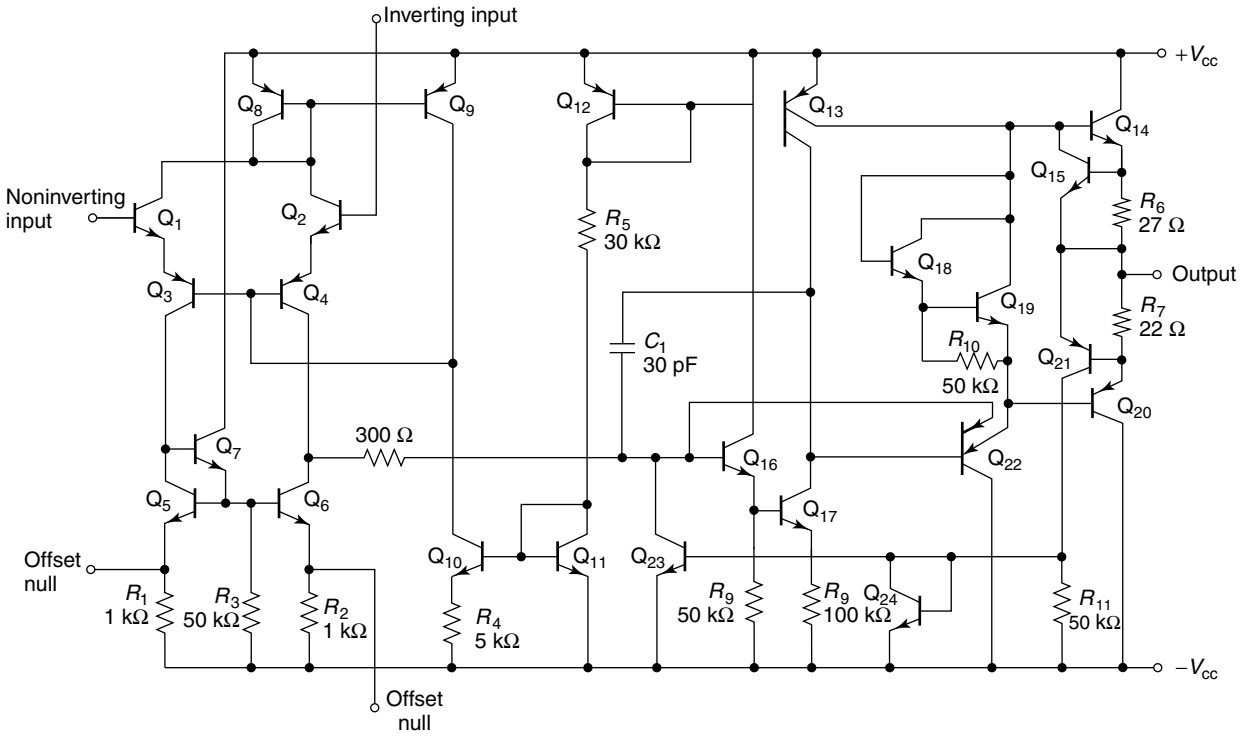


Figure 1. Schematic layout of circuitry of an integrated circuit amplifier.

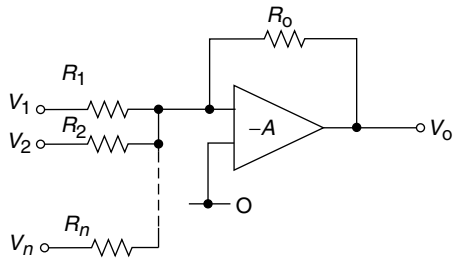


Figure 2. Generalized op-amp feedback connections.

draws negligible current, is given by

$$V_o = -\left(\frac{R_o}{R_1}V_1 + \frac{R_o}{R_2}V_2 + \cdots + \frac{R_o}{R_n}V_n\right) \times \left[1 + \frac{1}{A}\left(1 + \frac{R_o}{R_1} + \frac{R_o}{R_2} + \cdots + \frac{R_o}{R_n}\right)\right]^{-1} \quad (1)$$

This simplifies greatly if the gain A of the amplifier, in the open-loop state, is very much larger than unity, say at least 10 000. Then the above becomes simply

$$V_o = -R_o \sum_{i=1}^{n} \frac{V_i}{R_i} \quad (2)$$

This expression is, therefore, independent of the actual amplifier gain and shows that the gain of the feedback-connected op-amp is decided only by the ratio of the input and feedback resistors. These can be made very stable and accurate in value. As the gain of the op-amp is not a critical parameter, it can be seen that variation of the open-loop gain of the amplifier does not alter the feedback connection gain, this leads to increased stability by the use of feedback. The gain of the op-amp is made very large to allow for closed loop operation.

The gain stability and selectivity features are but one virtue of op-amp feedback usage. It can also be seen from the above expression that input voltages are summed according to a multiplier factor for each input channel. Furthermore, a single input signal appears at the output with the opposite polarity. This is called an *inverting arrangement*. Modern op-amps provide both inverting and noninverting inputs. Their output can swing around a zero voltage level by a typical ±10 V. They are rarely used in open-loop due to the existence then of extremely high gain (100 000 is typical), which has associated with it instability that rises in magnitude as the gain of the feedback system arrangements moves toward the open-loop gain value.

By suitable application of feedback connections, the op-amp can be used to mathematically integrate, differentiate, multiply, divide, summate, subtract, convert from voltage to current source, provide buffer amplifiers of many kinds, act as an analog signal filter, and more. The number of possible uses for linear circuit operation is enormous and

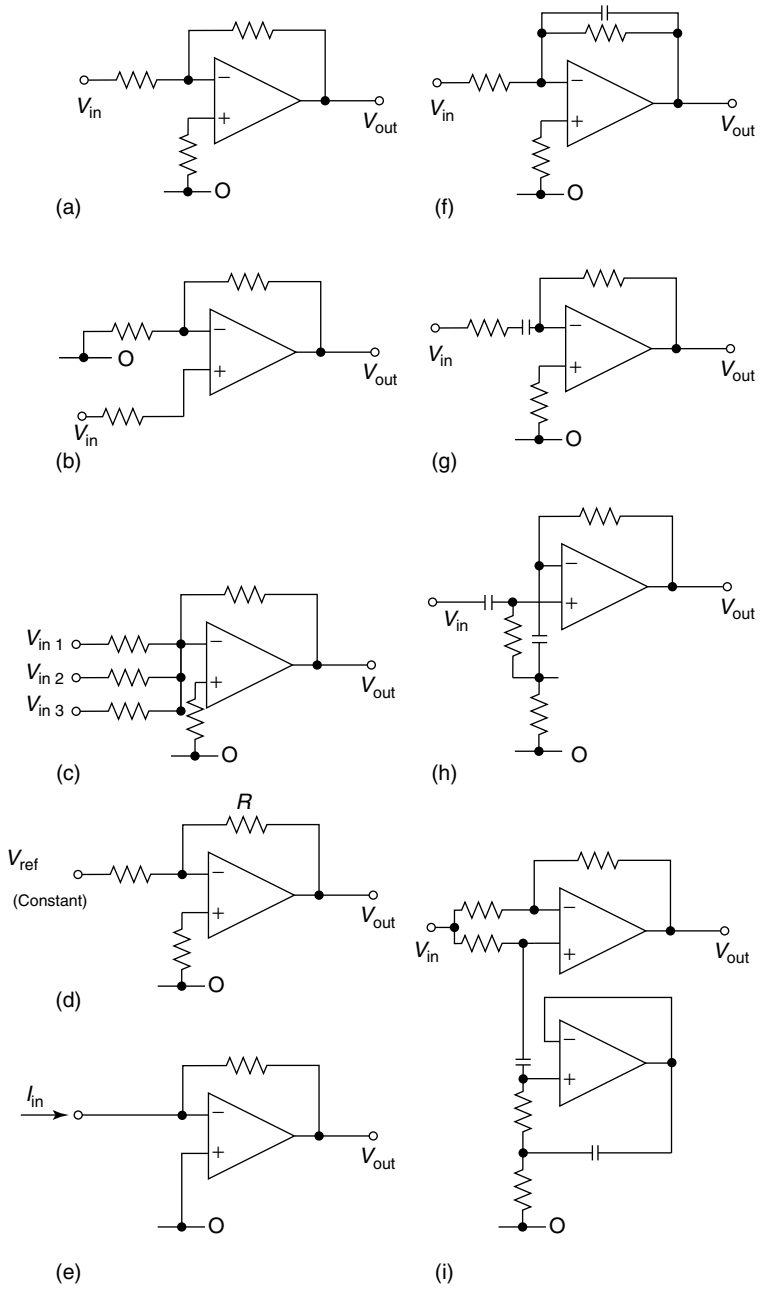


Figure 3. Some selected op-amp-based functional blocks (power supply connections not shown) (a) voltage amplifier, inverting; (b) voltage amplifier, noninverting; (c) voltage summation with amplification, inverting; (d) V_{out} proportional to R; (e) current to voltage converter; (f) integrator; (g) differentiator; (h) AC, voltage amplifier; (i) notch frequency filter.

grows continuously. Figure 3 shows some commonly used examples that illustrate this versatility of use.

Op-amps are usually selected and designed according to application notes provided by their makers and to designs published in the, now many, texts on their use. Application notes are generally obtained on request from the op-amp marketing agencies. Suitable texts concentrating on their use are Burr-Brown (1964, 1966), Clayton (1971), Connelly (1975), Faulkenberry (1977), Graeme and Tobey (1971), Huelsman (1977), Melen and Garland (1978), and NS Corporation (1976). Most general electronic texts that originated after around 1965 include articles on op-amp use. Several manufacturing companies specialize in high performance op-amp production; they also provide relevant users guides and design assistance.

The actual manufacture of integrated circuits is discussed in most introductory texts about general electronics. It is continuously changing in refinement and a bewildering

range of terms have been coined to describe the conceptual basis of the new technology. As it is not particularly important to the majority of electronic systems generation, for measurement applications it will not be discussed here. The user should not be daunted by a lack of ability to understand just how an integrated circuit of any kind operates internally. Schematic diagrams are usually published for ICs but these are of passing interest to the user, becoming important only when the information provided elsewhere on the data sheet and application note fails to satisfy the need. On the very rare occasion, it will be found that internal connections may give rise to unexpected operation when the IC is used in a system.

Op-amps and other amplifiers (buffers, servo, instrumentation, interface, and converter) form only part of the range of linear devices. Others are voltage and current regulators, analog-to-digital (A/D) and digital-to-analog (D/A) converter units plus sample-and-hold units all that are used in Data Acquisition Systems (DAS). Other linear modules include specialized radio, wireless, and audio integrated circuits; video amplifiers for television use arrays of transistors and diodes, timers, numerous different purpose units made for use in industry and automobiles (special experimental runs are sold for original equipment manufactures OEM to develop large volume products), display drives, phase-lock loops, and voltage level comparators. Some of these transcend the linear and digital boundary enabling conversion from one signal domain to the other, a very common practice.

3 DIGITAL UNITS

Information can also be processed using digital signals through application of logical principles. The foundation of these principles was laid by philosophers of the late nineteenth century, who had little inkling that one day their work would be applied to actual electrical hardware. Boolean algebra handles, in a mathematical sense, the interaction of logical situations.

These are typified by the use of simply two-state conditions, the most general and basic being connective statements, such as AND and OR. For example, a switch must be turned on if input A AND input B are energized, but not for either one of them. It could also be that the switch is on if A OR B is on.

This kind of thinking and problem solving was first developed to help philosophers conduct arguments on strictly objective logical lines. Their methodology was applicable to electrical switching networks and the first digital (computing) systems were formed using electromagnetic relays. These gave way to faster-acting thermionic valves and then to solid-state semiconductor devices, such as the transistor and the diode as the basic switching element.

The digital family of devices comprises relatively few basic units, but as can be expected they are made in many forms. In the 1970s, a considerable degree of standardization came about and two prime logic systems came to the fore. These are the ones using transistor–transistor logic (TTL) and those based upon complementary metal oxide semiconductor (CMOS) logic families of device. TTL systems are always used with high and low logic levels of nominally 5 and 0.5 V, respectively, CMOS, however, can be used at varying levels of supply. The range of digital devices also is often referred to as the *logic range*. Since then, all manner of logic types have been developed. Fortunately, to build digital systems one does not need such detailed internal knowledge. Basic digital building units are now described.

4 GATES

Systems that perform logical AND, OR and the complementary negated functions NAND, NOR plus some special cases, such as EXCLUSIVE OR, are called *gates*. As an example, two electrical switches in series will provide transmission of a signal if both are closed, that is, if switch A AND switch B are closed. This is denoted as A.B. Placed in parallel, A or B will give transmission; denoted as A+B. The negated NOT form results when the opposite switch state case is considered, the sense of the switches being reversed; denoted $\overline{A \cdot B}$, $\overline{A + B}$, $\overline{A}$, $\overline{B}$.

Modern practice no longer carries out gating functions with mechanical contact switches (some will be still encountered in special applications), the active semiconducting junctions being used as equivalent switches instead.

In use, logic ICs are provided with adequate buffering and suitable interfacing such that the user only has to first decide the logic gating needed and then connect the units accordingly. For large sets of gates – hundreds to thousands – more advanced gate-array methods are used. The internal design of logic units is, again, of little importance to the user. The number of gate functions available in a particular IC is decided by the number of pins available for connection to external circuitry, two being needed to supply power to the IC. For example, a 14-pin IC can provide 12 gate connections, which can be supplied by the maker as either four, two-input/one-output gates; three, three-input/one-output gates; or six, one-input/one-output signal inverters (these negate the logic polarity, but do not invert the signal polarity in the same way as does a linear signal inverter). Logic ICs usually contain only one kind of function on the IC chip; basic functions may be combined

together to form a higher-level function. The exception occurs when a much greater scale of integration (large-scale integrated circuits (LSI) to very large-scale integrated circuits (VLSI)) is used.

A basic theorem about logic gates shows that all digital logic systems can, in fact, be formed from just NAND gates. In practice, however, such methods are not cost-effective and makers subsequently offer a range of different ICs for appropriate selection. To make full use of its available logic elements, the gates on one physical IC unit are connected to different points of the circuit. Many of the functional units that are needed can be formed from gates alone, but again it is generally more practical to use purpose-designed units.

Two active elements can be so interconnected that there exist two inputs and two outputs with a large degree of positive feedback between them. The interstage passive element coupling used between the two elements decides the resulting function. This family was originally described as the multivibrator set. The three forms are the astable, bistable, and monostable. These provide three other basic digital system building blocks.

5 ASTABLE DEVICE

If the interstage coupling between two amplifying units (transistor) is capacitively, or AC, coupled to provide a large amount of positive feedback from one active element to the other, the system will oscillate from one side in the low state, with the other at the high state, changing the states cyclically to form a square-wave output source. The parameters of the interstage circuit largely control the frequency of the switching. These units are also called *clocks* (because they are used to pace in a digital computational system) or square-wave generators.

6 BISTABLE DEVICE

If the two stages are DC connected to provide the positive feedback to the other, the system will rest with one side high and the other side low. A change of the state occurs if one input is fed with sufficient energy to toggle the system to the other state. Such a unit can divide pulses by two as each output state only repeats a state after two input transitions. In early bistable designs, true short-duration pulses of energy were used; today, it is the transition energy of a signal going from a high to a low state (or vice versa) that operates these units. These are more commonly called *flip-flops, binaries, two-states,* and *memories*. They can be used to store a binary state for a chosen output that remains at a given state until toggled, or the power is disconnected. Several forms of flip-flop IC are available, some include gating into its inputs.

7 MONOSTABLE DEVICE

If one intercoupling is made capacitive and the other a DC link (combination of the astable and the bistable connections), the system will toggle over for an input transition returning to the first state after a time period set by the circuit's passive components. This is able to provide a small time delay to digital signals or to restore the squareness of incoming signals as a repeater for digital signal restoration. These are called *one-shots,* or *monostables*.

In each of the above multivibrator units, there are two outputs, one being the negated or complementary output of the other. Either output can be used depending on the logic sense needed. This often allows an inverter to be omitted to create the required logic gate system.

8 INVERTER DEVICE

Already mentioned is another basic digital unit, the inverter. This is used to invert the logic sense of a signal. The need for this often arises in digital system design.

Inverters also can act to buffer one stage from those connected to it. Digital systems often require many connections to an output. The ability of a unit to drive a given number of other inputs is expressed by its fan-out. In TTL systems, the number refers to standard TTL inputs; other uses have to be decided on their merits.

9 SCHMIDT TRIGGER AND COMPARATOR

When a varying level analog signal needs to be quantized into the two standard digital levels, the Schmidt trigger can be used. This unit provides a rapid square-edge output transition from a high to a low, or vice versa, state at a given level of input of analog signal. The trigger level for a rising signal may be higher than for a falling signal, the difference is called *backlash*. Backlash is often deliberately enhanced to give the trigger system a window in which it will not keep retriggering until the signal level has changed its analog level by more than a given amount.

A similar element is the comparator. In this, two voltage levels are compared. If one exceeds the other, the output toggles. By fixing one as a reference voltage, a comparator toggles at that level, as does a Schmidt trigger.

10 OVERALL FEATURES OF DIGITAL SYSTEMS

The above listed units form the basic building blocks of all digital systems. Figure 4 illustrates this point. Digital systems design rests on understanding their characteristics at the conceptual level. Where costs justify it systems formed of several of these units are manufactured as large-scale integrated (LSI) or even larger very large scale (VLSI) integrated circuits. With time, increasing numbers of basic functions are being made available as more sophisticated and powerful, higher level systems units. They are the subject of **Article 76, Electronic Systems Design, Volume 2**. For many projects, it is economical to design and make what are called *application specific integrated circuits* (ASICs).

Digital systems convey their information primarily as the timings of the level transitions; not by their signal levels. The actual voltage, or current, levels of the square waveforms are not of the same significance as they are in analog systems. For this reason, rise and fall times are kept as short as possible. Modern systems will make the transition in fractions of microseconds, that is, in nanoseconds. Circuit connections and component layout can degrade the switching edge. For this reason, high-speed circuits must be carefully laid out and interconnected inside their integration layout. Long interconnecting leads are especially bad unless designed appropriately. The fast rise times of the signals also lead to generation of stray signals. These can influence other parts of the circuit by electromagnetic induction. Furthermore, the energy transients within a stage can cause influencing transients on the power supply line. This would affect other units in the system. For this reason, it is usual practice to add decoupling filters to each small group of ICs to prevent signal unwanted transients disturbing the operation of others on the same mount.

The testing procedures for digital systems are quite different from those of analog circuits. First, information is only available about logic levels at various positions in the circuit. Display, using a logic analyzer, of numerous states is used to fault-find a digital system. Another feature is that the system can be stepped, or clocked on, one step at a time at any speed that is convenient. Analog systems must act in real time to provide proper operation of many of the circuit functions. Timing between stages is important in some digital designs and time delay may need to be considered in what is called a *race condition*.

The design of digital switching systems can be made very sophisticated using well-established logical mathematical and graphical procedures. Design is largely aimed at reducing the presence of redundant logic gates that generally occur in the first realized logic network design. In many applications, however, it is often less expensive overall to leave redundancies in the system; their location and elimination can be very time consuming and requires an in-depth understanding of switching theory.

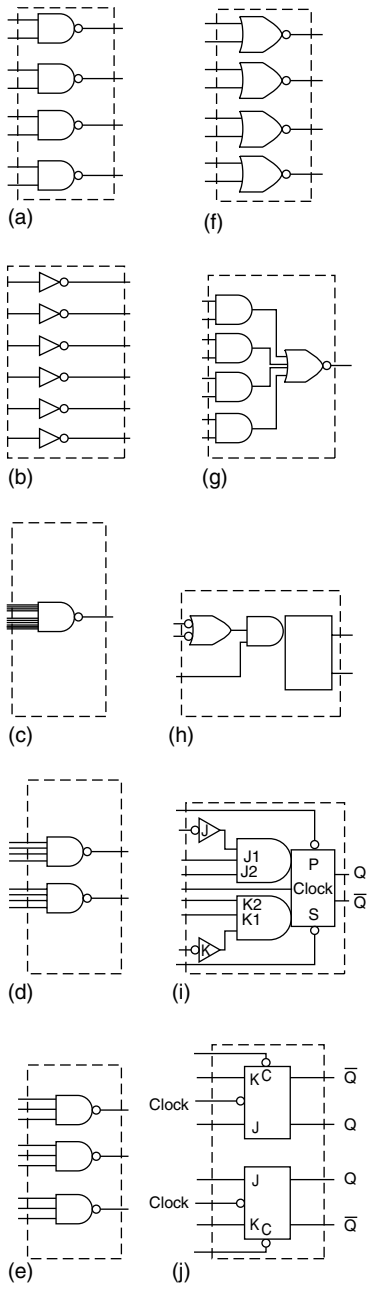


Figure 4. Schematic diagrams of some typical integrated digital circuits: (a) quadruple two-input positive NAND gates; (b) hex inverters; (c) eight-input positive NAND gate; (d) dual four-input positive NAND gates; (e) triple three-input positive NAND gates; (f) quadruple two-input positive NOR gates; (g) four-wide two-input AND–OR invert gates; (h) monostable multivibrator; (i) edge-triggered j–k flip-flops; (j) dual j–k master-slave flip-flops.

Digital systems units are covered in Blakeslee (1975), Bouwens (1974), Brophy (1977), Diefenderfer (1979), Heathkit (1978), Kostopoulos (1975), Lancaster (1976, 1977), Middleton (1977), Millman and Taub (1965), Millman (1972), Namgoster (1977), Sydenham (1979), Wakerly (1976), Frenzel (1997), Gibilisco (1997), Grob (1996), Horowitz and Hill (1989), Khalil (2001) and Mano (2001).

More detailed works exist on the various aspects, such as switching theory, Boolean algebra, and manufacture of digital systems.

RELATED ARTICLES

Article 36, Systems in the Frequency Domain, Volume 1; **Article 73, Overview of Electrical and Electronic Technique, Volume 2**; **Article 74, Basic Electronic Components, Volume 2**; **Article 122, Operational Amplifiers, Volume 2**.

REFERENCES

Blakeslee, T.R. (1975) *Digital Design with Standard MSI and LSI*, Wiley, New York.

Bouwens, A.J. (1974) *Digital Instruments Course* (several parts issued), N. V. Philips Gloeilampenfabrieken, Eindhoven.

Brophy, J.J. (1977) *Basic Electronics for Scientists*, McGraw-Hill-Kogakusha, Tokyo.

Burr-Brown (1964) *Handbook of Operational Amplifier Applications*, Burr-Brown Research Corporation, Tucson, AZ.

Burr-Brown (1966) *Handbook of Operational Amplifier Active R.C. Networks*, Burr-Brown Research Corporation, Tucson, AZ.

Clayton, G.B. (1971) *Operational Amplifiers*, Butterworths, Sevenoaks.

Connelly, J.A. (1975) *Analog Integrated Circuits*, Wiley, New York.

Diefenderfer, A.J. (1979) *Principles of Electronic Instrumentation*, Saunders, Philadelphia, PA.

Faulkenberry, L.M. (1977) *An Introduction to Operational Amplifiers*, Wiley, New York.

Frenzel, L.E. (1997) *Crash Course in Electronics Technology*, Butterworth-Heinemann.

Grob, B. (1996) *Grob Basic Electronics*, Glencoe McGraw Hill.

Gibilisco, S. (1997) *Teach Yourself Electricity and Electronics*, McGraw-Hill.

Graeme, J.G. and Tobey, G.E. (1971) *Operational Amplifiers – Design and Application*, McGraw-Hill, New York.

Heathkit (1978) *Electronic-Courses for Learn-at-Home: Course 1 DC Electronics; Course 2 AC Electronics; Course 3 Semiconductor Devices; Course 4 Electronic Circuits; Advanced Digital Technique Course; Microprocessor Course*. Available with experimental back-up and optional cassettes, Heath-Schlumberger, London.

Horowitz, P. and Hill, W. (1989) *The Art of Electronics*, Cambridge University Press.

Huelsman, L.P. (1977) *Active R.C. Filters*, Wiley, New York.

Khalil, H.K. (2001) *Nonlinear Systems*, Prentice Hall, New Jersey.

Kostopoulos, G.K. (1975) *Digital Engineering*, Wiley, New York.

Lancaster, D. (1976) *TTL Cookbook*, Howard W. Sams, Indianapolis, IN.

Lancaster, D. (1977) *CMOS Cookbook*, Howard W. Sams, Indianapolis, IN.

Mano, M.M. (2001) *Digital Design*, Prentice Hall.

Melen, R. and Garland, H. (1978) *Understanding IC Operational Amplifiers*, H. W. Sams, Indianapolis, IN.

Middleton, R. (1977) *Digital Equipment Servicing Guide*, H. W. Sams, Indianapolis, IN.

Millman, J. (1972) *Integrated Circuits: Analogue and Digital and Systems*, McGraw-Hill, New York.

Millman, J. and Taub, H. (1965) *Pulse Digital and Switching Waveforms*, McGraw-Hill, New York.

Namgoster, M. (1977) *Digital Equipment Trouble Shooting*, Reston, Reston, VA.

NS Corporation (1976) *Linear Applications Handbook*, (Vol. I, 1973; Vol. 2, 1976), National Semiconductor Corporation, Santa Clara, CA.

Sydenham, P.H. (1979) *Electronics – It's Easy*, (3 Vols), Modern Magazines, Sydney.

Wakerly, J.F. (1976) *Logic Design Projects Using Standard Integrated Circuits*, Wiley, New York.

76: Electronic Systems Design

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Approach to Electronic System Design 483
2 Counters 484
3 Frequency Counters and Timers 485
4 Registers 485
5 Digital-to-analog Converters 485
6 Multimeters 486
7 Microprocessor 486
8 Interconnecting Stages and Signal Transmission 487
9 Grounding and Shielding 488
Related Articles 489
References 489

1 APPROACH TO ELECTRONIC SYSTEM DESIGN

Originally, electronic instrumentation systems were devised using the basic active and passive elements, plus supplementary devices, to realize whatever basic functional units were needed. Almost all early electronic design was carried out using *discrete* components. Design then, therefore, first needed to form the functional building blocks, which were then interconnected to obtain the overall required performance.

As an example, to build a fast reversible digital counter in the 1960s, it was necessary to first develop a fast flip-flop using more active elements than were commonly used at the time for slower speed counters. A logical arrangement had then to be devised that would count as fast as possible, providing the output digital code needed for the display. It was also necessary to develop the display for the numerical output as a distinct project facet.

Today, the same detail of task no longer exists for the designer of measurement systems: complete counters like that mentioned above are now offered for sale with integral display. The most basic way to assemble the system now is to make use of standard integrated circuits driving an integrated digital display. The same unit is also available as a single LSI (Large Scale Integration) unit costing the equivalent of less than an hour of a technician's time.

Systems design is based on first having a sound conceptual understanding of the task, this being committed to paper in the form of block diagrams that lay down the signal flows and show the characteristics needed. These are then refined, until the first apparently workable system is ready to assemble. A prototype is then put together using temporary assembly, with soldered or plugged joints. A number of proprietary assembly methods are available; many avoid the need to solder. When the prototype is thought to be satisfactory (from actual tests) the circuit is made in the printed wiring, printed circuit, or hybrid integrated form. Large organizations will make use of advanced design support tool suites that can be used to create a design in the computer, which will be almost perfect when actually manufactured. While such tools do exist, they take time to install and learn. For this reason, many electronic systems created for measuring systems are built from components, the design being tested as it is developed.

It is generally the case that the first system design will not be entirely suitable. This occurs because design of any technological system relies on the designer making the correct assumptions about a system that he or she is not yet familiar enough to make. The actual act of assembly and testing the prototype considerably improves understanding

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

of the need. The degree of success in arriving at a suitable initial design relates to the complexity of the problem and the designer's experience. Where possible, proprietary modules are usually used to reduce the need for design knowledge. This eliminates many design decisions. If the modules are well tried and proven, then the whole is more likely to be reliable. Digital systems tend to be easier to develop than analog ones.

Interaction between stages is probably the greatest source of prototype development difficulty. Stages may excessively load the preceding one or transient signals can influence other parts of the system. Race conditions, in logic systems, can also occur if some part of the system acts faster than others where timing is an important parameter. Many of these problems cannot be dealt with by rigorous mathematical methods, unless a model-based tool is being used. The designer's ingenuity is called upon to ensure that as many of these likely problems as possible are circumvented, or allowed for in some way.

When a suitable prototype assembly has been developed, it must then be tested for operation in a real situation. Temperature, vibration, mechanical strain, humidity, and many more influence quantities may prevent proper operation. Because layout can be important, it will first be necessary, in some cases, to construct an assembly that is very close to the actual construction method to be used. The prototype should never be destroyed or changed as the next version is being made. It will often be found that the second, supposedly superior, system exhibits faults that were not seen before. A check of the first prototype may well show that the faults were also there in that model, thereby showing that the design is at fault, not the change in layout.

The design of integrated circuits is usually more rigorously achieved but even there, where it is possible to devote more effort to a specific problem, similar methods are often utilized to produce a layout for integration. Prototyping of this level only occurs in specialized groups.

A very common error is to regard a single, apparently satisfactory, assembly as being necessarily typical of a batch production run. Too often, production begins on the strength of inadequate testing and tolerance-spread checking. It is wise to expand from 1 to 10, to say, 100 systems, when very large numbers of an identical unit are needed. Specialized functions such as reliability and thermal behavior are often overlooked in low volume designs, with subsequent need for redesign.

This short article can only introduce electronic systems. In practice, the electronic part of a measurement system is usually but a portion of the whole, each module being considered in turn, and then the whole, on other grounds such as reliability, production needs, and sales and service requirements.

2 COUNTERS

Many modules are now available from which a system can be assembled. It is usually not necessary to build systems using the basic linear and digital integrated circuits with, perhaps, a few discrete components. This section considers a few of these larger subsystems; the number constantly grows as designs and demand appear, enabling integration to be used at a commercially viable level. A good set of maker's catalogs and Web links need to be developed to support chip availability information.

A cascaded chain of flip-flops will divide a time sequential string of digital logic levels according to the math of binary division. By suitable arrangement, the chain can be used to count up, or down, or be reversible according to an external direction control signal. If the signal passes from stage to stage in sequential order, the pulse *ripples through* taking a considerable time (up to microseconds) to pass through many stages. This form of counter is called *asynchronous* for state changes happening in any stage of the chain are not necessarily made all at the same time.

By suitable interconnection of the stages, the ripple-through time can be greatly reduced by causing an input pulse transition to set each of the stages simultaneously. These are called *synchronous* counters. They are more complex to interconnect, but, as they are rarely built from discrete components, this is not a problem for systems designers; it has been done in the manufactured IC. Counters can easily count at rates of 100 MHz and higher rates in the low GHz region, the reversible kinds reversing in as little as nanoseconds.

Binary counters count in a *binary* sequence. To change the contained binary code numerical value into the often more useful *decade* form, they must be decoded by a suitable conversion logic network.

It is sometimes more convenient not to count in pure binary, but to use what is called a *binary coded decimal* (BCD) system instead. In this system, four flip-flops are used to form a counting chain that passes through only 10 and not the binary's 16 states. Feedback between the counting stages is used to achieve this. The actual coding used within the four-element system decides the form of decoding network needed to drive the chosen decade number display. The individual BCD units are then cascaded using simple connections to obtain the number of decades display needed. Many of these design variables need not be known by the use of ready-made LSI counter/display modules. In many instances, their price is so low as to make any thought of new development quite unrealistic. They are very well tried and proven.

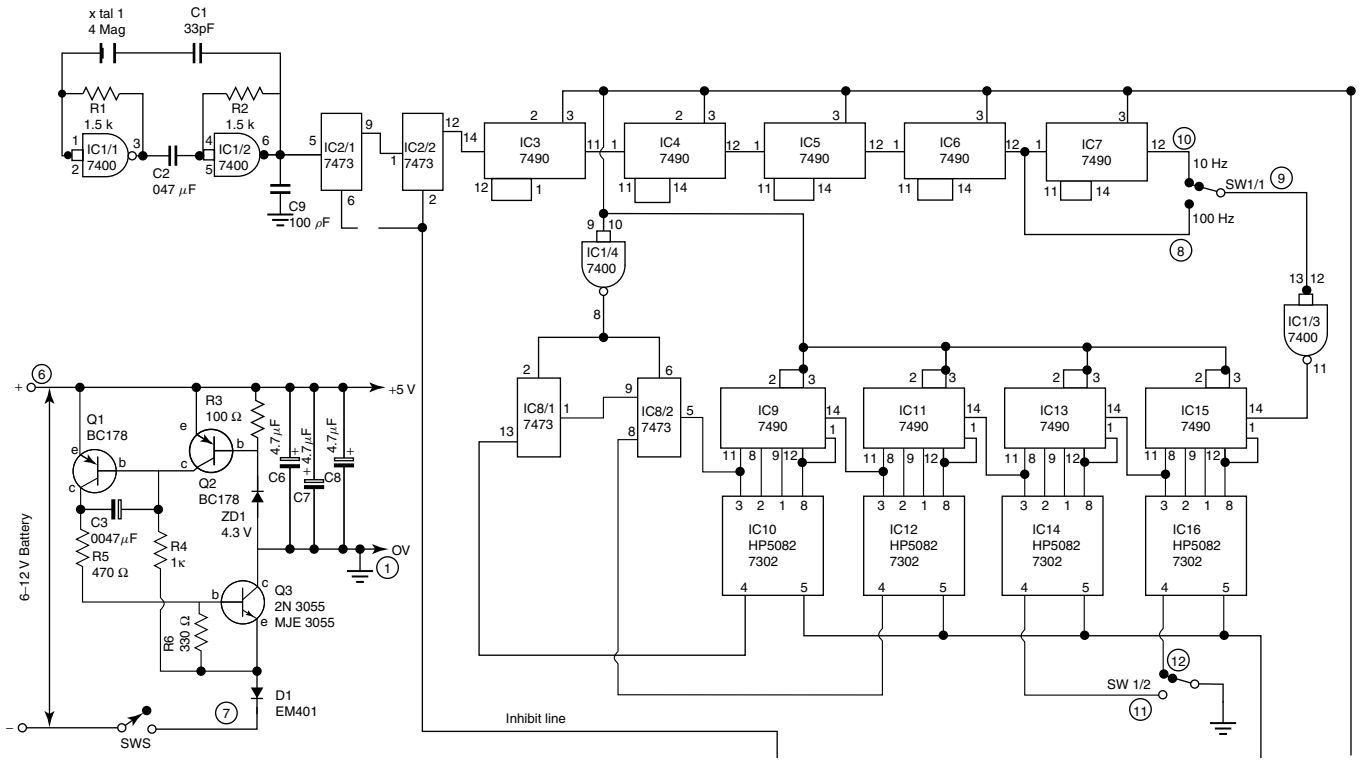


Figure 1. Schematic block diagram showing, as an example, how basic IC modules are used to form a general-purpose digital stopwatch.

3 FREQUENCY COUNTERS AND TIMERS

Addition of a *clock source* (a square-wave generator of known frequency) and a suitable turn on, and off, gate system forms a method for measuring the frequency of a signal. The gate is, in one version, operated from the clock allowing the counter to accumulate one count per cycle of the signal being measured, for a known period of time. This determines the frequency of the signal; the count is displayed as frequency. If a count from the clock is measured for the time during which the gate is held open by the cycle time of the signal, then a period will be measured. Figure 1 shows the IC modules used to build such a unit.

Similar principles are used to count objects as part of, say, a batch control system.

Timing is also possible using linear modules in which an output is held, say, high for a preset duration. In linear methods, mathematical integration is used to produce a time interval. Linear methods are not capable of the high timing precision afforded by properly designed digital counter/timers.

4 REGISTERS

A register is a form of cascaded flip-flops. A digital binary number is passed into the digit stages from one end (*serial input*) or to each simultaneously (*parallel input*); each flip-flop then records a respective digit of the number. Action of the clock-pulse input causes the number to remain intact, yet be stepped sideways in either direction. This is used as part of the hardware of several mathematical operations using digital binary numbers.

5 DIGITAL-TO-ANALOG CONVERTERS

A digital number is often needed in an analog equivalent form, for example, as might arise when an electronic digital calculation has been made and the resultant must operate an analog signal device. Several methods of conversion exist for this direction of conversion and for the reverse, where analog signals are converted into digital forms. These functions are rarely made any more from discrete components; the cost of ready-to-use modules being low enough to encourage selection for direct use in the larger system. Although designers do not need to develop these converters, it is very important that their internal operation is understood as the performance can vary greatly with application.

6 MULTIMETERS

As an example of an electronic instrument system, consider the apparently simple high quality multimeter. This instrument has undergone extensive change in internal design philosophy in recent years and exemplifies the general trend in electronic systems operations.

A multimeter has the purpose of allowing the user to measure voltages, currents, or resistances of a system and its components. Originally, these were always made from an electromagnetic meter indicating movement to which appropriate multipliers (series resistors – for voltage ranges), shunts (shunt resistors – for current ranges), and operational battery, and series resistor (for resistance) were switched into circuit as needed by the operator. A multimeter of better quality would normally have overload protection to safeguard the meter from accidental damage by inappropriate connection.

The entire unit used analog methods and could also incorporate a linear amplifier in the circuit to raise the terminal resistance to 1 MΩ or more in order to remove loading effects in use.

The alternative design, now almost always used, contains ICs and some discrete components. It operates internally almost entirely using digital, rather than analog, signals.

The user must still select which variable is to be measured, but need not select the range; systems are *autoranging*. The input signal is first converted into a digital equivalent using an analog-to-digital converter stage. The signal is then used with counters and digital display to show the value in digital format.

Another example is the use of digital methods to produce a very accurate *sample-and-hold* module, over a long period of time. In these, the analog storage capacitor that stores the voltage level in the linear analog system has *internal leakage* that degrades the voltage level with time. If the capacitor is used to sample the level and the value so obtained is immediately converted to a digital equivalent form, it can be stored for an indefinite time period. Recovery uses a digital-to-analog converter.

The relative cheapness, increased reliability, and often enhanced accuracy of digital alternatives easily outweighs the fact that internally their design is far more complex, and uses many more elements, than does the analog alternative.

7 MICROPROCESSOR

The concept of a totally general-purpose basic electronic digital building block was taken a step further with the introduction of the *microprocessor*. This is an LSI circuit, Figure 2, that contains the architectural features of a basic digital computer system. It can be set up internally, through the use of external *software* programming, to perform an incredibly wide range of circuit functions. It operates internally as a digital computing, but the interfaces used to connect it to the chosen application allow it to be connected to analog devices when needed. Its low cost enables it to be used as a most basic single component that is set by the user to perform specific tasks. It is set by software programming rather than by hard wiring.

In reality, a considerable amount of peripheral equipment may be needed and the cost of this can far exceed the cost of the microprocessor itself (less than a few man-minutes of time). Even so, the general-purpose nature of the module makes it a first choice, when seeking to design a complex control system. Rather than make connections with wires the user writes a program. Once set, the microprocessor-based system can be reset reasonably easily.

Use of microprocessors needs little understanding of the concepts, basic modules, and building blocks of electronics. Microprocessors have brought about a marked change in the design philosophy of electronic systems of the future. Some predict that the cost of the microprocessor will fall so low that it will find use in numerous low-cost applications. There is a running shoe with one in it. A modern automobile has dozens in its control and instrumentation.

With microprocessors, little electronic knowledge is needed to create the system: programming skills have been largely exchanged for network and system design expertise.

Despite this, peripherals still need to be designed and connected and in many situations, microprocessors and other marketed general-purpose computational designs are not able to match the performance possible from special-purpose designs.

The programmable logic controller (PLC) can be placed in the same class as microprocessors. These are systems in which the circuit wiring to provide the actual system task needed (basically switching networks) is again done by a form of direct programming into the computer-based system. Changes to the system operation are easily carried out without the need to enter the electronic level of operation, or to be an expert at programming a computer.

The operational power that can be provided by one, or more, microprocessors has, as did valves in their more primitive way, allowed the designer still greater freedom of action in implementing conceptual schemes.

It is now normal to build electronic systems that are self-checking, that can diagnose themselves, and that can make reconnections to overcome failures – these have been coined the *smart* systems. A useful level of artificial intelligence has already been achieved in electronic systems. The low cost, small space, and power needs

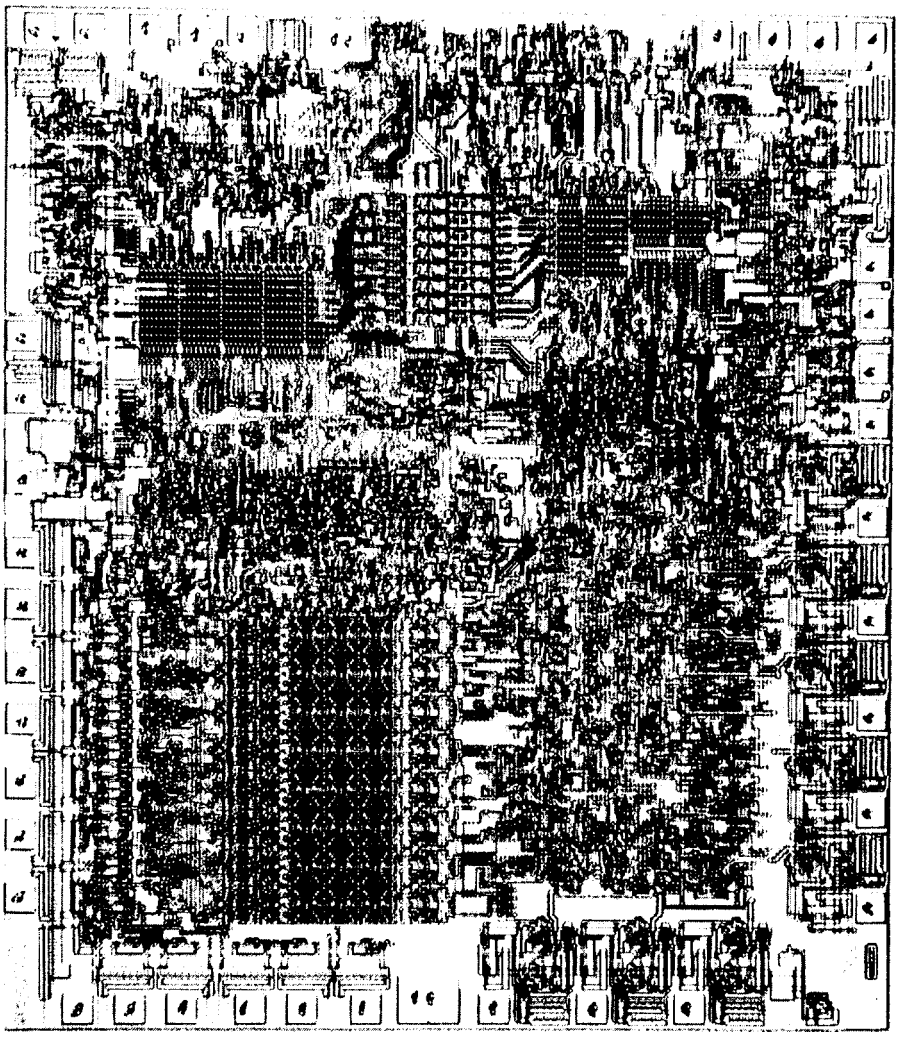


Figure 2. Internal arrangement of a typical microprocessor IC.

of the microprocessor enable very extensive strategies to be implemented on reasonable budgets of finance and time.

An original suggestion was that as the microprocessor was standard methodology, texts written on the microprocessor would be general purpose. To some degree this is true, but usually there is the need to access information on the one being used in a design. Too many variations have emerged from this very powerful technology.

8 INTERCONNECTING STAGES AND SIGNAL TRANSMISSION

Whereas the modern electronic system designer has less to consider now about the internal design of building blocks and modules than in the past, it is still as important as ever that they be interconnected properly. If the connecting links are considered as additional stages, then they are likely to receive the attention that they must be afforded. Even when modules are wired in close proximity, they must be connected appropriately.

It is necessary to consider the connections carefully, for the link itself may act as a frequency filter, alter the amplitude of signals, pick up unwanted signals, introduce phase shifts, or be mechanically unreliable.

Within closely assembled circuitry, simple printed and insulated wiring will generally suffice unless high frequencies are involved above around 1 MHz. If the distance between units exceeds a few meters, then it becomes necessary to select a more suitable form of signal transfer method. The *open-wire*, as seen on old telephone poles, is usable to around 10 MHz, but beyond that the higher frequency signals need other transmission methods. The open wires are very lossy and not particularly secure from damage.

An alternative is the *coaxial cable*. This is formed from conductors suitably placed with a separating dielectric. They are made in circular, flat, and strip sections each having its special application. They can convey signals to around 5000 MHz. Multicore insulated cables could be considered to be of this type, but they are not usually designed for high-frequency use, having excessive losses.

Above the usable range of signal frequency suited to coaxial cables, it is necessary to make use of *waveguides* and open-space *electromagnetic* radio links. Laser-based *optical fiber* links are also capable of conveying these high bandwidths and are used prolifically, see Wolf (1979) for details. They play a major role today as they are not prone to the same stray signal, pickup, effects as are electronic transmission systems.

Whichever kind of electrical communication line is used, certain basic generalizations can be made. First, a short line will possess small values of self, shunting, capacitance, and resistive losses. As the line length is increased, the capacitive effect becomes more dominant, the line then acts as a low-pass filter: this gives attenuation and phase shift to the signal being sent over the link.

The coaxial cable exhibits an interesting feature for reasonably high frequencies of use. There the impedance that can be seen by looking into the end of an adequately lossless line of any length can be regarded as being of fixed pure resistance. This is called the *characteristic impedance*, Z_0, of the line. Common values made are 600, 100, and 75 Ω. Thus, the actual length of the line matters little.

When high frequencies need to be sent over transmission links, a serious effect occurs if the terminating impedance is not purely resistive and equal to the characteristic impedance of the line used to make the connection. If mismatched, *reflections* of signal energy occur, and time-varying signals are generated in the links that feed on and back, at each end of the link. At lower frequencies, the mismatched line acts to alter the signal magnitude and phase; reflections produce distortion and loss of power transfer. It is, therefore, always important to properly match high frequency–signal terminations. This subject is covered in Johnson (1950) and in Sinnema (1979).

9 GROUNDING AND SHIELDING

Another aspect of stage coupling is the need to preserve the signal-to-noise ratio (SNR). Ideally, a coupling should not degrade the system, but in practice there will always be some reduction.

Analog signals are more prone to this, for their information is conveyed as an amplitude. Coding and the use of digital signals are often better alternatives to use.

Several connection techniques can be used to reduce stray pickup in wiring. The principles apply to both long and to even very short leads in low-level detection systems.

A first possibility is to transfer the information over the link at some frequency away from that of the expected noise. It is then theoretically possible to filter out the signal at the receiver. This method can be satisfactory, but all filters allow some level of signal through, and the price paid for sharp highly discriminating filters may not be acceptable; for instance, they tend to *ring* to rapid transient changes of signal level.

A better approach is to reduce the noise level at its source. If that is not possible, then the wiring can be arranged such that two parts of it pick up similar noise signals, which are out of phase by 180° at the detection stage. They then cancel to leave only the unbalanced contribution. This principle is known as *common-mode rejection*. A twisted cable pair does this, as each wire has an equal signal induced in it by common radiation coming from the noise signal source; the two being induced in opposite directions in the detector input circuit. To obtain the best from common-mode rejection, connectors, and the input stage of the active circuitry must also be differentially arranged in this manner. Common-mode rejection should not be used to remove noise that can be reduced at source, for the common-mode noise signals can require amplifiers with unduly large dynamic ranges to avoid serious distortion occurring.

Shielding is also a method used to reduce signal levels induced into connecting leads. Leads form antennae in which electromagnetic fields will induce voltages. Thus, the shorter the leads, the lower their impedance. The better they are shielded, the smaller will be the induced currents. Shields, and their necessary earthing connection, should not by themselves form a significant resistance closed electrical circuit, for the *earth loop* then produced will produce circulating currents that also induce unwanted signals elsewhere.

Radio frequency energy – electromagnetic interference – can also be bothersome. Often the problems are overcome by preventing it escaping from the generating source by careful scientific design of the enclosures. Alternatively, the detecting apparatus is screened.

Several specific texts are available that provide detail on long line–transmission systems, for example, Johnson (1950). Low noise systems, and how to couple them, are the subject of Morrison (1977), Motchenbacher (1973), and Ott (1976).

RELATED ARTICLES

Article 73, Overview of Electrical and Electronic Technique, Volume 2; **Article 74, Basic Electronic Components, Volume 2**; **Article 75, Electronic System Building Blocks, Volume 2**; **Article 78, Embedded Systems, Volume 2**; **Article 159, Designing and Building Software for Measuring Systems, Volume 3**.

REFERENCES

Johnson, W.C. (1950) *Transmission Lines and Networks*, McGraw-Hill, New York.

Morrison, R. (1977) *Grounding and Shielding in Instrumentation*, Wiley, New York.

Motchenbacher, C.D. (1973) *Low Noise Electronic Design*, Wiley, New York.

Ott, H.W. (1976) *Noise Reduction Techniques in Electronic Systems*, Wiley, New York.

Sinnema, W. (1979) *Electronic Transmission Technology*, Prentice Hall, London.

Wolf, H.F. (1979) *Handbook of Fibre Optics*, Garland STPM Press, New York.

77: Limits of Detection in Electronic Systems

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Sources of Fundamental Physical Limits on Detection 490
2 Johnson Noise 490
3 Shot Noise 490
4 Flicker Noise 491
5 Sources of Noise Resulting from Imperfect Design 491
Related Articles 492
References 492

1 SOURCES OF FUNDAMENTAL PHYSICAL LIMITS ON DETECTION

As in all of the energy regimes, electric energy systems also possess certain physical processes that prevent sensitivity from being increased to infinity. No matter how well contrived and constructed the apparatus is, the designer must face the fact that these fundamental effects cannot be eliminated. They provide an ultimate limit to detection for a given methodology. Three fundamental noise sources, and other controllable noise sources, are shown pictorially in Figure 1.

2 JOHNSON NOISE

Within the resistive components of any electric component, thermal energy causes the electrons to vibrate with increasing activity as the thermal energy increases. It is related to mechanical Brownian motion within matter. Thermal noise, also called *Johnson noise* after its prime investigator, has been quantified as

$$\overline{v}^2 = 4kTRB \tag{1}$$

where $\overline{v}^2$ is the mean-square voltage generated, k is the Boltzmann constant, T is the absolute temperature of the resistance, R is the resistance in question (in ohms), and B is the frequency bandwidth of the signal in the resistor.

This can also be expressed in terms of noise current and noise power. Unlike the Brownian motion noise of mechanical systems, which rarely is observed, this noise is relatively easily detected by electronic systems and is always a serious factor to consider when designing low-level (below microvolt) detection stages. Clearly, from the equation, reduction of bandwidth, temperature of the resistor, and magnitude of the resistor reduces the Johnson-noise level.

The frequency spectrum of this noise, in its purest form, is uniform, there being as much energy in a given bandwidth at any place in the frequency spectrum. If this condition applies, the noise is said to be white. Gray and pink noises are of this general form but they possess different frequency spectra. Gray, for instance, reduces in amplitude in a linear manner as the frequency rises. Such noise types are close approximations to different noise situations that arise in practice and, thus, can be used as test noise sources.

3 SHOT NOISE

Wherever electric current flows, the corpuscular nature of the electrons gives rise to discrete signal energy packets called *short noise*. If there are enough of them, their sum averages to a steady current and the effect is not generally

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

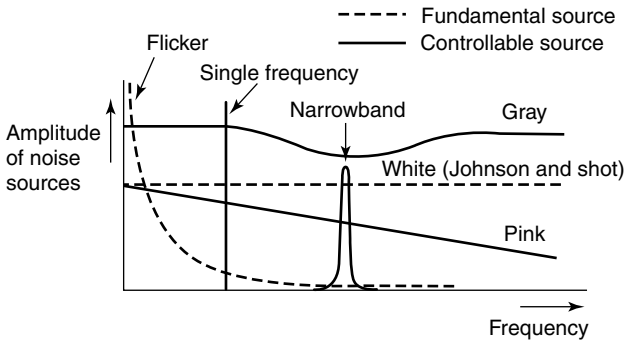


Figure 1. Frequency spectra of fundamental and other electronic noise sources. (Amplitudes of each type not shown to any strict scale.)

observable. If, however, each electron becomes an event in the detection system, as it does when very low-level signals are concerned, then this discrete randomness makes its presence detectable. This was first investigated for electron discharge in the former thermionic valves by using the assumption that the space charge was not filling the space available. The theoretical expression derived to explain the effect quantitatively was thus an estimate; in most applications, space-charge conditions alter the magnitude of the effect. The usual basic expression quoted is

$$\bar{i}^2 = 2eIB \tag{2}$$

where $\bar{i}^2$ is the mean-square current generated, e is the electron charge, I is the average current passing, and B is the frequency bandwidth of the signal passing.

However, valves, transistors, and other active electronic devices, each differ a little from this generalization. Suffice to say that this effect is also detectable by quite simple electronic apparatus and, therefore, must also be allowed for in the detection of stage design. Reduction in signal bandwidth and the through-current level reduces the level of this type of noise. Choice of detection device also has an important bearing on the magnitude. This form of noise is also spectrally white if it follows the given equation.

In both these cases, the formulations express the levels of noise signals generated within a component. The level that is received by the following stage depends upon the matching conditions between the two stages.

4 FLICKER NOISE

A third important fundamental noise source of electronic systems is called *flicker noise*. Since it varies approximately as the inverse of the frequency, it is often referred to as 1/f or *hyperbolic noise*.

Despite considerable research, it is still not possible to entirely quantify this noise form on the generalized basis. The physical reason for this in valves and in transistors is not proven. As a rule of thumb, an electronic system's flicker noise begins to become significant compared with the other fundamental limits at around 400 Hz, increasing below that frequency to be at infinity at absolute zero frequency. This is one reason for using carrier systems to provide good signal-to-noise ratios in detection equipment: the signal to be detected is generally modulated at well above 400 Hz, where this noise effect is lower than the other noise types.

Flicker noise has a direct counterpart in mechanical and thermal systems, for the greater the integration time that is used in an attempt to remove long-term, slowly changing, drift effects, the lesser the rate of improvement obtained.

The above-mentioned noise sources are the three dominant sources that cannot be reduced beyond certain limits. The use of bandwidth-reduction techniques reduces their level but at the usual expense of increase in response time. Cooling is often employed to reduce the operational temperature, including use of cryogenic stages. Electronic noise sources are well described in the literature (see Bennet, 1960; Connor, 1976; Delaney, 1969; Usher, 1974; Van der Ziel, 1954; Whalen, 1971; Wainstein and Zubakov 1962).

5 SOURCES OF NOISE RESULTING FROM IMPERFECT DESIGN

Assuming that fundamental limits do not restrict the design, then the system must also be assembled, connected, and shielded such that additional noise pickup is kept as small as is required. The common unit used to describe the degradation that always occurs from connection and stage performance is the noise figure (NF). It is given by

$$\mathrm{NF} = 10 \log \frac{\mathrm{SNR}_{\mathrm{in}}}{\mathrm{SNR}_{\mathrm{out}}} \,\mathrm{dB} \tag{3}$$

A perfect stage would add 0 dB, that is, nothing, but in practice, around 1.5 to 3 dB is the lowest addition obtainable. Shielding, layout, wiring, device selection, connection modes, matching, and other factors must be considered as possible means to reduce the NF of a stage. The order that stages with differing NF are connected in cascade is also an important factor. As a guiding rule, the stage with lowest NF should be used closest to the detection stage. Often, an initial stage is inserted that provides little useful gain but has a low NF that

assists the overall NF to be minimized. As practical measurement systems usually involve more than one energy regime, the noise sources of each regime must be considered.

Further information on noise in electronic signal systems is provided in the Section **Noise and Interference** and **Signal Recovery in the Presence of Noise**.

RELATED ARTICLES

Article 56, Transfer Characteristics of Instrument Stages, Volume 1; **Article 57, Static Considerations of General Instrumentation, Volume 1**; **Part 12, Elements: F – Signals and Noise, Volume 3**.

REFERENCES

Bennet, W.R. (1960) *Electrical Noise*, McGraw-Hill, New York.

Connor, F.R. (1976) *Noise*, Edward Arnold, London.

Delaney, C.F.G. (1969) *Electronics for the Physicist*, Penguin Books, Harmondsworth.

Usher, M.J. (1974) Noise and bandwidth. *Journal of Physics E: Scientific Instruments*, **7**, 957–61; (Reprinted in 1982, *Instrument Science and Technology*, Vol. 1, (ed. B.E. Jones), Adam Hilger, Bristol, pp. 110–115).

Van der Ziel, A. (1954) *Noise*, Prentice Hall, Englewood Cliffs, NJ.

Wainstein, L.A. and Zubakov, V.D. (1962) *Extraction of Signals from Noise*, Prentice Hall, Englewood Cliffs, NJ.

Whalen, A.D. (1971) *Detection Signals in Noise*, Academic Press, London.

78: Embedded Systems

Timothy Wilmshurst
University of Derby, Derby, UK

1 Definition of An Embedded System 493
2 Characteristics of Embedded Systems 493
3 Design of Embedded Systems 495
4 Programming the Embedded System 496
5 Testing of Prototype Embedded Systems 497
References 497

1 DEFINITION OF AN EMBEDDED SYSTEM

The term *embedded system* implies one computer or computers embedded within a system or product, which form an important part of that system, generally undertaking the system control. Related functions, for example, data processing, can also readily be undertaken. The embedded computer may be hidden from view, and the user may not even be aware of its existence.

Many definitions for embedded systems exist. The consulting company Dedicated Systems Experts (http://www.realtime-info.be) offer the following definition: An embedded system is a computer system that is enclosed in another system and makes an essential part of it.

Wilmshurst (2001) uses the following alternative definition, which explicitly evokes the control nature of most embedded systems: An embedded system is a system whose principal function is not computational, but which is controlled by a computer embedded within it.

Noting the characteristics of the embedded system, discussed below, Wilmshurst also uses the following more descriptive definition: An embedded system is a microcontroller-based, software-driven, reliable, real-time control system, autonomous, or human or network interactive, operating on diverse physical variables and in diverse environments, and sold into a competitive and cost-conscious market.

2 CHARACTERISTICS OF EMBEDDED SYSTEMS

The scale, complexity, and cost of embedded systems vary enormously. A typical system has some or all of the elements shown in Figure 1.

At the heart of the system lies an embedded microcomputer. Most commonly this is a microcontroller, although larger-scale computers (e.g. embedded personal computers) are also widely applied – *see* **Article 158, Microcontroller Systems, Volume 3**.

The embedded computer is itself made up of its hardware and software. Feeding into the embedded computer are input variables, taken from sensors and passing through input signal conditioning and a data conversion stage.

The computer may be networked with other systems, and it may have a user interface, although neither of these is essential to the concept.

The system is likely to generate output signals that are used to control system variables, for example, by driving actuators through a power electronic interface.

Figure 1 indicates the breadth of expertise that is required in the design and commissioning of an embedded system, including sensors, data acquisition, microcomputer hardware and software design, data communications and networking, human interface, control, and actuators and their drives.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

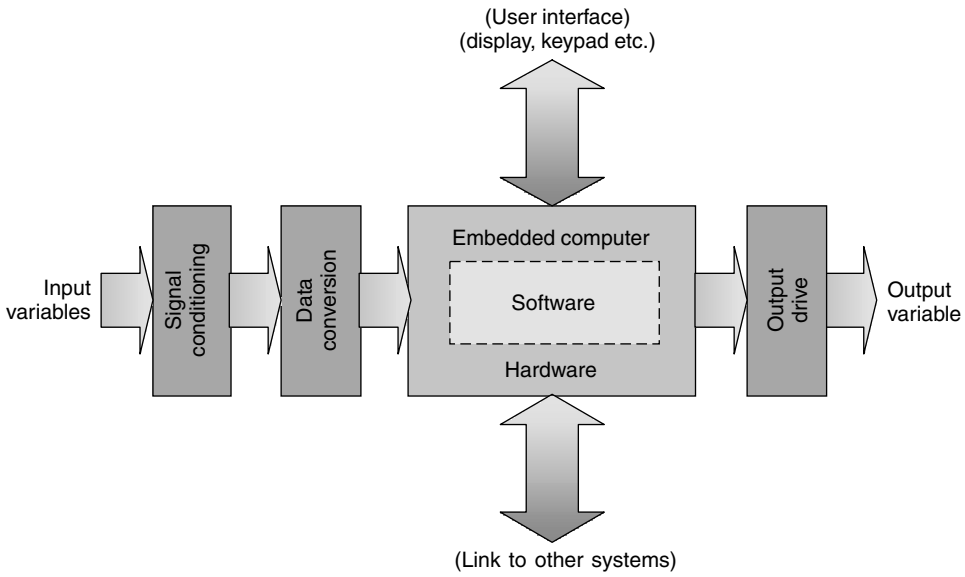


Figure 1. The essence of the embedded system.

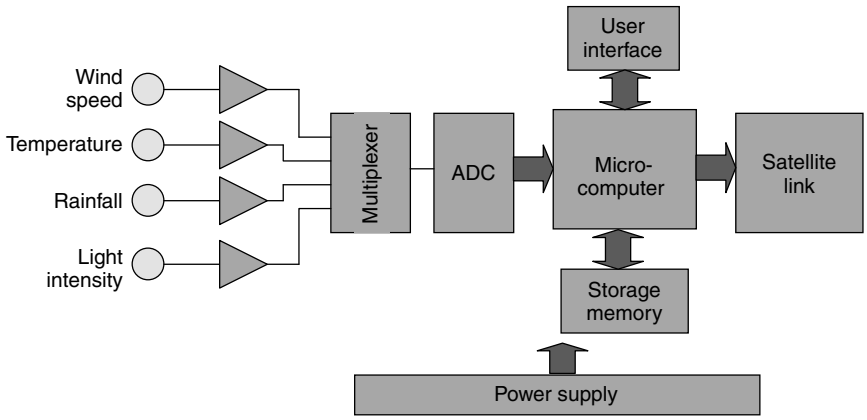


Figure 2. An example of an embedded system, a remote four-channel automatic weather station.

An example of an embedded system, a remote four-channel Automatic Weather Station is shown in Figure 2. Here four inputs are sampled periodically (not necessarily with the same period), and the resultant data processed and stored. It is also transferred via a telemetry link to a satellite. As it is designed for operation in the field, power supply is a critical feature of the system.

Further important characteristics of the embedded system are as follows.

2.1 Timeliness

Most embedded systems operate in an environment that is strongly time dependent. They must respond in an appropriate timescale to external events, or must generate time dependent outputs. This is generally known as operating in real time.

2.2 Efficiency of hardware/software utilization

Embedded systems are often sold in markets that are extremely cost sensitive. Alternatively, or as well, many such systems requirements are very space and/or power conscious. For these reasons, the embedded system is normally designed with restrictive hardware and software constraints.

The result of this, for example, is that the CPU may be only 8 or 16 bit, with limited computing power, and a restricted memory size may impose limits on program length or data storage capability.

2.3 Reliability

Many embedded systems now operate in situations having high demand on reliability. This includes, for example, many medical and motor vehicle applications. Appropriate techniques for hardware and software reliability are therefore applied.

2.4 Constraints of the operating environment

The operating environment of many embedded systems places further demands on its design. Many are battery powered, and so must be designed from a low power viewpoint.

Some operate in environments of hostile temperature, humidity, or vibration, for example, a motor vehicle engine management system.

These characteristics, taken in summary, form the final definition given in Section 1.

3 DESIGN OF EMBEDDED SYSTEMS

Pahl and Beitz (1995) have identified a design methodology for engineering products, which starts with

- *Clarification of Need*, and then moves through
- *Conceptual Design*, to
- *Embodiment Design*, and finally to
- *Detail Design.*

An adaptation of this process to the embedded system environment (Wilmshurst, 2001) is shown in Figure 3. This is a form of system life cycle – *see* **Article 63, Outline of Systems Thinking, Volume 2** and other articles of the subsection on 'Engineering a measuring system'.

This model provides a sequence of design phases. The principal characteristics of the system are identified in clarification of task, or requirements generation, that is then defined in the specification – *see also* **Article 69, Requirements Allocation, Volume 2**. From this, the principle hardware and software functions emerge. These then continue as two interlinked design strands. These strands should proceed in parallel, a process known as *codesign*.

Important aspects of the design process, whether or not this model is followed, are outlined below. Note however that software program design issues are mainly covered in Section 4 below – *see also* **Article 159, Designing and Building Software for Measuring Systems, Volume 3** and

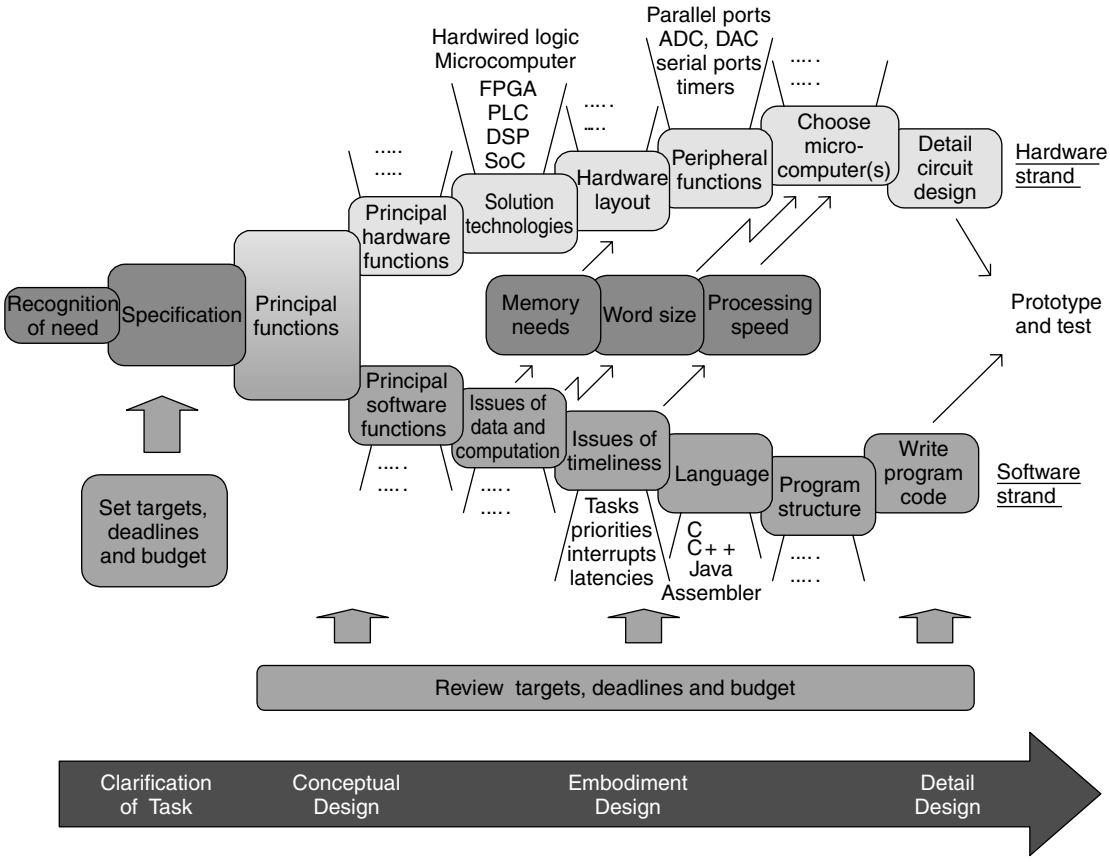


Figure 3. A design model for embedded systems. (Reproduced by permission of Institution of Electrical Engineers, with modifications.)

other articles of the subsection on 'Software aspects of measuring systems design'.

3.1 Solution technology

It is essential for the designer to be aware of the current technologies available for system implementation. The conventional approach, of designing a microprocessor or microcontroller into a hardware system, is to a large extent being replaced by the opportunities of ASIC (Application Specific Integrated Circuit), Field Programmable Gate Array (FPGA), and System on Chip (SoC) technologies. Thus, a microprocessor is not necessarily purchased as a piece of silicon, but as a piece of Intellectual Property, (IP) which can be downloaded onto an FPGA, along with other devices.

The SoC strategy takes this further, with CPU, memory, data buses, and an array of other system elements available, in user-configurable form, on the chip. Xilinx® (http://www.xilinx.com/) and Altera® (http://www.altera.com/) are important suppliers in this field.

Other distinctions, such as that between the conventional microprocessor and the DSP (digital signal processing) device (*see* **Article 146, Principles of DSP Hardware Design, Volume 3**), are also becoming blurred, with microprocessors and microcontrollers moving to take on DSP capability, and DSP devices taking their place in embedded systems.

As a complete alternative, Programmable Logic Controllers (PLCs) nowadays carry out sophisticated measurement, control, and network capability, and can themselves form the basis of embedded control.

3.2 Hardware layout

This process identifies the principal blocks of the hardware system, and their interconnection. It will include such fundamental decisions as whether there are one or more microcomputer elements, how data is transferred internally, what input/output techniques are used (and hence what peripherals are required), and so on.

3.3 Choice of microcomputer(s)

Having evaluated both the hardware and software demands that will be placed on the microcomputer(s), a device selection can be made. Alternatives include the

- ASIC/FPGA/SoC options, mentioned above;
- externally supplied microcomputer subsystem; and
- microcontroller designed into a custom piece of hardware.

Selection will be initially on device performance and cost. Other aspects, however, must be considered. These include cost of development tools (which can be significant and take time to learn), and the level and quality of support available for the chosen device.

There are a huge range of microcontroller suppliers, including Atmel® (http://www.atmel.com), Hitachi® (http://www.hitachi.com/products/electronic/semiconductorcomponent/index.html), Microchip® (http://www.microchip.com/), Motorola (http://e-www.motorola.com/webapp/sps/site/homepage.jsp?nodeId = 03M0ym4t3ZG), and Philips Semiconductors® (http://www.philips.semiconductors.com/products/standard/microcontrollers/).

4 PROGRAMMING THE EMBEDDED SYSTEM

The usual requirement in an embedded system is for a program with compact and reliable code, which executes efficiently, and which can interact in a direct way with the hardware resources.

Programming normally takes place using an Integrated Development Environment (IDE). This runs on the host computer, and develops programs for the target system (the embedded system under development).

4.1 Programming language

The requirements mentioned above make High Level Languages applied elsewhere less attractive.

One possibility is to use the native assembler of the processor. This leads to the most efficient code possible, and allows excellent interaction with the hardware. However, programming in Assembler is error prone, and time consuming.

An excellent compromise is the language C. This leads to compact and well-structured code, which can interact with the hardware, and which can be developed reasonably rapidly. Techniques adapting C to the embedded environment, including in high reliability applications, are widely documented (e.g. van Sickle (2001), Hatton (1994), MISRA (1998). C compilers are available for all microcontrollers, some with special extensions or restrictions, based on the controller architecture.

As larger-scale program memory (particularly with the wide-scale adoption of Flash memory) is now becoming available, C++ and Java are becoming increasingly used. Language subsets, removing the more memory-hungry

aspects of the languages, and targeted at the embedded world, include Embedded C++ (EC++) and Embedded Java.

4.2 Multitasking in real time

Most embedded systems are multitasking, and must meet real-time programming demands, that is, they have a number of distinct functions, which must be performed more or less in parallel, and which must meet certain deadlines. These may be of different relative importance, and may or may not be synchronous with each other.

A particular technique for dealing with this situation is to apply a Real Time Operating System (RTOS).

An RTOS determines which task is given CPU time, and for how long. This is called *scheduling*. It also controls task synchronization, use of system resources that may be needed by more than one task, and passing of data and messages between the tasks.

Various scheduling strategies exist. In Cyclic Scheduling, each task runs until completion, before the CPU turns to the next task. A Round Robin scheduler simply gives CPU access to each task for a fixed period in turn, irrespective of its priority. This is an example of preemptive scheduling, as task execution is interrupted before completion. A prioritized preemptive scheduler gives priority to tasks that have been designated as high priority; low priority tasks are not executed unless there are no high priority tasks waiting. Further details are given in many sources, for example, Simon (1999).

5 TESTING OF PROTOTYPE EMBEDDED SYSTEMS

Embedded systems can be a particularly challenging type of product to test. Signals and logic states may be deeply embedded within the system, and impossible to monitor. Worse still, the system is under the control of a fast-moving program, which may instigate fault conditions under circumstances that are then very difficult to replicate. A range of tools, of varying levels of sophistication, exists to address these challenges.

It is highly advantageous to avoid a situation in which untested software is run in untested hardware. This can be avoided by the use of software and hardware simulation. A software simulator is a computer package, usually part of the IDE, which attempts to replicate the action of a particular microcomputer under test, and which can run a trial program. It is low cost, and sometimes provided free of cost by microcontroller suppliers (e.g. Microchip's MPLAB™). Importantly, it holds values of all internal processor registers and memory locations, which can be examined by the user. It can run the program continuously or single-step.

MPLAB allows breakpoints to be set, which cause the program execution to halt at a specified instruction. Its Trace function maintains a record of register values up to the time of program halt. Although it does not operate in real time, instruction or machine cycles can usually be counted, from which program execution time can be deduced. It is possible to simulate the external environment in only a limited way. Despite these shortcomings, once a program has run on a simulator, there can be a fair degree of confidence that it will operate in the target hardware.

It is attractive to retain the facility of the software simulator when the program is downloaded to the target hardware. This can be achieved by using an In Circuit Emulator (ICE). This is a device that usually exists in the form of a pod, which is inserted in the target hardware in place of the microcontroller or processor. It remains connected to the host computer by an umbilical cable.

The ICE is designed to act precisely as the processor it replaces. However, it allows its internal registers to be interrogated by the host computer, and generally it has the breakpoint, trace, and other facilities of the software simulator. The ICE is not without its drawbacks. It is a costly tool, and committed to only one processor, or family of processors. There can be subtle restrictions on its use, for example, it may have clock frequency or power supply range constraints, or it may introduce awkward timing errors in high-speed systems. For these reasons, and because digital electronic devices are becoming ever more complex and compact, the ICE is slowly being replaced by other test methods.

Certain manufacturers now integrate diagnostic tools onto their chips. A good example is the Background Debug Mode (BDM) facility of many Motorola™ devices that allows values of internal registers to be read by the host computer via a dedicated serial link. As the diagnostic facility is built into the device, it is by definition noninvasive, and can have very valuable results.

Conventional electronic test tools, notably the oscilloscope and logic analyzer, remain useful, especially in simpler designs. Certain products, for example, from Agilent, combine the functionality of both instruments into one.

REFERENCES

Hatton, L. (1994) *Safer C*, McGraw-Hill International, ISBN 0-07-707640-0.

MISRA (Motor Industry Software Reliability Association) (1998) *Guidelines for the Use of the C Language in Vehicle Based Software*, McGraw-Hill International, ISBN 0-952415-69-0.

Pahl, G. and Beitz, W. (1995) *Engineering Design, a Systematic Approach*, 2nd edn, Springer-Verlag, London.

Simon, D.E. (1999) *An Embedded Software Primer*, Addison-Wesley, McGraw-Hill International, ISBN 0-201-61569-X.

van Sickle, T. (2001) *Programming Microcontrollers in C*, 2nd edn, LLH Technology Publishing, ISBN 1-878707-57-4.

Wilmshurst, T.J. (2001) *An Introduction to the Design of Small-Scale Embedded Systems*, Palgrave Publishers, ISBN 0-333-92994-2, Basingstoke.

79: Testing Electronic Systems

Patrick D.T. O'Connor
Consultant, UK

1 Circuit Test Principles 499
2 Manual Test Equipment 500
3 Automatic Test Equipment (ATE) 500
4 Special Test Equipment 502
5 Component Test 502
6 Design for Test 502
7 Test Software 503
8 Example of Manufacturing Test 503
References 503

1 CIRCUIT TEST PRINCIPLES

The term unit under test (UUT) is often used in the context of electronics testing to refer to the item or system being tested. The term fault is used for any defect or malfunction.

1.1 Analog circuitry systems

Testing analog circuit functions and parameters involves a range of measurement technologies and instruments. Some of the most important and frequently applied tests are:

- current, voltage potential, and resistance
- gain, impedance, waveforms, and other aspects of AC circuits
- signal characteristics, such as frequency, gain, power, distortion, jitter, and so on, for audio frequency (10 Hz–10 KHz), radio frequency (RF) (10 KHz–1000 MHz), and microwave (1000 KHz–100 GHz) and high-speed digital circuits.

1.2 Digital circuitry systems

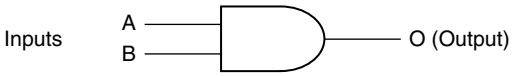
The basic approach to testing digital circuits is to check that the output logic is correct for each possible input state. As a simple example of combinational logic, consider a 2-input AND gate (Figure 1). The possible input conditions and the correct outputs for each of these is shown in the truth table. In principle, it is simple to test such a device for all possible input conditions and faults. The fault conditions in this basic approach are inputs or outputs stuck at 1 SA1 and stuck at 0 SA0.

In this case, four test conditions, or test vectors, are sufficient to test all operating conditions. For any logic network, the number of vectors necessary to show correct operation is 2^n, where n is the total number of possible input conditions. For a simple digital circuit, this presents no problem, but as device complexity increases, the time taken to perform such a test can become too long to be performed practically and economically in mass production.

In devices such as memories and processors, or complex circuits that include such functions, the outputs depend upon the inputs as well as on data flow and memory locations of stored data. This is sequential logic. Alternative test methods have been developed for these devices, such as pattern sensitivity tests for memories.

There are other practical limitations of the basic SA fault model approach to testing. There are other kinds of fault condition that can occur, such as a stuck at input condition, in which an output follows an input logic state. Fault occurrence can also be dependent upon other factors, such as timing, speed, interference, temperature, or voltage supply, as described in O'Connor (2001).

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.



Truth table:

A	B	O
0	0	0
0	1	0
1	0	0
1	1	1

Figure 1. Truth table for 2-input AND gate.

2 MANUAL TEST EQUIPMENT

Manual test equipment is used primarily during development since, at that stage, it is not important to minimize test time and greater flexibility is possible.

The main types of manual test equipment are:

- basic instruments, such as AVO meters (amps, volts, ohms) or digital multimeters (DMMs), power meters, and transistor testers;
- oscilloscopes, spectrum analyzers, and waveform generators to measure, compare, and analyze voltage and current levels and waveforms and to provide inputs so that circuit responses to input waveforms can be analyzed;
- logic analyzers, which are similar to oscilloscopes but specialized for displaying digital pulse streams;
- special instruments, such as radio frequency testers, distortion meters, high voltage testers, optical signal testers, and so on.

Figure 2 shows some typical modern manual test instruments.

2.1 Computer-based testing

Computer-based testing uses software that enables personal computers (PCs) to emulate test equipment. The PC is connected to the circuit to be tested via a data acquisition adaptor card. The PC screen becomes the instrument display, and the tests are controlled from the keyboard and mouse (Figure 3). Software such as LabVIEW™ – *see* **Article 105, Modeling with LabVIEW™, Volume 2** and LabWindows™ (National Instruments Corp. (URL: http://www.ni.com) (Johnson, 1997)) ease the task of test design. Computer-based testing can reduce the costs of test equipment since one PC can perform a range of test functions and test results can be easily stored, analyzed, and transmitted. The microprocessors in PCs also enable tests to be performed faster.

3 AUTOMATIC TEST EQUIPMENT (ATE)

Electronic equipment is tested as part of the manufacturing process to ensure that faults are identified and diagnosed

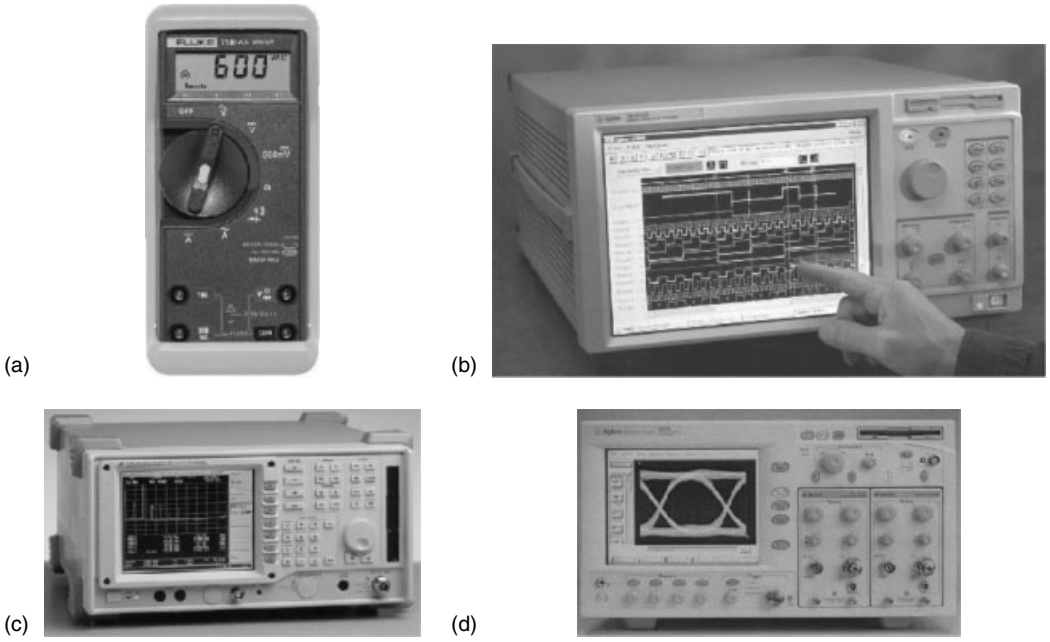


Figure 2. Manual test instruments: (a) DMM (copyright © Fluke); (b) logic analyzer (copyright © Agilent); (c) RF tester (copyright © IFR); (d) optical signal tester (copyright © Agilent).



Figure 3. PC-based oscilloscope with conventional instrument. (Copyright © National Instruments.)

and that shipped units are fault-free. In order to perform the tests economically, especially at high production rates and for complex circuits, most manufacturing testing is performed using automatic test equipment (ATE). The main types of ATE for electronic assembly testing are given below.

3.1 Vision systems

These refer generically to inspection systems that acquire an image and then analyze it. These include automatic optical (AOI) and X-ray inspection (AXI). They do not actually test circuits, but they have become part of many production test sequences because of the great difficulty of human inspection of the large numbers of components, solder connections, and conductor tracks on modern circuits.

3.2 In-circuit testers (ICT), manufacturing defects analyzers (MDA)

ICT (In-circuit testers) tests the functions of components within circuits on loaded circuit boards. It does not test the circuit function. The ICT machine accesses the components, one at a time, via a test fixture (sometimes referred to as a *bed of nails* fixture), which consists of a large number of spring-loaded contact pins spaced to make contact with the appropriate test points for each component. MDAs (manufacturing defect analyzers) are similar but lower-cost machines with capabilities to detect only manufacturing-induced faults such as opens, shorts, and missing components: justification for their use instead of ICT is the fact that, in most modern electronics assembly, such faults are



Figure 4. In-circuit tester (ICT). (Copyright © Agilent.)

relatively more common than faulty components. Figure 4 shows an ICT system.

3.3 Flying probe testers

Flying probe testers (also called *fixtureless testers*) access test points on the circuit, using probes that are rapidly moved between test points, driven by a programmed high-speed positioning system. The advantage over ICT is the fact that the probe programming is much less expensive and more adaptable to circuit changes than expensive multipin ICT fixtures are, which must be designed and built for each circuit to be tested. A consequent advantage is that they can be used for testing during development as well as in production.

3.4 Functional testers (FT)

Functional testers (FT) access the circuit, at the circuit board or assembly level, via the input and output connectors or via a smaller number of spring-loaded probes. Functional testers usually include facilities for diagnosing the location of causes of incorrect function. There is a wide range, from low-cost bench-top ATE for use in development labs, in relatively low complexity/low production rate manufacture, in-service tests, and in repair shops to very large high-speed high capability systems. The modern trend is for production ATE to be specialized and focused at defined technology areas, such as computing, signal processing, and so on.



Figure 5. Functional tester. (Copyright © Agilent.)

Some ATE for circuit testing during manufacture includes combined ICT and FT; the ICT functions are performed first, followed by FT. Figure 5 shows a typical modern test station.

Electronics ATE is a very fast-moving technology, driven by the advances in circuit performance, packaging and connection technology, design automation, and production economics; O'Connor (2001) and Brindley (1991) provide more information.

4 SPECIAL TEST EQUIPMENT

There is also a wide range of specialized ATE, such as for integrated circuits, power supplies, cable assemblies, communication systems, and so on.

4.1 Hot rigs

A 'hot rig' is a system that is used as a test facility for the subassemblies from which it is constructed. Any of the subsystems can be tested by being inserted into a cabinet that contains the other subsystems and then the system can be tested. 'Hot rigs' are often used for testing in the manufacture of systems, particularly when the quantities produced are relatively low so that test time is not critical. They are also used in repair depots for testing repaired items.

4.2 Electromagnetic compatibility (EMC) testing

Testing for electromagnetic compatibility EMC and electromagnetic interference EMI is described in **Article 17, EMC and EMI, Volume 1** and **Article 216, Test Facilities, Volume 3**.

5 COMPONENT TEST

Since electronic components are always tested by their manufacturers, and since modern manufacturing quality is very high, nowadays, it is very seldom appropriate or necessary for components to be tested by equipment or system manufacturers. We can consider electronic components to be in two categories from the test point of view: discrete components and integrated circuits or ICs.

5.1 Discretes

Components such as resistors, capacitors, connectors, coils, transistors, and other relatively simple types generally possess a relatively small number of performance parameters. Therefore, testing them is relatively easy, quick, and inexpensive, and they are fully tested by the component manufacturers. The manufacturing processes for these components are also relatively easy to control, resulting in high yields or low proportions of defectives. Therefore, testing by users is rarely cost-effective.

5.2 Integrated circuits

ATE for IC testing must be capable of testing all or most of the functions of the very complex circuits, at the maximum operating speeds. The problem of testing very large and fast ICs is one of the major challenges of modern electronics technology. New methods for testing very large and fast ICs are being developed in response to the challenges of complexity and cost. ATE for IC testing is very expensive, so testing by users is seldom cost-effective.

6 DESIGN FOR TEST

6.1 Tasks of an ATE

In order to test a circuit or system, the ATE must be able to perform the following functions:

- *Initialization*. The ATE must put the UUT into known initial conditions, for example, digital inputs and outputs at known logic states, parameters at known values, and so on.
- *Control*. Inputs must be generated to stimulate changes on the outputs.

- *Observation*. The ATE must be able to detect the changes in output states and values.
- *Diagnosis*. The ATE might be required to perform automatic or guided diagnostics to identify the origins or causes of faulty output states.
- *Partition*. It is often important that the ATE can test relatively small parts of complex systems separately to maximize test coverage, to enable effective fault diagnosis, and to minimize test time.

These capabilities are affected not only by the performance of the ATE but also significantly by the design of the UUT. Therefore, it is important that the system design takes account of the need for test. This is called *design for test* (DFT).

7 TEST SOFTWARE

Test software must be developed for any circuit that is to be tested on ATE. Automatic test program generation (ATPG) software is available to perform this task. The ATPG will create the test software from the circuit information on an electronic design automation (EDA) system and indicate the fault coverage. For some systems, particularly aerospace electronics (avionics) and military systems, the abbreviated test language for all systems (ATLAS) test language is used.

7.1 Scan design

For very large digital electronic circuits, internal circuits are often provided, which assist the ATE to perform the tests. The best-known approach is boundary scan. A boundary scan circuit receives test instructions from the ATE and then performs logical tests on the interconnections between the digital circuits that it controls. This capability is important for large IC packages containing many connections, particularly if they cannot otherwise be accessed for test by the ATE. The international standard for the boundary scan approach is IEEE1149.1. Most modern digital ICs include boundary scan capabilities. The boundary scan method is described in Parker (1999), and information is also available from JTAG Technologies (URL: http://www.jtag.com).

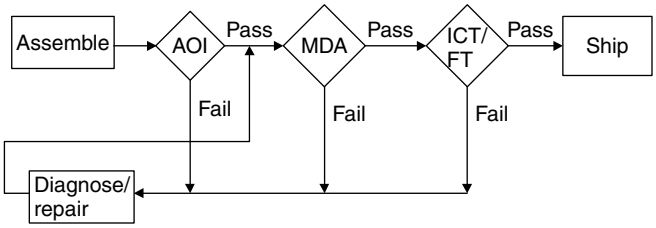


Figure 6. Typical electronic manufacturing test flow.

8 EXAMPLE OF MANUFACTURING TEST

For any manufacturing situation, the optimum test approach must be planned in relation to technologies, economics, and other factors. This topic is covered in Davis (1993) and O'Connor (2001). Figure 6 shows an example of a typical manufacturing test sequence.

REFERENCES

Brindley, K. (1991) *Automatic Test Equipment*, Newnes, London.

Davis, B. (1993) *The Economics of Automatic Testing*, 2nd edn, McGraw-Hill, London.

Johnson, G.W. (1997) *LabVIEW Graphical Programming*, 2nd edn, McGraw-Hill, New York.

O'Connor, P.D.T. (2001) *Test Engineering*, Wiley, Chichester.

Parker, K. (1999) *The Boundary Scan Handbook*, Kluwer Academic, New York.

80: Principles of Fine Mechanics – Kinematic and Elastic Designs

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Purpose of Mechanical Design in Instruments 507
2 Specification of Measurement Systems 508
3 Problems of Specification 510
4 Controlling the Mechanical Degrees of Freedom 511
5 Kinematic Design Basics 511
6 Elastic Design Basics 513
Related Articles 513
References 513

1 PURPOSE OF MECHANICAL DESIGN IN INSTRUMENTS

Mechanics is the subject that deals with the study and application of the motion and equilibrium states of material bodies. The aim, in the mechanical aspect of instrument design, is to create a specific mechanical system that provides the necessary controlled movements in such a way that predetermined moving and fixed relationships are preserved under reasonable conditions of operation.

Reuleaux, in his *Kinematics of Machines* of 1875, was apparently the first person to see machine design as more than a complete design that was modified until it worked. He defined a machine as 'a combination of resistant bodies so arranged that by their means the mechanical forces of nature can be compelled to do work accompanied by certain determinant motions'. He further expanded that motion of parts is simply determined by the geometry of the system of parts (Burstall, 1970). Reuleaux's generalizing definition applies equally well to mechanical instruments that are a form of machine.

Departure from the ideal concept will occur for many reasons, for example, unavoidable elasticity of component members, clearances and wear in joints, and 'second-order' influences such as thermal expansion of the material. Design is the process whereby a satisfactory compromise is made between the range of envisioned parameters to come to a single adequately satisfactory arrangement. The designer strives to produce a design statement that will lead to an article being made, which performs the set task within the stated specifications of performance without the need to remake the article after it is first assembled.

Mechanical design has two main stages of progression. The first is where the kind of necessary mechanisms are realized as a system. The following stage then selects a method of creating the chosen individual mechanisms as hardware that will suit the total task. For example, the mechanism seen to be needed may be a means of rotating one shaft at a different speed to that of the other while maintaining a strict gearing ratio. This is the first step. Followed by the study of the various ways of achieving this – toothed belts, friction gearing, spur gears, and others – to see which suits the case in hand.

No matter which type of mechanism is involved, the fundamental issues behind the choices of the first step are concerned with controlling the freedom of movement of selected parts and points. This will be expanded later.

The design of a measuring instrument, that a machine uses for handling informational rather than energy quantities, generally requires specific attention to the fidelity of the input–output and other relationships over extended

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

periods of operational life. There is less interest in efficiency of energy transfer (although in mechanical design, losses due to friction can cause thermal changes to components that give rise to error). Instrument design is generally dominated by attention to deflections of the system and its components rather than to their strengths.

A general introduction to mechanical machine design can be obtained from the widespread works describing the design of machines that can transfer energy, examples being Shigley (1988), Erdman, Sandor and Kota (2001), Juvinall and Marshek (2002), Mott (1998), Faires (1955), Beggs (1955) and Lent (1970). These accounts will, however, only take the user a limited distance into instrument design, for they generally lack specific detail about fine mechanism requirements.

Work which discusses the general systems level of design of instruments are few and generally now seen to be old by virtue of their date of publication. They are, however, most useful accounts, encapsulating considerable experience and know-how. These include Glazebrook (1922) Vol. III, Rolt (1929) Vol. II, Whitehead (1934), Strong (1938) and Braddick (1954). This material is provided to help overcome the difficulties in procuring information on this topic, now that it has little research in place.

Factors important to design include the methodology for controlling the design itself, cost of producing the design, building the instrument and servicing it, the degree of serviceability, the ease of measurement in production, the needs for calibration after manufacture, and ergonomics requirements (now often called the *man–machine system* – MMS). These and others must all be properly thought out so that the users get correct results. Many of these vitally important factors are often neglected in the initial design, being added by adaptation or modification at a later stage when the design has progressed so far into production that major change cannot be entertained. For example, a firm making fever thermometers only realized that their product must fulfill certain calibration criteria when they were ready to launch the product on the market place!

At present, instrument design is largely a case of some intelligent application of science plus a very large component of experience through years of actual involvement with similar products.

Design procedures for each of the different physical regimes of an instrument have their own peculiarities. Mechanical design must allow for the fact that a wrong choice is not as easily corrected as in, say, the electronic medium. There is, for example, no simple mechanical counterpart of the variable resistor of electronics, which can be placed into the system to find an optimum value with great ease and with little cost or performance penalty. To obtain adequate performance from a mechanism, usually very close tolerances are required. Dimensional parameters in mechanical systems will often require parts per million precision compared with parts per hundred in most electronic techniques. If adequate performance is to be achieved, it means that the various parts of a mechanical system must be made precisely, or at least be capable of being assembled to close tolerances. Control of wear is important since it may degrade performance.

Electronic design, in contrast, can more easily compensate for changes to the value of components, since they have been initially burned in to eliminate the few that will fail early.

To its credit, mechanical design is more intuitive and readily understood; wrong factors in a mechanical system often show up very easily without the use of sensory aids because mechanical systems interface more directly with the human sensors than concepts of other disciplines. Mechanical resonant systems can also yield very high-Q oscillations for low component cost.

An important feature of mechanical systems, which may be either advantageous or deleterious, is the inertia of components. The necessary inherent mass of mechanical parts is often the limiting parameter to dynamic performance when high frequency operation is needed. It can, of course, be a useful factor providing time integration properties. In a multiregime instrument, the mechanical parts will often dictate the upper frequency limit of operation. Generally, unless specifically and carefully designed to be otherwise, mechanical systems will be low pass in performance.

2 SPECIFICATION OF MEASUREMENT SYSTEMS

There exists a maxim '$pecify $anely or $uffer'. It sums up the importance of spending enough time on raising adequate and realistic specifications on which to base the design. Mechanical design does not differ from any other design in that to get started one must have a set of specifying statements that will outline, in a quite rigorous and meaningful manner, what the instrument must do. These statements are called *specifications*. Too often, instruments just happen with the specifications growing as the design proceeds. A proper set of 'specs' at the beginning of the design will aid the designer in realizing what is ultimately required. They should be set to paper *before* in-depth design begins, for once the designer has become immersed in the design, it is a fact of life that he or she can become very resistant to change.

Specification should not be raised by the designer or user alone but as a group effort. There should be a joint agreement between three parties – the system end user, the

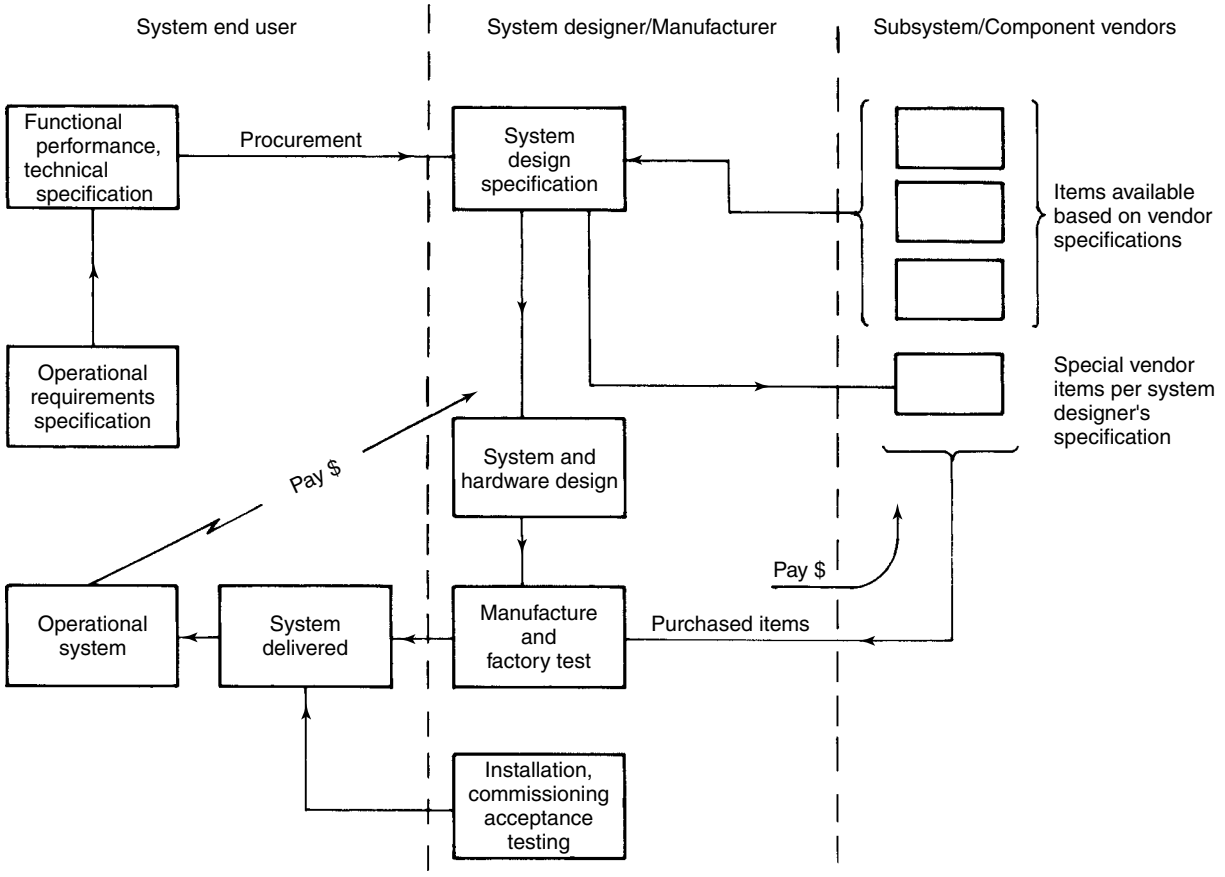


Figure 1. Vested interests in a typical system design process.

system designer and manufacturer (if they are of the same group), and the subsystems and component vendors who supply the parts that will be used, generally without modification. Figure 1 shows how Wheeldon (1974) depicted this process. His two papers expand the following points now given. Various kinds of specifications exist, each having a different purpose of definition.

Systems user and systems designer relationships

- Operational requirements specifications
- Functional and technical requirements specifications
- Design specifications (to implement and check design)
- Factory test specifications
- Acceptance test specifications (Factory)
- Documentation specifications
- Packaging and delivery specifications
- Installation specifications
- Commissioning specifications
- Acceptance test specifications (Field).

Systems designer and component vendor relationships

- Standard item specifications
- Modified 'Standard' items
- Special subcontract specifications
- Incoming inspection specifications (system designer)
- Incorporation of vendor specifications in system design specifications.

The designer is the person who has to provide specific detail to the factors presented by the various groups. A realistic producible design at a satisfactory cost must be achieved. Value judgments are needed about what the customer says is needed (which is often what is then *thought* to be needed); what the vendor can offer (or what is *said* to be available); the designer's own ability and facilities for design; and the permitted time and money available to complete the task.

Unfortunately, to specify a product completely before it is made requires powers of prediction and enormous effort in attempting this virtual impossibility. To attempt to create a total specification that gives the user complete assurance of being absolutely safe from faults and failures of compliance is an unreasonable expectation. The cost and time needed to get even close to this are prohibitive. A compromise is therefore struck that specifies the product in a short-form manner in reasonably standard 'shorthand' statements.

Specifications, therefore, attempt to portray what both the designer or maker and the purchaser or user expect to produce and receive, respectively. Owing to the small cost level associated with the smaller systems and the stand-alone instrument, the verbal interface is poorly developed and exploited. In larger extensive instrument systems, the complexity there makes better assessment necessary, but then the complexity still prevents absolutely adequate statements from being realized.

3 PROBLEMS OF SPECIFICATION

In practice, dissatisfaction often arises between the parties to the contract and, in cases where the financial implications are large, independent arbiters are called in to make judgments. The art and science of contract and specification writing are now extensive; specialists exist to cover the needs. It is beyond the scope of this article to go much deeper, but here are some simple rules to follow.

Try to use specifying statements that are the result of past experience. Examples are Codes of Practice and the various Standards specifications. For instance, reference to the BS code for dimensional tolerances will show if those called for are reasonable for the size of the part.

Always watch for 'minor' details in a statement that may be more significant than they are thought to be. For example, surface finish to be within a given limit may be specified for a pressure cell capsule, yet nothing is given about the fluid tightness of the material. Further, never leave important factors out because they seem so obvious. The converse also holds; if they are not given, check them out.

All statements in a specification should be carefully studied *before* signing an agreement; it is of little value to look into it afterwards. Look carefully for fundamental inconsistencies. Does the frequency response called for exceed that of any mechanism yet produced? Does the required or specified noise level fall below the thermal control errors? Suppliers cannot make an instrument that refutes the known physical laws; but they have been known to offer such for sale – unwittingly, of course!

Make clear what specifications apply to what, and at what stage. Care is needed to prevent them 'creeping' in the hands of enthusiastic clerks. There was a case of an optical shaft encoder that had to read to arc-minute accuracy when mounted on a radar tracking head. By the time the specifications were transferred through several stages of contract writing and then into the maker's plant, they had crept to stated need for arc-seconds performance. The manufacturer, not wishing to be caught out, went one better. The result was that the encoder cost about a hundred times the price of one that would have done the job. Furthermore, the full encoder accuracy could not be used because the radar head was not designed for better than arc-minute use.

Specifications can also come to the designer as the result of the so-called 'horse-camel' syndrome. The end user collects data sheets of the various products offered by several different sources. From each of these he picks out the individual good points, adding these together to raise his own specification list. This process entirely ignores the fact that these special features are often the result of the designer trading off one feature to allow another to be enhanced. The result of this combinational method is a specification of an instrument impossible to make. Unfortunately, there are many suppliers in the instrument manufacturing industry who lack enough experience to recognize that such tenders may be impossible to make, and they take them on only to find later (to their and the purchaser's regret) that they had been attempting a virtually impossible task. The use of independent expert consultant groups, a well-trained in-house team, study contracts, and part-stage costing may help.

At all times, it is vital to 'clear the air' between all groups involved. If the user believes the designer cannot do what is suggested, then it should be challenged very early. Similarly, the designer should state what is unclear early in the negotiations with the consumer. Change becomes progressively difficult as the designer is gradually forced to freeze the parameters in order to make concrete progress.

When writing specifications, considerable care must be given to the provision of statements that are meaningful and not ambiguous. Unfortunately, to write a reasonable specification requires the use of abbreviated statements. For example, statements of accuracy or linearity are used very loosely. Graphs could be used much more to define such parameters as frequency response, linearity, and the like. When in doubt, seek further detail about the meaning of a parameter.

There is no harm in verifying the self-quoted performance of subcontractors and vendors.

All too often, the user does not adequately think out the task for the instrument that is called for. The designer should be expected to ask poignant questions of the user, for the designer will often know the design problems more closely than the user. Very often, discussions show that the user really wanted something quite different from what was thought to be necessary. Take, for instance, the person who once asked for a thermometer to measure length changes of a component because he did not see that it could be measured directly as a dimension.

Each case stands on its own merits. In measuring instrument design, the spread of applications is so broad and the design becomes so complex that poor performance

often results. The ongoing state of instrument design and manufacture is such that it is suggested that instrument users must constantly be on their guard against the supply of faulty instruments – see Moss (1978) for a relevant report illustrating the potential situation.

Instruments, especially one-offs and those with mechanical elements, are still in a state of low inherent quality capability when compared with the performance that one now expects and usually gets when using electronic systems based on standard integrated circuits. If an IC works, it is going to be in specification and continue to operate. It will also be a very well designed component. In contrast, mechanical systems may continue to operate but with deteriorating fidelity.

4 CONTROLLING THE MECHANICAL DEGREES OF FREEDOM

Mechanical design is fundamentally concerned with providing a system of materials arranged to allow controlled movements, ranging from no movement in any axis, to all six degrees of freedom of an extended object within a defining space.

A theoretical point in space has three degrees of freedom – the three translations that are generally expressed as x, y, z in the Cartesian system; other coordinate methods can be used. The concept of rotation has no meaning for a theoretical point.

A point forming a position on an extended object, however, can have six degrees of freedom for, in addition to the three possible translations, there can be three axes of rotation about the point.

The essence of good mechanical design is to devise a system that restrains as many of the six freedoms as are needed while allowing freedoms. For example, the slide of a traveling microscope must allow one translation, and while doing so, constrain the other two translations and three rotations as perfectly as are needed. Lack of straightness of the slide will cause small translations and rotations to occur as it slides. The degree of precision needed dictates the necessary tightness in the tolerances of manufacture and the method by which this is accomplished in the face of wear, temperature effects, and other error sources.

Some other examples are the framework and sliding axes of a 3-D measuring machine (three translations allowed with rotations constrained) and the surveyor's theodolite (constrains all translations allowing the rotations).

An equally important case is where all degrees of freedom are to be constrained. This occurs when mounting a component to stay in a chosen position relative to another; for example, a method of locating a survey-leveling unit that is to be used repeatedly around a series of test lines.

At times, measuring machines are sometimes specified as, say, $3\frac{1}{2}$ axes, meaning that they have a rotation allowed within the movement provided along one translation axis. It does not mean that the rule of six degrees of freedom for an extended point has been broken.

Two extremes of design philosophy exist for realization of a suitable mechanical structural relationship. Each has its advantages and disadvantages and often the design makes do with a compromise between the two extremes. These extremes are called *kinematic* and *elastic designs*; in between them is the semi-kinematic arrangement.

5 KINEMATIC DESIGN BASICS

When all parts of a structure are held in such a way that they retain positional relationships without being in a state of variable stress except at their points of contact, the design is said to be kinematic. In other words, effects such as a change in size due to temperature or material creep will not stress the frame. Clerk-Maxwell in the *Handbook of the Special Loan Collection of Scientific Apparatus* 1878 wrote this:

> When an instrument is intended to stand in a definite position on a fixed base it must have six bearings, so arranged that if one of the bearings were removed the direction in which the corresponding point of the instrument would be left free to move by the other bearings must be as nearly as possible normal to the tangent plane of the bearing.

Strong (1938) defined the principle of kinematic design as being applied when a body has at least 6-n points in contact with a second reference body when it is to have n degrees of freedom.

A spherical ball sitting in a triangular hole, as shown in Figure 2(a), rests on three points of contact while it is under the force of gravity (which acts as the 'locator'). Thus, it has three degrees of freedom constrained with respect to the datum block (the translations) yet is free to rotate about three mutually perpendicular axes.

The classic example, demonstrating constraint of all six freedoms, is the so-called 'hole-slot-plane' or Kelvin-coupling arrangement. This is easily understood by reference to Figure 2(b). The hole constrains three translations (three contact points), the slot constrains two rotations (two contact points), and the plane constraints the other rotation (one contact point). Addition of any extra contact points, unless made perfectly enough (as in a ball-race bearing), would add a 'redundancy' causing elastic deformation of the frame due to 'overconstraint'.

It is easy to see that changes to the size of the top plate will cause the legs to move yet retaining the same stress

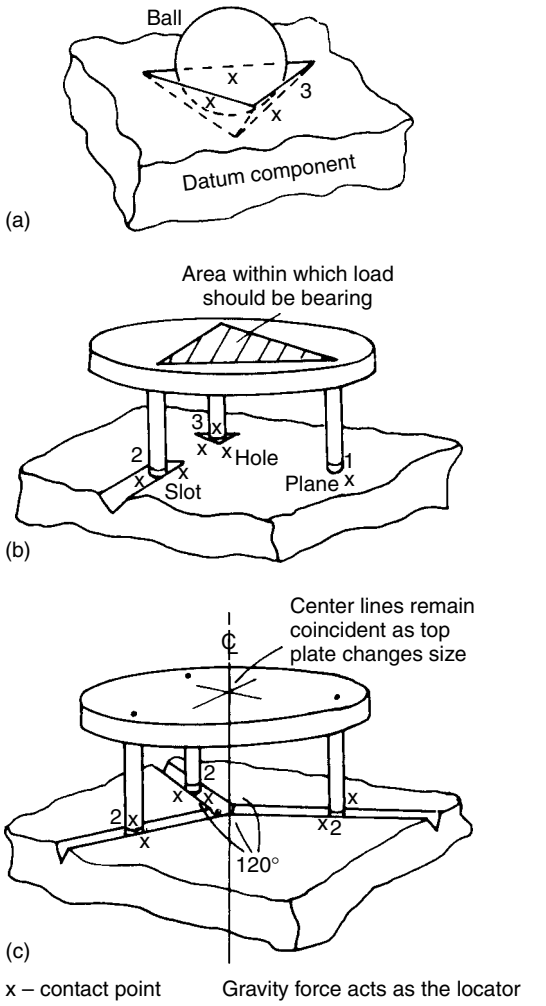


Figure 2. The degrees of freedom in a kinematic design of a coupling are equal to six less the number of contact points. (a) All three translations constrained. (b) Hole-slot plane in which all six freedoms are constrained. (c) Slot-slot-slot variation of case b.

relationships (provided the slotted plane contacts move accordingly). Note that the hole-slot-plane does not, however, retain the center of the top member coincident with that of the lower member should a relative dimensional change occur between the two members. A variation of the 'hole-slot-plane' that uses three radiating slots (Figure 2c) overcomes the center shift defect. Figure 3(a) shows a kinematic form in which one degree (translation) is allowed, this being proven because five contact points exist. Figure 3(b) shows how this form is modified with rolling bearings, to retain point contact while reducing wear. It is typical of the slide of a traveling microscope.

Many arrangements are possible depending on the need to allow or control a given number of degrees of freedom. Sketches of various couplings are given in Whitehead (1934) and Braddick (1954).

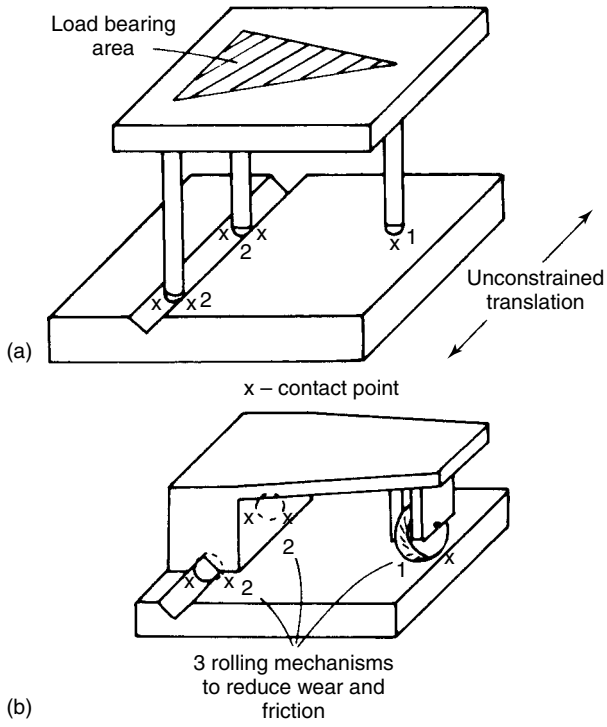


Figure 3. Kinematic design allowing one freedom of translation; (a) contacts subject to high wear; (b) contacts replaced with rolling devices.

The principle of kinematic (or other alternative) design is applied whenever two assemblies need to be coupled. It is necessary, therefore, to break down the whole system into its basic couplings or mechanisms, tackling the design of each in a hierarchical order to ensure that overconstraint does not occur unwittingly at a later stage by the addition of a link in an incorrect way.

Features of the kinematic method are that its parts can be made and installed or replaced without altering the positional system. Once made, the kinematic assembly will be reassembled with the same dimensional performance. Secondly, wear does not induce play in the coupling mode – it may, however, introduce dimensional change. Another feature is that the stresses exerted on all parts are well defined, and they can be designed virtually in isolation from the rest of the system once their contact loads have been decided.

Disadvantages are that there is a need for application of a locator force to hold parts together, such as gravity, magnetism, or small springs. The points of contact are also heavily loaded and consequently operate under conditions of high stress. The pure principle is often degraded by the use of extended contact surfaces to reduce the contact stresses – semi-kinematic design.

A further difficulty arises when it is seen that the systems result in triangular space frames when redundant links

are not present. The three-legged stool is a good example of the possible undesirable features of purely kinematic design. The stool always sits correctly on any reference surface without rocking, but place the load outside the central triangle defined by the leg positions and the linkage is soon lost. The locator should not be designed to withstand such forces.

6 ELASTIC DESIGN BASICS

The opposite philosophy is to join components rigidly, allowing them to flex within their elastic properties: a four-legged stool will still sit on any surface if the top is made flexible. The elastic philosophy deliberately contravenes the 'zero redundancy' law, doing so in a defined manner.

For this method to be designable, the system must be so constructed that the designer knows just what flexes and where. In cases where only limited translation and rotation is desired, the elastic method is probably the best. Purely elastic design is almost an academic abstraction. In practice, the method comes somewhere in between totally kinematic and elastic, being called *semi-kinematic*.

As an example of the two extremes, consider the need to swing a pendulum or the beam of a weighing balance. The pivot may be made using knife-edges that rock on a rigid flat (kinematic) or it may rotate on a single or a crossed flexure strip pivot (elastic). Another example might be the need to support a sensitive tilt meter on a given surface in such a way that the relative length changes, owing to temperature effects on the body of the instrument with no strain on the instrument body. This can be done with either the kinematic hole-slot-plane concept or by the use of three elastic flexible legs that bend a little as the body changes length (see Figure 4). It is important to make specific calculations of the degree of flexing that is needed. What may appear to be a stiff rigid member can be comparatively flexible when the forces are high or the movements small. The designer should regard all components at the onset as being compliant, altering this belief only after calculation of strains due to disturbances shows them to be adequately stiff or compliant as the case dictates.

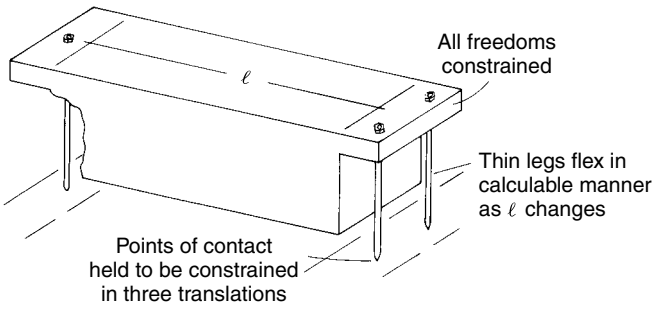


Figure 4. Semi-kinematic design – (ANAC tiltmeter support system).

The mechanisms of the two groups of design – kinematic and elastic – will be dealt with in more detail in later articles.

It is surprising that few texts on general machine design actually discuss these basic design differences. In some of the works that are specifically concerned with instrument design, these concepts are expanded. Greater depth and more examples on the topic can be found in Glazebrook (1922) Vol III, Whitehead (1934), and Braddick (1954). The choice of which philosophy to use and when is very much a matter of training and experience. Where elastic methods can be used, they usually offer considerable cost and assembly advantages. For example, locating one part relative to another with, say, one allowed translation of short stroke can easily be provided with a parallel-strip suspension. It will be virtually free of frictional effects. It will be cheaper to make but does suffer in that a changing deflection force is generated as the 'slide' moves.

RELATED ARTICLES

Article 64, Executing A Measuring System Design, Volume 2; **Article 69, Requirements Allocation, Volume 2**; **Article 81, Principles of Fine Mechanics – Systems Considerations, Volume 2**; **Article 85, Elastic Regime of Design – Design Principles, Volume 2**.

REFERENCES

Beggs, J.S. (1955) *Mechanism*, McGraw-Hill, New York.

Braddick, H.J.J. (1954) *The Physics of Experimental Method*, Chapman and Hall, London.

Burstall, A.F. (1970) *A History of Mechanical Engineering*, Faber and Faber, London.

Erdman, A.G., Sandor, G.N. and Kota, S. (2001) *Mechanism Design: Analysis and Synthesis*, Prentice Hall.

Faires, V.M. (1955) *Design of Machine Elements*, Macmillan, New York.

Glazebrook, R. (1922) *A Dictionary of Applied Physics*, Vol. III, Macmillan, New York.

Juvinall, R.C. and Marshek, K.M. (2002) *Fundamentals of Machine Component Design*, Wiley, New York.

Lent, D. (1970) *Analysis and Design of Mechanisms*, Prentice-Hall, Englewood Cliffs, NJ.

Moss, C. (1978) Quality of precision measuring equipment. *NCSL Newsletter*, **18**(1), 23–24.

Mott, R.L. (1998) *Machine Elements in Mechanical Design*, Prentice Hall.

Rolt, F.H. (1929) *Gauges and Fine Measurement*, Vol. II, Macmillan, London.

Shigley, J.E. (1988) *Mechanical Engineering Design* (McGraw-Hill Series in Mechanical Engineering), McGraw-Hill.

Strong, J. (1938) *Modern Physical Laboratory Practice*, Blackie and Son, Glasgow.

Wheeldon, R. (1974) *Specifications – The Identifying Facts*, Department of Cont. Education, University of New England, Australia.

Whitehead, T.N. (1934) *Instruments and Accurate Mechanisms – Underlying Principles*, Macmillan, London.

81: Principles of Fine Mechanics – Systems Considerations

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Wear-true Philosophy in Mechanical Design 515
2 Deciding the Right Size of a Fine Mechanical Instrument 515
3 Static and Dynamic Considerations for a Fine Mechanism 516
4 Mechanical Systems Modeling 517
Related Articles 519
References 519

1 WEAR-TRUE PHILOSOPHY IN MECHANICAL DESIGN

This article continues the discussion given in **Article 80, Principles of Fine Mechanics – Kinematic and Elastic Designs, Volume 2** by adding systems issues. Kinematic design is covered in **Article 82, Kinematical Regime – Members and Linkages, Volume 2**; **Article 83, Kinematical Regime – Fasteners, Bearings, Volume 2** and **Article 84, Kinematical Regime – Rotary Motion, Volume 2**. Elastic design is dealt with in **Article 85, Elastic Regime of Design – Design Principles, Volume 2**; **Article 86, Elastic Regime of Design – Spring Systems, Volume 2** and **Article 87, Elastic Regime of Design – Plates and Bimorphs, Volume 2**. Errors sources in both kinematic and elastic deisgn regimes are the subject of **Article 88, Error Sources in Fine Mechanics, Volume 2**.

In the Aeronautical Journal of July 1913, Horace Darwin, cofounder of the Cambridge Instrument Co., expounded a concept of design that states that as mechanical elements wear through use they retain the dimensional precision needed. He called it the *wear-true philosophy*.

The use of kinematic principles goes a long way toward this end, and by appropriate choice of proportion and arrangements, it is often possible to achieve a linkage that does not degrade in its accuracy as it wears or deforms.

Darwin illustrated this using the examples of the bearing arrangement of a wheelbarrow wheel. When the axle is fixed in the wheel, the axle ends wear round, and the bearing hole wears oval, yet still round enough to give good location. But fix the axle in the shafts and the shaft wears oval in the wheel and the wheel bearings wear very large and sloppy.

There seems to be no basic statable rule for achieving this and a situation that occurs often is when the designer attempts to arrange wearing surfaces to wear evenly on each contact face of the sliding surface. Another approach is to make the wear or deformation occur at a geometric position that has the least effect on the geometrical precision of the measuring component. Figure 1 shows two designs, one of which does not adhere to the 'wear-true' philosophy.

Darwin incorporated this principle in the design of his very successful 'rocking microtome' instrument of the 1885 period. It was capable of retaining the ability to cut microscope specimen slices to micrometer thickness for a half century of service – see Cambridge Instrument Company (1945) and Braddick (1954).

2 DECIDING THE RIGHT SIZE OF A FINE MECHANICAL INSTRUMENT

Unlike the situation occurring in electronic and in optical disciplines where the size of components usually dictates

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

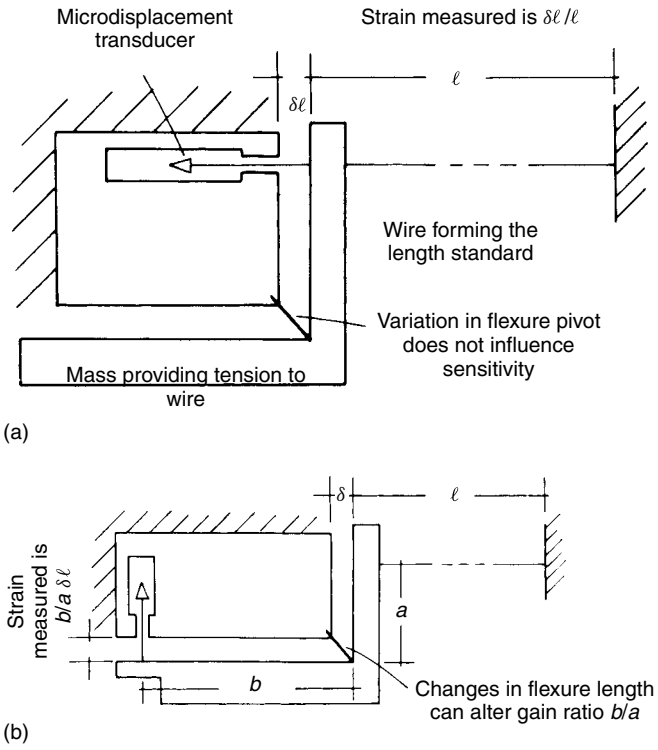


Figure 1. Two balances used to tension a wire length-standard in earth-strain meters; (a) uses wear-true design principle, (b) does not.

the overall size of the instrument system, mechanical instruments may or may not have clear-cut reasons for having any particular size. So, at the outset of design, a choice must be made if progress is to take place that does not need to be completely revised later. Here are a few considerations that may help decide the overall size of the instrument. Braddick (1954) provides specific case studies to illustrate some particular reasons for a given size.

Certain functions, such as a size dictated by a necessary component, may decide the basic size to be used. Examples are the objective size and focal length in a telescope, the length of translation of a measuring machine, the load and volume to be supported or contained, and the need to match a given existing product.

Portability may dictate low weight, but small size often means increased cost and difficulty with assembly and servicing. However, dynamic response rises as parts become smaller. The physical size of a high frequency response anemometer must be very small.

If energy must be stored in mechanical or thermal form, large rotating or static masses may be needed. Examples are the wheel of a gyroscope and the bed of a measuring machine that must integrate short-term temperature fluctuations.

Ergonomic needs dictate that the unit must have manual controls that suit man-size fingers or be mounted at heights that can be easily seen and reached. Whatever the reason, the designer should attempt to find objective bases for choosing a given size. If not, there will always exist the possibility of having to rescale the physical form at a later date.

3 STATIC AND DYNAMIC CONSIDERATIONS FOR A FINE MECHANISM

The static regime of a measuring systems is covered in **Article 57, Static Considerations of General Instrumentation, Volume 1**. The dynamic regime is introduced in **Article 59, Introduction to the Dynamic Regime of Measurement Systems, Volume 1**. The transfer function of a system block describes the input–output relationship at all relevant frequencies of operation. It comprises a DC gain term multiplied by a frequency dependent term. Although there exists in reality a continuum of performance from DC to the highest frequency of operation, descriptions of the performance are generally split into two domains, the first being the steady state or static characteristics and the other the dynamic behavior that is described by the frequency term.

Mechanical systems are no different from others in that they possess both static and dynamic characteristics.

The design of a direct-reading traveling microscope requires virtually no understanding of the dynamics of movements that the operator imposes on its sliding parts unless it is large, in which case the inertia of the sliding part can be great enough to impart excessive forces restricting the operator's manipulation response. However, if the readout is automated by closing the translation drive loop, the dynamics of the whole system become important if stable and rapid high-speed operation is to be obtained.

Apart from a very cursory statement saying that dynamics do occur in instruments, the majority of instrument design texts give little help with this aspect to the designer of modern multiregime instrumentation. Instead, the designer must read much more widely in order to become familiar with the basics of dynamical systems and then apply the general rules to instrument design where it is needed.

Definition of static performance involves specifications being stated about such parameters as resolution of readout sensitivity to a given variable input, linearity, accuracy, precision, repeatability, and many more concepts. These concepts, which are unfortunately not totally standardized in terminology, are described in most works on instrumentation and therefore need not be developed here. Doebelin

(1975), Foxboro (1972), and Bell and Howell (1974) are examples. They are terms generally describing only the static behavior, the results of tests measured after the systems come to rest. During transients, the performance can be very different indeed. In many cases, the manufacturer provides no statement of dynamic behavior.

In the dynamic mode, the designer must be aware of the effect of friction in its various forms, of the effects that the various kinds of slop between couplings produce: backlash, dead zones, and other nonlinear operation. Where the best acceleration is needed from a motor driving load, the load inertia must be matched to that of the driving source. A gearbox is often needed to obtain this match. The transient performance of electric motors is generally compared by the magnitude of their ratios of torque^2/inertia. This is called the '*power-rate*' (West and Leonard, 1955, and Rothwell, 1966). The same term can be applied to any form of motor (spring motors, gravity feeds, and other mechanical motors), in a similar manner (Sydenham, 1969), giving the designer a chance to assess the relative merits of various mechanical or electrical drives. Where high acceleration is concerned, the diameters of rotating parts should be kept as small as possible; rotational inertia increases rapidly with an increase in diameter, sometimes as the fourth power of the radius. The problem of matching mechanical loads to sources is identical to the problems existing in electrical power transfer.

When resonance is to be invoked, or appears unwanted, the designer needs to know how to enhance or reduce it. The role of damping must be understood. In some instances the designer attempts to design for the highest Q conditions so that the system constantly 'rings' by noise excitation (a relevant example being one form of gravity wave detector that uses large freely supported blocks of aluminum as detecting resonators). This provides increased sensitivity to signals at or near the resonant frequency. The dimensional changes, transduced to electrical signals, are then inverse-filtered to remove all but the varying amplitude of the frequency at which the system is resonating. This variation contains the signal to be detected. In this case, damping is deliberately kept to a minimum.

In many applications, however, the aim is to introduce damping deliberately to produce a certain response to input excitation. This is done by adding damping to the resonant system with suitable mechanical, or other regime, dissipation elements. Consideration of the means to reduce the effects of mechanical hysteresis suggests that the addition of damping is not always the best approach to use, because vibration to rest can often improve the static precision of such elements as bellows, springs, and diaphragms.

Thorough study of the dynamic performance of an instrument is far more complex and time consuming than is the case for the relatively simple static domain. Certain mechanical parts may not be stiff enough or may possess too much mass when under the rapidly changing movements existing in velocity, acceleration, and jerk conditions. The first step to designing for the dynamic mode in a machine linkage begins with studies that provide the designer with maximum values of velocity, acceleration, and jerk. The kinematics of linkages is a topic well covered in texts on the design of mechanism, see, for instance, Beggs (1955) and Mabie and Ocvirk (1958). This leads to the stress levels that occur from which the deflections can be calculated (although often with considerable difficulty, for the maximum deflection and dimensional change points are not always in obvious positions). Inertia force considerations calculated from knowledge of the moments of sections of components, coupled with the velocity and acceleration data, lead to other dynamic considerations.

4 MECHANICAL SYSTEMS MODELING

The above description of the static and dynamic aspects of instruments is not based upon an overall unified systematic understanding of the topics but is typical of the ad hoc approach that many designers use. Each of the above observations can be better explained and calculated using a common fundamental mathematical basis for dealing with the performance of physical systems. The mathematical description of mechanical linear systems is the same as that for electrical, thermal, and fluid regimes, *see* **Article 59, Introduction to the Dynamic Regime of Measurement Systems, Volume 1**. By the use of appropriate coefficients, the performance equations of one regime can be interfaced with those of another. Figure 2 is an example showing how the moving coil meter can be expressed in a common electrical regime, which can then be studied as an electric circuit to obtain its dynamic behavior. This means that electromechanical systems can be represented by a single mathematical expression. The actual performance will closely be as predicted from this model (in so far as the real situation matches up with the model used).

This understanding is represented in the systems approach to physical system – see Finkelstein (1977) for its use in measurement. This is not to be confused with the much more widely applied 'general systems theory' that attempts to take in far more experiences, including such topics as biological and sociological phenomena. The latter is similar in concept but has to operate without the same degree of mathematical rigor and formal representation that is present in the former.

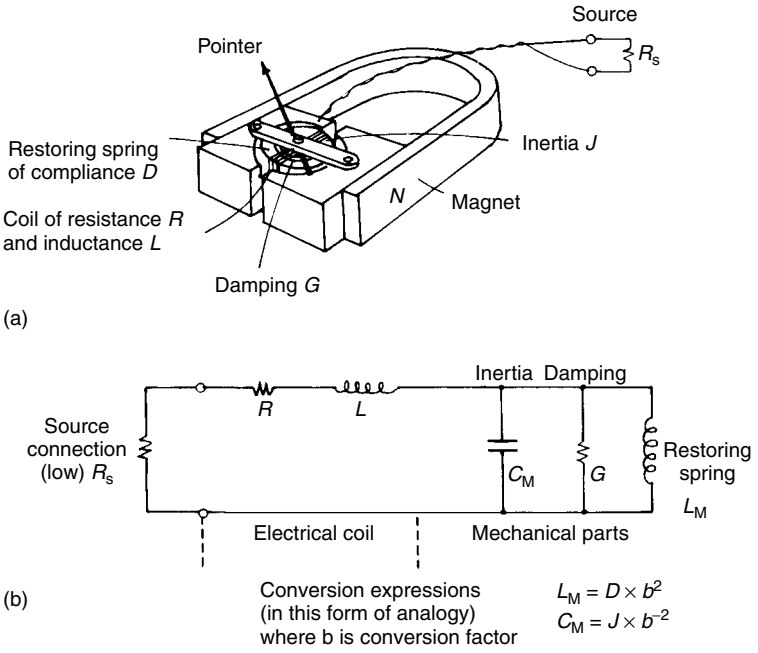


Figure 2. Unified representational model, made in the electrical regime, of electromechanical moving coil indicator; (a) physical electromechanical system; (b) electrical model.

The systems approach for physical systems has been steadily developing from roots emanating from the early radio era when microphones and loudspeakers were first being developed. Various papers were written on what were later seen to be constituent parts of this integrated topic. Eventually the basics were assembled together as a single methodology in such books as Olson (1943), where the 'analogies' were presented as a systematic approach to the design of linear dynamical systems. Initially, the approach was developed for the design of electromechanico-acoustic devices. It gradually found its way into the description of electromagnetical machines in general; see Fitzgerald and Kingsley (1961) and Koenig and Blackwell (1961). The material was later presented as a method to cover the dynamics of all physical system regimes, Shearer, Murphy and Richardson (1971) being a well-presented and extensive work; it also lists other books of relevance.

Each of those above mentioned works was in general written for the study of energy conversion devices. Instruments are a class of those specifically concerned with the conversion of information into knowledge, that is, using an energy conversion transmission path to convey meaning between points in a system. The foregoing theory of physical systems can be condensed somewhat when instruments are the prime interest.

Modern finite element CAD tools are now able to support sophisticated design development, moving from basic element representations, stress and strain calculations, and graphs, onto CNC software files for manufacture. These tools, however, still require the user to have a well-developed sense of the metrology aspect of the system being developed.

Mathematical models, *see* **Article 71, Modeling Methodology, Volume 2** are only as good as the numerical values that give the workable boundaries to the system of equations used. These are invariably obtained from experimental procedures; thus, there will always be a decision to make about whether it is more expedient to generate a mathematical model or to build the device and test it.

Models may also take the form of scaled physical versions or analog simulations in another regime. For example, an application required the development of a small linear vehicle for robotic research. Testing its linear movement performance to check out the circuitry, the braking, acceleration, and control circuits was more conveniently done using the drive system to operate, as the final stage, a rotary equivalent of the linear mass to be driven. Digital simulation is another choice for modeling the system.

A word of caution must be given, however, on the use of models. It is all too easy to become used to the idea that the model that has taken so much hard work to develop is a true and entirely faithful representation of the final instrument. Models are rarely more than just adequate because the effort needed to improve them beyond a certain point is not seen to be further productive in low production rate instruments. The user often decides to compromise model precision in order to make progress.

A model, although it may well simulate the correct input–output relationships, may not do this by the same internal mechanisms as the original system. The field is covered by the discipline of 'systems identification' and 'simulation'. This point is demonstrated by thinking about the fact that there are many technological ways to achieve an end: fuel gauges in a car can be electronic or entirely mechanical. Models must be so constructed that any internal data that are desired are available to the user of the model.

The united 'analogies' approach to instrumentation also shows that the performance of an instrument system can be tailored by the use of any appropriate discipline. In setting up an instrument design, the degree of mechanical proportion needs consideration in the light of what can be achieved using alternative electrical, optical, and fluid regimens. For example, several years ago the design of gauges for dimensional measurement generally required the designer to incorporate as much mechanical sensitivity as possible in order to obtain an adequate mechanical signal output. Today, the electronic alternative allows virtually unlimited dimensional sensitivity: fractions of nanometers can be routine (Sydenham, 1972). Mechanical stages no longer have that constraint imposed on them; the trading off of sensitivity usually allows higher frequency response.

A second example of the use of alternative regimes arose when a radar-tracking turret was assembled and found to have an unexpected resonance in a drive shaft. It was more easily overcome by adding electronic compensation circuitry to the control systems than by replacing the shaft with a stiffer design. Instrument design is no longer the matter for a single approach, nor does electronics always offer the best solution.

Design is the procedure whereby a mass of general facts are gradually sifted to sort out which combination will achieve the specification. There is seldom a unique solution and the designer is constantly required to make choices, knowing that a whole range of other possibilities are then omitted from future considerations. Finkelstein (1977) and Bosman (1978) discuss this in relation to instruments.

RELATED ARTICLES

Article 56, Transfer Characteristics of Instrument Stages, Volume 1; **Article 59, Introduction to the Dynamic Regime of Measurement Systems, Volume 1**; **Article 80, Principles of Fine Mechanics – Kinematic and Elastic Designs, Volume 2**; **Article 84, Kinematical Regime – Rotary Motion, Volume 2**.

REFERENCES

Beggs, J.S. (1955) *Mechanism*, McGraw-Hill, New York.

Bell and Howell (1974) *The Bell and Howell Pressure Transducer Handbook*, CEC/Instruments Division, Pasadena, CA.

Bosman, D. (1978) The Systematic Design of Instrumentation Systems. *Journal of Physics E: Scientific Instruments*, **11**, 97–105.

Braddick, H.J.J. (1954) *The Physics of Experimental Method*, Chapman and Hall, London.

Cambridge Instrument Company (1945) 50 Years of Scientific Instrument Manufacture, *Engineering*, **159**, 361–363, 401–403, 461–463, 501–502.

Doebelin, E.O. (1975) *Measurement Systems: Application and Design*, McGraw-Hill, New York.

Finkelstein, L. (1977) Instrument Science – Introductory Article. *Journal of Physics E: Scientific Instruments*, **10**, 566–572.

Fitzgerald, A.E. and Kingsley, G. Jr. (1961) *Electrical Machinery – The Dynamics and Statics of Electromechanical Energy Conversion*, McGraw-Hill, New York.

Foxboro (1972) *Process Control Instrumentation*, Pub.10SE, Foxoboro-Yoxall, Surrey.

Koenig, H.E. and Blackwell, W.A. (1961) *Electromechanical Systems Theory*, McGraw-Hill, New York.

Mabie, H.H. and Ocvirk, F.W. (1958) *Mechanisms and Dynamics of Machinery*, Wiley, New York.

Olson, H.F. (1943) *Dynamical Analogies*, Van Nostrand, New York.

Rothwell, E. (1966) The Characteristics of the Printed Motor. *Control*, **10**(93), 136–139.

Shearer, J.L., Murphy, A.T. and Richardson, H.H. (1971) *Introduction to Systems Dynamics*, Addison-Wesley, Reading, MA.

Sydenham, P.H. (1969) *Manual and Automatic Large-Scale Metrology*, PhD thesis, University of Warwick, UK.

Sydenham, P.H. (1972) Microdisplacement Transducers. *Journal of Physics E: Scientific Instruments*, **5**, 721–733.

West, J.C. and Leonard, J.L. (1955) The Necessary Torque Requirements for a Servo-motor. *Journal of Physics E: Scientific Instruments*, **32**, 30–32.

82: Kinematical Regime – Members and Linkages

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 General Information on Kinematic Components in Fine Mechanism 520
2 Members in Fine Mechanism 521
3 Columns and Shafts Collapse Strength 521
4 Second Moments for Common Members Forms 521
5 Use of Shafts and Pins in Fine Motions 521
6 Linkages 524
7 Design of the Ideal Linkage 524
Related Articles 525
References 525

1 GENERAL INFORMATION ON KINEMATIC COMPONENTS IN FINE MECHANISM

Mechanisms act as links, called *pairs*, that couple parts of a system together. Although not a classification generally used, it is convenient (and logical) to divide mechanisms into two groups: those based mainly on kinematic concepts and those that operate via elasticity. For our purposes, links that are strictly semikinematic because of their bearing area being greater than a point contact, including those which are manufactured to perfect enough tolerances, will be placed in the kinematic class forming the first group. Both types do use elasticity to spread the initial load but do not rely on compliance to allow subsequent deformations to take place. For example, a bolted joint deforms as it is tightened but does not flex to the extent as would an elastic cantilever beam that it could be holding. Ball bearings also deform at the points of contact, but the degree of yield is generally smaller than the component they might be supporting.

The various kinds of mechanism that might be employed include such devices as members, columns, bearings, linkage arrangements, wobble pins, gears, belts, cams, fasteners, flexures, shaft couplings, and clutches. There are very many of these, and in the past they have always been ordered into logical groups. Today, it seems from the published material, no satisfactory classification has been achieved. Authors still differ widely in the way that their works are organized.

In this account, they are ordered to reflect the design process: how the designer moves from one stage to the next, gradually completing the detail of the system being created.

Kinematic and semikinematic mechanisms are divided into members, linkage arrangements of members, fastenings that join members rigidly, bearings that are used at the joining points of linkages to allow motion, and the transference of motion across a linking point where rotation is required at a given gearing ratio. This article addresses members and linkages.

Texts that contain information on the various types of mechanisms are from the time when there was considerable interest in fine mechanism. It is apparent from inspection of the interiors of such equipments as printers and copy machines, that this design expertise still resides somewhere, but not in recent texts. The useful source books include Beggs (1955), Braddick (1954) for which a 1966 edition exists, Faires (1955), Mabie and Ocvirk (1958), which was reprinted in 1975, Peatfield (1951), Doughtie (1966), Reliance Gear (1972 and subsequent years), and Trylinski (1971). The many publications of Geary (1955, 1961, 1962, 1964) may also be of use in

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

the general selection of mechanism. Beggs' work contains a 'repertory of mechanism' that is provided as a source of ideas. Of these books, Trylinski's work is the most theoretical general work that is specifically on instrument mechanism. The date of publication of works on mechanism is not very significant, for mechanisms have been known for centuries. Little has changed in the principles used. They have just become more sophisticated in implementation. Very few new mechanisms, in fact, have been invented in recent decades. A key paper by Furse (1981) is relevant to kinematic design.

2 MEMBERS IN FINE MECHANISM

Here, the discussion centers on three-dimensional materials that are fashioned to have significant length over radius of gyration ratios. Included are columns, beams, plates, and shafts.

A slender member can be used in tension with the dominant failure mechanism being when it yields beyond its elastic limit in tension. The way it is fastened to another member is another design issue to be allowed as a connection (*see also* **Article 83, Kinematical Regime – Fasteners, Bearings, Volume 2**).

The elasticity and the strength of a slender member are calculated from the usual strength expressions relating Young's modulus of elasticity to compliance and the material parameters that define the limits of tensile load. Compliance remains the same for all conditions of a material; tempering and other treatment does not alter the elasticity, only the ultimate point at which yielding begins; where Hookean elasticity ceases.

There is an exception to this rule; that is, when the material is formed as an extremely slender fiber; the mechanical properties may change significantly. It is good to remember that the end connections may need to be of a much larger section than the member to keep bending stresses low. In instrument design, where dimensional change is important, the compliance of apparently adequately strong members may be too high and they need to be made well over the strength needed. Calculations should be made to verify that it is adequate from the viewpoint of its compliance.

3 COLUMNS AND SHAFTS COLLAPSE STRENGTH

In tensile use of a member, the longitudinal forces on it act to stretch the member; in this case the force system is stable. If, however, the force is reversed so that it now compresses the slender member, the first failure mode to appear will not be compressive yield but collapse due to the column buckling as it bends to form one of several possible modal shapes of a bent member.

This form of collapse is quantified by use of Euler's column formula (Timoshenko and MacCullough, 1959). The formula gives the value of load at which the fundamental mode, for the given end connections, begins to appear.

Euler's point of buckling depends upon the end conditions of constraint that are applied to the two ends of the beam. These may be so-called *pin-connected* (that allows rotation at the end in any direction) or be firmly clamped so that perpendicular rotation is not possible at the point of connection. The first case is also referred to as *simply supported*, the latter as *built-in* or *fixed-ends*. Another condition of Euler buckling can also occur when the end is not constrained in translation, being free to move sideways under the load.

An interesting feature about the two sets of key beam formulae (Eulers' buckling and deflection) is that although each has a different basic expression, within each group the expressions vary only by their numerical coefficient. The stiffness of a column or beam increases as the ends are made more constrained and the end point loading made more central. Figure 1 shows the various cases of collapse that can occur.

As a guide, Euler collapse begins to occur when the slenderness ratio rises to reach around 100. The value must be calculated for the smallest radius of gyration when the member has an asymmetric cross section.

4 SECOND MOMENTS FOR COMMON MEMBERS FORMS

Figure 2 provides information about the moments of typical sections of structural member as they are then needed. These formulae are needed for many mechanical member designs. They assist selection of cross-sectional shapes for suitable column and elastic members.

5 USE OF SHAFTS AND PINS IN FINE MOTIONS

Rotating shafts also have a performance failure that can be likened to the column collapse characteristic. A shaft that is free to turn in bearings that do not constrain its rotation at the bearings will reach a critical speed at which it begins to whirl – the central region attempts to move outwards owing to centrifugal forces. It takes up a bent profile while it rotates. This, too, depends upon the elasticity of the material and the slenderness ratio of the shaft. The expression giving the lowest whirling speed is given in Figure 1. Attachment of loads along the shaft alters this value (Faires, 1955; Mabie and Ocvirk, 1958).

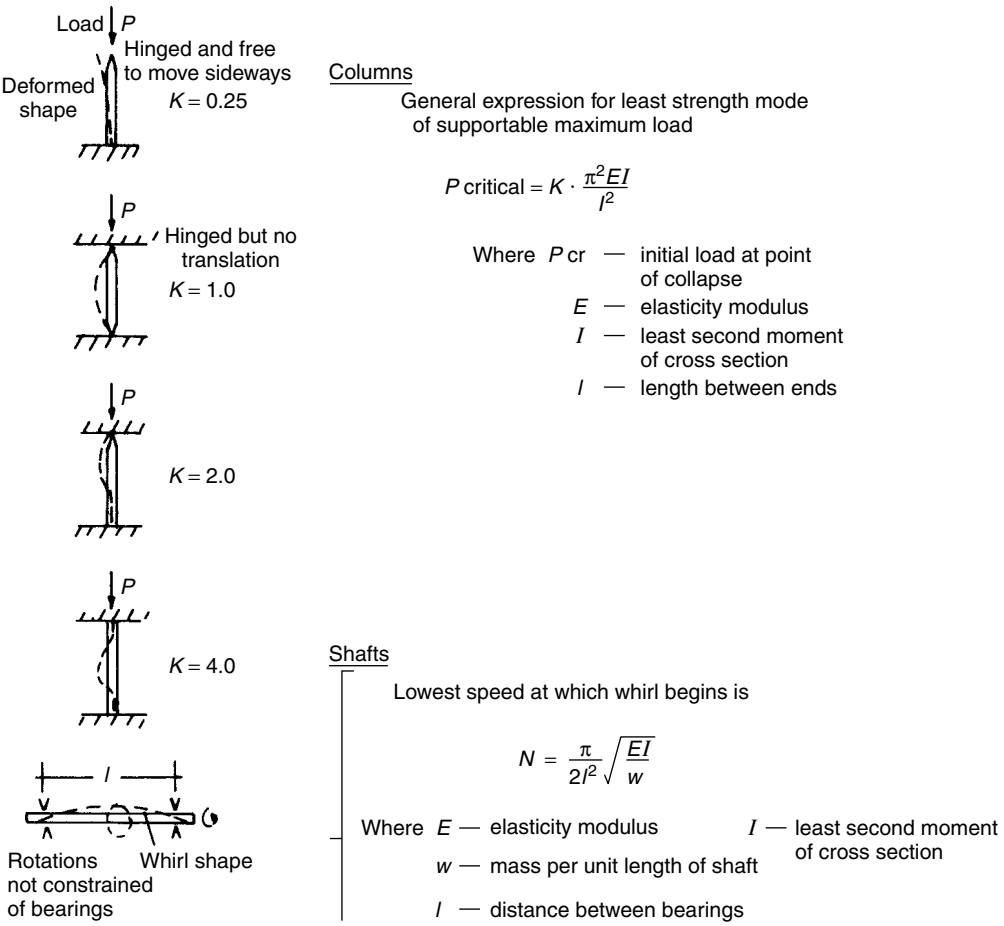


Figure 1. Columns and shafts – collapse point.

Slender members might be used as shafts coupling motion, or as parts of a linkage to produce a given motion. They can also form wobble pins that transfer linear motion without forcing any constraint on the freedom of translation that occurs at the end of the members forming either side of the drive.

Wobble pins, Figure 3, can only be used to transmit compressive force, since the compressive force keeps the point contacts mating. Ends are usually formed to be small radii hemispheres. Another use for slender members is where thin legs are used to support a tiltmeter while allowing freedom of expansion of its body. If in tension, drive can be transmitted by relatively thin wires or rods, for in that situation the Euler's collapse mode cannot occur.

Slender members (often called *beams*) used in the horizontal direction to support loads are the usual subject of texts on the strength of materials; origins are largely for civil engineering structures analysis. The various cases of point load at a given position along the beam, a uniformly distributed load or of a mixture of loads are each well developed for the various free- and clamped-end connections.

The formulation extends to cover the two-dimensional plate having freely supported and built-in edges to which is applied a central point load or a total area distributed load. These latter cases cover the needs of most diaphragms in pressure gauges. Intentionally designed deflection of shaped diaphragms is discussed in Trylinski (1971).

No difficulty should be experienced in obtaining the various formulations. In the case of very complex loading, it may be necessary to develop the expression from the basic approaches that are used to obtain the quoted equations; see Timoshenko and MacCullough (1959) for a full treatment of the equations of beams and other components. Texts containing tables giving the common expressions are Braddick (1954), Faires (1955), Neubert (1975), and Timoshenko and MacCullough (1959).

Beams that are curved and have to carry linear tensile or shear loads, such as a hook, will experience high bending stresses in the region of the bend. Areas around rivet and pin connections can be unexpectedly weak near

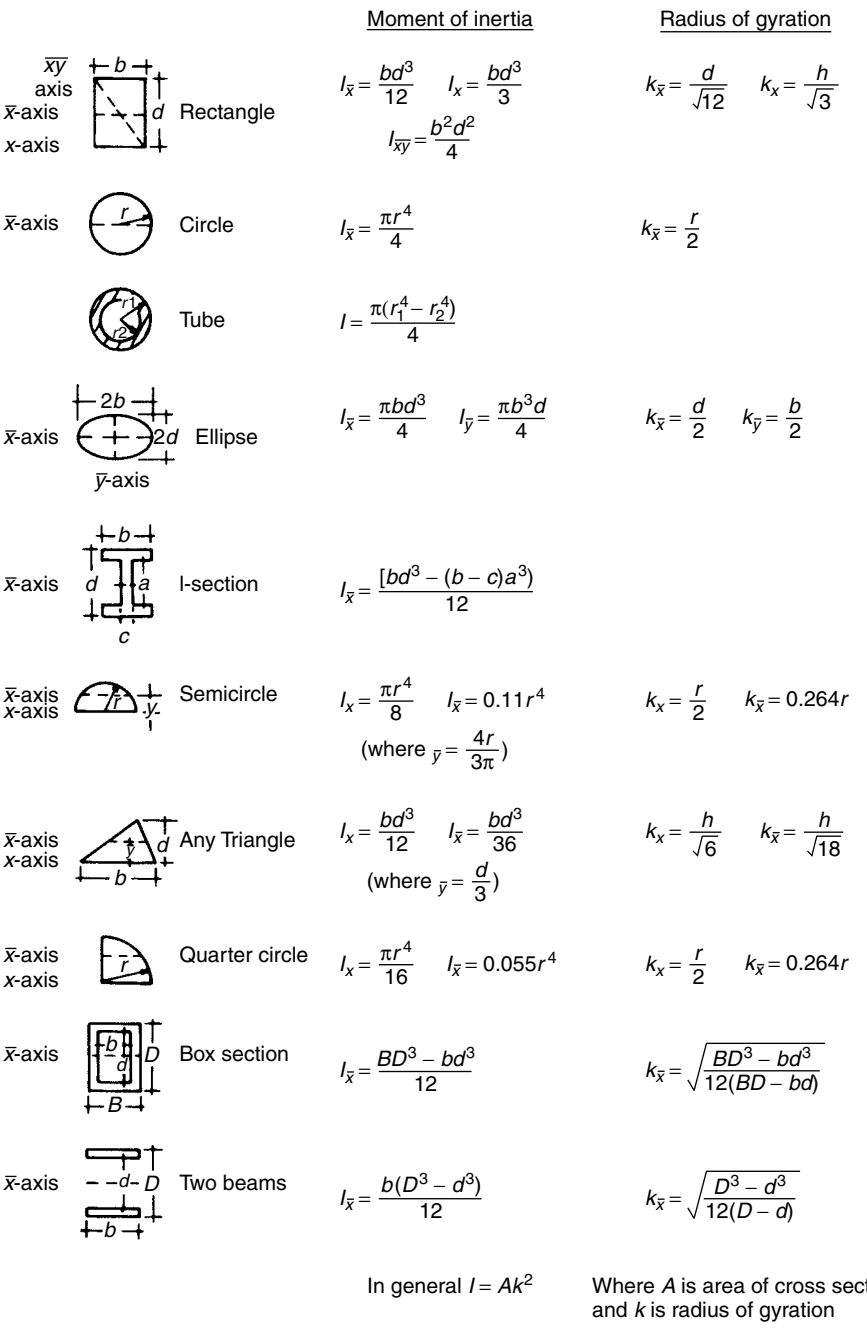


Figure 2. Second moments of commonly met sections (composites can be built up by addition of these sections).

the joint if care is not taken to increase the section size.

Beams find use in instruments as structural members and as sensing members; they may, therefore, need to be appropriately inelastic or elastic. Successful design rests on choice of material, design of the member shape and size, and the connections used.

Shafts also can be made elastic or rigid depending upon the need. They should be large in section and short in length to increase the frequency of resonance when they transfer rotary motion in closed-loop systems.

The torsional elasticity of a shaft is calculated from the torsional or shear modulus for the material, the process of calculation being analogous to that for analyzing linear elasticity of a member.

Dynamic balancing of rotating parts is needed where vibration levels must be kept low.

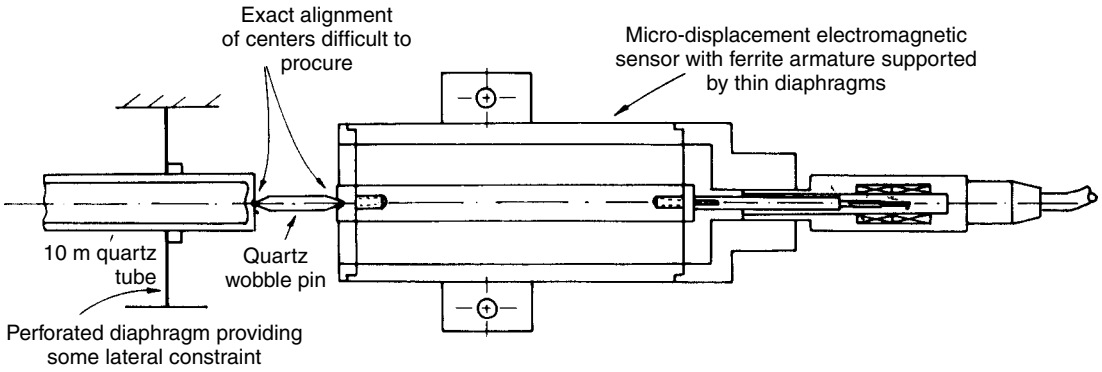


Figure 3. A wobble pin can be used to couple a moving member to a more constrained point. It will buckle if too slender for the compression force (Sydenham, 1974).

6 LINKAGES

Members connected together can be used to provide conversion of motion of various kinds – from linear to rotary, from linear to linear, and from rotary to rotary. The range of possible linkage functions is vast.

Linkages, Figure 4, may form a mechanism for geometrical size conversion such as in the pantograph, or as the method by which a balance scale can be made insensitive to the position of the load in the scale pan (seen in the so-called *static enigma* by Roberval in 1669, which appears to contravene the laws of the lever). Linkages have found extensive use in the classic typewriter, pneumatic controllers, weighing scales, in mathematical calculating instruments, in recorder pen drives, and numerous other applications.

The method of analysis of linkages can also be used to study the action of solid components, such as cams; the instantaneous contact points form geometry that can be considered to form hypothetical members. Linkages become structures under certain conditions of constraint.

7 DESIGN OF THE IDEAL LINKAGE

The mathematical analysis of linkages has fascinated academics for centuries. Many of the defined fundamental problems are still to be solved in theory. The rise and development of the kinematics of linkage is described in Ferguson (1962). To a large extent, analysis, the method of breaking down the action of a given linkage and motions of its members, has succumbed to methodology that can be used to ascertain the positional motion of linkages of reasonably simple nature. As the number of members in the linkage rises, the problem becomes more difficult. The four-bar linkage begins to present a significant theoretical challenge.

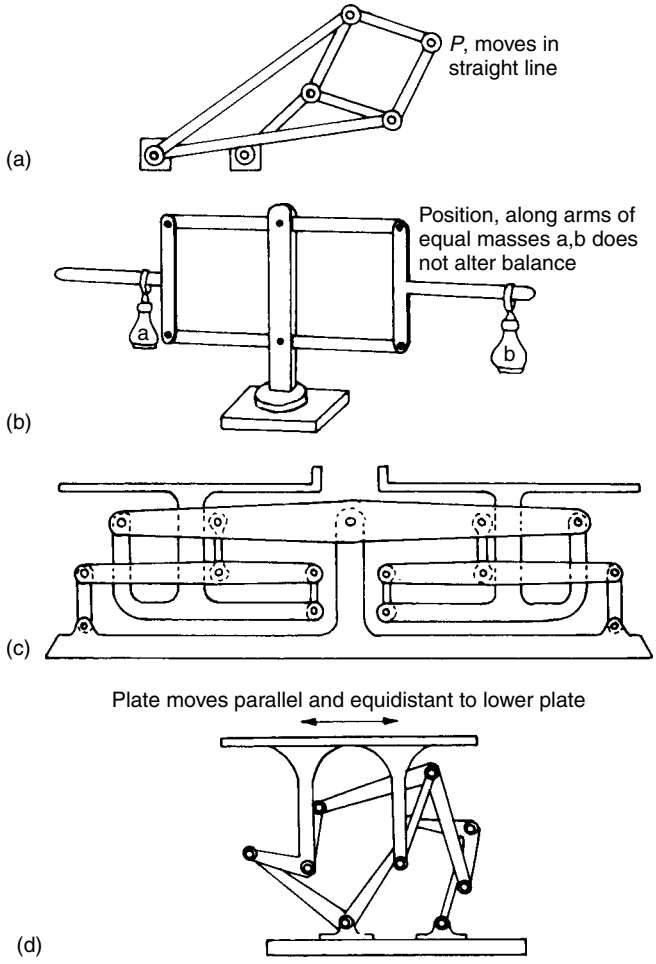


Figure 4. Some linkages used in instruments: (a) Peaucellier's straight-line generator one form; (b) Roberval's static enigma; (c) Beranger's use of Roberval's static enigma in a scale; (d) Sylvester–Kempe parallel translating linkage.

As there does not seem to be a simple general theory that can be applied to decide the motion at the various parts of a linkage system, the problem is often tackled by the use

of graphical procedures in which the motion is generated in incremental steps. To assist the designer, several authors over the centuries have published hard copy catalogs that show the motion of variously constrained linkages. Today, appropriate CAD tools can carry out this function.

What designers would rather have is the means to systematically synthesize – rather than analyze – a linkage system, in order to obtain a required motional relationship. This is, however, still in development; only relatively simple cases can be handled. A computer solution of the equations involved has extended the use of synthesis methods. Usually, the designer resorts to the use of models (in the computer or real), makes use of the above-mentioned catalogs, or iterates toward a suitable linkage arrangement using analysis of each choice.

It is sobering to realize that the theoreticians sought for many years, the theory by which a perfect linear action can be produced by some point in a system of rotating links. Unexpectedly, out of this and other similar interests has emerged useful mathematics for other disciplines (such as the Chebyshev functions now used in electronic filter design, but first compiled for steam engine linkage behavior). This linkage was finally produced by Peaucellier (Figure 4a) in the 1870s.

Many useful systems for instrument work have been invented. The 'static-enigma' balance mechanism made simple-to-use, accurate, top access scales feasible (Figure 4b, c).

The 'Sylvester–Kempe translating linkage' (Figure 4d) allows one platform to move parallel and equidistant from a second surface.

Johnson's 'converting motion' provides two revolutions from a single reciprocating translation.

Other systems related directly to linkages are the Geneva stepping mechanism, the Scotch yoke, the Oldham coupling, and many more.

Ability to analyze a linkage might be necessary to establish the clearances needed as the members move, to assess the velocities and accelerations of the various parts and points, to predict the input–output functions, and to study the action of systems like cams and gears.

The methods of analysis of linkages are discussed in Beggs (1955), Faires (1955), Mabie and Ocvirk (1958), Zimmerman (1962), and Lent (1970). Trylinski (1971) uses the procedures in his discussions of the kinematics of various mechanisms used in fine mechanics. Ferguson (1962) includes several references to linkage theory. The journal *Precision Engineering* (first issue, January 1979) is a source of papers. Burstall (1970) discusses the history of linkage development giving several references to early works. This book contains illustrations of linkages that were invented during the history of man's involvement with technology. Pollard (1929), on the design of couplings, also might help understanding of the procedures.

RELATED ARTICLES

Article 56, Transfer Characteristics of Instrument Stages, Volume 1; **Article 57, Static Considerations of General Instrumentation, Volume 1**; **Article 70, Measuring System Design Methodologies, Volume 2**; **Article 80, Principles of Fine Mechanics – Kinematic and Elastic Designs, Volume 2**; **Article 81, Principles of Fine Mechanics – Systems Considerations, Volume 2**; **Article 88, Error Sources in Fine Mechanics, Volume 2**.

REFERENCES

Beggs, J.S. (1955) *Mechanism*, McGraw-Hill, New York.

Braddick, H.J.J. (1954) *Mechanical Design of Laboratory Apparatus*, Chapman & Hall, London (1st edn, 1954).

Burstall, A.F. (1970) *A History of Mechanical Engineering*, Faber and Faber, London.

Doughtie, V.L. (1966) *Elements of Mechanism*, Wiley, London.

Faires, V.M. (1955) *Design of Machine Elements*, Macmillan Publishing, New York (1st edn, 1934).

Ferguson, E.S. (1962) Kinematics of Mechanisms from the Time of Watt, *United States National Bulletin 228*, Paper. 27, Smithsonian Institution, Washington, DC, (pp. 185–230).

Furse, J.E. (1981) Kinematic Design of Fine Mechanisms in Instruments. *Journal of Physics E: Scientific Instruments*, **14**, 264–272.

Geary, P.J. (1955) *Knife Edge Bearings*, Sira Institute, Chislehurst.

Geary, P.J. (1961) *Instrument Ball Bearings*, Sira Institute, Chislehurst.

Geary, P.J. (1962) *Fluid Film Bearings*, Sira Institute, Chislehurst.

Geary, P.J. (1964) *Magnetic and Electric Suspensions*, Sira Institute, Chislehurst.

Lent, D. (1970) *Analysis and Design of Mechanisms*, Prentice Hall, Englewood Cliffs, CO.

Mabie, H.H. and Ocvirk, F.W. (1958) *Mechanisms and Dynamics of Machinery*, Wiley, New York (3rd edn, 1975).

Neubert, H.K.P. (1975) *Instrument Transducers – An Introduction to their Performance and Design*, Clarendon Press, Oxford.

Peatfield, A.E. (1951) *Engineering Components and Materials*, EUP, London.

Pollard, A.F.C. (1929) *The Kinematical Design of Couplings in Instrument Mechanisms*, Adam Hilger, London.

Reliance Gear. (1972) *Catalogue RG5*, Reliance Gear Co Ltd, Huddersfield. (Also similar catalogues by PIC Design Corp, New York).

Sydenham, P.H. (1974) 2000 hr Comparison of 10 m Quartz-tube and Quartz-catenary Tidal Strainmeters. *Geophysical Journal of the Royal Astronomical Society*, **38**, 377–387.

Timoshenko, S. and MacCullough, G.H. (1959) *Elements of Strength of Materials*, Van Nostrand, NJ. (Numerous editions and printings exist since the 1st in 1935).

Trylinski, W. (1971) *Fine Mechanisms and Precision Instruments – Principles of Design*, Pergamon Press, Oxford. Translated by A. Voellnagel from Polish original.

Zimmerman, J.R. (1962) *Elementary Kinematics of Mechanism*, Wiley, London.

83: Kinematical Regime – Fasteners, Bearings

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Fastenings in Fine Mechanism 527
2 Permanent Joints 527
3 Detachable Joints 528
4 Instrument Bearings and Pivots 529
5 Reduced Friction Pivots 530
6 Surveys of Instrument Bearing Information 531
Related Articles 532
References 532

1 FASTENINGS IN FINE MECHANISM

The purpose of fastenings is to join components together in a manner that will provide a firm constraint of all degrees of freedom at the joint. Available designs vary greatly, the choice largely depending upon the magnitude of the forces and torques involved, whether the joint is permanent or needs to be detachable, the degree of vibration experienced, and the level of dimensional accuracy needed. The design of joints is as important as the other factors of a fine mechanical system.

An introductory, nonmathematical catalog of fastenings can be found in Peatfield (1951). A thorough treatment of many of the possible types is given in Trylinski (1971), this being a key work of reference. Machine-design texts such as Faires (1955), Lent (1970), Shigley (1988), Erdman, Sandor and Kota (2001), and Juvinall and Marshek (2002) generally treat only the joints used in power machines, that is, bolted, riveted, and welded arrangements.

In some applications, the dimensional stability of the joint can be very important. One view suggests that in such cases, it is better to avoid using joints, making the whole unit from a single piece of material. Examples are the manufacture of flexure pivots in which the flexure part is formed by thinning down the piece of material. There is, however, very little substantive evidence to prove that this is necessary. To the contrary, it has been found that remarkably crude joints, such as knotted metal wires, rings placed over a shaft and in tension, small magnets holding a part to a surface, and bolted joints (Jeffery and Sydenham, 1973), will each hold dimensional relationships across the joint to less than micrometer level. Couplings working at the very small tolerances of nanometers or less may well exhibit instability, but it is far from easy to verify where such a drift actually occurs. Errors of this magnitude can come from so many sources. As an example of the many forms of spring-gravimeter that have been made, units using bolted metal arrangements have shown as small a drift rate as others with mostly continuous parts.

Joints forming part of a mechanical resonance loop can, however, seriously alter the Q-factor because of the damping inherent in the jointed faces. Joints should be avoided where high-quality factors are needed.

In instruments, the joints must provide adequate strength but also they must usually be stiff in compliance. Because of the small sizes of the metal work in instruments, many ingenious joints can be used that would not be practical for large-size machine construction. Joints divide into permanent and detachable methods.

2 PERMANENT JOINTS

Figure 1 illustrates examples of some methods of making a permanent joint. A commonly used permanent joint is formed by passing the metal of one part through a hole

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

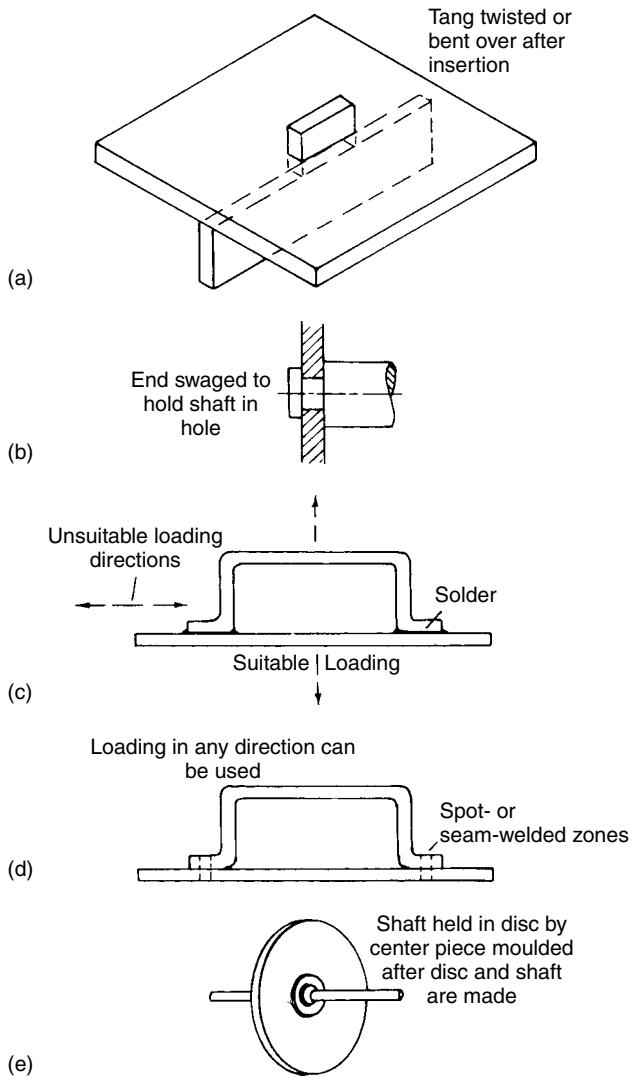


Figure 1. Some permanent fastening methods; examples of each type: (a) bent tangs; (b) swaged after insertion; (c) soldered or glued; (d) resistance welded; (e) molded in place.

in the other and by swaging or twisting it to lock the two parts together. The tang can be twisted, bent over, swaged flat, center punched, or crimped. Trylinski (1971) gives illustrations of many types.

Force-fit joints that rely on friction and a certain degree of cold welding as the parts are forced into unison can also be used. For these, the tolerances of the mating parts must be carefully controlled. These give good dimensional control.

Welded joints are often used, the smallest possible area of welding being at micrometer size. Fine pressure capsules are often welded using focused laser radiation. Gas or electric welding is also used. Distortion due to the heating must be allowed for. Not all materials can be welded.

Adhesives can be used to form permanent joints, epoxy formulations having considerably changed the role of adhesives in instrument construction. Work at the former Sira Institute in England has resulted in several publications on the use of adhesives, Shields (1976) being a substantial account. The slow setting rate of cements and adhesives can make them unsuitable for mass production methods. In some cases, using cements, those substances that fill a gap without stressing the parts that are held together, may be the most suitable method of securing components, for example, holding a lens in a holder.

When using glued or cemented joints, it is necessary to ascertain whether the joint will have adequate heat transfer or electrical conduction across the joining layer that is interposed between the parts. Cements and adhesives are often hygroscopic; they may also creep in dimension. In some instances, however, gluing may be the only satisfactory method of making a fastening. A good design attempts to form a fillet of glue over a large contact area. The glue should not be used to form a dimensional link but should be used to adhere two creep-free materials in close contact. Trylinski (1971) contains a large bibliography on the subject.

3 DETACHABLE JOINTS

Figure 2 shows some examples of detachable methods of jointing. These can be made using screw threads, which hold flanges together or which are formed on the end of rods to be jointed; using pins holding a shaft in a hold (tapered, parallel, roll-pin, shear-pin); using keys that hold a shaft in a hole; and as press-fit combinations. Where the jointing forces are low, such as the holding of a dust cap over a socket, the use of the lids and caps held by simple turn-lock devices can be used. Trylinski also covers these in detail.

When bolts are used to hold a flange, the designer has the choice of either making the bolt hole with close-fit tolerance to ensure the correct dimensional relationship or using some other method of location, using the bolt only as a clamping device. Locating methods in use are spigots or taper and parallel cylindrical pins. It is not good practice to use a large clearance hole in which the actual position is set up on assembly. This may suffice for a one-off prototype, but in conditions of vibration or where service personnel may not understand what is needed, the position can be lost. Where positional accuracy is vital, the joint can be drilled and pinned after assembly. A spigot is used where the location is predetermined. Where the bolt is used to retain position, its shaft should define the location, not the thread diameter. To ensure that this is done, the thread is made with a smaller outer diameter than the shaft next to the head of the bolt.

Screw threads are used extensively and numerous forms of thread profile and diameter are found in applications. The

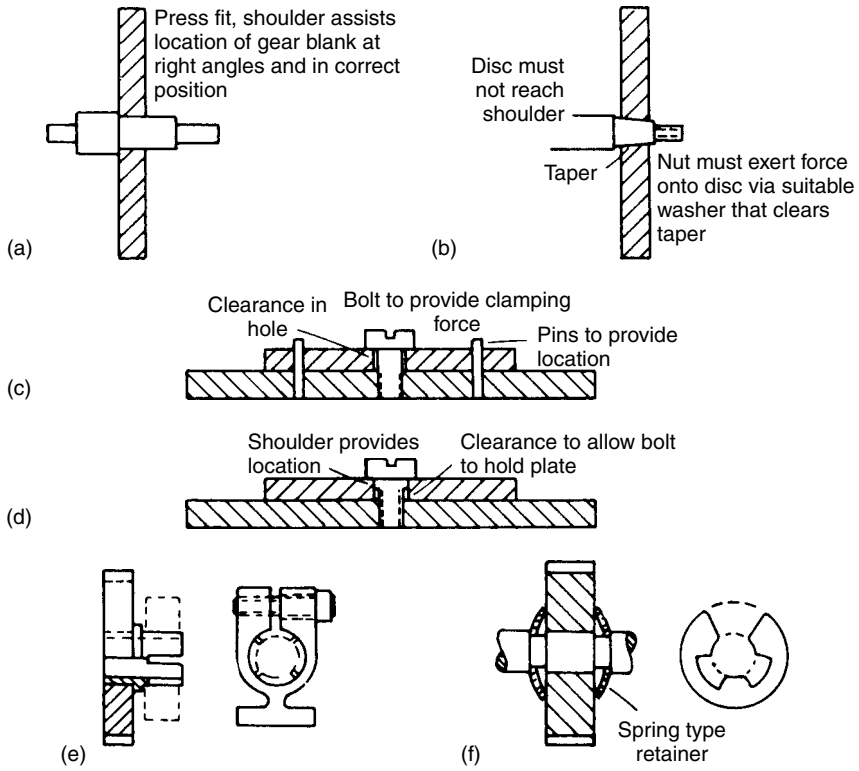


Figure 2. Some detachable fastening methods – examples of each type: (a) press fit; (b) taper shaft; (c) pinned; (d) screw thread; (e) clamp-hub; (f) retainer ring.

designer of a production instrument should strive to use a standard thread where possible. Different disciplines have their own threads – cameras and microscopes being examples. Multiple start, special thread profile, taper threads, and others are sometimes used in instruments. Threads can be locked with some form of lock or tab washer or can be pinned. Where the joint must be secured against unauthorized opening, the shaft outside of the nut is drilled to take a small wire that can be sealed. Special proprietary fluids are available to lock threads; paint is sometimes used. Here again, Trylinski (1971) is a definitive reference work.

4 INSTRUMENT BEARINGS AND PIVOTS

Joints between two parts that allow motion, such as rotation or sliding in translation, are generally called *bearings*. Trylinski calls them *supports*. They are also referred to as *pivots*.

The relative motion may combine both rotation and translation through the bearing axis. They may be included either to limit end-float in a thrust bearing or to allow an arm to turn about a fixed shaft pin. Figure 3 shows a collection of bearings that can be used in instruments.

Selection of bearings begins by a definition of the role that the material support has to play. All bearings are imperfect and will add some measure of freedom at the pivot-bearing point. System sketches realize the degrees of freedom that are needed; study of the system's operation decides the tolerances that are permissible. Other factors to be considered will be the relative speed of the two surfaces involved, the kind of lubrication that can be used, the coefficient of friction, the life needed, assembly requirements, cost and market availability, effect of temperature changes, and the degree of stiffness needed from the joint formed in the bearing. It is quite possible to produce a usable bearing (using preloaded rolling balls) that is as stiff and as backlash-free across the joint as is the rest of the structure.

Supports that allow movement by virtue of elasticity of a flexure are discussed in **Article 85, Elastic Regime of Design – Design Principles, Volume 2**; **Article 86, Elastic Regime of Design – Spring Systems, Volume 2**; and **Article 87, Elastic Regime of Design – Plates and Bimorphs, Volume 2**.

Bearings for rotating shafts have been in use since the earliest times. Rawhide was used by the Egyptians around 1500 B.C. Bronze bushes were used in wagons of Roman times. Crude ball bearings date to the same period. Caligula

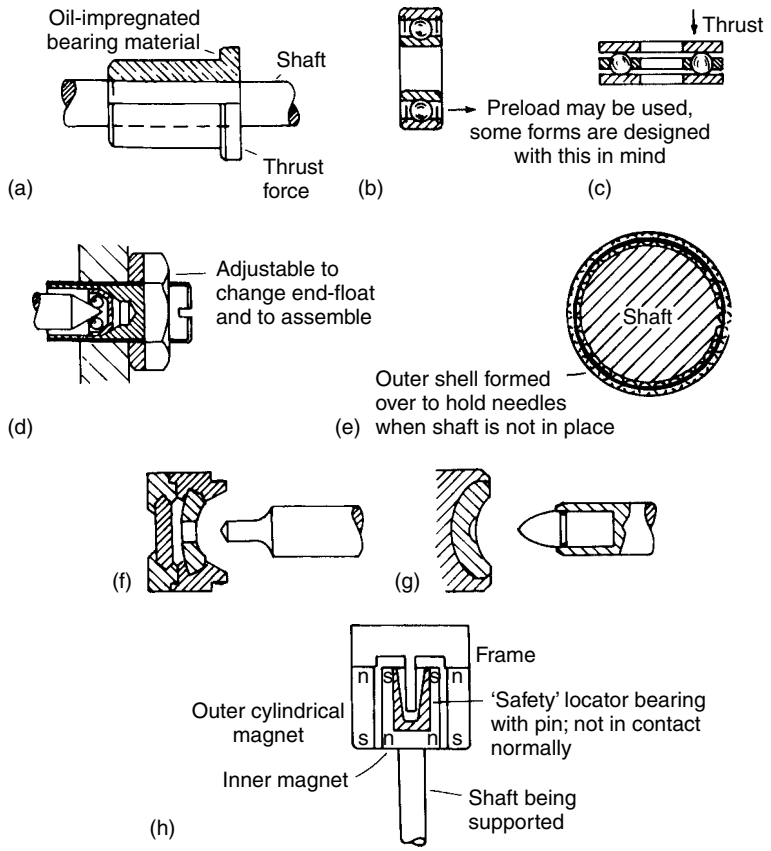


Figure 3. Selection of passive bearings used in instruments: (a) plain bush; (b) radial ball race; (c) ball thrust race; (d) pivot type ball; (e) needle roller; (f) jewel with thrust; (g) center support; (h) magnetic.

had a ship with turntables running on large bronze balls. These early ball systems were not strictly freerunning, for they were constrained so that the balls also slid while turning; Burstall (1970) gives more detail.

Around 1770, Count Carburi used a system of metal tracks and balls to move a 1.5×10^6 kg granite mass across the frozen Russian land (Turner, 1973). The tracks were laid end to end as the statue was moved.

5 REDUCED FRICTION PIVOTS

Friction was recognized as the limiting phenomenon in many an elegant mechanical apparatus of the early times. In those times, ball races, as we know them today, could not be made to the degree of accuracy and quality control that was needed. Many methods were used to minimize the effects of friction in bearings. A particularly interesting one is the so-called *frictionless bearing* arrangement attributed to Sully (1680–1728). This used a system of wheels and shafts running on the periphery of overlapping support wheels held in the prime bearings (Figure 4) effectively to increase the torque that is placed onto the bearing areas. The constraining moments of the bearing have less effect on the output in this arrangement. A number of well-known instruments of that time used this bearing – Atwood's falling machine, Michell's and Nairne's dipneedles being examples (Turner, 1973; Sydenham, 1979).

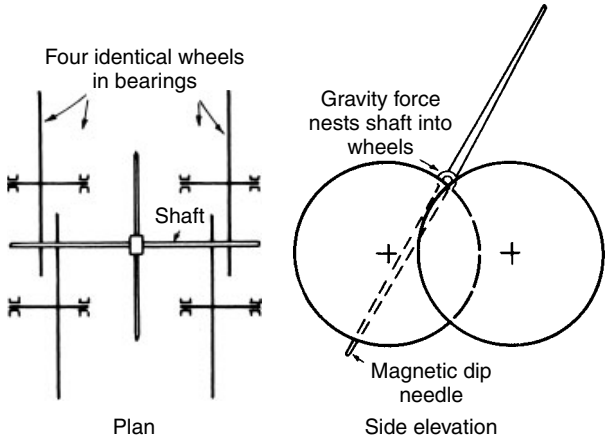


Figure 4. The Sully (*c* 1700) frictionless bearing arrangement as used by Michell in a late eighteenth century dip circle instrument.

6 SURVEYS OF INSTRUMENT BEARING INFORMATION

The design of bearings is a very well developed aspect of instrument construction. Considerable theory now exists to complement the highly developed fine mechanical production capability. Low-friction precision bearings can be as small as a millimeter in outside diameter. The designer who wishes to be sure that his or her choice is justified and is within the requirements of the task has a wealth of theory and experimental data to fall back on.

Manufacturers provide considerable design material to assist application. Permissible load, run-out, clearances, speed, friction force, life, and assembly arrangements are all well documented.

Since the various forms of bearings are generally familiar, this account will only introduce each kind, giving references to where in-depth information can be found. An excellent work on each type is Trylinski (1971). It provides information about the small and miniature-sized bearings that is not generally given in machine-design texts. It covers jewel bearings; direct in the base bearings used in clock mechanisms; conical and tapered supports, which use a tapered shaft end running in a hole in a suitable bush; spherical ends; hydrostatic designs; center supports in which a spherical shaft end runs against a concave plate; cover-plate type thrust bearings, including the shock-resistant Incabloc arrangement; mercury supports; knife-edges; rolling-contact bearings, such as the very easy to use cup-and-cone miniature and subminiature ball bearings; and wire track bearings in which the rolling contact ball is used to form its mating track by being pressed into soft metal wires. The chapter in his text on 'supports' covers friction and bearing area stress calculation, errors in location, and many other theoretically based explanations of behavior. Eighty-two references are included. A further chapter on guides covers the design of sliding arrangements, which are usually of rectilinear form.

Further in-depth works on instrument bearings include the series compiled by Geary of the Sira Institute (then called the British Scientific Instrument Research Association, BSIRA). His second booklet (Geary, 1955) was a bibliographic survey of knife-edge bearings. It gives an illustrated technical introduction, followed by 166 references. Geary concluded then that knife-edges should be considered as blunt or rounded edges, for, in practice, deformation occurs, resulting in a less than ideal knife-edge. He also wrote that there was a general dearth of research and objective data available, especially in the English language, and that the best-written works were of German origin. He suggested that some of these works should be translated. (Trylinski's work was later translated, helping to fill this gap in available knowledge.)

Geary's fourth report of the series on instrument parts (Geary, 1961) covered instrument ball bearings in a similar way, providing a specific account on the small series of precision-quality ball bearings. It is a key document where information on the use and choice of small ball bearings is required. Harris (1966) is also relevant.

Applications requiring zero clearance and a large measure of stiffness across the ball-bearing joint can use specially designed ball races – the angular-contact series. In use, the shaft and housing dimensions are chosen to force the balls to ride against one side of the race. This is known as *preload*. Ball nuts, those drive nuts in which balls run in contact with the lead screw thread, are often used together in pairs, one being rotated against the other to preload the nut into the shaft. Ground leadscrew shafts can provide remarkably stiff, yet resistance-free, joints.

For nonmagnetic applications, beryllium–copper ball races can be used. Stainless-steel bearings are also sold. Bearings capable of 300 000 rev min^{-1} are used in dentist's drills.

Each application dictates careful choice from within the various classes of bearings if best operation is to be achieved. The fifth report (Geary, 1962) covered fluid-film bearings. This provides a general introduction to this bearing type and gives details of liquid-lubricated bearings using external pressure generation, of gas-lubricated bearings, and of less well known bearings such as fluid-squeeze films, mercury globule, and acoustic radiation pressure supports.

The frictional force of the different kinds of bearings varies considerably, ball-bearing systems exhibiting considerably less frictional torque levels than simple liquid-fluid bushes. As an indication of the magnitudes involved, oil-filled gyroscope ball bearings having five balls running at 4-mm diameter might exhibit a breakaway torque of 8×10^{-6} Nm. Running torque is generally about 30 to 50% less. Very small, three-ball pivot bearings breakaway at 1.0×10^{-7} Nm torque. Torque figures increase with load on the pivot. In contrast, a small plain bronze bush needs about 6 to 8×10^{-4} Nm to move from rest. A small radial ball race will need around 3×10^{-5} Nm torque before it begins to rotate. These figures can only be used as a guide – the actual value varies greatly with bearing form, preload, shear load, and shaft size. They must be calculated or measured for the support to be actually used. Misalignment can introduce considerable frictional loss, unless designed as a self-aligning arrangement. Suspensions for electrical indicating instruments use knife-edges, jewels, bushes, and elastic bands. Drysdale and Jolley (1924), Part 1, include an extensive discussion of these, most of the material given being about jewel bearings. Geary (1964) covers magnetic and electric suspensions. Permanent magnetic suspensions

for instrumental use were surveyed by Fremery and Doden (1978).

RELATED ARTICLES

Article 57, Static Considerations of General Instrumentation, Volume 1; **Article 58, Description of Accuracy, Linearity, and Drift, Volume 1**; **Article 59, Introduction to the Dynamic Regime of Measurement Systems, Volume 1**; **Article 80, Principles of Fine Mechanics – Kinematic and Elastic Designs, Volume 2**; **Article 81, Principles of Fine Mechanics – Systems Considerations, Volume 2**; **Article 85, Elastic Regime of Design – Design Principles, Volume 2**.

REFERENCES

Burstall, A.F. (1970) *A History of Mechanical Engineering*, Faber and Faber, London.

Drysdale, C.V. and Jolley, A.C. (1924) *Electrical Measuring Instruments – Part 1 Commercial and Indicating Instruments*, Ernest Benn, London.

Erdman, A.G., Sandor, G.N. and Kota, S. (2001) *Mechanism Design: Analysis and Synthesis*, Prentice Hall.

Faires, V.M. (1955) *Design of Machine Elements*, Macmillan Co, New York (1st edn, 1934).

Fremery, J.K. and Doden, K. (1978) Active Permanent Magnetic Suspensions for scientific Instruments. *Journal of Physics E: Scientific Instruments*, **11**(2), 106–113.

Geary, P.J. (1955) *Knife Edge Bearings*, Sira Institute, Chislehurst.

Geary, P.J. (1961) *Instrument Ball Bearings*, Sira Institute, Chislehurst.

Geary, P.J. (1962) *Fluid Film Bearings*, Sira Institute, Chislehurst.

Geary, P.J. (1964) *Magnetic and Electric Suspensions*, Sira Institute, Chislehurst.

Harris, T.A. (1966) *Roller Bearings Analysis*, Wiley, New York.

Jeffery, G.J. and Sydenham, P.H. (1973) Stability of Strainmeter Mounts. *Geophysical Journal of the Royal Astronomical Society*, **32**(2), 185–193.

Juvinall, R.C. and Marshek, K.M. (2002) *Fundamentals of Machine Component Design*, Wiley, New York.

Lent, D. (1970) *Analysis and Design of Mechanisms*, Prentice Hall, Englewood Cliffs, CO.

Peatfield, A.E. (1951) *Engineering Components and Materials*, EUP, London.

Shields, J. (1976) *Adhesive Handbook*, Newnes-Butterworth, London.

Shigley, J.E. (1988) *Mechanical Engineering Design (McGraw-Hill Series in Mechanical Engineering)*, McGraw-Hill.

Sydenham, P.H. (1979) *Measuring Instruments – Tools of Knowledge and Control*, Peter Peregrinus, London.

Trylinski, W. (1971) *Fine Mechanisms and Precision Instruments – Principles of Design*, Pergamon Press, Oxford. Translated by A. Voellnagel from Polish original.

Turner, G.L.E. (1973) *Van Marum's Scientific Instruments in Teyler's Museum – Descriptive Catalogue*, Noordhoff International Publishing, Leyden.

84: Kinematical Regime – Rotary Motion

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Transference of Rotary Motion in Fine Mechanism 533
2 Couplings 533
3 Gears 534
4 Cams 536
5 Flexible Connecting Drives 537
6 Belt Drives 537
7 Chain Drives 537
8 Electromechanical Gearing 538
Related Articles 539
References 539

1 TRANSFERENCE OF ROTARY MOTION IN FINE MECHANISM

As in electrical systems, mechanical information machine counterparts require means to transfer signals between subunits of the system. These signals may have to be transmitted at the direct unity ratio or be operated upon to alter this ratio up or down. The motion may need to be converted into such motions as intermittent, quick return, and others.

Transferring signals and forces over large distances in mechanical equipment is more difficult than with electrical signals. Generally, in practice, we convert mechanical information into the electronic form for long-distance transmission. This was not always so, and there are still some applications where direct mechanical transfer is more secure, more reliable, and more easily serviced.

This article discusses those components that, while not generally being thought of as computing elements, nevertheless, might be used in that manner and should be regarded in that light. It is important to see coupling systems in this way, for there are many traps to fall into because of imperfections in components and mechanism concepts. Covered here are the shaft coupling devices, gears of various kinds, cams, flexible belt-type drives, and electromechanical gearboxes. The general principles of mechanical design are to be found in **Article 80, Principles of Fine Mechanics – Kinematic and Elastic Designs, Volume 2**. The use of kinematic design methods for members and linkages is covered in **Article 82, Kinematical Regime – Members and Linkages, Volume 2** and for fasteners and bearings in **Article 83, Kinematical Regime – Fasteners, Bearings, Volume 2**. The design method using the elasticity of materials to obtain controlled motion is covered in **Article 85, Elastic Regime of Design – Design Principles, Volume 2** and **Article 86, Elastic Regime of Design – Spring Systems, Volume 2**. Errors of all kinds is the subject of **Article 88, Error Sources in Fine Mechanics, Volume 2**.

The form and design of these components are well covered in the general machine design texts such as: Peatfield (1951), Faires (1955), Beggs (1955), Mabie and Ocvirk (1958) and Lent (1970). Information specifically relating to instruments is given in Trylinski (1971) and Reliance Gear (1972). Because of the extensive use of gears and other couplings in industry, there are many national standard specifications.

2 COUPLINGS

When coupling two shafts together as a direct drive, the factors to be considered are the amount of misalignment that must be tolerated, the constancy of the velocity and

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

positional form transmitted within a revolution, the torque level, the distance between supports to reduce shaft whirl, damping of vibration, stiffness needed in closed-loop control, allowance for end-float and more severe longitudinal movement, ease of assembly, and such factors as availability and cost. A sample of the range is shown in Figure 1.

Constant velocity couplings began with the design of Hooke reported in 1674. The single Hooke joint, however, does not give a constant velocity (or position) ratio between the input and the output as the coupling rotates through a turn. However, two units can be used together in series so that the modulation imposed by one is canceled by the other. Other more compact forms of constant velocity joints were later developed for the automotive industry, examples being the Bendix–Weiss, Rzeppa, and Tracta joints. These, in general, have not been produced at instrument size. They all have the advantage of providing constant velocity from a single coupling joint. The above coupling stages require the two coupled shafts to be positioned such that their extended center lines intersect in the coupling center. A double Hooke joint can, however, also take up translational as well as angular misalignments.

Where two shafts are parallel but not longitudinally in line, the Oldham coupling can be used. This allows for misalignment by a sliding, cross-shaped connector.

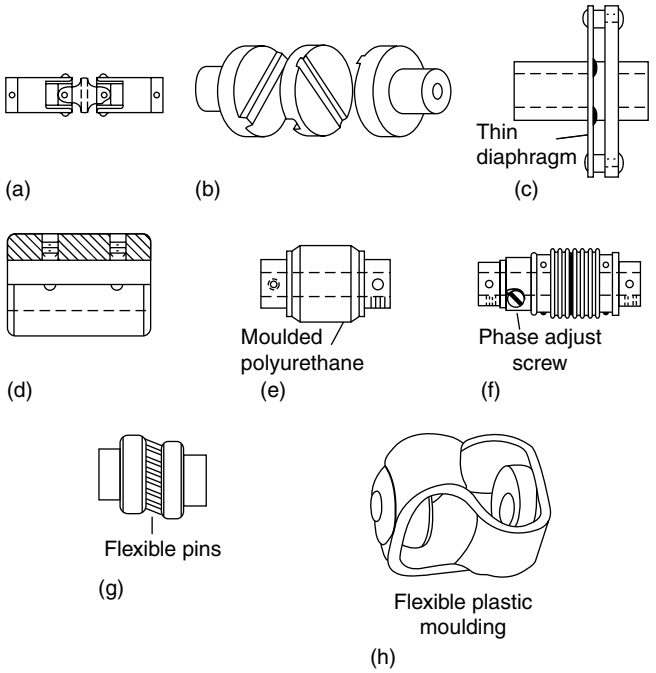


Figure 1. Selection of couplings used in instruments: (a) double Hooke joint; (b) Oldham; (c) membrane; (d) rigid; (e) molded flexible; (f) bellows with phase adjust; (g) Flex-thane; (h) Kudriavetz.

A solid or rigid coupling can only be used where the two shafts are aligned to exacting tolerances. If they are not, the coupling will 'work' on its screws or bend the shafts, causing looseness or failure to occur.

Coupling drives come in many forms; see Reliance Gear (1972) and Trylinski (1971) for illustrations of the various kinds of instruments. The latter definitive work presents the geometrical theory for many designs. The following names are used in the literature for various kinds of couplings; slip joint and slip couplings, membrane, spring and self-locking spring, antibacklash, bellows, molded plastic flexible, universal lateral, Kudriavetz, Flex-thane multiple polyurethane pins, multijaw, sleeve, slip clutch, overrunning clutch, and phase adjust (a type where the relative shaft phase can be slightly adjusted by a small screw rotation). There is no shortage of design possibilities; each has advantages and disadvantages and must be matched to the application. It is important to appreciate the motional and stress characteristics of each in any given application.

Where the motion deliberately is not to provide constant velocity or positional adherence to the input form, intermittent drives may be used. These include the quick return, Geneva, intermittent gear drives, and the various forms of clockwork escapement mechanisms. Many couplings make use of elasticity to allow for misalignment. In these instances, it is unlikely that they will be of use if information fidelity or smooth unmodulated drive is needed through the link. Today, industrial drives tend to use electronic control to implement such motions but there is still place for mechanical solutions due to their lower cost for low energy coupling situations.

3 GEARS

Although the toothed gear was in use in ancient times, it was not until more recent times that it could be used extensively because of the relatively low cost resulting from mass production and good quality control of the form and size of the two mating parts involved.

Gears are used to alter (or maintain) the ratio of speeds between two turning or partly rotating shafts. Many forms enable the directions of the two shafts to be placed at any required angle and displacement. The main limitation of gears is that the shafts involved must be reasonably close together; this can be a serious shortcoming. Idler gears and unity ratio stages can be used to couple stages that are not close together. A combination of gears is called a *gear train*; it is housed in a gearbox.

The simplest form of gearing uses two smooth disks running in contact driving by friction force of the contact. Friction gears are necessary in many applications, for example, in converting the linear motion of running strip into

rotation to drive an encoder, as parts of various mechanical computing devices, to subdivide a circular table, to drive paper through a chart recorder, or to drive a record turntable or tape recorder capstan – see also Sydenham (1968), and Jain and Sydenham (1980). Trylinski (1971) includes a chapter on friction gearing. Experience of Sydenham and Brandwijk (1966) indicates that friction gearing (Figure 2) can hold positional division of a 16:1 ratio step-down to the order of one part in 20 000. The variation in ratio is apparently random in nature, and absolutely fixed ratio gearing and, thus, synchronism cannot be maintained over a large number of revolutions. Best results are produced when the drive wheel has a rounded profile running on a flat surface, when the pressure is as light as possible to provide enough driving torque, and when the two materials are very hard, such as steel. High friction coefficient drive surfaces, like rubber, give good drive adherence but poor constancy of gear ratio.

Where absolute synchronism or high torques are needed, the use of toothed gears becomes necessary. Spur gears are made in numerous forms and are generally familiar.

In instrument work, the design of instrument gears closely follows general gear practice – with a few exceptions. One is the need to obtain precision operation. Gear trains in instruments and precision, small-size, closed-loop servo units will possess a certain degree of clearance between the teeth. This small level of backlash can introduce an error of measurement or instability of a servo loop. It can, however, be reduced by the use of antibacklash gears. These use two gears running adjacently that are spring loaded to cause the two halves to clamp around the mating teeth of the other gear. Whereas antibacklash gears do eliminate backlash, this form of device can introduce a degree of elasticity through the train, possibly resulting in a more compliant servo-loop response. The backlash spring must hold the gears in contact against the friction of the train. It is often better to use higher grade gears for which backlash can be reduced by closer tolerances of manufacture.

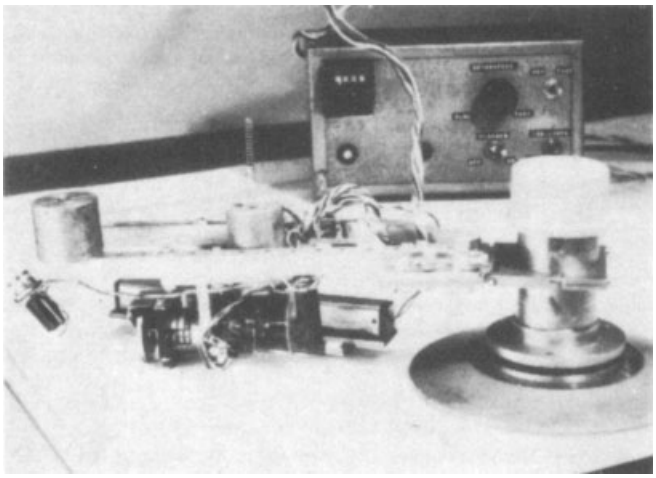


Figure 2. Friction gearing used to subdivide a circle in the automatic manufacture of radial metrology gratings.

The inherent inertia of rotating parts may be significant; it must be considered when designing gear trains for dynamic use. Inertia rises as the fourth power of the gear radius and each stage adds inertia as the reflected inertia seen looking into it, multiplied by the gearing ratio. Multiple stage step-up gears soon accumulate significant inertia. The inverse obviously applies; step-down gear boxes see little inertia. High-ratio step-down boxes, however, will have increased backlash and the wind-up of the train becomes more significant as the elasticity of the gears, shafts, and bearings combine to form a more compliant through-link. As a general rule, the best response servos and the lightest action manual instruments result from the use of a low gear ratio and low tolerance gear wheels.

The choice of more stages of lower ratio to obtain a given overall ratio may be better than using a single large ratio step. Friction of step-down trains is generally of little consequence, but step-up units may load the drive excessively because of the multiplier effect of the train ratio.

A gear train is only as good as the bearings and box structure allow it to be. Considering that high quality gearing runs with around 5 μm runouts of concentricity and with pitch errors of similar magnitude, it can be seen that the bearings and machining must be of the highest quality and the box be very rigid.

Gears are graded according to many codes, examples being the AGMA (old and new – American Gear Manufacturers Association) Admiralty, British Standard Specification (BSS), Mil Specs, and more. Specifying odd, nonstandard, gears can be expensive. It is better to choose from standard ranges of diametral pitch and quality classes wherever possible. Reliance Gear (1972), Trylinski (1971), Mabie and Ocvirk (1958), Beggs (1955), Merrit (1975), and Michalec (1966) each cover the theoretical and practical design of the various forms of gears. CAD tools are for supporting gear designs in power gear trains, but are not as useful for the specialism of fine mechanism gears.

The designer can usually make use of the marketed instrument fine mechanism products that have been properly designed to obtain efficient and smooth operation. Manufacturers provide tables and other information to assist the designer to call up the correct centers, clearances, the lubrication to use, and how to design gear trains with specified backlash. Reliance Gear (1972) contains a practical step-by-step guide to the design of fine-pitch gear trains, suggesting the following hints be followed.

The accuracy of gears can be gradually relaxed as the sensitivity to error reduces. The slowest speed gears are the most sensitive. Use spur gears where possible to make the most of the production precision that is available. Keep the number of meshes to a minimum to preserve transmission fidelity. Preload bearings to reduce backlash caused by bearing slop, which is not eliminated with antibacklash gears. Do not use pinned hub gears for high accuracy work. Sometimes gears can be matched by correct phasing to reduce the overall eccentricity error.

The finest pitch gears, theoretically at least, must produce the most precise gear train but practice limits their production. As a guide, avoid teeth finer than 120 DP or the finest above this that is consistent with the load requirements. Avoid mixing pitches in a train. Cleanliness of assembly and in use are vital.

High ratios are obtained in relatively small volume spaces using worm and wheel gears. These, however, may be inappropriate where the gear train has to transmit motion in both directions, as can occur in some servo drives. They can also be damaged easily in instruments where there are shafts that can be unwittingly forced by the operator or service man. Where this might occur, a friction clutch can be used to allow excessive torque to be dissipated without harm.

The use of gears is not restricted to full-circle wheels only. Part gears can be used where the motion needed is only part of a revolution or where intermittent motion is needed. The Leibnitz carry mechanism used in classic mechanical calculators and mechanical automobile odometers is an example. Where precision length measurement or positioning is needed using gears, a rack and pinion is generally adopted. The pinion must run with little backlash and the guides must be carefully manufactured to keep the system aligned. The pieces of rack must by phased properly when butted together to extend the length range. Precision racks can provide relative readout or positioning to within one part in 100 000 (or a little better) uncertainty.

4 CAMS

This is a two-part fundamental mechanism, in which one part, the cam, imparts motion to the second part, the follower, that stays in contact with the surface of the cam. Cams take the shape of disks or plates turning on shafts, as curved slots cut around a shaft, as eccentrics within a yoke; they can also be formed as linear alternatives in which a follower rises and falls (Figure 3) in response to the shape of a linear plate profile moving relative to it. They are used to provide rotary to rotary, or to linear motional conversion, or linear to linear action. Cams are characterized by the feature that the output motion is only generated over a small relative angular magnitude. Cams cannot provide large strokes or rotations unless made equal or larger in size as the amplitudes needed. They find use in timing, function generation, for compensation, and as a means of producing motion from a source.

Fundamentally, cams are continuous surface forms of linkage. At any instance or position, the point of contact between the two parts and the support forms a linkage. The design of cams is well developed and they can be synthesized as well as analyzed using graphical and computational methods. Shigley (1988), Erdman, Sandor and Kota (2001), Juvinall and Marshek (2002), Sclater and Chironis (2001) Parmley (2000) and Mott (1998) are also sources to use. Mabie and Ocvirk (1958) give a good coverage of cam design; Beggs (1955) and Faires (1955) provide less extensive accounts. Rothbart (1956) is a text devoted to cam design. Surprisingly, because cams are used extensively in instruments, Trylinski (1971) does not cover the subject.

Cam design basically attempts to realize the correct conversion action from the two components, but it is complicated by the fact that high-speed cams will impart significant velocity, acceleration, and jerk to the follower. This can cause the follower to leave the cam surface and produce excessive wear. In other words, the correct functioning of the cam and its follower is lost unless the shape is appropriate for the dynamics involved. Various cam shapes provide different velocity, acceleration, and jerk characteristics. Typical forms used are parabolic, modified parabolic, and simple harmonic for relatively slow actions. For higher speed applications where jerk becomes significant, the forms used are cycloidal, half-cycloidal, harmonic,

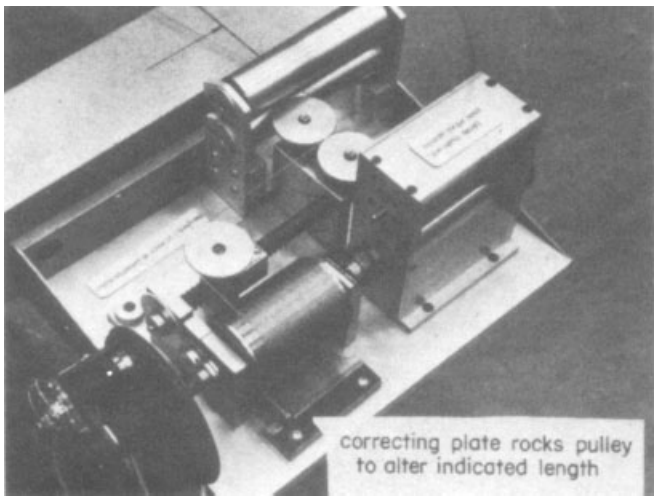


Figure 3. Linear corrector plate used to provide automatic compensation for calibration factor, in-scale nonlinearities and possibly thermal errors in a wire length measuring device (Sydenham, 1969).

and eight-power polynomial. Each offers certain desirable features of operational acceleration and jerk magnitudes; see Mabie and Ocvirk (1958) for more detail.

Cams may run directly against the follower and be constantly in sliding contact. This situation can be designed for by the use of hardened steels and other suitable materials. Alternately, a roller or hydraulic follower may be used. The follower can be a hinged rotating or sliding translating component. The cam in disk or plate form can be shaped to form a function required for applications of compensation or function generation. In these applications, velocity, acceleration, and the higher derivatives generally give no problems. Cams are also used to generate a shape that is 'followed' by optical means. This is used in furnace and other temperature controllers where the temperature must change with time in some specified manner.

Clearly, rapidly changing functions are not readily accommodated by cams because the follower has a finite size and will introduce, by virtue of its aperture width, some degree of integration. Cams are often required to rotate only in one direction. A reversed motion can become locked when the follower comes against a rise rate to which it cannot respond. Cam shaped gear systems also exist that generate specific functions, but these are strictly gears, not cams.

Cams are also made in which the follower responds to the resultant of two inputs. These are referred to as *3-D* or *space cams*. These need a spherical ended follower that is allowed to move in two translations as it is directed by the cam. Space cams are generally of complex 3-D shape and can be made by sintered forming methods or be cut by numerically controlled machine tools. These cams are unable to impart much force into the follower so they are often used with a servo that slaves to the follower's motion.

The dynamic behavior of cams can be complex because of the high value accelerations that can be present. These can cause the cam to appear elastic and can induce resonance in the cam body and in the follower and its spring. Beggs (1955) reviews this problem.

5 FLEXIBLE CONNECTING DRIVES

This classification is given by Trylinski (1971) to that group of motional transference mechanisms that use flexible elements to transmit motion. They may do this with or without conversion from rotary to linear regimes. Some links use elasticity to provide the freedom of bending; others use kinematic, pivoted principles such as those found in a link chain.

Elastic types will not fail because of the bending stresses being excessive in their outermost 'fibers' if they have slenderness ratios of at least 300 : 1 (ratio of radius of bending to the thickness of the flexure in the direction of bending). This figure is only a guide. A calculation should be made of the stresses in the outer layer when the member is bent over the smallest radius, in order to see if the material is not near to the yield point. There may be a jockey wheel or a guide pulley added to keep the system taut; often these are made too small in diameter. The stress level is as important as designing for adequate drive forces because excessive bending stresses will eventually cause the drive material to fatigue and break.

6 BELT DRIVES

There are many alternatives available, each having particular advantages and disadvantages. Flat belts made of a material having adequate tensile strength and friction coefficient can be used with flat pulleys (actually the pulleys should be slightly curved, rising in the center, to cause the belt to track into the middle). More usually, instruments would use small section vee-belts or continuous round section belts made of reinforced plastics. O-rings make excellent belts; as they are made to tight tolerances, they provide smooth drives. If the forces to be transmitted are high, the pulleys and belts should be designed after calculations are made to decide the ability of the drive to operate with a realistic tension that will transmit the force without breaking. Round and vee-profile pulleys provide increased friction drive force without needing high belt tensions. Their design is laid down in texts on general machine design and is well defined. Trylinski (1971) is a satisfactory source on flexible instrument drives. Figure 4 gives a range of flexible drives used in instruments.

The friction drive belt depends upon the normal force exerted between the belt and the pulley surface in order to develop enough grip. These forces can sometimes be prohibitive and a better alternative may be needed. Furthermore, the drive may need to maintain a strict synchronized positional relationship between the two pulleys. For these cases, friction belt drives are not satisfactory.

7 CHAIN DRIVES

Synchronized flexible chain drive can be obtained by the use of toothed flexible belts running on appropriately toothed wheels. These are generally available only for comparatively large-size drives, the smallest being about 10-mm wide. Metal roller link chains also can be used but these, too, are not generally made for the size drives needed in the smaller instruments. Smaller special chains

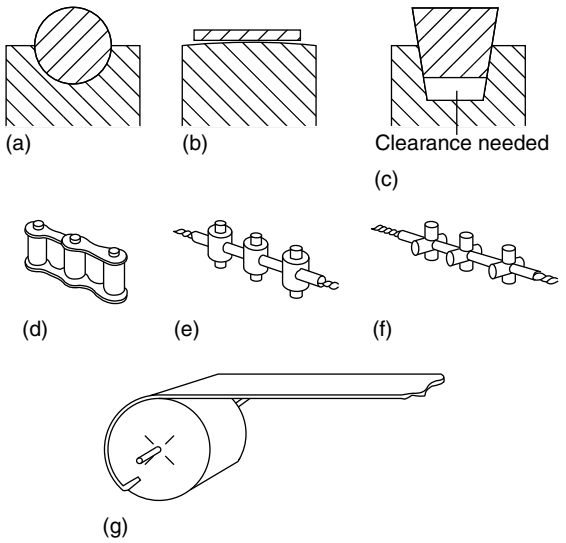


Figure 4. Cross sections of flexible connecting drives used in instruments: (a) round, with or without reinforcement; (b) flat (on slightly domed pulley); (c) vee; (d) metal link roller chain; (e) polyurethane molded – pinned belt; (f) as for 3-D pins; (g) metal strip.

and toothed belts are made but their availability is not widespread. The user will have to search hard for a supplier (see Reliance Gear, 1972, and the British Machine Chain Co, Malden). Examples are the miniature oval-link chains that use specially shaped drive wheels in which the chain links lie keeping its crossed link shape, and the hook-link chains that use sprockets whose teeth pass through the links to provide drive. Fine metal chains can give backlash-free coupling at the force level needed.

Typical working forces for small chains are 1 N, with links of 3-mm pitch. Small, plastic, positive-drive belts are also manufactured. One system uses stainless steel or cord-reinforced polyurethane molded belts, which comprise a round belt having bars formed across it protruding to each side. Another form has its sides shaped like a ladder. The bars nest into a single (or double) row of teeth on the drive pulleys. One Dacron-reinforced belt will take 1 N forces, the unreinforced belt only 0.05 N. Some designs will carry a 15-N tensile load. There is also a toothed plastic belt that will drive sprockets set at 90° to each other without the need for belt twist. For illustrations of the above see Figure 4.

When only a small distance of angular motion is required, the belt need not be endless; further, it can be fastened to the two pulleys to obtain synchronized drive. Examples are the mechanical drive of a tuning dial of a radio in which a cord connects the tuning control with the translating station indicator. Thin steel strips can also be used to couple wheels. As these can only drive in tension, a continuous strip or a biasing spring are needed. Trylinski (1971) gives several illustrations of the possibilities.

Flexible drives can offer extended distances between drive centers and the ability to provide gearing ratios; they can absorb vibration, attenuating it through the link (but they may lack stiffness if this is so); they can allow for misalignment between shafts, they can be designed to allow slippage if the load is too great (or the converse), and they can also be designed to give a continuously variable gearing ratio by the use of tapered pulleys and changing radius pulleys. They are virtually backlash-free and will have relatively low inertia compared with gears trains because they have less mass in motion. They can, however, introduce vibrations from the unsupported belt, which can modulate the drive throughput of force or information.

8 ELECTROMECHANICAL GEARING

The present low cost of digital electronics and the general availability of very reliable and inexpensive optical and other shaft encoders has given the designer the ability to make more use of electromechanical gearing and shaft motion transference.

Combination of a shaft digitizer (Figure 5) that generates electrical positional signals as the shaft rotates and a motor that steps in synchronism with the generated signals forms the basis of an electromechanical gear box.

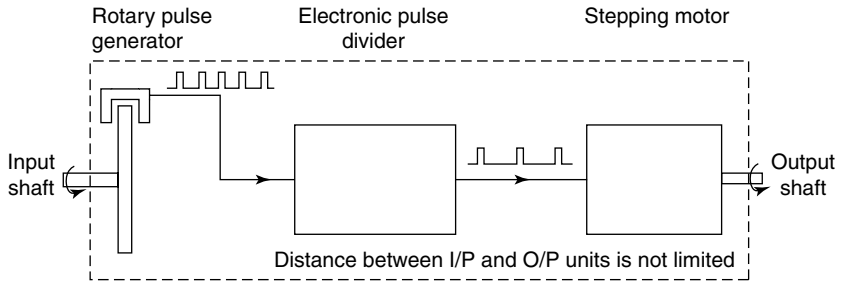


Figure 5. Diagrammatic representation of an electromechanical gear box.

Illustrating the method with the use of the optical radial gratings, the disc attached to the input shaft can conveniently generate around 1000 to 10 000 pulses per turn at rates rising to at least 10^6 pulses s^{-1}; that is, shaft speeds varying from 0 to 1000 rev s^{-1} for low resolution or to 10 rev s^{-1} for high resolution can be accommodated. The digital signals are then passed through a digital divider arrangement to form the required divided pulse rate to suit the speed needed from a stepping motor. Dividers are available as integrated circuits (IC) in which preset switches or electronic signals can be used to set any division number needed. Two ICs can give ratios from unity to 255.

The advantages of this alternative form of 'gear box' are that the input and output shafts need not be in close proximity or in any particular orientation, the power levels of the two sides of the link can be at any level needed, more than one shaft signal can be combined or driven, and the division ratio is very easily changed under computer control. The ratio can be varied smoothly from one extreme to the other. Accuracy can be at any level desired and backlash is not present.

This form of system, however, can be prone to electrical interference, but it is quite possible, with circuitry effort, to make the error rate as small as is desired by the use of coded digital communication transmission methods.

The cost of the electronic parts will be reasonably small and the mechanical parts needed are few – a shaft carrying an encoder disk to couple to the output of a mechanism and an output shaft of the stepping motor. Such gear boxes are cheaper than conventional mechanical units, especially where high precision is needed. These will, of course, need a source of electrical power, which may exclude them from some applications.

RELATED ARTICLES

Article 57, Static Considerations of General Instrumentation, Volume 1; **Article 59, Introduction to the Dynamic Regime of Measurement Systems, Volume 1**; **Article 80, Principles of Fine Mechanics – Kinematic and Elastic Designs, Volume 2**; **Article 81, Principles of Fine Mechanics – Systems Considerations, Volume 2**; **Article 85, Elastic Regime of Design – Design Principles, Volume 2**.

REFERENCES

Beggs, J.S. (1955) *Mechanism*, McGraw-Hill, New York.

Erdman, A.G., Sandor, G.N. and Kota, S. (2001) *Mechanism Design: Analysis and Synthesis*, Prentice Hall.

Faires, V.M. (1955) *Design of Machine Elements*, Macmillan Publishing, New York, (1st edn 1934).

Jain, P.K. and Sydenham, P.H. (1980) Radial Metrological Pattern Generating Engine. *Journal of Physics E-Scientific Instruments*, **13**, 461–466.

Juvinall, R.C. and Marshek, K.M. (2002) *Fundamentals of Machine Component Design*, Wiley, New York.

Lent, D. (1970) *Analysis and Design of Mechanisms*, Prentice Hall, Englewood Cliffs, NJ.

Mabie, H.H. and Ocvirk, F.W. (1958) *Mechanisms and Dynamics of Machinery*, Wiley, New York, (3rd edn 1975).

Merrit, H.E. (1975) *Gear Engineering*, Wiley, New York.

Michalec, C.W. (1966) *Precision Gearing: Theory and Practice*, Wiley, New York.

Mott, R.L. (1998) *Machine Elements in Mechanical Design*, Prentice Hall.

Parmley, R.O. (2000) *Illustrated Sourcebook of Mechanical Components*, McGraw-Hill.

Peatfield, A.E. (1951) *Engineering Components and Materials*, EUP, London.

Reliance Gear. (1972) *Catalogue RG5*, Reliance Gear Co Ltd, Huddersfield. (Also similar catalogues by PIC Design Corp, New York).

Rothbart, H.A. (1956) *Cams: Design, Dynamics and Accuracy*, Wiley, New York.

Sclater, A. and Chironis, N.P. (2001) *Mechanisms and Mechanical Devices Sourcebook*, McGraw-Hill.

Shigley, J.E. (1988) *Mechanical Engineering Design (McGraw-Hill Series in Mechanical Engineering)*, McGraw-Hill.

Sydenham, P.H. (1968) Linear and Angular Transducers for Positional Control in the Decameter Range. *Proceeding of the IEE*, **115**, 1056–1066.

Sydenham, P.H. (1969) *Manual and automatic large-scale metrology*, PhD thesis, University of Warwick.

Sydenham, P.H. and Brandwijk, J.F. (1966) Production of Radial Line Gratings Using a Mechanical Photographic Apparatus. *Journal of Scientific Instruments*, **43**, 380–382.

Trylinski, W. (1971) *Fine Mechanisms and Precision Instruments – Principles of Design*, Elsevier, translated by A. Voellnagel from Polish original.

85: Elastic Regime of Design – Design Principles

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Place of Flexure Elements in Instruments 540
2 Basis of Designing Flexure Elements 541
3 Design Parameters of the Flexure Element 541
4 Carrying out the Design of a Flexure Element 543
5 Specialized Formed Flexure Elements 543
6 Commonly Found Beam Design Formulae 543
7 Types of Flexure Elements 544
8 Information on Flexure Systems 546
Related Articles 546
References 546

1 PLACE OF FLEXURE ELEMENTS IN INSTRUMENTS

Mechanisms used to assemble instruments are divided into two groups, namely, kinematic and elastic. Kinematic elements are made sufficiently inelastic to be unchanging in shape as they are stressed by mechanical forces. Here, the general principles of the elastic group that deliberately makes use of elasticity of components to achieve a given mechanism function are discussed. Flexing action produced by kinematic mechanism, such as a roller chain or belt, is covered elsewhere. Detailed discussion of uses of elastic elements is found in **Article 86, Elastic Regime of Design – Spring Systems, Volume 2** and **Article 87, Elastic Regime of Design – Plates and Bimorphs, Volume 2**. Implementation in microchip form is covered in descriptions of Microelectro Mechanical Systems (MEMS), **Article 162, Principles of MEMS, Volume 3**; **Article 163, Uses and Benefits of MEMS, Volume 3**; and **Article 164, Principles of MEMS Actuators, Volume 3**.

By appropriate choice of compliance, component shape, and method of use, spring elements (often called *flexures*) can provide many of the functions provided by the kinematic forms. Elastic design deliberately makes use of elastic properties of materials to allow controlled movement where needed.

In many instances, a flexure method is to be preferred over a kinematic arrangement. This generally occurs when only limited movement between two parts is needed. For example, a crossed-strip flexure hinge provides a limited range angular hinging action and offers negligible backlash and frictional force. It does not need lubrication and is not affected by dust or dirt. It is very simple to make and often does not require tight manufacturing tolerance.

Flexure elements can be devised in numerous shapes and forms. Good design is always based on proper understanding of the compliance of a flexure element in its various directions of bending and compliance values of other associated elements. In some cases, it needs to have a high ratio of compliance in one direction with respect to another; in other cases, it needs to be the same. The addition of later items, such as protective covers, additional small loads, and further forces from sources other than those designed for, could considerably alter the final performance of an elastic design.

Flexure elements appear to be deceptively simple. They are easy to devise and install and they will often provide quite adequate performance, with only the barest of theoretical understanding being applied. However, a thorough appreciation of the factors involved will assist the user to understand their limitations; they are not perfect elements.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

The necessary relevant basic theory for carrying out elastic flexures design pertains to that of the deflection and stresses of beams and plates – commonly called *strength of material* or *beam theory*.

To this, the knowledge of imperfections of real materials must be added because these introduce such effects as damping, friction, hysteresis, fatigue, fracture, and nonlinear deflection versus load characteristics – subjects not normally covered in beam theory texts.

Fundamental material (which is of civil and mechanical engineering structures design) is to be found in such works as Timoshenko and MacCullough (1959), Timoshenko and Gere (1961), and Roark and Young (1975). Faires (1955) provides a shorter but basically adequate account of the strength of materials theory. Such texts, being written for structure designers and not for instrument applications, generally place greater emphasis on stress than on the deflections that are so important to instrument design.

2 BASIS OF DESIGNING FLEXURE ELEMENTS

Flexure design studies are usually based upon the assumption that the force-deflection characteristics of the materials used are linear. Such a basis can, indeed, give the designer a very wide choice of workable ideas, but on occasion, it might be beneficial to make use of nonlinear materials and processes. Eisley (1963) provides an entry point into the literature in this area of knowledge, giving about 90 references to published papers. Schmidt and DaDeppo (1973) may also be useful in this area.

A considerable amount of published design information is available but it is from earlier times with little being published recently. This should not be taken to imply that the information has become outdated but merely that this field matured several decades ago. There is little likelihood of a basically new flexure element being invented, most having been devised by now. This article provides a comprehensive overview of the many possibilities for elastic design. Combination of several flexure mechanisms leads to more options, many of which are yet to be applied.

The apparent simplicity of flexures can hide the fact that they are not perfect elements. As an example, a parallel strip support for a translating table can easily support a mass of many kilograms, providing translational sensitivity and stability of nanometer order without the backlash being apparent. However, temperature effects may alter the inbuilt spring rate quite considerably, causing errors in other ways. Choice of the wrong material can introduce significant hysteretic backlash.

As flexure systems deliberately introduce spring compliance into a system where mass is involved, it is obvious that linear second-order and higher dynamic systems may be produced (refer to **Article 189, Basic Principles of Flow Measurement, Volume 3**). The natural frequency of such systems becomes an important parameter where the system has to follow dynamic inputs in a faithful manner (Sydenham, 1982, 1983). Damping may be deliberately introduced to prevent resonances from building in amplitude or to make the dynamic response at least critical so as to exhibit response characteristics of first-order nature. It is feasible, by appropriate design and ingenuity, to alter the natural frequency without altering the compliance. For the same reasons, it may be necessary to keep the supported mass to a minimum so that the natural frequency is kept as high as possible.

3 DESIGN PARAMETERS OF THE FLEXURE ELEMENT

Before considering practical flexures and their application, it is appropriate to discuss several basic design parameters of flexure elements since these apply to all shapes and forms of system built with them.

Flexures are formed by the use of a mechanical member (or possibly, other spring form such as electromagnetism), where design deliberately allows the member to deflect in specific directions. It is usually imperative that deflection occurs only within the elastic region of the material's deformation (Figure 1). Although texts on materials give the general impression that the Hookean elasticity law applies in practice, there, nevertheless, is always a small amount of hysteresis; a strained member does not come back to exactly the same strain condition for a given stress excursion. Special materials, those with a reduced mechanical hysteresis, exhibit a lower level of this effect. In many applications of flexures in instruments, however, extreme reproducibility is not important. Tovey (1974) discusses the materials used to form flexure elements.

It is therefore important that the flexure members cannot be deflected beyond the chosen allowable elastic limit. As a rule, the smaller the deflection, the smaller the residual strain error. Plastic deformation, if allowed to occur, will introduce permanent set, with subsequent zero shift errors arising. For this reason, design often calls for mechanical stops and other devices to be incorporated to prevent overdeflection. Assembly and transport can often give rise to the greatest deflections that might be encountered. As a rough rule of thumb, bending flexure elements should not be deflected into a radius of curvature smaller than around 300 times its thickness in the direction of bending. Beyond this level, the outer fibers will be stressed beyond their elastic limit, causing surface cracks to

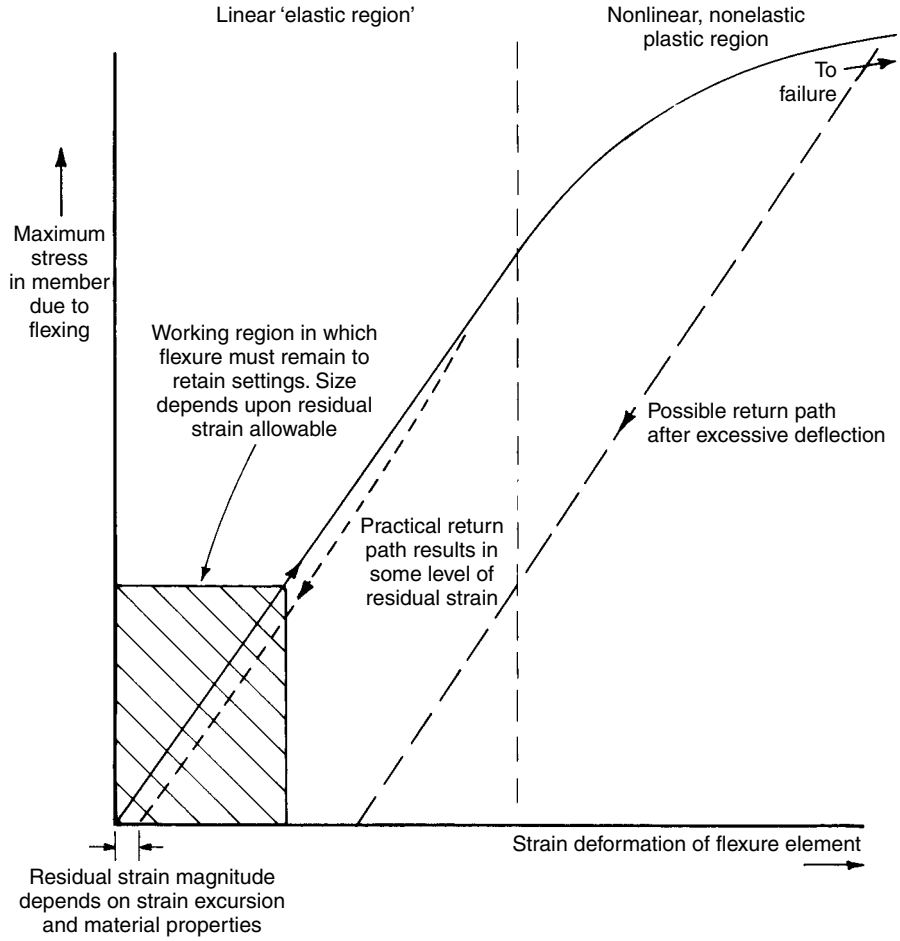


Figure 1. Stress–strain relationships for materials used to make flexure elements.

begin. If this is important, the bending moment caused by expected flexing levels should be calculated for the material used.

As a flexure deflects, it will provide a force reacting with the direction of motion. This is the flexure system compliance. It is produced fundamentally by the elasticity of the material, the effect of which is modified by the total flexure system arrangement. As will be mentioned below, it is possible to devise flexures that provide positive, negative, and even zero forces, as well as nonlinear displacement characteristics.

Using strips in tension will increase the stiffness according to the load (Eastman, 1937). Conversely, they can be used in longitudinal compression quite satisfactorily, provided Euler buckling does not occur. The basic elements still undergo positive or negative deflections toward or away from the elastic limit, but the net effect of the whole flexure system is one of these options.

The fundamental design parameters of a flexure material are its elastic modulus, shear modulus, hysteresis level, elastic limit, endurance limit to constant flexing (especially when in conditions of simultaneous corrosion effects), thermal coefficient of elastic modulus, and such other parameters in any other energy domain. Rolling friction is sometimes a parameter for which more information is published in Minato, Nakafuku and Takemura (1969). Gohn and Fox (1965) discuss stress relaxation in springs. Carson (1937) is about measurement of elastic drift in springs.

When very small displacements are needed, a method used by Jones (1967) might be applicable. It rests upon the principle that an elastic material that is squeezed by a clamp will exhibit a small length change. Poisson's ratio links this action. This method gives fine control to clamping screws on threaded adjustments such as micrometer heads. The same principle was employed by Baird (1968) to temperature-compensate an Invar rod with a steel wire cover soldered to its outer surface.

When flexures are made from thin hardened and tempered strips, they can tear easily at connection points. Careful edge finishing is needed to reduce this failure mode. The material must also be cut to suit any directional grain that exists in the flexure material.

Materials used for flexures are generally the metal carbon springs, stainless steels, copper, and nickel–iron alloys. In some applications, fused-silica (fused-quartz is a term that is usually synonymous) forms the flexure element. In less exacting applications, plastic, rubber, paper, and fiber elements might be suitable.

The above parameters are concerned with the properties of the material used to form a flexure. Having selected the material, it is then necessary to consider the properties of the flexure system as a whole. Some factors that may be relevant include the force-deflection characteristic, range of motion, required parallelity of pseudosliding action, stability of a center of rotation with angular deflection, and effect of load on the system properties because the system may stiffen up, with increasing load being supported. Another important design parameter is to be sure that the system can be manufactured to the tolerances needed and that it can be assembled without damaging the flexure element in some way. Preform in flexure elements can be bothersome at times. Elements may need to be heat-treated to stabilize them before use and be assembled using jigs to prevent overstressing.

4 CARRYING OUT THE DESIGN OF A FLEXURE ELEMENT

Design of a flexure system, like that of a kinematic support system, should always consider the sources of error that might arise in normal operation.

At present, and perhaps it will always be so, the design of a flexure system is initially approached as an art on the basis of experience, with the possibilities open to the designer. A possible arrangement is first selected that may fulfill the need. This is then analyzed as a science to establish the sensitivity, load-carrying capacity, limits of required action, and so on. If the design is not quite suitable, it is then modified. It would be an obvious advantage if computer-aided design programs were available for use in optimizing standard design configurations (of which there are several kinds) in the direction that a designer requires.

Design equations for most types have been reasonably well developed, but the designer usually must spend considerable time in locating the appropriate study. Many reports on flexure elements usage do not include a theoretical account, their design often being apparently raised by intuition and cut-and-try prototype building procedures. Of the references to general accounts given above, those including a well-developed basis of theory to support a design are Chironis (1961), Eastman (1937), Andreeva (1966), Geary (1954, 1960), Thorpe (1953), Trylinski (1971), Wahl (1963), and Weinstein (1965).

5 SPECIALIZED FORMED FLEXURE ELEMENTS

It is stated by some authors and users of flexures that the applications requiring the greatest dimensional stability should use flexures that have end mountings made in the same piece of material. (The usual method is to clamp the flexure end between rigid plates.) This assertion has not been adequately proven by rigorous experimental procedures and it must be left to the user to decide whether this additional cost is justified.

A similar argument exists as to whether metal spring elements give better performance than fused-quartz springs. Experience with both has shown that poor reproducibility of the zero reading with time is more likely to be a factor of overall design than of the material used. It is possible to find evidence that suggests that either material group is the better to use for precision instrument springs. As the problem reduces to one concerning the hysteresis and temperature properties of the materials, it will be found that certain metal alloys are far superior to fused-quartz in this respect.

A key parameter that tends to be overlooked in such comparisons is the level of flexure excursion. Small movements of a spring element, such as those occurring in a tidal gravimeter or a tidal tiltmeter, will give rise to much smaller levels of hysteresis when compared to the performance of an exploration gravity meter or a borehole tiltmeter. In the latter case, the flexure element must range over larger excursions and can be expected to be more severely handled.

6 COMMONLY FOUND BEAM DESIGN FORMULAE

Figure 2 provides the usually needed deflection and stress expressions for a few of the commonly found elements. Note that point load and distributed load cases each follow the same general laws.

For flexures to be theoretically designable, it is necessary that their stress levels be determinable for all parts of the system. In the kinematic design procedure, forces are arranged so that they act at known, well-defined points of interaction. In flexure design, deflection takes place over a distributed region.

To assist design, it is often expedient to structure the flexure elements of the system as members having dramatic changes of size at given positions, so the compliance alters very significantly. It can then be assumed that the flexure member is deforming elastically, with the end mounts being adequately rigid.

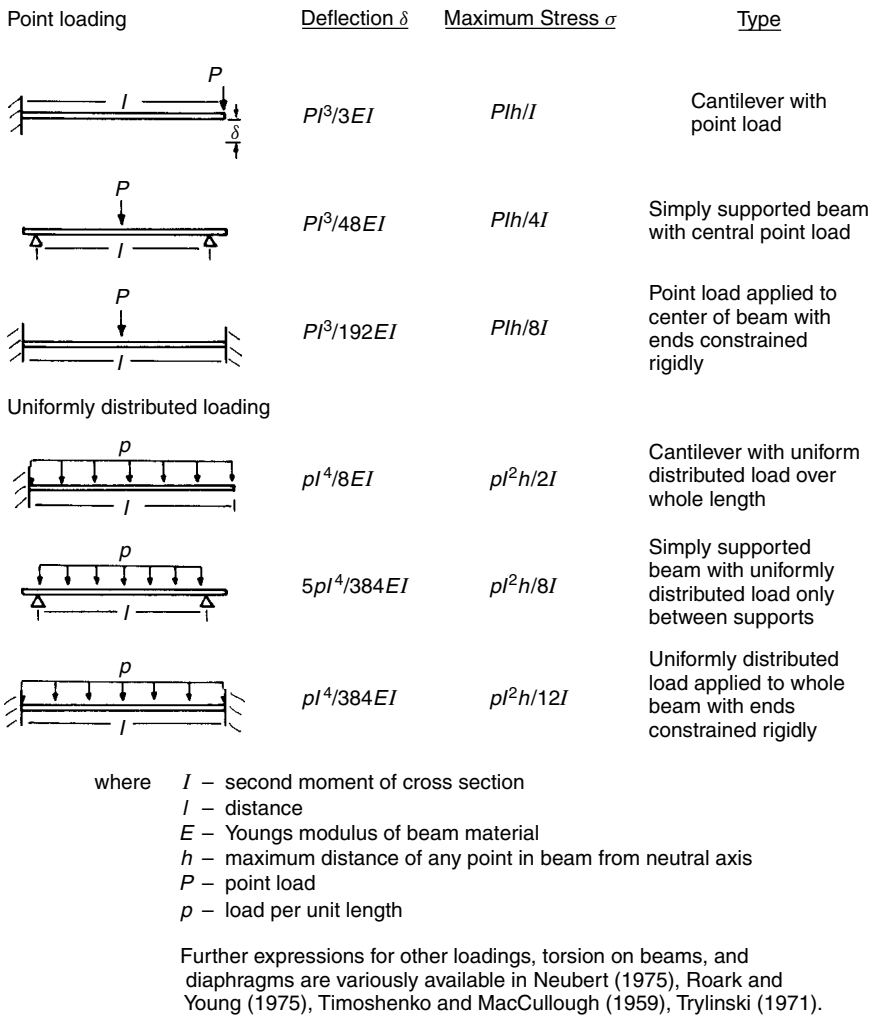


Figure 2. Expressions for commonly found beams.

From the viewpoint of fatigue strength, however, such sudden geometrical changes in shape can give rise to locations in the flexure where stresses can be raised. Flexure designs needing high numbers of highly stressed cyclic movements will require a more sophisticated treatment to allow for graded compliance changes along the flexure element.

7 TYPES OF FLEXURE ELEMENTS

A flexure can provide controlled movement in a system of flexure elements, using one or more of three basic modes of flexure (Sydenham, 1984).

It can be deflected in pure transverse bending, in pure axial torsion, or it might be deformed elastically in its longitudinal compressive or tensional mode (Figure 3a). In practice, combinations of these three actions occur to a lesser or greater extent depending upon the flexure system arrangement. Too often, designers (and quoted formulae for flexure arrangements) do not recognize this combinational feature. For example, a simple single flexure strip being used at the top end of a clock pendulum to allow it to swing in one preferred direction also supports the mass of the pendulum. The flexure element is, therefore, simultaneously stressed in both bending and tension. The tension component can considerably stiffen the rotational compliance in the bending direction.

The helical coil spring is a torsional arrangement in which torsional flexure is converted to provide linear spring-force action. Torsional flexure systems can make use of forms other than strip or other solid members. Two wires, placed apart, can produce a similar action – these are called *bifilar suspensions*.

The third class, wherein longitudinal tension or compression is used without transverse bending action of the flexure, is applicable when stiff flexures are needed. Simple examples would be the use of a bolt to hold two

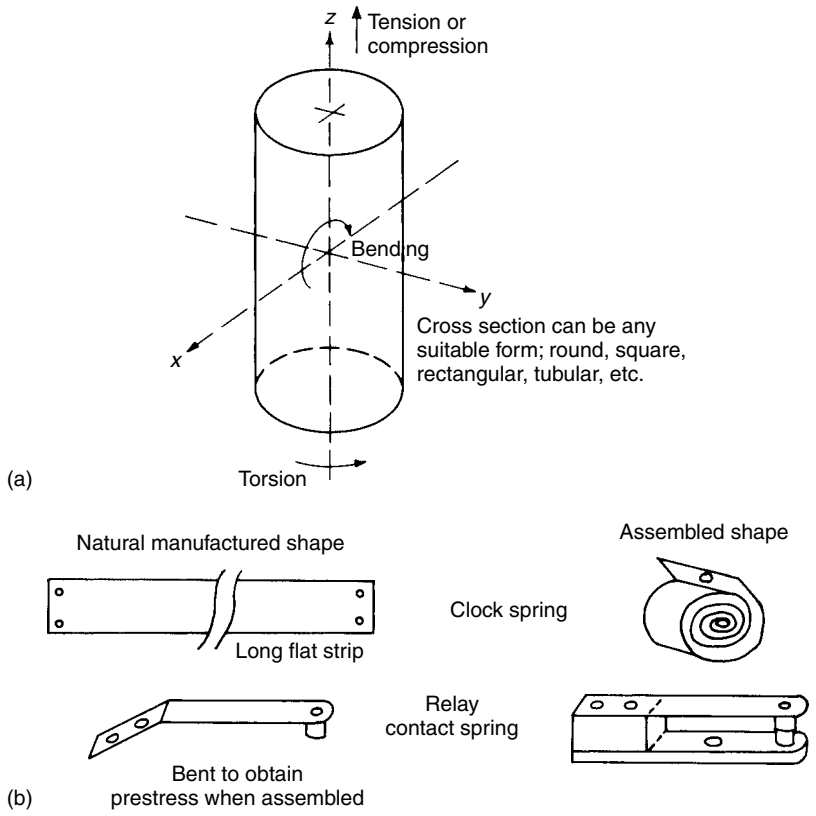


Figure 3. Fundamental flexure design options: (a) modes of flexure deflection; (b) natural shapes used either in some manufactured shape or preformed.

parts together yet allowing some additional elasticity to enable it to cope with excess tension loads or a rubber isolation mount. Torsional systems often stress the individual members this way. Such methods may dictate the use of materials with low values of Young's modulus so that they provide reasonable compliance from practical cross sections of material. Their use is, of course, dependent on the requirement and materials' suitability.

With the basic elastic kind of action being decided, the designer then has a further fundamental option available. Many flexure systems use the element in a stress-free state when the system is at its rest position. For example, a helical spring may be coiled so that the turns just nest together and with no extension, the spring material is in a relatively stress-free condition (see Figure 3b). Elastic energy is then stored as the spring is tensioned.

Another example is the use of clock spring to provide drive energy for a clock. In its unassembled state, the natural shape of the flexure is a long flat strip. As it is wound up, it takes on the curved shape seen in the assembled clock.

The alternative spring flexure arrangement is to prestress the spring material into a natural shape that operates in the reverse manner to the obvious situation exemplified above. A torque motor (that is needed to drive a mechanical clock) can be produced by precurving the strip so that its natural shape is a small radius, that is, the unassembled spring rests as a tight coil that is then unwound as it is deformed to store energy. Springs holding electrical contacts in relays are usually preset to some extent so that their assembled state holds the contact in the correct position, applying a force to the other contact.

Yet another distinction can be drawn between the flexure system that is purposely intended to store energy in order to produce a force on demand and that which uses flexure action as a mechanical linkage. In the latter case, it is often necessary to design for minimum stored energy as deflection occurs, storage and its resulting force-producing action being an unwanted parameter. Often, the flexure system simultaneously provides both mechanical linkage and energy storage functions; for example, a cantilever flat strip spring provides return action and guidance.

Through the use of a suitable design of spring, it is possible to obtain mechanical magnification of small displacements. Such springs are generally of the type in which a small change in length of a tensional spiral spring produces a rotational motion in the center or at the end of the spring element. The rotation can then be used to operate a

suitable length output arm or the mirror of an optical lever. Torsional magnifiers are discussed in Geary (1960).

In practice, there are a very great number of possibilities, so it is not feasible to list every combination and arrangement that has been reported.

8 INFORMATION ON FLEXURE SYSTEMS

Collectively, the literature (when reviews included within several of the citations are also recognized) provides a very large range of diagrams and applications of flexure elements in the design of instruments. What constitutes an instrument is a matter of personal definition in this area, for flexure elements find as much use in power-supporting applications as they do in fine mechanisms.

Young (1989) is a definitive source of formulae for mechanical situations. The design of springs elements, which commonly concentrates upon helical and flat power springs (*see* **Article 86, Elastic Regime of Design – Spring Systems, Volume 2**), as opposed to instrument springs, is dealt with by Chironis (1961) and Wahl (1963). Spring suspensions and flexure devices, for which the emphasis is upon fine mechanism, are covered in Eastman (1937), Geary (1954, 1960), Andreeva (1966), Trylinski (1971), and Sydenham (1984). The definitive reference work is Howell (2001). It covers the wide range of topics introduced here.

Geary (1964), Frazier, Gilinson and Oberbeck (1974), and Fremery and Doden (1978) cover another type of spring suspension: that provided by magnetic and electric fields. Dorien-Brown (1971) is one review of air springs.

Many other works include mention of flexure arrangements in sections ranging from chapter length down to a few paragraphs. These include Rolt (1929), Whitehead (1934), Thorpe (1953), Braddick (1954), Beggs (1955), Weinstein (1965), Jones (1967, 1968), and Neubert (1975). Hugill (1978) compares several spring-mass systems used in gravimeter designs. Equations of sensitivity are given in each case. Gast (1974) includes discussion on flexures. Dratler and Block (1972) is an interesting report. Sclater and Chironis (2001) is a useful source of ideas.

RELATED ARTICLES

Article 80, Principles of Fine Mechanics – Kinematic and Elastic Designs, Volume 2; **Article 81, Principles of Fine Mechanics – Systems Considerations, Volume 2**; **Article 82, Kinematical Regime – Members and Linkages, Volume 2**; **Article 86, Elastic Regime of Design – Spring Systems, Volume 2**; **Article 87, Elastic Regime of Design – Plates and Bimorphs, Volume 2**; **Article 88, Error Sources in Fine Mechanics, Volume 2**.

REFERENCES

Andreeva, L.E. (1966) *Elastic Elements of Instruments*, Israel Program for Scientific Translations Ltd, Israel. Translation from Russian.

Baird, K.M. (1968) Compensation for Linear Thermal Expansion. *Metrologia*, **4**, 145–146.

Beggs, J.S. (1955) *Mechanism*, McGraw-Hill, New York.

Braddick, H.J.J. (1954) *The Physics of Experimental Method*, Chapman & Hall, London.

Carson, R.W. (1937) Measuring Elastic Drift. *Proceedings of ASTM*, **37**, 661–675.

Chironis, N.P. (1961) *Spring Design and Application*, McGraw-Hill, New York.

Dorien-Brown, B. (1971) Air Springs for Vibration Isolation, *Harold Armstrong Conference on Production Science in Industry*, August, Institution of Engineers, Melbourne, (pp. 213–225).

Dratler, J. and Block, B. (1972) A Wide-band Horizontal Accelerometer with Preliminary Earth Normal Mode and Seismic Investigations. *Geophysical Journal of the Royal Astronomical Society*, **27**, 337–367.

Eastman, F.S. (1937) The Design of Flexure Pivots. *Journal of Aeronautical Sciences*, **5**(1), 16–21.

Eisley, J.G. (1963) Nonlinear Deformation of Elastic Beams, Rings and Strings, (Bibliography, 92 refs). *Applied Mechanics Reviews*, **16**(9), 677–679.

Faires, V.M. (1955) *Design of Machine Elements*, Macmillan, New York.

Frazier, R.H., Gilinson, P.J. and Oberbeck, G.A. (1974) *Magnetic and Electric Suspensions, Types of Suspensions Developed in the Charles Stork Draper Laboratory*, MIT Press.

Fremery, J.K. and Doden, K. (1978) Active Permanent Magnet Suspensions for Scientific Instruments. *Journal of Physics E: Scientific Instruments*, **11**(2), 106–113.

Gast, T. (1974) Vacuum Micro Balances, their Construction and Characteristics. *Journal of Physics E: Scientific Instruments*, **7**, 865–875.

Geary, P.J. (1954) *Flexure Devices – Pivots, Movements, Suspensions*, Sira Institute, Chislehurst.

Geary, P.J. (1960) *Torsion Devices*, Sira Institute, Chislehurst.

Geary, P.J. (1964) *Magnetic and Electric Suspensions*, Sira Institute, Chislehurst.

Gohn, G.R. and Fox, A. (1965) Stress Relaxation in Metallic Springs and Conductors. *Electrotechnology*, **16**, 85–92.

Howell, L.L. (2001) *Compliant Mechanisms*, Wiley, NY.

Hugill, A.L. (1978) *Gravimeter Design*, Int Rep DSS/AH/171, September, Systems Science Department, The City University, London.

Jones, R.V. (1967) The Measurement and Control of Small Displacements. *Physics Bulletin*, **18**, 325–326.

Jones, R.V. (1968) More and More about Less and Less. *Proceedings of Royal Institution of Great Britain*, **43**(202), 323–345.

Minato, K., Nakafuku, C. and Takemura, T. (1969) Rolling Friction of Metals. *Japanese Journal of Applied Physics*, **8**(10), 1171–1179.

Neubert, H.K.P. (1975) *Instrument Transducers – An Introduction to their Performance and Design*, Clarendon Press, Oxford.

Roark, R.J. and Young, W.C. (1975) *Formulas for Stress and Strain*, McGraw-Hill, New York. (Editions since 1938).

Rolt, R.H. (1929) *Gauges and Fine Measurements*, Vol. 2, Macmillan, London.

Schmidt, R. and DaDeppo, D.A. (1973) Variational Formulation on Nonlinear Equations for Straight Elastic Beams. *Journal of Industrial Mathematical Society*, **23**(2), 117–136.

Sclater, A. and Chironis N.P. (2001) *Mechanisms and Mechanical Devices Sourcebook*, McGraw-Hill, London.

Sydenham, P.H. (1982, 1983) *Handbook of Measurement Science*, Wiley, Chichester.

Sydenham, P.H. (1984) Elastic Design of Fine Mechanism in Instruments. *Journal of Physics E: Scientific Instruments*, **17**, 922–930.

Thorpe, A.G. II. (1953) Flexure Pivots-design, *Product Engineering*, **24**(2), 192–200.

Timoshenko, S.P. and Gere, J.M. (1961) *Theory of Elastic Stability*, McGraw-Hill, New York.

Timoshenko, S. and MacCullough, G.H. (1959) *Elements of Strength of Materials*, Van Nostrand, NJ. (Numerous editions since 1935).

Tovey, F.M. (1974) *Material Properties in the Design of Transducer Flexibles*, Ai Research Manufacturing Co, Phoenix, AZ. (plus illustrations to accompany above as separate).

Trylinski, W. (1971) *Fine Mechanisms and Precision Instruments – Principles of Design*, Pergamon Press, Oxford. Translated by A. Voellnagel from Polish original.

Wahl, A.M. (1963) *Mechanical Springs*, McGraw-Hill, New York.

Weinstein, W.D. (1965) Flexure Pivot Bearings, *Machine Design*, Part 1 June, 151–157, Part 2 July, 136–145.

Whitehead, T.N. (1934) *Instruments and Accurate Mechanisms – Underlying Principles*, Macmillan, London. (Also Dover Pubs, New York, 1954).

Young, W.C. (1989) *Roark's Formulas for Stress and Strain*, McGraw Hill, London.

86: Elastic Regime of Design – Spring Systems

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Helical Springs 548
2 Spiral Springs 549
3 Flat-strip Springs 550
4 Managing the Degrees of Freedom in Elastic Systems 552
5 Buckling Effects in Flexure Elements 552
6 Special Shapes of Flexure Elements 554
7 Examples of Spring Flexure Elements 554
Related Articles 554
References 554
Further Reading 555

1 HELICAL SPRINGS

Elasticity of mechanical components is used to obtain necessary degrees of freedom in the elastic regime of fine mechanical design. The general principles of elastic design are given in **Article 85, Elastic Regime of Design – Design Principles, Volume 2**. Comparison with the alternative kinematic methods is made in **Article 80, Principles of Fine Mechanics – Kinematic and Elastic Designs, Volume 2** and **Article 81, Principles of Fine Mechanics – Systems Considerations, Volume 2**. Spring element design has regained interest in MEMS applications, *see* **Article 162, Principles of MEMS, Volume 3**.

The design of spiral (helical, coil) springs is well covered in numerous commonly available sources because they form part of the training of engineers in the general design of machines. For example, each of the following authors Chironis (1961), Andreeva (1966), Faires (1955), Geary (1960), Trylinski (1971), and Wahl (1963) can be consulted for the theoretical design of the familiar cylindrical coil spring. Howell (2001) is the definitive source in elastic mechanism. Such springs are generally intended for use as pure tension or pure compression flexure elements (Figure 1a and b), but they can also be used as torsional or bending elements.

The above references discuss the key spring design parameters to obtain a spring with a given spring rate (compliance, stiffness), adequate range of deflection, adequate fatigue life, and given maximum stress levels in the material, each applying for the range of materials normally available. The more general machine design texts, however, do not usually consider such factors as the effect of temperature on the spring rate or the need to reduce the inherent mechanical hysteresis effects – two factors that are vital to instrument design.

Figures 1 and 2 show several of the many shapes and forms that fine mechanism elastic spring elements may take.

Helical torsion springs (Figure 1c) are formed into shapes that are similar to the coil spring, but the forces are applied to the coil as a torsion, rather than as longitudinal extension or compression. In this use, the basic mode of flexure of the spring material is not torsional, as it is in the tension helical spring, but occurs as bending action.

In this use, the turns are required to rotate with respect to each other; practical design has to allow for any friction (the binding effect) that may occur when the coils slide against each other. Such springs are often made with spacing between turns. Another factor to be considered is that as the spring is twisted tighter in torsion the effective radius reduces. Long torsional helical springs may buckle; guides are often employed to support the spring, in which case the added effects of friction must not be forgotten in sensitive system design. Wahl (1963) suggests that buckling is

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

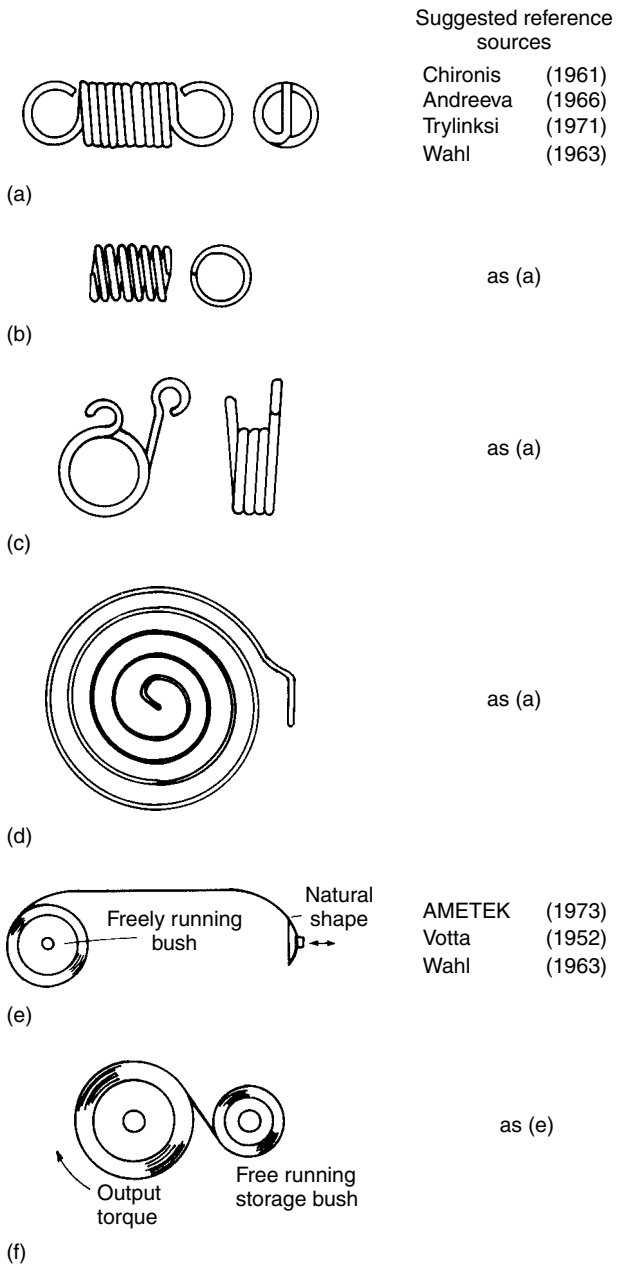


Figure 1. Examples of coiled springs: (a) helical tension; (b) helical compression; (c) helical torsion; (d) spiral torsion; (e) linear Neg'ator™; (f) torque Neg'ator™.

avoided if the spring is not turned at the torsional input by more than 0.87 of a turn for length to diameter ratios that are short, the input rising to 1.25 turns for longer springs having the same diameter. This can only be taken as a guide; references cited by Wahl should be consulted for more detail.

2 SPIRAL SPRINGS

Spiral springs are those wound springs in which flat material is used (Figure 1d) that is curved in a decreasing radius and where all turns lie in a flat plane. The clock and electromechanical indicator restoring spring are in this group of springs. These are generally referred to as *hairsprings*; when the turns do not come into contact and are particularly fine in dimension. Hairsprings can also buckle at high deflections.

It is in this class of springs that the Neg'ator spring design fits. This preformed spring was invented and reported in 1952 by Votta (1952). It has the interesting and somewhat hard to visualize characteristic that it can provide linear tensioning action, exhibiting constant force with displacement (as opposed to most springs in which tension increases with displacement). By choosing a suitable bush size to support it (Figure 1e), it is possible to tailor the design to provide very slight negative or positive spring rates on the basis of unusually high base loads. It is particularly useful as a means to counterbalance gravity-acting masses in instruments. Arranged as a torque producing device, Figure 1(f), instead of giving purely linear action, it forms a constant-torque drive motor with a large range of output turns. It is marketed as the Tens'ator™ spring. Typical commercial design data is available in AMETEK (1973).The dynamic performance of this spring has been analyzed as a dynamic motor in Sydenham (1969), where the commercially offered width to thickness ratios and suggested spool diameters that lie far from the optimum values needed to obtain minimum rotational inertia are shown. Given the optimum spring material thickness and width, the use of small diameter spools in such motors can provide very responsive torque sources; energy stored can be dangerous while in use, and protection is needed to prevent damage, should a spring break or come away from its end constraints.

The design of spiral springs is covered in Andreeva (1966), Geary (1960), Trylinski (1971), and Wahl (1963). Each of these sources covers some of the same material, but additionally it provides a somewhat different design information. It is necessary to select the appropriate work for the task in hand. None of them appears to give a coverage that caters for all likely needs; there are just too many spring options in existence.

Trylinski (1971) concentrates on fine mechanism springs, that is, those including hairsprings and main springs for instruments such as clocks and watches. Geary (1960) is useful for its general introduction and for the bibliography provided. It provides the theory of torsional systems using bifilar and multifilar suspensions. Wahl (1963) and Andreeva (1966) cover in general all types of springs.

Several highly sensitive scientific instruments make use of a spiral form of spring, in which the turns are not in contact and lie in a helix. Figure 2(a) gives an arrangement used in one gravity meter to suspend the proof mass. Figure 2(b) is that of a somewhat similar use, but where

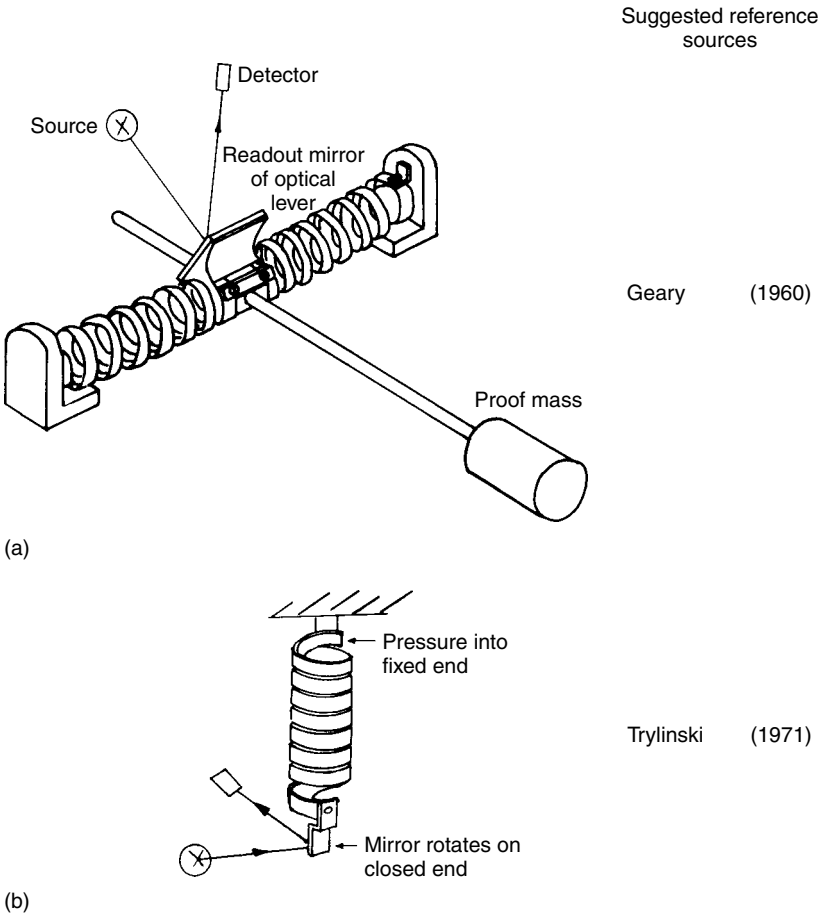


Figure 2. Examples of coiled spring: (a) helical torsion in gravimeter; (b) helical Bourdon tube.

the form of a spiral quartz Bourdon tube is used for ambient pressure measurement. Reference sources are given in the figures.

3 FLAT-STRIP SPRINGS

Flat strips used in bending and torsional modes of basic flexure, as with previously treated springs, can also be used as energy storage elements or as supportive mechanisms. In many applications, they serve both purposes together.

With so many options reported, it appears most expedient, in this brief introductory review, to present schematic illustrations of a selection of options describing some of them in detail to bring out certain design features. Figure 3 provides various views of flat-strip springs used to provide energy and support. Several applications use the flat strip in a coiled fashion; it is not possible to entirely separate spring forms into unique groups.

The first type shown in Figure 3(a) is the use of a strip in which one end is secured so that rotation at the fastening is prevented. Deflection progressively occurs along the strip, increasing in amplitude as the force moves outward. The force-deflection characteristic depends upon the shape and thickness of the strip. It is nonlinear for a parallel, uniform strip. To linearize the spring rate, it can be tapered in thickness and width. Leaf springs used in vehicle suspensions are actually a very tapered cantilever spring that is cut into sections and stacked in a pile. Their sliding contact also provides damping. Early expensive vehicles further tailored the characteristic by tapering the spring material in thickness as well as in width. The design of these is straightforward. The strip can be regarded as a beam with one end (or two ends if needed) fastened rigidly. Another option for altering the rate characteristic is to let the step progressively lean onto a shaped support as it deforms (Beggs, 1955). Figure 3(b) shows the strip used in the same way as a beam with both ends supported. In this option, the ends can be simply supported so that rotation is allowed at the ends. Deflections are not linear with load unless the beam is suitably proportioned in thickness by the means mentioned above.

When cantilever springs are used as small rotation locating mechanisms, the effective pivot point, called the

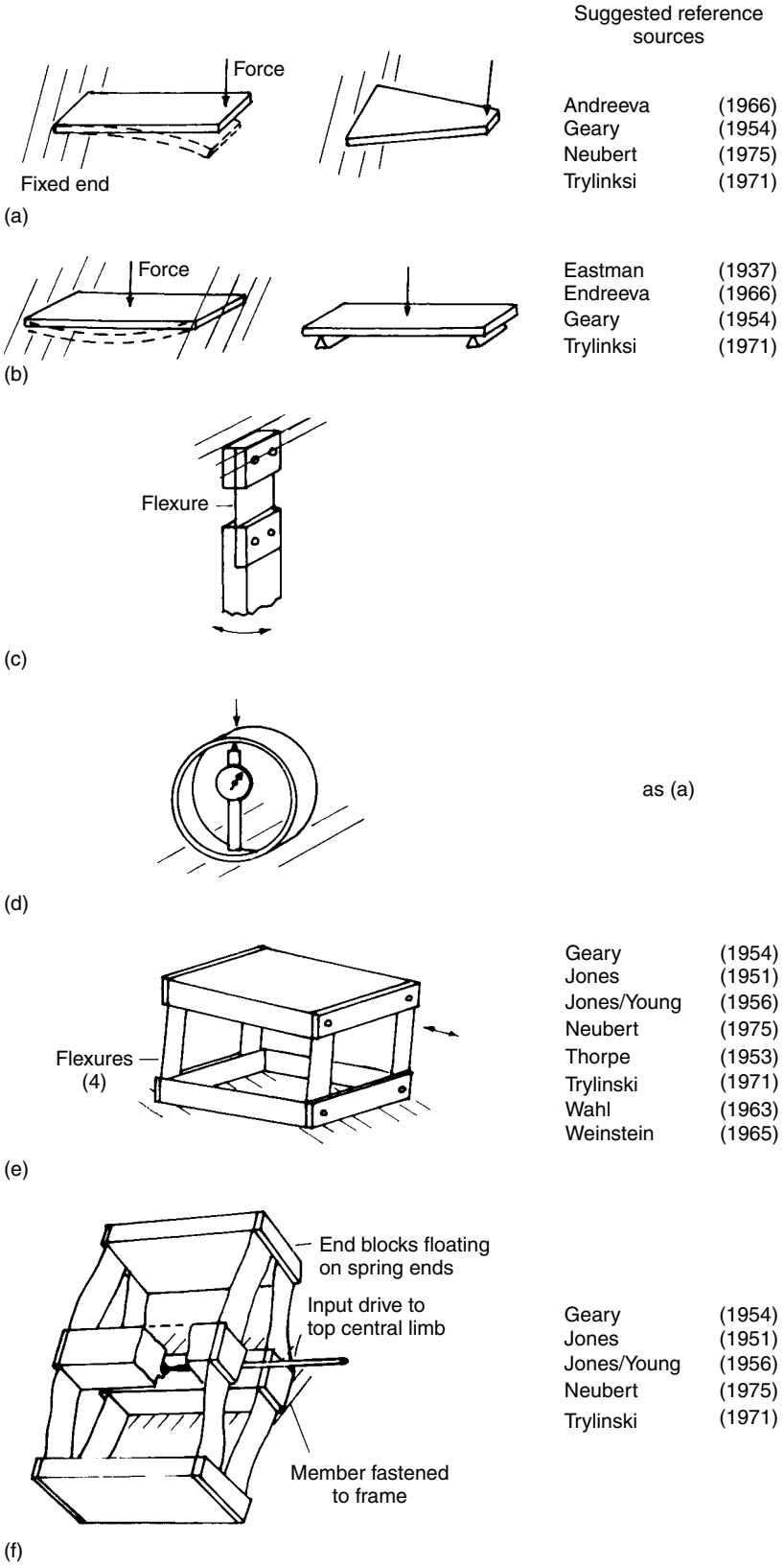


Figure 3. Examples of flat-strip springs: (a) cantilever; (b) beam; (c) single flexure pivot; (d) load cell; (e) parallel-strip slide (uncompensated); (f) parallel strip (one form of compensation).

instantaneous center of rotation, moves in a curved path. Cantilever flexures can also be used with the end of the beam being placed in torsion. A tensioned beam used this way with long length and narrow width is called a *taut band suspension*.

Figure 3(c) shows a strip that bends while supporting a load. As has already been mentioned, the bending stiffness increases as the load supported is increased. Many expressions given in texts do not allow for this. The pivot action provides an instantaneous center of rotation that moves in a circular path, as does a cantilever support. Geary (1954) should be consulted for reference sources on this design. Trylinski (1971) and Eastman (1937) are also relevant.

Proof rings, which form the force to displacement conversion element of a load-cell type, are formed of shaped rings, tubes, rods, and more, to which are attached a displacement sensor. The sensor might be a simple mechanical indicator or an electrical output device. In either case, the design aims to obtain linear or other specific conversion characteristics by suitable choice of mechanical force load-cell element. Figure 3(d) shows a circular proof ring formed from a thick strip by machining it from a solid or by forming a flat plate.

4 MANAGING THE DEGREES OF FREEDOM IN ELASTIC SYSTEMS

Displacement of a table in one axis, while the other five degrees of freedom are substantially held in constraint, can be obtained by use of a parallel-strip suspension shown in Figure 3(e). These take several forms of which the simplest variation is shown. Study of a given four-strip suspension will reveal that, as the amplitude of displacement rises, the table being supported does not follow a perfect linear translation but traces out a curved path in which it rotates and rises and falls. These effects reduce as the length of the strips increase with respect to the amplitude of movement needed.

Introduction of a symmetrical but oppositely applied suspension, Figure 3(f), produces a differential error compensating system that can provide almost pure translation. These are called *eight-leaf* (or *strip*) *parallel strip guides*. Jones (1968), Trylinski (1971) and Neubert (1975) show alternative physical arrangements of the compensated method. Jones (1951) and Jones and Young (1956) are basic reading about parallel-strip suspensions.

It will be apparent that the suspension produces a reacting force as the table is displaced from the central rest position. The force to deflection ratio is constant over small ranges of travel for both the four- and the eight-strip suspensions. The ratio follows cubic relationships for thickness and spring length and is linear with width of spring material and elasticity modulus. Designs sometimes use two adjacent strips that are merged across into one strip. If the suspension has to bear a downward acting load, the flexures can be stiffened with plates (to avoid Euler's buckling) added to the central length of the strip (Geary, 1954).

Parallel-strip suspensions will deflect according to the inverse of their stiffness values. Thus, a series connection of a stiff unit driven by a more compliant unit acts to reduce motion in a positive, accurate, and stable dividing manner. Figure 4(a) shows the layout of strip flexures used to obtain displacement reduction of 1000:1, with negligible friction and force requirements. It was designed for displacing an optical grating to a discrimination level of 0.08 μm, (Sim, Ross and Taylor, 1971). This application is typical of many in which ingenuity is applied to obtain very effective mechanical mechanisms using flexures.

Having replaced the kinematic slide with a flexure equivalent, at least for small amplitudes, it may seem obvious that there might also exist a limited action replacement for the rotating bearing. This does exist and is known as the *crossed-strip pivot* (there are numerous name variations with 'crossed' used in each). Figure 4(b) shows the four-strip hinge. This may take on the alternatives of three strips when two are combined inside the center of the other two, or a two-strip form in which they do not cross within their physical length, doing so at the theoretical extension of the strips (see Geary, 1954).

The springs are generally arranged to cross at 90°, but this is not vital. Considerable design options are available with these. The instantaneous center of rotation moves such that the support can provide positive restoring or collapsing moments (negative restoring force) when excursion reaches given points.

Added to this feature is the possibility of preforming the springs, altering the crossing point from the center to higher or lower positions between the two surfaces, and by forcing the springs into nonflat shapes by their clamping end conditions. Haringx (1949), Geary (1954), and Trylinski (1971) give in-depth treatment of the crossed-strip pivot. It would appear that only one company has marketed off-the-shelf crossed-spring assemblies as OEM (Original Equipment Manufacturer) products, see Bendix Corp (1975). It is necessary to design and manufacture them to order.

5 BUCKLING EFFECTS IN FLEXURE ELEMENTS

So far it has been suggested that care is needed to prevent buckling through the application of a load greater than

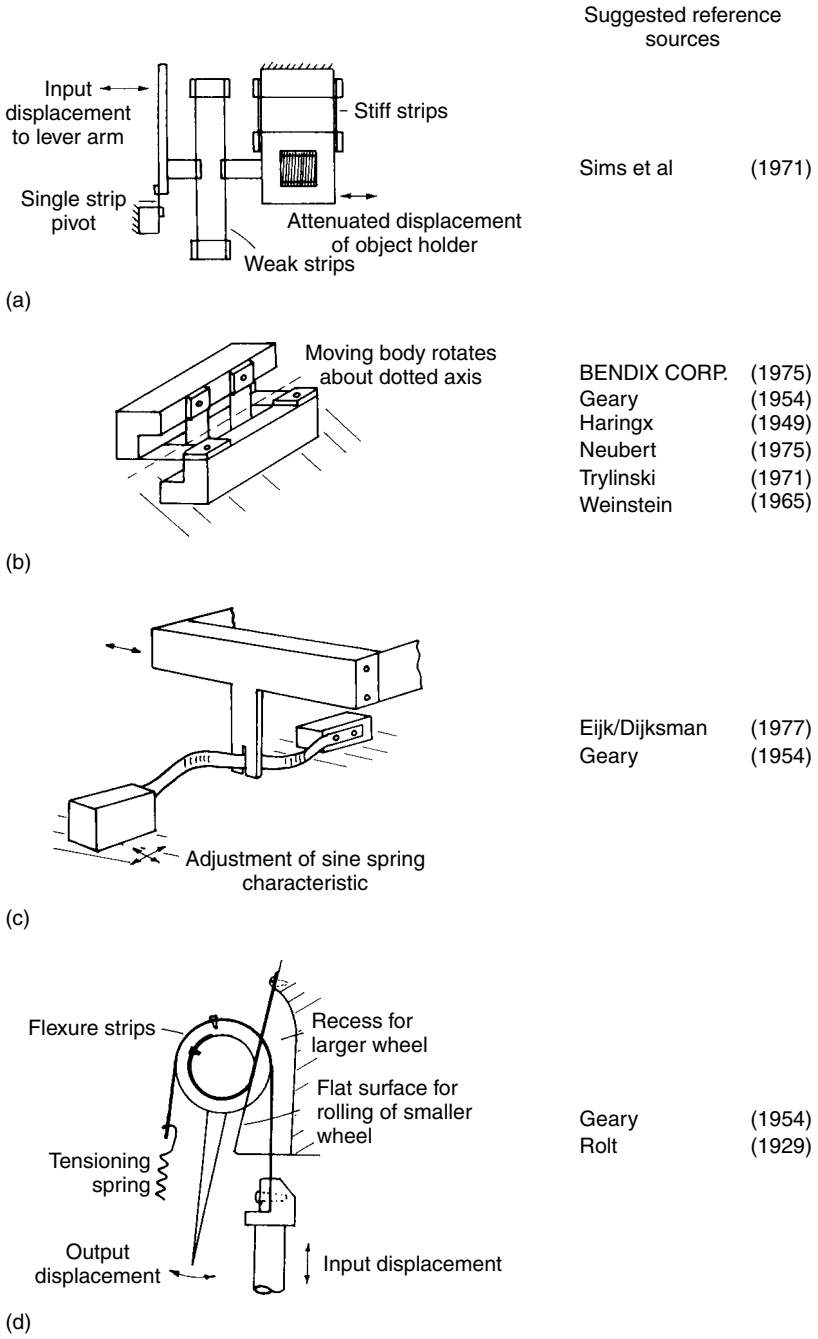


Figure 4. Examples of flat-strip springs: (a) displacement attenuator using strip suspension; (b) cross-strip pivot; (c) sine spring (used here to reduce the restoring force of parallel-strip suspension); (d) rolling suspension.

the Euler buckling limit. This effect, however, can be put to good use to form yet another spring element that can support considerable loads with very small spring rates (including negative) without incurring the penalty of a large initial deflection that occurs with a simple helical spring. Figure 4(c) shows how Eijk and Dijksman (1977) used the buckled-beam spring strip (also called a *sine spring*) to provide a negative restoring force to cancel the force provided by a parallel-strip support; Geary (1954) also discusses these.

Geary (1954) describes a buckled strip arrangement used by the staff of the National Physical Laboratory. Their application made use of the feature that a buckled beam provides a rotating surface on which a mirror can be mounted. As the end of the beam is buckled further by translation, the mirror rotates to form an optical lever.

6 SPECIAL SHAPES OF FLEXURE ELEMENTS

A strip can also be employed to form a rolling suspension. Figure 4(d) shows an example adapted from Rolt (1929) (see vol. 1, p. 333 for greater detail). The strips here act to hold the roller onto a flat surface while allowing the system one rotational degree of freedom. One spring strip must be tensioned (by a spring or a gravity mass) to keep the system taut. As is illustrated, the linear motion of the input is converted into a rotation that also can be given gain by use of different diameter input and output rollers.

The bifilar suspension has already been mentioned; Figure 5(a) shows a form of this. These take numerous forms in practice, ranging from the two round section wire filaments shown to systems using multiple filaments. As a first approximation, these can be considered to have completely flexible filaments, but a more detailed design would be needed to consider their elasticity. Trylinski (1971) and Geary (1960) discuss their design in detail; numerous scientific instruments have been built using such suspensions.

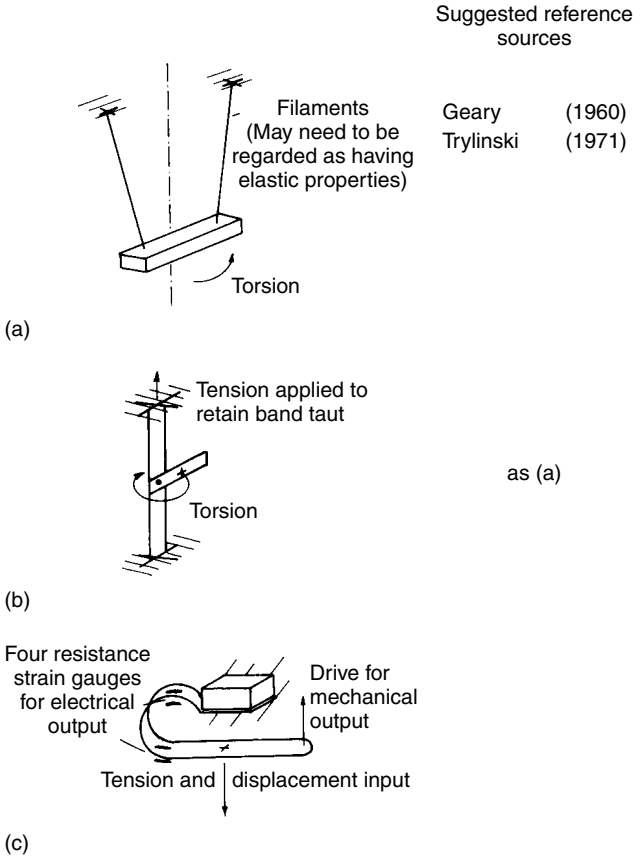


Figure 5. Examples of flat-strip springs: (a) bifilar suspension; (b) taut band torsion; (c) special design used in a relative humidity meter.

A similar torsion action can be obtained using a tensioned taut band Figure 5(b).

7 EXAMPLES OF SPRING FLEXURE ELEMENTS

Finally, in this short review of the innumerable possibilities of spring systems it must be stressed that special shapes can often be devised for special purposes. Figure 5(c) shows one case employed in a low-cost commercial humidity meter to simultaneously provide tensioning force to a stretched hygroscopic membrane and, as the membrane forces the flexure to bend, to provide load-cell action by driving four resistance-strain gauges mounted to its external curved surfaces.

Flexure elements can play a very important part in instrument design and offer an alternative to kinematic design, that is not always the best option to choose.

RELATED ARTICLES

Article 80, Principles of Fine Mechanics – Kinematic and Elastic Designs, Volume 2; **Article 81, Principles of Fine Mechanics – Systems Considerations, Volume 2**; **Article 85, Elastic Regime of Design – Design Principles, Volume 2**; **Article 162, Principles of MEMS, Volume 3**; **Article 164, Principles of MEMS Actuators, Volume 3**.

REFERENCES

Andreeva, L.E. (1966) *Elastic Elements of Instruments*, Israel Program for Scientific Translations Ltd, Israel. Translation from Russian.

AMETEK. (1973) *Design Data for Negator Constant Force Springs*, Hunter Spring Division, Hatfield, PA.

Beggs, J.S. (1955) *Mechanism*, McGraw-Hill, New York.

Bendix Corp. (1975) *Free Flex Flexure Pivot Engineering Data*, Pub No 00U-6-752B, Bendix Corp, Electric and Fluid Power Div, New York.

Chironis, N.P. (1961) *Spring Design and Application*, McGraw-Hill, New York.

Eastman, F.S. (1937) The Design of Flexure Pivots. *Journal of Aeronautical Sciences*, **5**(1), 16–21.

Eijk, J. and Dijksman, J.F. (1977) *Plate Spring Mechanism with Constant Negative Stiffness*, Internal Report, Fine-mechanics Group, Technische Hoschule, Delft.

Faires, V.M. (1955) *Design of Machine Elements*, Macmillan, New York.

Geary, P.J. (1954) *Flexure Devices – Pivots, Movements, Suspensions*, Sira Institute, Chislehurst.

Geary, P.J. (1960) *Torsion Devices*, Sira Institute, Chislehurst.

Haringx, J.A. (1949) The Cross-spring Pivot as a Constructional Element. *Applied Scientific Research*, **A1**, 313–332.

Howell, L.L. (2001) *Compliant Mechanisms*, Wiley, New York.

Jones, R.V. (1951) Parallel and Rectilinear Spring Movements. *Journal of Scientific Instruments*, **28**, 38–41.

Jones, R.V. (1968) More and More about Less and Less. *Proceedings of Royal Institution of Great Britain*, **43**(202), 323–45.

Jones, R.V. and Young, I.R. (1956) Some Parasitic Deflexions in Parallel-spring Mounts. *Journal of Scientific Instruments*, **33**, 11–15.

Neubert, H.K.P. (1975) *Instrument Transducers – An Introduction to their Performance and Design*, Clarendon Press, Oxford.

Rolt, R.H. (1929) *Gauges and Fine Measurements*, Vol. 2, Macmillan, London.

Sim, P.J., Ross, H.A. and Taylor, R.J. (1971), Photographic Gratings for Linear Measurement, *Harold Armstrong Conference on Production Science in Industry*, August, Institution of Engineers, Melbourne, (pp. 57–68).

Sydenham, P.H. (1969) *Manual and Automatic Large-scale Dimensional Metrology*, PhD Thesis, University of Warwick.

Trylinski, W. (1971) *Fine Mechanisms and Precision Instruments – Principles of Design*, Pergamon Press, Oxford. Translated by A. Voellnagel from Polish original.

Votta, F.A. Jr. (1952) The Theory and Design of Long-deflection Constant Force Spring Elements, *Transactions of ASME*, **74**, 439–450.

Wahl, A.M. (1963) *Mechanical Springs*, McGraw-Hill, New York.

FURTHER READING

Thorpe, A.G. II. (1953) Flexure Pivots – Design, *Product Engineering*, **24**(2), 192–200.

Weinstein, W.D. (1965) Flexure pivot bearings, Machine Design, Pt 1, June 151-7, Pt 2 July, 136–145.

87: Elastic Regime of Design – Plates and Bimorphs

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Flat-plate Flexure Elements 556
2 Pressure-bellows Diaphragms 557
3 Capsules for Pressure Sensing 557
4 Nonmechanical Elastic Systems 558
5 Bimorph Flexure Systems 559
6 Multimorph Flexures 560
Related Articles 560
References 560

1 FLAT-PLATE FLEXURE ELEMENTS

Relatively thin plates can be used to provide precision guidance of the center point of the plate; the plate can also simultaneously act to contain pressure of a fluid.

The simplest arrangement (Figure 1(a)) uses a flat, thin, flexible sheet of material that is fastened around the periphery either as a fixed or a simply supported edge in which edge rotation is allowed. In the center, the member to be guided is fastened. As these provide a rapidly rising restoring force with deflection, they are generally only applicable to guide very small motions. Trylinski (1971) provides expressions that show that those provide a linear force–displacement relationship for small movements.

Two basic methods may be employed to obtain a linear force-displacement characteristic over a larger range of movement with a reduced spring rate; that is, those that are more easily deflected.

The first method is to slit the membrane in some appropriate way so that it can deflect further. Figure 1(b) shows several possible shapes. As a very approximate guide to estimating the compliance, designs can often be regarded as providing the net effect, summed in the center, of one or more cantilever beams. The spiral case shown can be mentally unwound to form a long thin beam supported only at one end. In some designs, the plate is not supported around the periphery but at points within the plate. There appears to be no published analytical design information on their design. Trylinski (1971) and Neubert (1975) devote space to these guides but do not provide analytical expressions for other than the simple continuous plate with point or distributed loading. Gibbs (1957) appears to be one of the few reports specifically devoted to the subject, but it contains little more than is given in Trylinski (1971).

Thin plates used for this purpose are called *slit membrane guides* or *diaphragm spring guides*. They are used extensively in such applications as loudspeaker moving-coil supports, geophone suspensions, vibrator suspensions, and anywhere that relatively small amplitude, frictionless translation is desired. By careful design, they can operate for millions of cycles at high frequency, providing high stiffness across the diaphragm and low stiffness in its plane.

They do not appear to be marketed as a stock item and like most flexures have to be manufactured as needed. Slits are formed by photomechanical etching methods, by spark machining, or by plain stamping where quality does not need to be very high and the production run is large.

Slits may well provide greater compliance and range but they remove the property of the plate to retain fluid pressure. In cases in which guidance alone is needed, this may be an advantage to prevent air pressure pumping. Some applications may deliberately use a closed membrane to exclude moisture and dirt.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

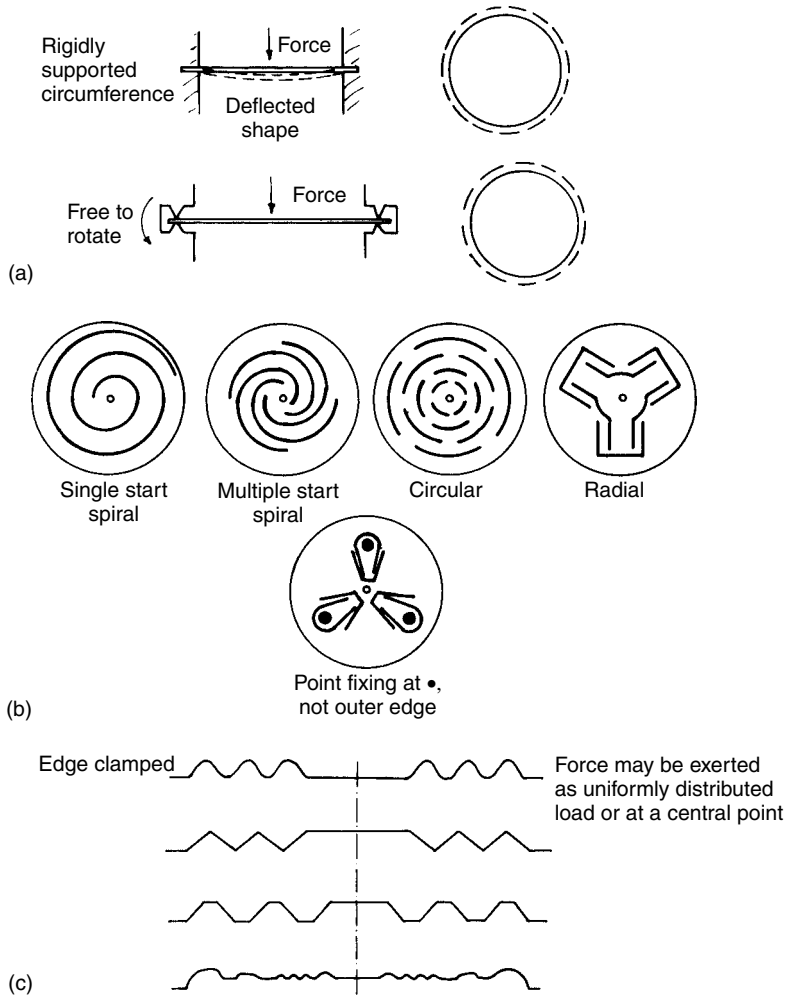


Figure 1. Flexure membranes: (a) solid diaphragms, (b) slit diaphragms, (c) corrugated plates and capsules.

2 PRESSURE-BELLOWS DIAPHRAGMS

If pressure-tightness is essential, as it is in pressure gauges, some other method of making the diaphragm more elastic is needed. It can be obtained by pressing corrugations into the plate. Several commonly used cross-sectional shapes are shown in Figure 1(c).

Sealed diaphragms are also called *membranes*; Trylinski (1971) calls them *pressure controlled elastic elements*.

Presumably, because sealed diaphragms are used so extensively in pressure sensors, a product line that forms a large part of process instrument sales, there has arisen the extensive need of pursuing their design in considerable depth. Design is reviewed in analytical terms in Trylinski (1971). CDA (1953) provides practical detail of application as do Beggs (1955) and BELL and HOWELL (1974). Computer-aided procedures developed in the Department of Applied Physics of The City University, London have made the design procedures of compliant plates considerably simpler to implement; work previously needing much actual prototype manufacture and testing has been bypassed by the use of rigorous mathematical modeling of the profiles needed.

3 CAPSULES FOR PRESSURE SENSING

To obtain greater deflection for a given force, the diaphragms may be joined to form capsules and stacks of capsules, as shown in Figure 2(a). Bellows also provide a similar increase in rectilinear flexibility (Figure 2(b)). As the characteristics of the materials that must often be used – because of temperature limits and the required high strength levels – do not always have the desirable low thermal coefficient of elasticity, pressure systems often need some form of thermal compensation built into the stack assembly. Aneroid barometer bellows and capsules are sealed with internal pressure; the choice of fluid, pressure, and quantity

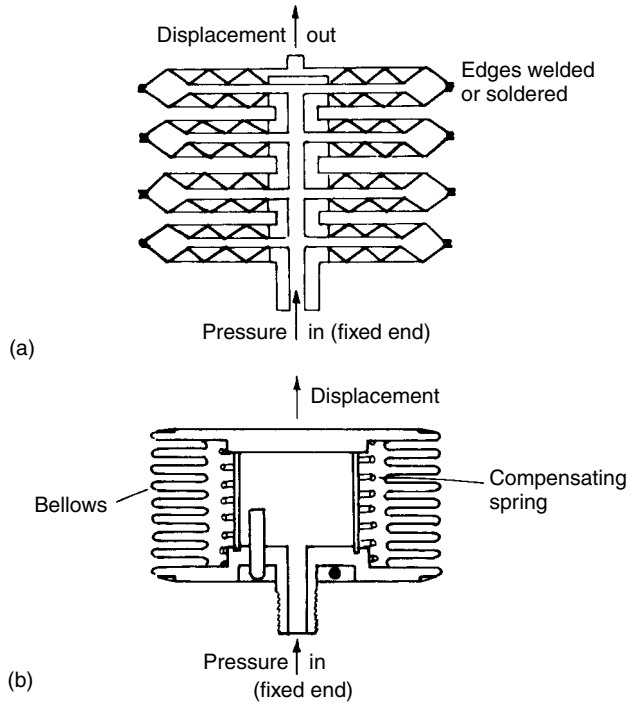


Figure 2. Flexure membranes: (a) capsule stack and (b) bellows.

can be made to assist compensation; internal helical springs are sometimes used.

In pressure sensors, diaphragm deflection is transduced by direct linkage to a mechanical device or a sensor might be actually attached to the membrane. In the latter case, membrane design must allow for this.

Although not made from plates, the subject of this section, it is appropriate to include mention of Bourdon tubes. These are flexible tubes formed such that pressure differences between the inside and the outside cause the tube to deflect. Trylinski (1971) provides design formulae for these. Deflection is harnessed to drive a pointer or an appropriate electric sensor.

4 NONMECHANICAL ELASTIC SYSTEMS

Electric and magnetic fields have been used to provide variable spring rates and so have columns of air. Fluids that are easily compressed are more suitable (where springiness is needed) than the virtually incompressible fluids such as water and oil. Thus, an oil hydraulic bearing is very stiff, having little compliance, while an air bearing will be softer.

Magnetic suspensions are reasonably common in instruments. They can be formed from permanent magnets arranged in a suitable manner using additional electromagnetic coils so that spring rate and stability can be controlled. In either case, it will be clear that the force of attraction rises sharply as the two magnetic poles of unlike kind come closer together. Magnetic suspension systems need some form of stabilizing arrangement to make them statically stable. Earns Haw, in 1839, showed theoretically that the inverse square law for force and distance fields (magnetic and electric fields are such) cannot inherently support a fixed magnitude pole in a stable manner (Geary, 1964). Automatic control is employed to alter the force with distance as a closed-loop system. Further stabilizing mechanisms may be needed to ensure that one pole does not slide away from the desired position. This can be visualized when attempting to sit (levitate) a pole over the top of another pole of like polarity.

An advantage of magnetic springs is that they can be made to act through sealed walls of containers. An example of an application might be as the connection and weighing medium for a sensitive chemical weighing balance. The process of measurement can be achieved by altering the field strength of an electromagnet so that the balance returns to a given datum in the manner depicted in Figure 3. The electric current needed to achieve the balance state is related to the weight of the mass that is supported.

Electric fields also exert forces so they can be used as elastic elements in the design of an instrument. This method has been used in at least one gravimeter to restore the balance position.

Further detail on magnetic and electric suspensions is to be found in Geary (1964). Weighing balance applications are detailed in Gast (1968), Gast (1974), and Wolsky and Zdanvk (1980).

As the name implies, an air spring is a compliant movement obtained by the use of air, contained in some way, to support a given load. The automotive pneumatic tire is the most obvious example of air spring being used in that application to assist isolation of vibration.

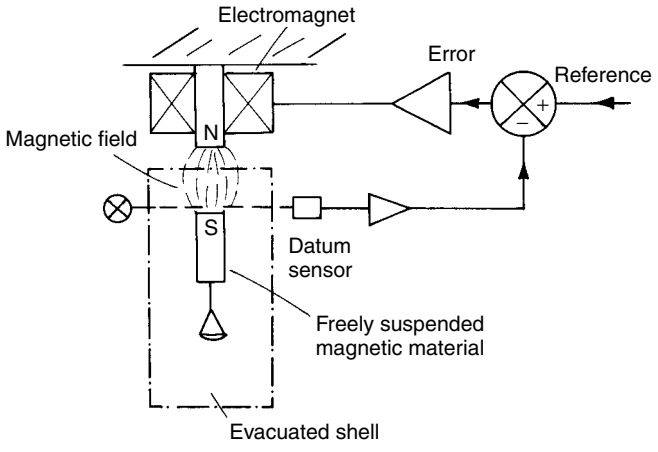


Figure 3. Schematic of force-balance, magnetic suspension weighing technique.

Air springs are capable of substantial load bearing without large deflections, and they can easily provide low natural frequency, second-order suspension systems. They have been used extensively in vibration isolation of machinery. The spring usually takes the form of a round flat flexible bag that is pumped up to a suitable pressure.

Air springs have no resonant effects within the spring itself (as do mechanical coil springs). Their rate can be made to stiffen with increasing load, being nonlinear devices. This means that second-order systems formed with them remain at substantially the same natural frequency with changing load. They are, however, affected by ambient temperature, and they require a relatively large amount of maintenance work. They are generally regarded as being suitable for only heavy load situations. The reader should consult Jones (1967) and Dorien-Brown (1971) for further references and an introduction to this form of elastic system.

5 BIMORPH FLEXURE SYSTEMS

If two materials (Figure 4(a)) having different thermal expansion rates are placed side by side with one end of the parallel combination fixed securely, temperature changes to the element so formed will bend traversely due to differential expansion between the two elements. Such elements are called *bimetallic devices* if made from metals. A similar device can be formed (Figure 4(b)) using two pieces of piezoceramic that are fastened together and have deposited electrodes placed along their outer edges and the central interface. These are termed *bimorph-benders*, or twisters.

Bimetals bend with temperature, converting temperature change into displacement. Bimorphs convert electrical energy into displacement. Bimorphs involve a reversible process – that is, deflection will generate electricity – but bimetal elements are irreversible.

Bimetals are made by soldering or welding thin plates of suitable material together, the final shape being stamped as strips, disks, or whatever shape is needed. The side having the higher thermal coefficient of expansion is called the *active side*; the reverse is known as the *passive side*. As no material has an absolutely zero expansion characteristic and metals generally possess only positive coefficients, the net effect is always that of a combination of two positive coefficient expansions. The forces of this expansion are considerable, and bimetal elements can often directly drive an electrical contact or dial indicator.

To obtain the greatest deflection effect, the metals used need to be, for the passive side, one with the lowest practical coefficient (which is usually Invar), and on the active side, one with a high coefficient (stainless steels, brasses, and special alloys).

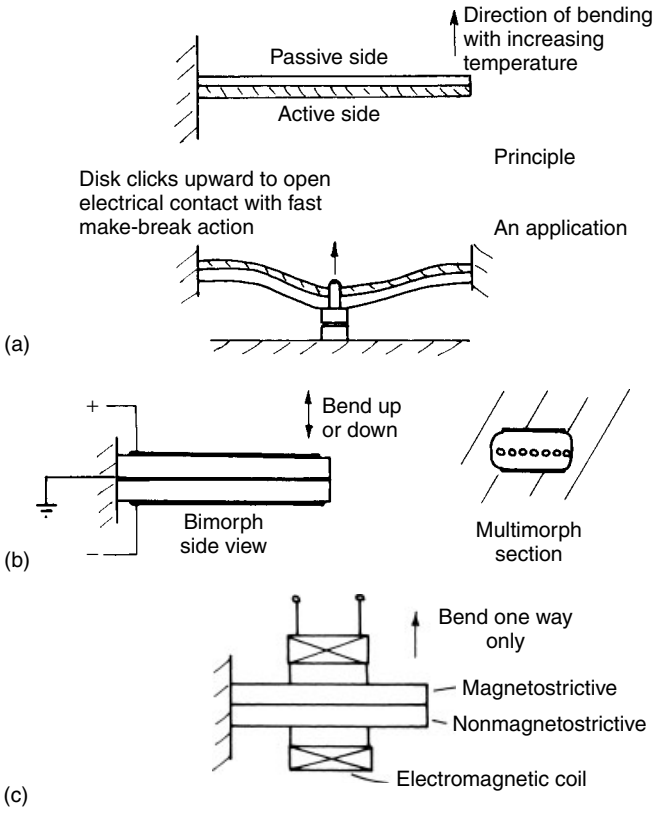


Figure 4. Differential expansion bending and twisting devices: (a) bimetals, (b) ceramic bimorph and multimorph, (c) magnetostrictive bimorph.

Traverse deflection can be converted into rotation through the use of spiral spring forms. These also can be used to obtain greater deflection. However, as the deflection per unit thermal change parameter is increased, it will usually be found that the stiffness of the bimetal element reduces, making it less capable of driving a load without distortion. Preloading the bimetal element against a stop (prop) provides a means of adjusting the datum at the temperature position at which the element begins to respond.

The design of bimetal elements is covered in detail in Timoshenko and MacCullough (1959) and in Trylinski (1971).

Bimorph devices are less common than bimetal elements and information about them is a little harder to locate. Neubert (1975) gives a short account of how piezoelectric sandwiches can be made to provide bending, twisting, and beam deflection action, as well as bending a plate across the diagonal or lifting the free corner of a three-corner mounted plate.

Design information is usually available from the suppliers of the materials, Brush-Cleveite being one source of application notes.

6 MULTIMORPH FLEXURES

A variation of the bimorph is the multimorph. These can provide greater elasticity and are said to be easier to manufacture. Instead of sandwiching two strips together, a single strip is made with a row of fine holes formed through the centerline. These are filled with conductive material to form the inner electrode. The outer faces are coated with electrode material in much the same manner as for the true two-slice sandwich bimorph.

Other forms of differential bending devices could be formed on this principle but they do not appear to have been used. For example, a nickel strip sandwiched (Figure 4(c)) to a stainless steel side would bend in the same direction with an increasing magnetic field of either polarity; the nickel, but not the stainless steel, is magnetostrictive. This is the magnetic equivalent of the ceramic bimorph. It would also give reversible characteristics.

Bimetal and bimorph devices can be used to provide compensation for the physical attribute that suits their input energy form. Bimetal elements are often used, for example, to correct diaphragm-type pressure gauges. They are also used to trip circuits when the ambient temperature produced by the direct true ambient or by the use of a heating coil or direct heating causes the strip to deflect. **Article 88, Error Sources in Fine Mechanics, Volume 2** on error sources and their reduction, provides more information about the use of thermal differential methods.

The dynamic response of such differential devices may be important. Bimetal devices are relatively slow to act since they must transfer thermal energy to obtain deflection. They will be substantially first-order systems in response because the thermal storage and dissipation rate combination is much slower than the rate at which the strain energy transfers. Clearly, the smaller the element the faster the response, and physical size limits would generally provide these elements with time constants in the order of seconds. Miniature thick and thin film devices are able to respond much faster but at the expense of less power drive capability.

Bimorphs are inherently high frequency devices. As the rate of energy change for the electrical and the strain energies transfer is of the same order, these will provide second-order response with high natural frequencies in the kilohertz and higher region; low damping is a feature. They are not, of course, a replacement for the bimetal since the two devices operate with different input energy forms.

RELATED ARTICLES

Article 57, Static Considerations of General Instrumentation, Volume 1; **Article 58, Description of Accuracy, Linearity, and Drift, Volume 1**; **Article 59, Introduction to the Dynamic Regime of Measurement Systems, Volume 1**; **Article 80, Principles of Fine Mechanics – Kinematic and Elastic Designs, Volume 2**; **Article 86, Elastic Regime of Design – Spring Systems, Volume 2**

REFERENCES

Beggs, J.S. (1955) *Mechanism*, McGraw-Hill, New York.

BELL and HOWELL. (1974) *The Bell and Howell Pressure Transducer Handbook*, Bell and Howell, CEC/Instruments Div, Pasadena, CA.

CDA. (1953) *Copper in Instrumentation*, Pub No 48, Copper Development Association, Radlett.

Dorien-Brown, B. (1971) Air Springs for Vibration Isolation, *Harold Armstrong Conference on Production Science in Industry*, August, Institution of Engineers, Melbourne, pp. 213–225.

Gast, T. (1968) Conversion of Measuring Values for Mass, Density and Flow into Proportional Frequencies with the Aid of Free Magnetic Suspension, *Industrial Measurement Techniques for On-line Computers*, In IEE Conf Pub 43, June, Institution of Electrical Engineers, IEE, London.

Gast, T. (1974) Vacuum Micro Balances, their Construction and Characteristics. *Journal of Physics E: Scientific Instruments*, **7**, 865–875.

Geary, P.J. (1964) *Magnetic and Electric Suspensions*, Sira Institute, Chislehurst.

Gibbs, D.F. (1957) Spring Diaphragms. *Journal of Scientific Instruments*, **34**, 34–35.

Jones, R.V. (1967) The Measurement and Control of Small Displacements. *Physics Bulletin*, **18**, 325–326.

Neubert, H.K.P. (1975) *Instrument Transducers – An Introduction to their Performance and Design*, Clarendon Press, Oxford.

Timoshenko, S. and MacCullough, G.H. (1959) *Elements of Strength of Materials*, Van Nostrand, NJ. (Numerous editions since 1935).

Trylinski, W. (1971) *Fine Mechanisms and Precision Instruments – Principles of Design*, Pergamon Press, Oxford. Translated by A. Voellnagel from Polish original.

Wolsky, S.P. and Zdanvk, E.J. (1980) *Ultra-microweight Determination in Controlled Environments*, Wiley, New York.

88: Error Sources in Fine Mechanics

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Fundamental Limits in Fine Mechanical Design 561
2 Creep and Hysteresis in Mechanical Mechanisms 562
3 Geometrical Error Factors in Fine Mechanism 562
4 Sine and Cosine Errors, and Centering Error in Fine Mechanism 563
5 Influence Parameter Errors in Fine Mechanism 564
6 Temperature Error in Mechanical Systems 565
7 Reducing Thermal Errors in Fine Mechanisms 565
8 Thermal Compensation Methods in Fine Mechanism 566
9 Temperature Error in Elastic Elements 567
10 Miscellaneous Errors in Fine Mechanism 568
Related Articles 568
References 568

1 FUNDAMENTAL LIMITS IN FINE MECHANICAL DESIGN

Having covered the overall design issues relevant to the operational requirements of fine mechanical systems in **Article 80, Principles of Fine Mechanics – Kinematic and Elastic Designs, Volume 2**; **Article 81, Principles of Fine Mechanics – Systems Considerations, Volume 2**; **Article 82, Kinematical Regime – Members and Linkages, Volume 2**; **Article 83, Kinematical Regime – Fasteners, Bearings, Volume 2**; **Article 84, Kinematical Regime – Rotary Motion, Volume 2**; **Article 85, Elastic Regime of Design – Design Principles, Volume 2**; **Article 86, Elastic Regime of Design – Spring Systems, Volume 2** and **Article 87, Elastic Regime of Design – Plates and Bimorphs, Volume 2** it is now appropriate to discuss the error regime of designs.

In any instrument design, there exist certain design factors that cannot be eliminated and must be worked within. All physical processes inherently contain physical phenomena that provide a limit to sensitivity or stability, which ever is the issue. In the majority of mechanical designs, this is not likely to be of great significance. Research material on the detailed nature of fine mechanical errors reached its highs well before the advent of electronic forms of publication. Much of the research material was published before electronic solutions overtook mechanical means from the 1950s. Thus, sources may be hard to locate – but they are often worth finding for they provide a wealth of practical information on less popular topics.

In mechanical systems, the prime fundamental limit to sensing detection is Brownian movement, a random motion of the molecules forming the substance of the parts concerned. The motions of the molecules are averaged by the mass of the part, so larger components exhibit less overall dimensional change due to this source than smaller parts. It has been extensively studied in relation to electrical indicating galvanometers, in the period before the 1950s, and became of interest again in more recent times in the hope of detecting gravity wave radiation. Chapter 17 in Bleaney and Bleaney (1957) provides a basic introduction to the theoretical considerations. Braddick (1966) discusses noise sources in experimental setups. Further references are given in Sydenham (1979) where the importance of this

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

limit in the evolution of electromechanical instruments has been reviewed.

The bibliography given in Logan (1973) contains references to many of the published papers on gravity wave detectors. Jones (1967, 1968) contains useful practical accounts of the limits of measurement sensitivity.

To provide a rather crude estimate of the level of dimensional noise that Brownian motion effects will generate in a mechanical component, some published examples are quoted. These, however, must be accepted only in the full context as given in the quoted reports. Jones (1967, 1970) estimates that the root-mean-square length change of a 10-mm brass block of 1 mm^2 area is likely to be about 10^{-16} m when averaged over a 1 s integration time. This level of change is fine for most requirements, but it will be a limiting issue for those mechanical sensing systems that are taken to the limit of possibility.

As a second example, Usher (1973), a review of seismometer designs contains an estimate that a seismometer with a mass of 10 gm, a natural period of 1 s, a mechanical quality factor, Q, of 100, and operated at room temperature will possess a noise equivalent acceleration of 3×10^{-11} g. He shows that the factor for reducing this in that design is the damping, which should be kept low. The natural frequency of the spring-mass system is not a factor in determining the amplitude of the Brownian motion noise. In these instruments, this noise source must be considered in order to create satisfactory instruments.

Showing up in yet another form, in a sensitive electromagnetic galvanometer, the source there might give rise to noise equivalent currents of around 10^{-12} A. Further examples are to be found in Barnes and Silverman (1934).

2 CREEP AND HYSTERESIS IN MECHANICAL MECHANISMS

Although not strictly a truly fundamental limit, the effects of creep and hysteresis of mechanical materials often act as an effective fundamental limit. There are practical levels to which any amount of better design, external influence parameter control, and material processing cannot better. Despite considerable research effort seeking to quantify the fundamental mechanisms of creep and hysteresis in materials, there does not yet appear to be available an adequate model of the underlying process. Many texts are available that summarize the current state of knowledge. Design to reduce hysteresis and creep is generally based upon experimental data about the material used, its strain history and the experience of the designer.

It is suggested that hysteresis can be reduced by limiting the mechanical stress excursion to the smallest allowable cycle amplitudes and by the use of low hysteresis materials.

Creep is a still more elusive parameter. Clearly releasing locked-in stresses by annealing and other stress relief methods helps. Weathering for prolonged periods is also used where time permits, for it is known that the creep rate generally reduces with time. A part considered to be creep-free, however, may change its creep characteristics quite markedly if remachined or worked in some way. Sydenham (1973) is concerned with the long-term stability of tensioned Invar and fused silica.

The difficulty faced by the instrument designer is that the vast majority of creep and hysteresis data is compiled for highly stressed parts, that is, those operated at high temperatures and large stress excursions, such as are found in aircraft turbine blades. Instrument designers, on the other hand, are generally interested in comparatively low stresses and well controlled, around ambient temperatures. The extreme difference between the two cases makes extrapolation from one zone to the other a very unreliable procedure.

As a guide, it is generally good practice to stress materials (those that must be stable) to no more than a tenth of their elastic limit. Brasses are considerably better than steels in as far as creep is concerned in instrument parts. It is not possible to provide any really worthwhile rule of thumb guides for these two parameters.

Both of these factors are given a very poor account in instrument design texts, very little information is provided on both these factors, presumably because of the lack of general knowledge about them. Trylinski (1971) discusses hysteresis in instruments, but not in numerical terms.

Creep is often finally accepted as a factor that cannot be reduced to the levels really desired. Being largely (an often incorrect assumption) systematic and adequately linear with time, corrections can be made in the measurements using a sensing mechanism that creeps. A prime example is the exploration gravity meter in which an established, return to start, routine is set up with respect to time and position of readings; the drift can be assessed and removed from the data taken in the field.

3 GEOMETRICAL ERROR FACTORS IN FINE MECHANISM

Attention is now turned to these features of a mechanical design that can be controlled in order to improve the performance and accuracy of the measuring apparatus.

A clearly identifiable group are those sources of error that relate to the geometrical features of the mechanical assembly. These are almost all parameters that can be kept in control by careful design of the assembly and through adequate selection of components, materials, and sizes.

Mechanical design of instruments invariably rests upon the assumption that there exist points in the mechanical space framework that are invariant in position with respect to each other. They are often merely drawn as crosshatched lines in schematic diagrams of the system.

In reality, all materials are elastic and change dimension to some degree with force exerted on them. It is, therefore, important to ensure that such elements are, indeed, sufficiently strong, inelastic, and stable.

Stresses occurring in manufacture, testing, transport, and assembly are often far in excess of those experienced by the component in its working state. For example, there is need to finally align a measuring machine in its application, for test bed results cannot be relied upon to be the same. Bolting down stresses can change the geometric shape. Transportation might have imposed distortion of the frame.

A second form of geometric error sources is the dimensional relationship change that will occur because of the actual dimensions of individual parts lying in an error band given by the tolerances to which the parts have been made. Tolerances should be selected for the various surfaces of a component according to the assembled geometry required. There is an art in tolerancing; it comes largely with experience. Whereas, to overspecify may well control the assembly dimensions more rigidly, it makes the instrument far more expensive.

Standards, such as BS 1916 on limits and fits for engineering, provide the starting point to the selection of realistic tolerances for given situations. This aspect is generally included in courses on mechanical engineering. Tolerancing is usually a given feature of mechanical CAD tools.

Related to tolerance error is the gradual loss of accuracy that might occur if a wear-true philosophy has not been adopted. Wear in pivot points and along slides will change the effective geometry of the instrument, thereby possibly degrading the performance.

Such matters, plus other general aspects of design, are gained largely by direct experience. Several authors in the past have endeavored to pass on that experience as descriptive guides to general instrument design. These include Braddick (1966), Glazebrook, Darwin and Mason (1923), Rolt (1929), and Whitehead (1954).

Rather more specific are errors caused by assuming that location points, such as a center or a sliding perpendicular, are provided in a perfect manner by a bearing or a slide, respectively.

3.1 Abbe's principle

Abbe's Principle, although well known among practitioners, does not seem to have been defined in English language texts about instrument design. Basically, it states that a measurement is generally best performed by applying the measuring device to the object in as direct a manner as possible. For example, in the field in which this was first expounded – mechanical dimension measuring equipment – it is preferable to determine the length of a part using a comparison method that introduces as little error due to the comparison process as is practicable. Figure 1 contrasts the use of a scale to measure the length of a part with the method in which a traveling microscope head slides along a scale. Where the scale markings and the interval that must be measured can be placed in actual contact, the reading cannot include any error from the comparison process. (Of course, it could contain error of the scale or from its misreading.) In contrast, the traveling microscope arrangement can introduce error by way of the slide not being straight, causing the microscope to project the scale incorrectly on to the unknown.

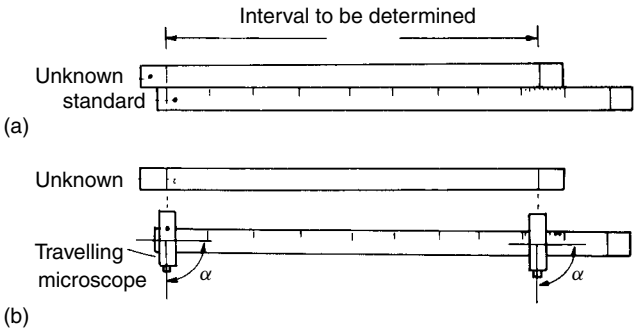


Figure 1. Abbe's principle. In (a), the unknown and the standard are compared in a direct manner but in (b), errors may exist in the projections of the scale by the microscope.

Whereas Abbe's Principle might be the ideal to aim for, it is often not practicable to adhere to it. If not, then other geometrical errors must be considered.

4 SINE AND COSINE ERRORS, AND CENTERING ERROR IN FINE MECHANISM

Angular variations introduced by the mechanism embodied in the traveling microscope example given in Figure 2 need consideration. It depends upon the effect on the length being determined or required to be held constant.

If the same microscope is used at each end of the length and the slide is straight (Figure 2a), any fixed angular error, α, in the mounting of the microscope head will not cause error in the length determination because it is equal at each end. If, however, the microscope mount is perpendicular to the slide at one end and the slide causes the head to rotate

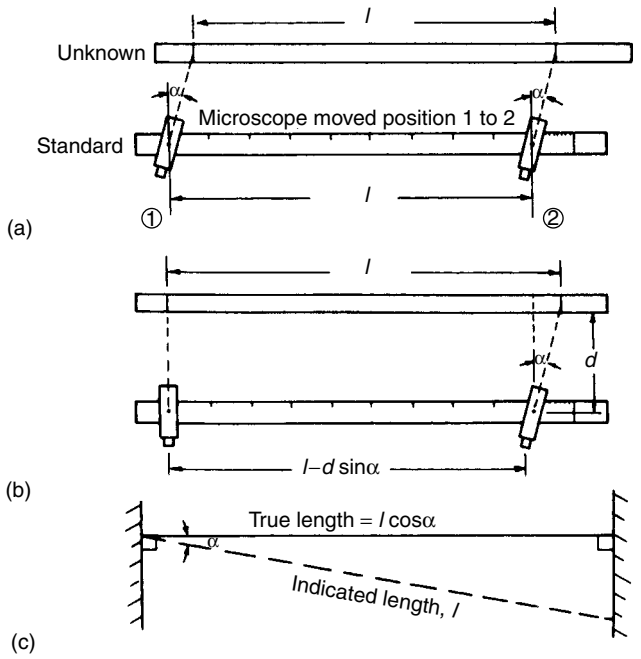


Figure 2. Sine and cosine errors (a) error of microscope mounting not significant, (b) sine error occurs at one end due to rotation of microscope, (c) cosine error.

by angle, α, at the other end (Figure 2b), an error equal to $d\sin\alpha$ will arise. As this error is directly additative, and because sines of small angles are significant quantities in such cases, quite small angular changes can give rise to relatively important error magnitudes.

When measuring a length (Figure 2c) between two end intervals instead, as arises with an inside screw micrometer (the above sine error case occurs in vernier calipers), then a second kind of error source can occur. If the part to be measured is not aligned with the central axis, the length will be $l\cos\alpha$, not l as indicated. This is not as significant as the sine error of the previous case for if the angle α is small, which it usually is, the $l\cos\alpha$ is close to the same as α. Another example where this error arises, is in the setting up of a distance measuring ranger. The actual alignment need not be very accurate, because the error caused by misalignment will not be significant.

Whereas cosine errors are generally easy to control, they must, nevertheless, be calculated to ensure that the approximation holds to the degree of accuracy that is required. Sine errors, similarly, must be controlled by ensuring that the angles remain within tolerable limits under practical use. As d is reduced, in Figure 2b, it will be seen that Abbe's Principle is approached.

Errors of this form are discussed but only briefly under the principle of alignment in Rolt (1929) and in Braddick (1966).

4.1 Centering error

Where the measurement of location depends upon a bearing around which motion takes place, there may arise errors of centering. It is common knowledge that such errors can be reduced by reading the scale on opposite sides, taking the average of these as the correct value. Whitehead (1954) devotes space to this problem, pointing out that it is not always necessary to use the best possible bearing tolerances. He provided an analytical study of the magnitude of errors. Many instruments are made to be read at least at diametrically opposite places. Indeed, some arrange more than this number of averaging processes to be carried out in provision of the output value. Whitehead, however, only considered the use of a centering bearing for a radial scale. In linkages, the error effect must be considered from other viewpoints. Braddick (1966) also provides material on centering errors.

Although the above error sources have been illustrated here in terms of the traditional, manually read instruments, it will be found that these error sources arise frequently in the design of the mechanical features existing in transducers of the modern form.

5 INFLUENCE PARAMETER ERRORS IN FINE MECHANISM

The very best of internal designs can only provide satisfactory operation if variations in external influences either lie within tolerable limits or can be handled by the internal design arrangement.

All systems are influenced by changes in ambient conditions to some extent. It is the relative magnitude of these effects that matters. Temperature, ambient air pressure, relative humidity, magnetic and electric fields, acoustic noise level, mechanical vibration, chemical attack, and more can each possibly give rise to error in operation of the installed instrument.

It is common and established practice to test an instrument toward the end of the design cycle for its tolerance to such effects. Calibration is often accomplished taking into account given levels of vibration of influence parameters. It is, however, often overlooked that the test conditions may not be reproduced in the final use situation, where yet more influence parameters may be in existence.

The designer of an instrument must know which of these influence quantities will exist in use and how to allow for them. There are two approaches. One is to control the internal environment of the instrument by the use of a suitable enclosure; for example, the internal evacuation in a vacuum weighing balance that avoids

errors of buoyancy and provides some degree of thermal isolation. The alternative is to design each element of the system, such that the overall system can tolerate significant enough influence parameter variations. An example here would be to use low thermal expansion materials plus certain mechanical structural arrangements (that will be discussed below) to reduce thermal differential expansion error to within tolerable limits for the expected design temperature variations.

In practice, one or the other, or a combination of these two alternatives is used. Probably the hardest part of the task is probably to properly and adequately decide what influence quantities are relevant and their magnitudes of change.

6 TEMPERATURE ERROR IN MECHANICAL SYSTEMS

Without doubt, the most prevalent influence parameter in the mechanical design of instruments is temperature change. Basic materials with absolutely zero thermal coefficient expansion do not exist. Furthermore, materials with the lowest values are not always suitable for the manufacture of instrument parts. It is, however, possible to produce instruments from the relatively high coefficient materials, which are reasonably temperature insensitive. Before going into the methods that might be used, it is instructive first to examine the general thermal design of a mechanical instrument.

Thermally controlled jacketing and mechanical compensatory mechanisms are often used to overcome temperature error. The value of each, generally, rests on an underlying assumption that the temperature is uniform within the instrument. This is often far from the case. It is not adequately recognized that thermal heat flow is not an instantaneous process. Any change of forcing temperature on some part of an instrument will give rise to changes in temperature of all of its parts at varying rates. If, in exacting designs, the instrument is regarded as being a thermally dynamic rather than static assembly, then better design should result.

If the external temperature does not change, then only internal sources of heat, such as arise from electronic circuitry enclosed in the housing or from friction between moving parts, can cause thermally produced differential dimensional errors. Heat source changes due to adiabatic action on air pressure may be relevant. Design, therefore, should aim to keep internal and external sources of temperature change as small as necessary. They, however, cannot ever be held at an absolute stability. Internal heat transfer mechanisms should be arranged so that temperature uniformity results as speedily as is required. In practice, these simple rules are very difficult to adhere to.

Consider first the problem of thermal uniformity. The materials that are most suited for construction, such as the steels, aluminums, and brasses, do not have very high thermal conductivities. They also possess comparatively high thermal expansion coefficients. Lower coefficient materials generally have still lower thermal conductivities.

Heat is transferred by conduction, convection, and radiation. Knowing the constraints imposed by the materials that have to be used – good thermal conductors are not the best of structural materials – it is possible to arrange for heat transfer to be improved through the use of design layout, shape, cross-sectional size, and similar design possibilities.

7 REDUCING THERMAL ERRORS IN FINE MECHANISMS

As a guide, design for good thermal uniformity will aim to provide good conduction inside the instrument with a high thermal barrier attenuating external changes through the use of some form of integrating enclosure. Enclosure design can also be most sophisticated. Reference will be made to that later.

To obtain good internal uniformity, all parts should be made to maximize conduction through the use of high thermal conductivity materials, large cross sections, good thermally conducting joints, and optimal arrangement to keep conduction paths as short as possible.

Conduction and convection can often be improved by the use of a stirring arrangement such as a fan or stirrer paddle. The internal fluid, in such cases, will be a suitable liquid or gas that will assist heat transfer. The transfer of thermal energy is, however, often decided more by the heat transfer film coefficient existing at the interfaces than it is by the fluid's internal properties.

Sometimes the design aims to prevent convection taking place, thereby aiming to maintain the temperature constant, if not uniform, within the instrument. It is a physical fact that the smaller the convecting cell size, the lower the rate of convection. Below a certain cell size and shape it ceases completely. Because of this, it is often helpful to fill interspaces of an instrument with loose thermal insulation so that the internal convecting spaces are made smaller.

Heat transfer by radiation can be increased by suitable use of surface treatment, blackening being the most used. The actual surface treatment used will be decided by the thermal wavelengths involved. Within the system, parts are blackened and the whole is then surrounded by a thermally reflecting enclosure. The silvered vacuum flask is a good example of this.

It will be appreciated from this account of the possibilities that the thermal design of an instrument can be very complex. Perhaps that is why this aspect of design of instruments has not been adequately researched or organized. Braddick (1966) gives brief mention to the problem. Considerable practical information exists in primary journal literature on temperature controllers, but the basic principles, theory, and design do not appear to have yet been ordered in a didactic manner.

Basic texts on heat transfer, such as Eckert (1950) and Jakob and Hawkins (1957), provide general design principles. Publications relating to cryostat design (White, 1959) and use also delve in these problems. Abachi, Molenat and Malbrunot (1979) provide a useful bibliography of reported designs. Designers of microbalances have given thermal effects considerable attention, Wolsky and Zdanuk (1980) being a key review work. Wolsky *et al*. (1967) discusses how gold plating a quartz balance arm can reduce thermal expansion errors through subtle heat transfer mechanisms. Sporton (1972) is probably the best available paper to consult when beginning a design of an enclosure.

Thermal enclosure design is often needed in the development of the total instrument system. Some applications require the instrument to operate within temperature limits as small as a millikelvin variation. Design aims to form an enclosure in which the walls seen by the contained instrument are uniform in temperature and of the lowest practicable thermal impedance.

Simpler designs generally use an almost point source of heat, which is controlled from a single point on–off temperature sensor. Clearly, this does not come close to the rule of maintaining uniformity throughout and, therefore, only provides crude control. Far better are the designs in which the heater and the sensor are distributed throughout the wall area of the enclosure. Proportional control is generally assumed to provide the best control loop but this may not always be the case.

Parameters of thermal enclosure design are the external variations and frequency to be controlled, the size and shape of the instrument (clearly smaller instruments are easier to control to a given limit than large structures), the limits of control variation required, the internal heat sources, the enclosure design (sandwiched heat conducting and insulating walls), plus judicious use of air spaces have proved very effective (Sporton, 1972).

In many applications, the aim is to integrate the external thermal forcing function to an adequately small internal variation. Thus, it seems an obvious step to use an enclosure that has an adequately long thermal time constant. In practice this can be, with passive methods at least, less effective than might be expected. The thermal time constant (that for a 64% change if the system is assumed to approximate the first-order linear response) for a vacuum flask is around 2 to 3 h; a solid house brick box is similar. When using such methods to average out daily variations, they are barely effective. Daily variation smoothing requires several meters of rock cover.

Active systems, those in which temperature is actively controlled, can be far more effective. Long time constant passive surrounds also suffer in that they must be given time to stabilize before the internal instrument can be used in its intended state. As thermal mass is the dual of thermal insulation, it is possible to construct first-order thermal enclosures from either conducting or insulating material. Each has its advantages. Use of both kinds together raises the order of the enclosure's thermal time response and, therefore, if designed properly, can provide better results.

In the practical construction of thermal controllers, it is far easier to produce heat energy than it is to remove it (compare electric heating elements with Peltier cooling cells, for instance). For this reason, thermal surrounds often are operated at some margin above the ambient temperature so that natural heat loss acts to provide the heat reduction mechanism.

It is the author's experience that a contributing factor to the mechanical instability of materials is the stress brought about by temperature variations of the part. It appears that the more controlled temperature variations are, the better the ultimate stability of creep exhibited. This is further supported by the practice of thermostating gravity meters with a temperature-controlled jacket that is kept energized even when the instrument is not being used.

8 THERMAL COMPENSATION METHODS IN FINE MECHANISM

The alternative method for reduction of thermal mechanical error is the use of compensatory mechanisms.

Most constructional materials possess a positive coefficient of thermal expansion; that is, they expand in length with increase in temperature. Owing to the shortage of suitable, large enough, negative coefficient materials (they do exist – carbon fiber and the silicas), it is not generally possible to counteract the positive expansion with a negative part as shown in Figure 3. This kind of component, however, using two different positive coefficient materials, can provide an adjustable positive coefficient rod by

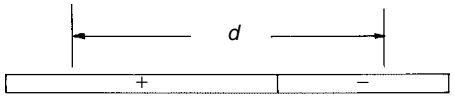


Figure 3. Thermal compensation by cascaded bars.

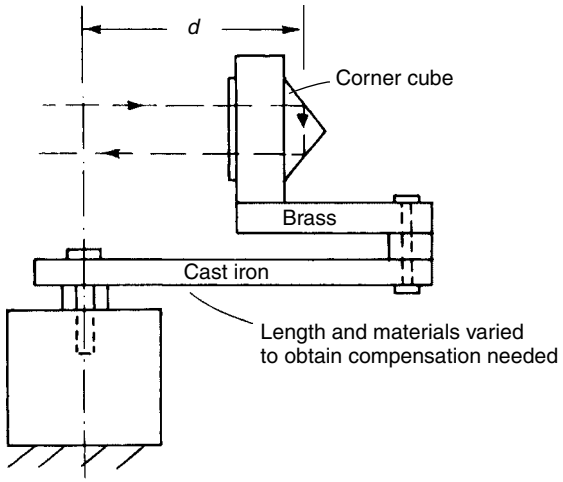


Figure 4. Thermal compensation by reentrant bars.

clamping it with distance, d, spanning the ratio of the two lengths required.

This problem can be completely overcome (within limits) by the use of two positive coefficient materials arranged as a reentrant compound structure. Figure 4 shows the mounting arrangement used to maintain the reference arm distance, d, in a Michelson optical interferometer constant with temperature change to the mount and corner cube. As the cast iron expands outward with temperature rise, the brass expands in the reverse direction at a larger rate, giving a net change of zero to distance, d. Concentric tubes can also be used to form a zero length change bar. Bimetal strips, in which the two materials are joined, can also be used to change dimension in some way through a forcing action.

The choice of materials being limited means that the design of reentrant arrangements must be proportionate to the relative lengths in order to obtain the zero (or other) state using the small range of thermal expansion values that will be found to be usable. In the above mount, the design also allowed for the expansion of the glass of the corner cube.

The assumption that both parts of a reentrant link change (track) temperature together may not hold; during temperature transients the compensation may not operate satisfactorily.

A second method, one that is not used as commonly as the reentrant arrangements (because, it seems, it is not well publicized) is to make use of the stresses set up by temperature change in an element. Baird (1968) used this principle (Figure 5), to compensate a bar of Invar.

In this method, steel wire was wrapped around an Invar rod. It was then soft-soldered to it, so that axial stresses in the steel wire, caused by temperature changes, are coupled into the circumference of the Invar bar. With temperature rise, the steel attempts to form a larger diameter ring around

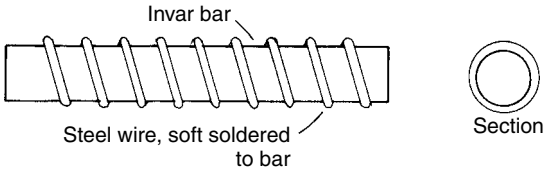


Figure 5. Thermal compensation by counterstressing.

the Invar, causing it to be tensioned outward. This, in turn, via the Poisson ratio shape–volume change effect, causes the Invar bar to reduce in length. The reverse holds true. He provided the theoretical considerations for obtaining the degree of compensation needed for a pair of materials. Clearly, this method can be applied by plating layers or by pressing tubes over the base bar. The thickness and the external material used are the key design factors.

A third method for compensating thermal errors is to measure the temperature change in a part, applying a generated appropriate length change where needed. Many methods exist for the provision of small length changes under electric control. They include piezoelectric crystals, magnetostrictive rods, and thermal heating of rods and electromechanical devices. Measurement of the temperature of a part can be difficult because of transient thermal effects. The control loop, however, may also introduce phase shifts that distort the required behavior.

9 TEMPERATURE ERROR IN ELASTIC ELEMENTS

Temperature changes to elastic elements will give rise to length changes according to the thermal coefficient of length; additionally, however, the elastic modulus will change.

The combined effect is to alter the elastic rate (scale error) of the spring element and also to alter its free length, giving rise to a zero position shift (zero error). If the component is designed to be a flexing spring, then these combined effects can be very serious. Exactly how the spring's combined scale and zero behavior varies with temperature change depends upon the spring's form and geometry. Producing a spring that has negligible temperature error does not occur merely by selecting a zero elasticity modulus material, nor will it occur by using a zero thermal coefficient of expansion material.

It can be shown (see Gitlin, 1955a) that for the common cantilever, helical, and torsional forms of spring the error of spring rate change (scale error) is not a factor of the physical dimensions but it is dependent upon the thermoelastic and linear expansion coefficients and the magnitude of the temperature excursion. Thus, once the material is chosen,

a given level of scale error exists for all designs with that material. Beryllium copper, for example, possesses a scale error of −0.03% per degree Centigrade for an increase in temperature, that is, the spring softens with temperature rise.

Shift in zero with temperature is dependent upon the spring's form. The three articles by Gitlin (1955a, b, c) are key reference accounts about the thermal errors in instruments. Trylinski (1971) includes discussion of this problem.

9.1 Error due to ambient pressure

The mechanical components of an instrument change dimension according to the external pressure exerted upon them by their ambient environment. This has a very small effect, but on occasion it can be significant. It is calculated from the Poisson ratio for the material, which expresses the longitudinal change in length for transverse stress variation. As a guide, metals will lengthen by around 1 part in 10^8 for a typical daily ambient barometric pressure variation.

Error from this source can be reduced by the use of ambient pressure control, which need not necessarily be at the evacuated level as long as the pressure remains constant enough. Alternatively, it can be compensated in some way through the use of some form of pressure-sensitive device. Sealing an instrument inside a container does not necessarily provide constant pressure inside the container; temperature changes to the gas inside it will give rise to significant pressure changes.

10 MISCELLANEOUS ERRORS IN FINE MECHANISM

The above discussion has dealt with sources of errors occurring in mechanical components. Also of importance might be those cross-coupled, mechanically produced effects that transduce mechanical energy into another energy domain. An example is thermoelectrical current, which is generated by the mechanical materials used in combination with thermal effects. Piezoelectric and magnetostriction are electrical resistance-strain changes that are each brought about by mechanical input parameters. They must, therefore, be considered in the mechanical design stages of an instrument as possible error sources in the total system.

It can be seen that there are many sources of error in, or caused by, the mechanical components of an instrument. If good mechanical performance is to be obtained, these must be considered and reduced to satisfactory levels of magnitude. This point needs to be stressed because this aspect of mechanical instrument design is sadly neglected in texts, in courses, and in practice. Scientists involved in the manufacture of their own experimental and measurement apparatus often reach these barriers and they, therefore, are aware of the errors that can arise. Other designers, however, often do not realize that their designs also may suffer from such effects; it is necessary to eliminate them as design parameters by reasonably rigorous study before they can be honestly regarded as insignificant. They are often easily reduced at the initial design stage.

The main error source is temperature and it affects just about every component. It always needs consideration before any measurements are made or instruments designed.

RELATED ARTICLES

Article 16, Reduction of Influence Factors, Volume 1; **Article 48, Common Sources of Errors in Measurement Systems, Volume 1**; **Article 80, Principles of Fine Mechanics – Kinematic and Elastic Designs, Volume 2**; **Article 81, Principles of Fine Mechanics – Systems Considerations, Volume 2**; **Article 82, Kinematical Regime – Members and Linkages, Volume 2**; **Article 85, Elastic Regime of Design – Design Principles, Volume 2**

REFERENCES

Abachi, H., Molenat, J. and Malbrunot, P. (1979) Continuously Variable Cryostat Giving Accurate Settings at any Temperature between 77 and 300K. *Journal of Physics E:Scientific Instruments*, **12**, 706–711.

Baird, K.M. (1968) Compensation for Linear Thermal Expansion. *Metrologia*, **4**, 145–146.

Barnes, R.B. and Silverman, S. (1934) Brownian Motion as a Natural Limit to all Measuring Process. *Review of Modern Physics*, **5**, 162–192.

Bleaney, B.I. and Bleaney, B. (1957) *Electricity and Magnetism*, Clarendon Press, Oxford.

Braddick, H.J.J. (1966) *The Physics of Experimental Method*, Chapman & Hall, London, (1st edn, 1954).

Eckert, E.R.G. (1950) *Introduction to the Transfer of heat and Mass*, McGraw-Hill, New York.

Gitlin, R. (1955a) How Temperature Effects Instrument Accuracy. *Control Engineering*, April 70–78.

Gitlin, R. (1955b) What's Available for Compensating Instruments for Temperature Changes. *Control Engineering*, May 70–77.

Gitlin, R. (1955c) How Temperature Compensation can be used. *Control Engineering*, June 71–75.

Glazebrook, R., Darwin, H. and Mason, C. (1923) Instruments, Design of Scientific, *A Dictionary of Applied Physics*, Vol. III, Macmillan, London, 445–450.

Jakob, M. and Hawkins, G.A. (1957) *Elements of heat Transfer*, Wiley, New York.

Jones, R.V. (1967) The Measurement and Control of small Displacements. *Physics Bulletin*, **18**, 325–336.

Jones, R.V. (1968) More and More about Less and Less. *Pro R Instn Gt Br*, **43**, 323–345.

Jones, R.V. (1970) The Pursuit of Measurement. *Proceeding of the IEE*, **117**, 1185–1191.

Logan, J.L. (1973) Gravitational Waves – A Progress Report. *Physics Today*, **26**(3), 44–52.

Rolt, F.H. (1929) *Gauges and Fine Measurements*, Vol. II, chapter XIV, Macmillan Publishing, London.

Sporton, T.M. (1972) The Design of a General-purpose air Thermostat. *Journal of Physics E:Scientific Instruments*, **5**, 317–321.

Sydenham, P.H. (1973) Nanometre Stability of Invar and Quartz Suspended in Catenary. *Journal of Physics E:Scientific Instruments*, **6**, 572–576.

Sydenham, P.J. (1979) *Measuring Instruments, Tools of Knowledge and Control*, Peter Peregrinus, London.

Trylinski, W. (1971) *Fine Mechanisms and Precision Instruments – Principles of Design*, Pergamon Press, Oxford (Translated from Polish by A. Voellnagel).

Usher, M.J. (1973) Development in Seismometry. *Journal of Physics E:Scientific Instruments*, **6**, 501–507.

White, G.K. (1959) *Experimental Techniques in Low-temperature Physics*, Oxford University Press, Oxford.

Whitehead, T.N. (1954) *The Design and use of Instruments and Accurate Mechanism*, Dover, New York (1st edn, 1933).

Wolsky, S.P. and Zdanuk, E.J. (1980) *Ultra Micro-weight Determinations in Controlled Environments*, Wiley, New York.

Wolsky, S.P., Zdanuk, E.J., Massen, C.H. and Poulis, A. (1967) On Gold Coated Quartz Arms to Reduce Thermal Expansion. *Vacuum Microbalance Technique*, **6**, 37.

89: Optical Materials

Pak L. Chu
City University of Hong Kong, Hong Kong, China

1 Blank Materials 573
2 Abrasives 575
3 Tools for Optical Manufacturing 575
4 Blocking Materials 575
5 Polishers 576
6 Cleaning Agents 576
7 Optical Cements 576
8 Material Properties 576
References 578

1 BLANK MATERIALS

Optical technique has long been important for measuring and observing instrument design, making its first impact as the telescope and microscope of the sixteenth-century Galilean times. Combined with fine mechanics development, it provided a vast array of mechanical-optical and, since the early twentieth century, electro-optical instruments. They are vital to modern telecommunications in optical fiber technology, in database storage devices, audio recording and playing equipment, and so on. While the pace of development of the technologies involved is truly amazing, the basics of the optical parts remain as it has been for centuries. Over time, the advances have been in understanding the optical theory and in using computers to support design.

Optical methods made significant strides when machine computation reached a state where the previously needed person-years of design effort have been reduced to hours. A further boost came when the traditional optical-mechanical industries combined their expertise with twentieth-century electronics. In this emerging century, microminiature optics will feature strongly.

The eventually realized laser source, in the 1960s, provided more capability for optical methods to be used in more kinds of measurement. In the solid-state form, the laser source has revolutionized many measuring instruments.

Several texts can be useful here – see Musikant (1985), Ahmad (1997), Hobbs (2000), Smith (2000), Weber (2003).

Optical instruments have been extensively used for varied applications in diverse areas and thus have immensely contributed to the science of measurement. During the latter part of the twentieth century, the interest in this branch of physics has further been stimulated and sustained with the availability of high power–coherent light sources and detectors. This era has thus also witnessed many original approaches toward developing sophisticated optical instruments and systems and associated technological advancements in meeting intricate measurement needs for different sectors of the economy. The quite relative newcomer, the miniature optical-array detector, has also opened up remarkable possibilities for faster and better measurement.

This article primarily aims at elaborating the fundamental issues connected with optical technology and understanding the basic principles underlying the construction of important optical measurement instruments and their applications. Mass-produced optical elements still need to use many of these centuries' old methods. Small-scale production, typical of measuring systems, has to employ them as a handcraft needing considerable skill and know-how.

Raw materials used in the optical shop are considered under the categories blank materials, abrasives, optical

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

tooling, blocking materials, polishers, cleaning agents, and optical cements.

The stock material used to form optical elements is known as a *blank*. Glass is the most commonly used blank material for the production of optical elements but other materials, including plastics, are also used.

Glass is a transparent dielectric substance, in which the molecular structure is characterized by an extended network lacking in periodicity and symmetry, but with atomic forces comparable to those present in crystals. Glasses of optical grade, commonly known as *optical glass* are used for quality optics.

As optical materials are commonly used to diffract radiation passing through an element, the effects that will cause distortions of images transmitted need to be identified and kept within limits. Mirror elements, where the radiation is reflected from the surface, are less impacted by defects but they still need to keep their form and shape as residual stresses relax over time or with thermal cycling in operation. Optical glass must be extremely homogeneous, must not absorb gases and liquids and must have negligible residual strain, bubbles, and striae. Striae are optical lines seen in glass due to variations of its refractive index.

The specifications of A grade optical glass state that it should not possess any visible striae or cord, with birefringence less than $10\,\mu m\,cm^{-1}$, that is, no color beyond light gray should be seen in the crossed Nicols test and it should be totally free from bubbles larger than 0.01 mm in diameter. This kind of glass is refined by heating and cooling many times to remove the defects.

Optical glass is characterized by its refractive index and the constringence or Abbe number (or simply ν value). This is defined as $(N_d - 1)/(N_F - N_C)$, where N_F and N_C are the refractive indices of the material corresponding to the F and C lines of the spectrum of hydrogen having wavelengths 4861 and 6563 Å respectively, and N_d pertains to the helium yellow line at wavelength 5876 Å.

The physical significance of ν is that it relates to the dispersion caused by the material: the higher the ν value, the lower is the dispersion. In simple terms, this means that the higher the ν value is the more the radiation is bent as it changes from one index to the next. Glasses with higher ν are customarily called *crown glass* and those with lower ν are known as *flint glass*. Crown glasses, therefore, have considerably less dispersion when compared to flints. This combination is useful for it allows corrected lenses to be formed by the use of two elements placed together.

Optical glass contains silica, which, when mixed with the oxides of potassium, calcium, barium, lead, and more in different proportions, yields a wide range of optical glasses of different properties. Table 1 gives some typical compositions.

A wide choice of optical glasses varying in refractive index and ν value is available commercially. The end use and design considerations govern the choice of appropriate material.

Testing of optical glass involves evaluation of refractive index, ν value, striae, strain, and bubble contents. The test methods are simple and can easily be adopted in actual practice by following standard specifications.

Some other blank materials employed to make best use of the different radiation bands of the EM spectrum are quartz (both crystalline and amorphous varieties), rock salt (NaCl), sylvine (KCl), fluoride (CaF_2), calcite, KBr, ZnS, calcspar, Iceland spar, solid solutions of thorium bromide with thorium iodide, Corning 7940 glass, Mylar, sapphire, and Irtran. All these possess quite different transmission and reflection properties making them necessary for different wavelength operations.

Curves illustrating transmission of light through the atmosphere, as also through some selected materials, can be found in manufacturers data sheets and in works such as the classic work by Hardy and Perrin (1932) and Weber (2003).

The spectral characteristics of optical glasses render them largely unsuitable for incorporation in infrared or ultraviolet imaging systems: other special materials are available for use at these wavelengths.

Plastics of many types are frequently used in optical technology particularly for making inexpensive and not too critical optical elements. Polymethyl methacrylate, in short PMMA or commercially called *Perspex*, has properties like crown glass, and polystyrene is equivalent to flint type. The principal drawbacks in their usage are the nonavailability of materials with wider choice of refractive index and ν value, lack of homogeneity, lower softening point, fragile nature, and proneness to scratching.

Table 1. Percentage composition of some typical optical glasses.

Glass type	SiO_2	B_2O_3	Na_2O	K_2O	CaO	BaO	ZnO	PbO	Al_2O_3	Fe_2O_3
Borosilicate crown	71.0	14.0	10.0	–	–	–	–	–	5.0	–
Crown	74.6	–	9.0	11.0	5.0	–	–	–	–	–
Light flint	62.6	–	4.5	8.5	–	–	–	24.1	–	–
Barium flint	45.2	–	–	7.8	–	16.0	8.3	22.2	–	–

2 ABRASIVES

To create optical elements, the blank must be shaped into flats or curves. The traditional method has been to grind blanks with abrasives and a form. It is possible to cut the material into shape using computer-controlled machine tools with micrometer tolerances, but they are still not capable of replacing grinding in most applications. Plastics can be molded as is done in spectacle lenses but again a mold shape is needed and the method is not capable of producing geometry of adequate accuracy for exacting optical systems.

Glass is quite a hard substance and for working on its surface, special abrasives are used. Obviously, an abrasive must be hard compared to the surface being ground. On Moh's scale, the hardness numbers of diamond, ruby, topaz, and quartz are, respectively, 10, 9, 8, and 7.

Carborundum (SiC), emery, in natural as well as artificial forms (fused Al_2O_3, aloxite, alundum), and boron carbide (B_4C) are some of the abrasives commonly used in optical workshop practice. These are mostly used in powder form suspended in a water medium. Diamond powder is available in almost all grades and is used for impregnation on working tools particularly when large stock removal is necessitated.

The average grain sizes of abrasives used in the optical shop range between 3 to 300 μm. Rough work requires a faster cutting rate, which can be achieved by using coarser grains. Silicon carbide or carborundum powder is most suitable at this stage. Once the desired shape and size have been given to the blank, finer grades of abrasives are progressively employed to obtain better finish. Aluminum oxide or emery is normally used for this purpose.

Grain size grading of abrasives is done in several ways. An example is the elutriation time taken, in minutes, for the emery to settle through a column of water contained in a vessel 30 cm in diameter and 1-m high, using the sieve or mesh size through which that particular grade can pass.

Table 2 gives the commonly used nomenclature of particle size grading and the corresponding average grain size.

The quality of uniformity in grain size is often indicated by numbers 0, 1, 2, 3, . . . , 10; 0 being perfectly graded and 10 being badly graded.

For the polishing operations, extremely fine-grain hard powder free from impurities is used. Rouge (Fe_2O_3), cerium oxide (CeO_2), putty powder (SnO), chromium oxide, and diamond paste are some of the common polishing agents.

Tripoli, diatomaceous earth, chalk, talc, slate, and pumice powder are other abrasive materials. Garnet optical powders have also been used for the particular purpose of ophthalmic (spectacles uses) machines employing autofeed slurry pumping systems.

Table 2. Abrasive gradation.

Mesh size	Other designation	Elutriation time in minutes	Approximate average size (μm)
60	M 60	–	290
100	M 100	–	150
180	M 180	1	85
220	M 220	–	75
700	M 302	5	20
850	M 302½	10	16
1000	M 303	20	13
1300	M 303½	40	10
1800	M 304	60	5
3000	M 305	–	3

3 TOOLS FOR OPTICAL MANUFACTURING

Turning, smoothing, and polisher forming tools are normally made of cast iron. Molten iron treated with meehanite (CaSi) gives a fine structure free from porosity and other defects. Polisher-holder tools may be made of aluminum. Gunmetal has also been used.

A radius turning attachment is used for providing the desired curvature on the tool surface. As the turning process is quicker compared to lapping, it is desirable to produce, as accurately a curved surface from the tool as is possible by turning.

Templates and gauges are generally made of brass, 2-mm thick, the edge reduced to about 0.5 mm. Collets for holding the job during curve generation on automatic machines may be made of brass (aluminum is also used). Clamping bells required in centering and edging operations are normally made of brass.

Test plates, test spheres, polygons, standard angles, optical flats, and the like are made to the required tolerance and durability requirements, from specified materials such as white plate glass, ophthalmic glass, optical glass, or low thermal expansion coefficient materials such as Pyrex, fused silica, CerVit, and Corning ULE.

4 BLOCKING MATERIALS

After smoothing, several optical elements are normally blocked over a tool and later operations performed upon them in situ. This requires a material that holds the optical elements rigidly during working. The blocking material should be hard enough to withstand the pressure and friction of the smoothing and polishing tool, yet be soft enough for easy extraction of the secured optical elements whenever needed.

Pitch is the most common blocking material as it is soft when warm, but hardens on cooling. Normal tar can be hardened to the desired degree by mixing with it additives such as sawdust and cotton wool. Blocking material should be cloth filtered to eliminate undesired hard foreign particles. Beeswax, plaster of paris, and hydrated lime are used for blocking optical flats, prisms, and similar elements. Low melting–point alloys are also used as blocking material.

5 POLISHERS

A polisher is an accurately shaped cast iron tool upon which the polishing medium remains mounted. Polishing of optical surfaces is achieved by lapping the job with a preformed polisher of desired shape and size and applying polishing powder. The accuracy of the surface quality generated on the finished product largely depends upon the quality of the polisher. A polisher material should be hard enough to retain its form during polishing, yet it must be amenable for alteration as and when needed. Pitch, as stated earlier, possesses these properties: a polisher is formed of warm pitch of the desired shape and size. On cooling, it becomes sufficiently hard to retain its character.

Polisher pitch is softer than blocking pitch. Mixing turpentine with hard pitch produces polisher pitch. Polishing grade pitch is commercially available. As the hardness of pitch depends upon the temperature, the viscosity and hardness of blocking and polishing pitch should be maintained in accordance with the ambient temperature of the shop floor.

Felt, cloth, and wax are some other materials used in preparation of polishers. Wax polishers cause less streaks and scratches but are slow compared to pitch and do not produce accurate surfaces so readily. For hand polishing, wood pitch is considered to be best.

For better results, polisher pitch might be prepared to a consistency such that it is readily but not deeply indented by the thumbnail. Thus, it should be neither too soft nor too viscous. Loaded polishers are made by addition of wood flour, cotton wool, or yellow felt to the pitch. These help maintain the shape of polishers. Addition of beeswax to pitch reduces its tendency to scratch the working surface.

Teflon has also been tried with success as a polisher. Different types of polishing pads are also available commercially, and are mainly used for medium grade flat work.

6 CLEANING AGENTS

At various stages of the work, the optical surfaces are cleaned to remove grease, pitch, or wax. Teepol solution is very useful for degreasing. Methylated spirit, isopropyl, and anhydrous alcohol are commonly used for general purpose cleaning of optical surfaces. Benzene, benzol, and trichloroethylene dissolve pitch and are used during the deblocking operation. Kerosene dissolves beeswax efficiently. Acetone is an excellent solvent and is used for cleaning polymerized HT cements.

7 OPTICAL CEMENTS

Optical cement must be transparent, reasonably color free and must not develop much strain on the cemented surfaces. These are normally nonadhesive, but join two glass surfaces after some physical or chemical operation has been performed upon them. Decementing may or may not be possible, depending upon the type of cement used. Some cements set on heating to a particular temperature followed by curing at another temperature.

Canada balsam, a natural oleoresin with refractive index 1.52, is commonly used in optical cementing work. Its unique characteristic is that it has very little tendency for granulation or crystallization on drying from a solution. Cellulose caprate (refractive index 1.47–1.49), which is another that is widely used, is a thermosetting cement.

Thermosetting cements readily release the cemented components upon heating. Many other types of commercially available optical cements, on exposure to ultraviolet light, are set by a polymerization process. These are extremely fast drying, but once cemented, the reopening process becomes quite tedious.

8 MATERIAL PROPERTIES

In an optical material, its transmittance is an important characteristic to be known because it measures how much light can pass through the piece of material. Technically, it is defined as the ratio of the transmitted optical power to the incident optical power when an optical beam is launched into the piece of material. Obviously, the difference, instead of the ratio, between the input power and the output power, accounts for the loss or attenuation of the material. The attenuation consists of reflection, scattering, and absorption of light by the material. The magnitudes of these quantities depend on the wavelength of light in operation. Since the optical material is made of molecules such as silica dioxide (SiO_2) in glass and CH_4 in PMMA polymer, these molecules can be regarded as spheres with certain diameters. If the wavelength of light is smaller than these diameters, it will see the molecules as spheres and when it impinges onto the sphere, it will be scattered into directions

other than the original direction of propagation. This kind of loss is called *scattering loss* and is typically described by Rayleigh's law of scattering:

$$L_{\text{scattering}} = \frac{K}{\lambda^4} \tag{1}$$

where K is a constant dependent on the diameter of the sphere and λ is the wavelength of light. Thus, we can see that as the wavelength shortens, the scattering loss increases dramatically. Of course, if the wavelength of light is greater than the diameter of the sphere, the propagating light will not see individual spheres in the optical material. In this case, the Rayleigh type scattering will be zero. However, if the material is not homogeneous, the resulting nonuniformity will also cause scattering if its period of variation is smaller than the wavelength of light.

Now, the reflection of light at the surface of the optical material also depends on its smoothness. The simplest relation describing the reflection is the reflectance defined as the ratio of the optical power reflected by the surface and the optical power incident to the surface. Mathematically it is defined as

$$R = \left(\frac{n-1}{n+1}\right)^2 \tag{2}$$

This formula assumes that the interface is formed between air and the optical material with refractive index n. It is also assumed that light enters the interface perpendicularly. In fact, if light enters at an angle, the reflectance can be calculated by means of a complicated formula normally found in textbooks of optics such as Hardy and Perrin (1932).

As light passes through the optical material, it is not only scattered but can also be absorbed by the material. The absorption can be divided into two parts: the ultraviolet absorption and the infrared absorption. For ease of explanation, let us consider the absorption of a pure silica (quartz). It is made up of molecules of SiO_2. The Si atom can oscillate at several wavelengths, that is, its resonant wavelengths. Two of them are particularly important to the absorption loss of glass, that is, 0.068 and 0.116 μm. Light with these wavelengths entering into the glass will be completely absorbed to give rise to the oscillations of the Si atoms. Not only so, light with wavelengths close to the resonances will also be absorbed although the intensity of absorption decreases as the difference between them increases. The wavelength range of visible light varies from 0.4 to 0.8 μm. Thus, the influence of the Si atom resonant absorption is significant. This is called *ultraviolet absorption*. On the other hand, the molecule SiO_2 also has several resonant wavelengths. An important one is at 9 μm. It also has harmonics at 4.5 and 2.25 μm and so on. Its long absorption tail extends well into the visible wavelength range causing extra losses. This is called *infrared absorption*.

Roughly speaking, the absorption spectrum of an optical material can be represented by a V-curve as shown in Figure 1.

The left arm of the V represents ultraviolet absorption and the right arm of the V represents infrared absorption. The apex of the V therefore represents the minimum absorption loss and the wavelength at which this occurs. For silica glass, this minimum is 0.15 dB km^{-1} at 1.55 μm. It forms the basic material for communication-type optical fiber. Figure 1 also shows that fluoride glass has a much lower absorption loss of 0.01 dB km^{-1} at 2.7 μm, while the KCl glass has a still lower loss of absorption at 0.001 at 4.7 μm. However, these glasses tend to absorb water and are not strong mechanically.

Figure 2 shows the transmittance of several infrared transmitting glasses.

Another important property of optical material is the wavelength dependence of its refractive index. The

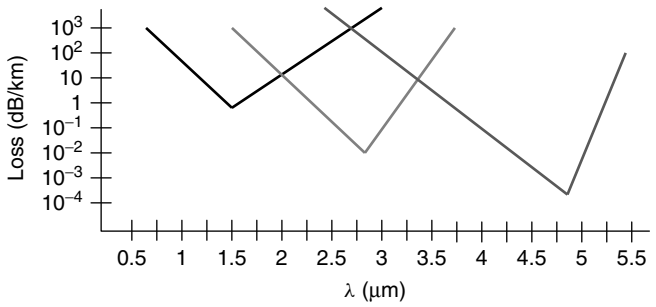


Figure 1. Absorption spectra of SiO_2, fluoride, and KCl glasses.

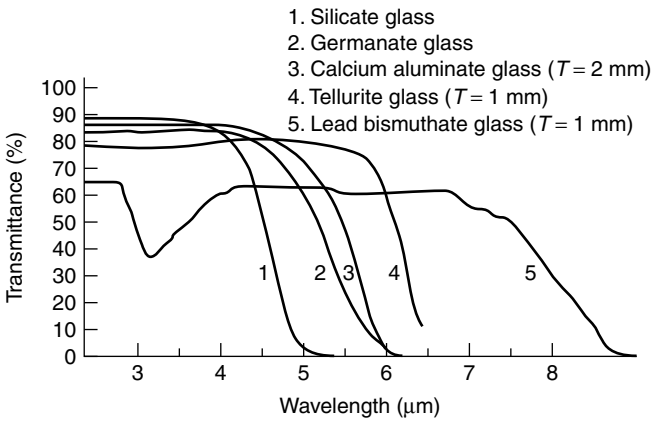


Figure 2. Transmittance spectra of infrared glasses. (Reproduced by permission of Marcel Dekker.)

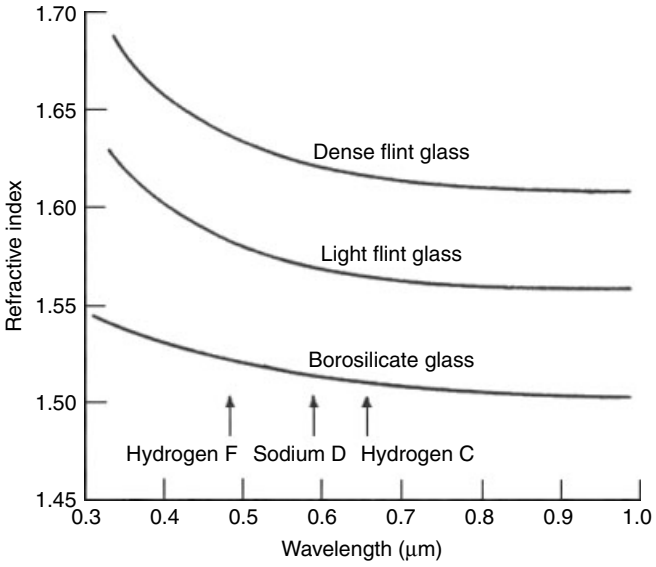


Figure 3. Refractive index spectra of common glasses. (Reproduced by permission of Marcel Dekker.)

dispersion of the material is defined as the rate of change of its refractive index as a function of the change of wavelength, mathematically,

$$(dispersion)_{\lambda} = \left(\frac{\mathrm{d}n}{\mathrm{d}\lambda}\right) \tag{3}$$

Because of the dispersion, a white light passing through a block of glass will spread into many colorful beams. In fact, this is a serious problem in optical fiber because the dispersion reduces the information carrying capacity of the fiber. Figure 3 shows the refractive index spectrum of several common glasses.

REFERENCES

Ahmad, A. (1997) *Handbook of Optomechanical Engineering*, CRC Press, Boca Raton, FL.

Hardy, A.C. and Perrin, F.H. (1932) *The Principles of Optics*, McGraw-Hill, New York.

Hobbs, P.C.D. (2000) *Building Electro-Optical Systems: Making It All Work*, Wiley-Interscience.

Musikant, S. (1985) *Optical Materials*, Marcel Dekker, Inc., New York.

Smith, W.J. (2000) *Modern Optical Engineering – the Design of Optical Systems*, McGraw-Hill Professional, New York.

Weber, M.J. (2003) *Handbook of Optical Materials*, CRC Press, Boca Raton, FL.

This article is based in part on a chapter by J. Prasad and G. Mitra originally published in Handbook of Measurement Science Vol 2, P.H. Sydenham, Wiley, 1983.

90: Optical Elements

Pak L. Chu
City University of Hong Kong, Hong Kong, China

1 Refracting Elements 579
2 Reflecting Elements 581
3 Gratings and Filters 581
4 Optical Fibers 582
5 Polarizers 583
6 Modulators 583
7 Optical Geometric Element Design 584
8 Interferometry 585
9 Optical Transfer Function (OTF) 588
10 Holography 590
References 592

1 REFRACTING ELEMENTS

According to the ray concept, a beam of light travels in rectilinear path inside any homogeneous medium. The optical elements interposed in the light path cause a change in the course of the original path by means of

- refraction
- reflection
- diffraction.

On the above basis, the optical elements that can be regarded as the building blocks of any optical system can be broadly considered under three categories:

- refracting elements
- reflecting elements
- diffraction gratings.

Many other miscellaneous optical elements, for example, graticules, filters, fiber-optic elements, polarization components, and optical modulators, are also employed for varied purposes in the optical system design.

This article deals with basic understanding of the function of the various types of basic optical elements.

Refracting elements are made up of an optically transparent homogeneous medium bounded by two or more surfaces. When a ray of light crosses from one medium to another, the path of propagation of the ray is altered in accordance with the well-known Snell's law, which essentially forms the fundamental basis for developing designs of the varied types of refracting elements.

The lens and the prism are two basic types of refractors.

1.1 The lens

The purpose of a lens is to form an image of an object, that is, a lens is essentially an image-forming element. The image may be

- real
- virtual
- formed at infinity.

A lens that forms a real image is called a *positive lens*; similarly, a virtual image–forming lens is called a *negative lens*.

When a lens forms the image at infinity, it is called a *collimating lens* since it transmits a parallel beam after refraction.

A single lens element is characterized by its two radii of curvature, axial thickness, diameter, and type of glass.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

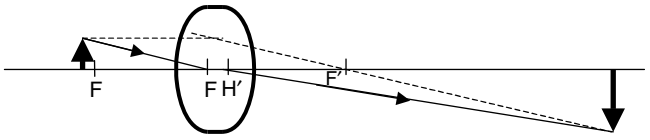


Figure 1. Lens definitions.

The line joining the two centers of curvature is called the *optical axis* of the lens.

Associated with each lens are three pairs of points called *cardinal points*, which are of immense value in analyzing its behavior. These are the

- focal point
- principal point
- nodal points.

When an incident parallel beam of light traverses through a lens, it either converges to an axial point (in case of positive a lens) or appears to be diverging from a point (in case of negative a lens).

This situation can be modeled both for a bundle of rays traveling from left to right, or vice versa (Figure 1). These convergence points of parallel incident rays are called *focal points* F and F′. The planes perpendicular to the optical axis passing through the first and second focal points are known as the *first* and *second focal planes* respectively. Obviously, rays passing through focal points become parallel to the optical axis after refraction. The surface generated by the locus of the point of intersection of the incident and refracted rays is known as the *principal plane*. Like focal planes, there are two principal planes, and their axial intersection points are designated as first and second principal points H and H′.

The effective focal length f (or simply focal length) is defined as the distance from the focal point to the corresponding principal point, while the distance from nearest lens surface to the focal point is called the *back focal length*. Nodal planes are a pair of conjugate planes corresponding to unit angular magnification. Nodal points are the axial intersection points of these planes. Nodal planes coincide with the corresponding principal planes when the lens is immersed in air.

1.2 The prism

A prism is essentially a non–image-forming optical element used for two main purposes, namely, to deviate the light path to any desired direction and to create wavelength dispersion of incident light into its constituent colors.

The incidence angle corresponding to which the refraction angle becomes 90° is termed the *critical angle*, I_c.

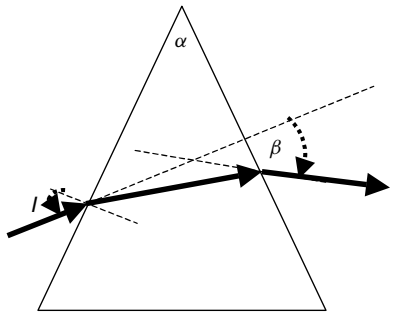


Figure 2. Ray deviation in prism.

From Snell's law, it follows that

$$I_c = \sin^{-1}\left(\frac{N'}{N}\right) \tag{1}$$

where N and N' are the refractive indices of the incident and refracting media respectively. This equation forms the basis of design of many deviating prisms.

The angular deviation β suffered by a ray traversing through a prism (Figure 2) immersed in air, with vertex angle α, can be shown to be given by

$$\beta = I - \alpha + \sin^{-1}[(N'^2 - \sin^2 I)^{1/2}\sin\alpha - \cos\alpha \sin I] \tag{2}$$

where I is the incident angle and N' is the refractive index of the prism material.

In the minimum deviation position ($\beta = \beta_m$),

$$N' = \frac{\sin\left[\frac{1}{2}(\alpha + \beta_m)\right]}{\sin(\alpha/2)} \tag{3}$$

which is a very suitable equation for refractive index measurement.

As the refractive index corresponding to each color (wavelength) is different, the deviation suffered by each of them is different; this causes wavelength dispersion of the incident beam into the constituent colors after refraction. The resolving power of a prism is a measure of its capability for separating adjacent spectral colors.

For erection of an inverted image without angular deviation, the prisms commonly used are Leman, Goerz, Hensolt, Porro types 1 and 2, and Abbe (or Konig). The Abbe prism does not cause lateral displacement. The Schmidt prism erects an inverted image and deviates it by 135° (or 45°). Penta and Amici prisms cause a 90° deviation. A special feature of the penta prism is that the transmitted beam remains stationary even when the prism is turned. Hence, it is also called an *optical square*. A Dove prism inverts without deviation, and on rotation the transmitted image rotates synchronously.

The actual configurations of the various types of prisms can be seen in standard textbooks such as Ahmad (1996),

Jacobs (1943), Levi (1968), Smith (1966), Smith (1978), Smith (2000), Hobbs (2000), and Weber (2002).

In this field, much of the detail was defined a considerable time ago. At that time, more detail was often published than what is easily available today. For this reason, older titles often contain useful information not available elsewhere.

2 REFLECTING ELEMENTS

Reflectors are those elements that reflect back the incident ray in the original medium; they are called *mirrors* or *reflectors*. A mirror may be flat, concave, or convex. Concave and convex mirrors act like positive and negative lenses respectively regarding their imaging properties.

In many specialized applications, particularly in astronomical telescopes, reflector elements of parabolic, elliptical, or hyperbolic shape are used. Parabolic mirrors are also extensively employed in lighthouse and other projection systems.

Beam splitters, which permit only a specified fraction of the incident light to pass through and maintain a known definite intensity ratio between the reflected and transmitted beams, can also be considered as a class of reflectors.

Hot, cold, partial, pellicle, and multilayer dielectric metal mirrors are some other useful reflecting components employed in optical system design.

3 GRATINGS AND FILTERS

Diffraction gratings essentially consist of a large number of closely spaced, equidistant, lines or rulings formed on plane or concave optical surfaces like a prism or filter; it causes spectral separation of the incident light beam.

The efficiency of a grating is determined by the amount of energy diffracted in any desired direction. This is enhanced by control of the groove shape. Sawtooth groove patterns of different step angle enable light concentration in the preferred direction (called *blazing*).

Linear gratings with straight apertures are traditionally produced by mechanically ruling precise parallel grooves, with a diamond tool, on a suitably worked blank. They can then be replicated from a master ruled grating by molding methods using plastics. These may be of transmitting, or reflecting type.

Concave reflection gratings are frequently used for efficient spectral separation. Spectroscopic grade gratings carry extremely fine grooves (of the order of 400 lines per millimeter).

The resolving power of a grating, that is, its ability to separate adjacent spectra, is given by the relationship mN, where m is the order of diffraction and N is the total number of lines.

In recent years, holographic gratings have also become commercially available. Their development greatly extended grating applications and performance.

Another form of manufacture has been developed – the holographic grating. Holographic gratings are produced by recording the optical interference patterns created between two parallel laser beams obliquely set relative to each other. Relative obliquity of the interfering beams enables control of line density. Compared with the ruled grating the holographic grating remains free from periodic error.

With proper choice of the recording medium, blazed holographic gratings can also be made. Other advantages of holographic gratings include a greater number of grooves being possible, up to 6000 grooves per millimeter; large dimension, up to 400-mm diameter; no ghosting; very low level of stray light; recording being possible on concave, toroidal, and aspheric blanks; and a possibility existing for making aberration correction.

A grating formed with concentric circular grooves (or obstructions) is called a *Fresnel zone plate*. It has the property of a lens but with several focal lengths for the various wavelength components.

Metrological gratings, used to determine length, carry identical opening and obstruction widths and are usually much coarser (tens of lines per millimeter). The constructional details of a versatile, low-cost, radial pattern-generating engine suitable for the manufacture of medium accuracy masters of metrological patterns as scales, line gratings, absolute multitrack encoder disks, spirals, circles, zone plates, and variable density and variable width circular tracks has been reported (Jain and Sydenham, 1980).

The Fiber Bragg grating is an important optical signal processing component. It finds applications in optical communications and in optical sensing. In this grating, the refractive index of the fiber core experiences a periodic variation with a period Λ. Any light with a wavelength λ satisfying the following relationship

$$\lambda = 2N\Lambda \tag{4}$$

will be reflected by the grating, where N is the average refractive index of the fiber core. The reflectivity increases with the index variation and length of grating. A reflectivity close to 100% can readily be obtained in single-mode optical fibers. The creation of this grating depends on the photosensitivity of the glass making up the core of the fiber. Boron- or phosphorus-doped germanosilicate glass is normally photosensitive, whereas a simple germanosilicate glass is not. However, the latter can be made photosensitive when hydrogen is introduced into it. To create the grating,

a phase mask with rulings of period Λ is placed close to the fiber and the mask is illuminated by an optical beam of appropriate wavelength at which the core glass of the fiber is photosensitive.

3.1 Graticules

Graticules are fine measuring scales or well-defined patterns placed in the focal plane of an optical instrument. These are employed for determining size, distance, direction, position, or any other attribute of the object that is viewed coincident with them. Graticules are also often denoted by terms such as diaphragm and web or in accordance with their use, examples being sighting scales, cross-lines, (optical) micrometers, and stage micrometers.

Since a graticule is usually required to be viewed in the instrument under high magnification, tolerance requirements in the manufacture of graticule blanks are generally very severe. For example, scratch width and dig or pit diameter should not exceed 0.01 and 0.05 mm respectively; surface accuracy should be within two fringes and sphericity should be within 0.5 of a ring; thickness should be within 0.05 mm; and decentration should not exceed 0.5 to 1 min (Prasad and Singh, 1970).

Graticules are made by several different techniques, namely, ruling and etching, ruling followed by vacuum deposition of chromium or kanthal, photographic processes, and photo etching. Details of the procedures followed in graticule manufacture and compositions of some photoresists suitable for this work are given elsewhere (Horne, 1974).

Health and safety regulations now prohibit the use of many traditional photoresists, so it is important to keep up with the latest information.

3.2 Filters

Filters types fit into the three broad categories:

- absorption or color filters
- interference filters
- spatial filters.

Absorption filters derive their performance from bulk interaction between light and the filter media; the transmission generally being a smoothly decreasing function of thickness. Such filters are made in a variety of base materials: gelatin, glass, and plastic are commonly used.

Selective color transmission is obtained primarily by ionic absorption or by selective scattering. Gelatin filters are low in cost and provide a wide color choice. Glass filters are widely used in photography and colorimetry and also in the production of interference filters for blocking extraneous passbands. Plastic filters are available both in sharp-cutoff and intermediate bandwidth types.

Thin film interference filters operate in the same manner as a Fabry–Perot interferometer. These are usually designed for normal incidence but may be constructed for specific nonnormal applications. Nonnormal incidence results in a shift of the passband towards shorter wavelengths. Two basic types of interference filters are available – metal-dielectric and all-dielectric. These are manufactured by thin-film, vacuum deposition techniques using special manufacturing plant. The experimental techniques adopted are described in Horne (1974), Macleod (1969), Smith (2000), and Weber (2002).

Under spatial filters are included components, which, in a functional sense, are used to modify the spatial characteristics of the incident light. A typical example is the fine pinhole (about 5–10 μm in diameter) used in a Fourier system for filtering out the unwanted portion of diffraction caused by the circular aperture of the microscope objective lens when used in a laser beam.

4 OPTICAL FIBERS

Fiber-optic methods are based on the ability of thin smooth strands of transparent materials to convey incident signal by the total internal reflection process, without interference from neighboring fibers. It is achieved by drawing thin (of the order of 10–100 μm), coated optical fibers having a core and cladding made of different refractive index materials.

Fiber-optic elements can broadly be grouped into two types, noncoherent and coherent. Noncoherent fiber bundles consisting of thousands of thin fibers not arranged in any order are used for transporting light and are often called *light guides* or *light cables*. In coherent fiber bundles, the fibers at the end faces are arranged in the same manner so that optical image can be transmitted without suffering distortion. The fiber bundles are usually made of fibers with relatively high loss. Hence the length of each bundle is about 1 m.

The flexible coherent fiber bundle is the basic component of many endoscopic medical and industrial inspection instruments used in remote examination of internal cavities. The rigid form of a coherent fiber bundle, for example, a faceplate, is the essential component of image intensifiers, enabling high optical efficiency in image transfer. These components have also been used as windows of cathode-ray tubes built for direct recording of high-speed computer printout or other data recording.

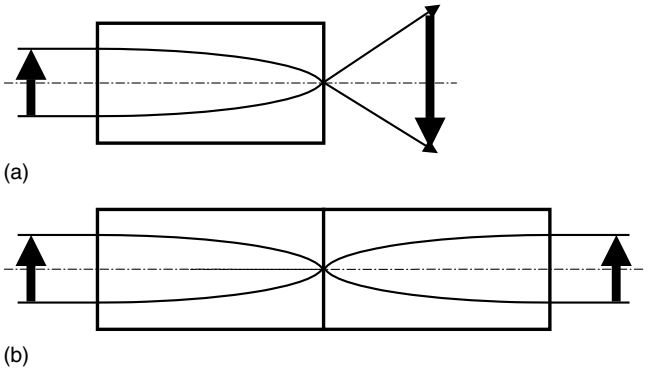


Figure 3. Grin Lens (a) Quarter Pitch Grin Lens, (b) Half Pitch Grin Lens.

Low loss (attenuation approaching $0.2\,\mathrm{dB\,km^{-1}}$ at a wavelength of 1550 nm) single-mode fibers, for use in optical communication, have also been developed.

Another fiber-optic component is the *gradient index lens*, also called *grin lens* (Figure 3). It is essentially a short length of graded index fiber (about 5–10-mm long) with large core diameter (between 1 and 2 mm). The imaging properties of this lens depend on its length. A quarter pitch lens forms an inverted, real, and magnified image of an object at infinity on the opposite end of the lens. A half pitch lens simply inverts the image of an object from one end face to another without magnification.

5 POLARIZERS

Polarizers are used for the production, control, and analysis of light in special polarization states. Their operation depends on the properties of birefringence (double refraction), dichroism (a special case of birefringence), and change in polarization characteristics upon reflection (at, and near, Brewster's angle).

Retardation plates (quarter, half, and full wave), the Savart polariscope, Soleil–Babinet compensator, dichroic polarizers (sheet type), laser-polarizing beam splitters, and Cornu depolarizer are some of the important components.

6 MODULATORS

In communication, display, data recording, and measurement, it is often desirable to control the amplitude, phase, frequency, and state of polarization of the light beam. Optical modulators are the devices variously used for these purposes. They are generally based on basic electro-optic (Pockels, Kerr), magneto-optic (Faraday, Cotton–Mouton) or acousto-optic (Debye–Sears, Raman–Nath) effects.

With the advent of the laser source of radiation, optical modulators have assumed much importance (Hartfield and Thompson, 1978). Table 1 gives some basic information about the various types of optical modulators. A popular high-speed optical modulator that finds application in optical fiber communication is the electro-absorption modulator. It is a reversed biased pn junction similar to a photodetector. Light impinging onto the junction is absorbed and the absorption varies in accordance with the bias voltage. Hence the modulating signal is the electric bias voltage and the light leaving the device is intensity modulated. An electro-absorption modulator is rarely used freestanding because of the high loss associated with the device. Instead, it is part of an integrated optical transmitter, that is, a semiconductor laser and the modulator are built together on the same chip, with the output of the laser being fed directly into the modulator. In this configuration, the additional loss caused by the modulator is only of the order of 1 dB or so.

Table 1. Types of optical modulators.

Type	Principle	Typical materials	Typical applications
Electro-optic	Induced birefringence on application of external electric field	ADP, KDP, $LiNbO_3$, ADA	Amplitude, phase, frequency and polarization modulation, high-speed shutter, laser Q-switching, mode locking, cavity dumping
Magneto-optic	Induced optical activity on application of external magnetic field	Fused silica, dense glasses, H_2O, CS_2, P	Deflection, scanning, spatial modulation, information processing
Acousto-optic	Diffraction and scattering of light waves by sound	Fused silica, GaAs, YAG, YIG, TiO_2	Scanning, optical delay line, heterodyning
Mechanical	Chopping of light beam by rotating sector disc	–	Signaling, IR detection
Electro-absorption	Light absorption in going through a reversed biased pn junction depends on biasing voltage	Semiconductor material such as InGaAsP/InP, dependent on the operating wavelength	High-speed optical communication systems

7 OPTICAL GEOMETRIC ELEMENT DESIGN

In geometrical optics, diffraction effects are neglected and the ray aspect of light is considered. Cophasal surfaces orthogonal to a set of rays are defined as geometrical wavefronts (or simply wavefronts).

An ideal lens produces an image exactly alike the object. However, an ideal lens is only a theoretical postulate since, in reality, there are certain parameters inherent in the refraction process that cause aberrations and deviate the image structure from its ideal shape. The purpose of optical system design is to work out technical specifications of each of its constituent elements, tailoring the geometry such that the entire system gives an adequately aberration-free acceptable image.

7.1 Aberrations

There are five monochromatic and two chromatic aberrations. Spherical aberration, coma, astigmatism, petzval curvature, and distortion fall in the first category, with longitudinal and lateral (or transverse) chromatic aberrations forming the second.

Spherical aberration arises due to the fact that rays originating from an axial object point and intersecting the refracting surface at different heights do not meet at a single point after refraction. In other words, the paraxial focus is different from the marginal focus; the axial distance between them is called *longitudinal spherical aberration* ($\Delta l'$).

When rays from an off-axial object point pass through the lens, the paraxial and marginal bundles again have different foci. A comet-like flare is then seen in the image plane, which is generated by the partial overlapping of circular light patches of increasing diameter. This off-axial aberration is called *coma*. It causes asymmetry in the image.

Astigmatism occurs because rays from an off-axial object point lying in the meridional and sagittal planes focus at different points. The distance between these two foci is called *astigmatism*.

An image of a straight object perpendicular to the optical axis is formed on a curved surface instead of a plane. This is due to petzval curvature.

Owing to unequal lateral magnification throughout the field, the appearance of an image changes from the original object; this occurs even when the above aberrations are not present. Pincushion distortion means that the outer zones are more magnified compared to the inner; the reverse manifestation is called *barrel distortion*.

Any incident polychromatic beam suffers dispersion due to the wavelength dispersive nature of the material of the refracting element. Thus, with incident white light, a point object has different image locations for its different colors. The axial and transverse distances between the violet and red foci are called *longitudinal* and *lateral chromatic aberration* respectively. General accounts of lens types and aberrations are available in Levi (1968) and Smith (2000).

7.2 Elements of lens design

It is possible to reduce aberrations of an optical system by choosing suitable combinations of two or more lens elements. Consistent with the stipulated focal length, f number, image definition, and other requirements, the selection of proper lens combination, their glass types, radii of curvature, thickness, axial separation, and stop specifications falls within the purview of lens design. Detailed analyses of lens design are available in several standard treatises (Buchdahl, 1954; Conrady, 1957, 1960; Cox, 1964). The relationship between ray and wavefront aberrations and their evaluation has been extensively covered in Hopkins (1950). Zoom lens are covered in Clark (1973). Today, much of this foundational theory is embodied into lens and optical system design tools.

The analysis of an optical system requires tracing a large number of rays through it and consequently determining the residual aberrations, which become the basis for further refinement. Depending upon the computational technique adopted, several ray tracing schemes have been developed, Smith (1966) being an example.

Ray tracing schemes require numerous calculations, of simple nature, the whole process being very time consuming and tedious by traditional methods. Electronic computation has been applied to this with great effect.

In either case, considerable experience is required to design even a simple corrected lens system. For this reason, an instrument designer would normally call upon the services of a lens designer or make use of commercially available proprietary lens systems when lens design is a critical feature of the whole system.

7.3 Stop and pupil

To control the illumination and field coverage, certain diaphragms are used in optical systems. An aperture stop determines the maximum cone of light that passes through the system; its location has a marked influence upon off-axial aberrations of the system, though spherical aberration and longitudinal chromatic aberration remain independent. A field stop is the diaphragm used to control the image size.

Entrance and exit pupils are defined as the images of the aperture stop formed by the elements placed on the left and right of the stop respectively. The ratio of the pupil diameters represents system magnification.

7.4 Field angle

Field angle is defined as the angle subtended by the object at the entrance pupil; it specifies the maximum object size that can be imaged by the system. As the field angle increases, the effects due to off-axial aberrations become more pronounced, with consequent increased difficulties in the design process.

7.5 Effects of non–image-forming elements

Non–image-forming elements, such as plane parallel plates and prisms, also introduce aberrations when placed in non-collimated light beam. Such elements are often employed as a corrector for compensation of residual aberrations of the system. In Prasad, Mitra and Jain (1975), the aberrations produced by a single plane parallel plate immersed in air and inclined at an angle with the vertical are discussed. They provide equations for calculating the lateral displacement, tangential coma, and astigmatism of the system.

A parallel plate can also be used to convert large angular displacements into small translational equivalents that can be used on a position sensitive photodetector to sense movement. This has been used in metrology (Sydenham, 1969).

7.6 Performance analysis

Performance of the system can be expressed in terms of residual aberration characteristic curves corresponding to different field angles or in terms of a spot diagram depicting the image plane intersection points of several rays passing through various zones of the system. The resolving power test is based on Rayleigh's criterion of resolution; the Strehl intensity ratio and, as is discussed later, the optical transfer function (OTF) technique are other methods of image quality assessment.

7.7 Optical modules

Illustrations of configurations of often met optical systems are given in Figure 4. Their design clearly becomes progressively more complex as the number of elements rises. Zoom lenses, for instance, were only made economic by the advent of electronic computer methods of design.

8 INTERFEROMETRY

Interferometric measurements serve as a very powerful measuring tool yielding a high degree of precision (within fractions of a wavelength of the radiation used).

Under suitable conditions, two light beams interfere with each other, producing a resultant pattern consisting of alternate bright and dark bands (or fringes). Shape and location of these bands depends upon the difference in total optical paths traveled by the two interfering beams. Accordingly, length and phase variables may be measured directly in an interferometer, while refractive index, displacement, velocity, and more can be evaluated by converting them into corresponding optical path difference OPD.

In the following, we first outline the fundamentals of interferometry, followed by a brief description of common types of interferometers.

8.1 Conditions of coherent radiation interference

Consider two light waves A_1 and A_2 of the same angular frequency ω, given by

$$A_1 = a_1 \sin(\omega t + \phi_1) \tag{5}$$

$$A_2 = a_2 \sin(\omega t + \phi_2) \tag{6}$$

where a_1, ϕ_1 and a_2, ϕ_2 are the corresponding amplitude and phase respectively, at any instant of time t.

Straightforward mathematical analysis shows that superimposition of A_1 and A_2 produces a resultant amplitude A given by

$$A = a \sin(\omega t + \theta) \tag{7}$$

where

$$a = [a_1^2 + a_2^2 + 2a_1 a_2 \cos(\phi_1 - \phi_2)]^{1/2}$$

and

$$\theta = \tan^{-1}\left[\frac{a_1 \sin\phi_1 + a_2 \sin\phi_2}{a_1 \cos\phi_1 + a_2 \cos\phi_2}\right] \tag{8}$$

From this, it can be seen that the resultant intensity pattern is maximum ($I_{\max}$) or minimum ($I_{\min}$), when $\phi_1 - \phi_2 = 2n\pi$ or $(2n+1)\pi$ respectively, where n is zero or an integer.

Fringe visibility (or contrast) is defined as $(I_{\max} - I_{\min})/(I_{\max} + I_{\min})$. For best visibility (or contrast), a_1 and a_2 should be equal or nearly equal.

Summarizing the conditions of interference,

1. the two interfering beams must be of the same wavelength monochromaticity;

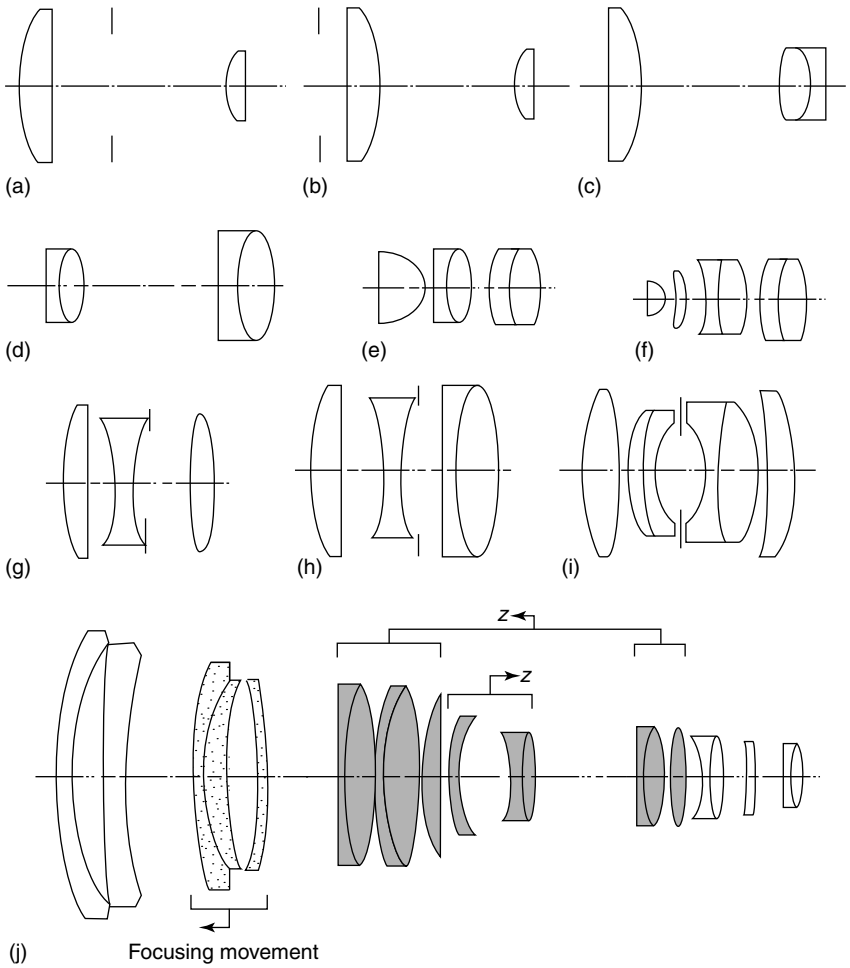


Figure 4. Configuration of some optical systems: (a) Huygen eyepiece; (b) Ramsden eyepiece; (c) Kellner eyepiece; (d) Lister-type microobjective; (e) Amici-type microobjective; (f) oil-immersion microobjective; (g) Cooke triplet; (h) Tessar lens; (i) double Gauss; (j) zoom television lens-Taylor Hobson (Clark, 1973). (Reproduced from *Zoom Lenses*, Clark A.D., Copyright (1983) © Adam Hilger.)

2. the phase difference between them must remain closely constant over the period of observation (otherwise, the fringe pattern would dance around, causing a blurred patch – the coherence condition); and
3. the state or plane of electromagnetic vibrations in both beams must be the same Fresnel–Arago law.

To satisfy the above conditions, light sources used in interferometry generally have a high degree of monochromaticity and are highly coherent so that a large path difference can still produce a meaningful interference pattern. The so-called coherence length is a measure of how long the beam length difference can be to retain coherence.

Further, all interferometers, to a greater or lesser degree, remain sensitive to stray mechanical vibrations. Adequate precautions are needed during their installation to minimize these effects. This makes use in severe vibration conditions difficult.

The principle of interferometry is used in many measuring instruments. Such use is discussed in Dyson (1970), Ahmad (1996), and Smith (2000)

8.2 Michelson interferometer

The Michelson interferometer (Figure 5) is used for the measurement of wavelength of light, slow air drift, calibration of a standard length, and other scientific applications.

Light from an extended source S (white light or monochromatic) is divided into two paths by a beam splitter B (Figure 5). The end mirrors M_1, M_2 reflect back the respective coherent beams 1, 2, which recombine to form fringes in the viewing telescope T. In one of the arms, a glass plate P identical in construction to the beam splitter is introduced to compensate for the inequality of glass path traversed by the interfering beams. In practice, corner-cubes

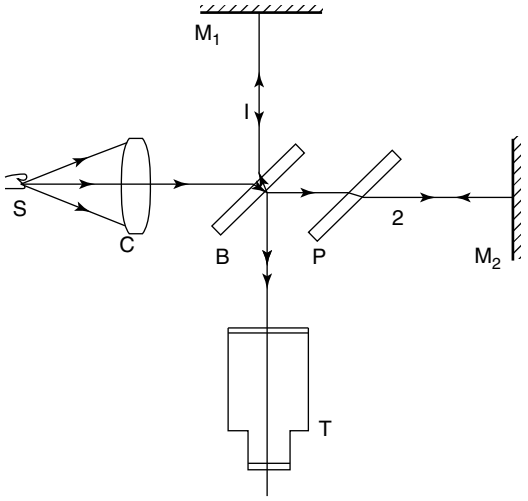


Figure 5. Michelson interferometer: S, source; C, collimating lens; B, beam splitter; P, compensator; M_1, M_2, mirrors; T, telescope.

can be used instead of flat mirrors; this makes adjustment considerably easier.

The Twyman–Green interferometer is a versatile instrument that is particularly suited to optical shop-floor applications. It is similar in construction to the Michelson interferometer. Here, the incident light is formed as a well-collimated monochromatic beam and the compensator plate is dispensed with. The optical job under test is placed in one arm of the interferometer, while the other arm provides the reference beam. In cases of a lens, prism, and the like, the beam passes twice through the test piece; hence the fringe pattern depicts twice the amount of total error, which is inclusive of surface defects, material inhomogeneities, and aberrations.

The invention of the laser source enabled development of an unequal path Twyman–Green interferometer of different geometry that is suitable for testing large optics in a production line.

8.3 Fizeau interferometer

A Fizeau interferometer is suitable for qualitative as well as quantitative assessment of surface quality. In its classical form, it is used (Figure 6) for flat surface testing. A collimator C sends a parallel beam of light from source S onto the test surface T. A good quality plane parallel glass plate F, whose lower surface is of a high degree of flatness (master surface), is placed over the test piece with an air gap. The collimator focuses the return beams onto the observation point O that is suitably located using a beam splitter B. The interference pattern is formed between the rays reflected from the master and test surfaces. Departure

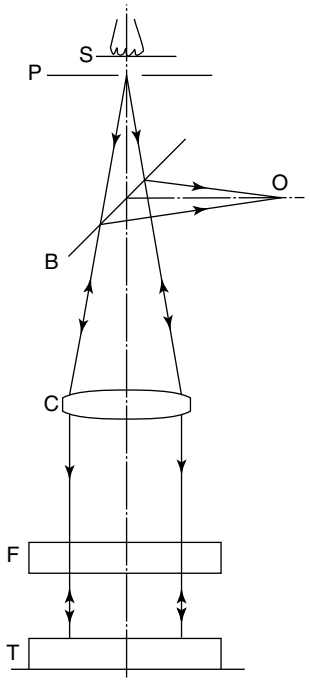


Figure 6. Fizeau interferometer: S, source; P, pinhole; B, beam splitter; C, collimating lens; F, master surface; T, test surface; O, observation point.

of linearity of the fringes expresses the surface error in terms of fractional fringe width (each fringe corresponds to $\lambda/2$ departure).

Numerous modifications of this classical instrument exist. Concave and convex surfaces are tested in a diverging and converging beam respectively, the master flat then being replaced by the complementary master surface. Using a laser source, a spherical surface may be tested against a master optically flat surface. In this interferometer, a converging beam is focused at the center of curvature of the test surface. Another set of fringes is obtained when the beam is focused at the vertex of the test surface, and as such, the radius of curvature can also be measured.

8.4 Rayleigh interferometer

The Rayleigh interferometer is a very useful instrument for the measurement of the refractive indices of gases and liquids and for the control of their composition. Light from a slit source is collimated and made to fall upon two tubes placed side by side, one containing the sample under test and the other a standard sample. Fringes are viewed through a focusing lens and a magnifier. Fringe displacement is measured by tilting a compensator plate placed in the light path passing through one of the tubes. White light sources may be used enabling easy measurement of the fringe displacement.

The Michelson stellar interferometer employs two widely separated slits whose interdistance can be varied. Each slit selects different parts of the wavefront. From the change in visibility of the fringes seen through the telescope, a stellar diameter can be evaluated.

8.5 Jamin interferometer

Used in refractometry, a Jamin interferometer has a thick plate that splits the incident beam into two using reflections at the front and back surfaces. An identical plate recombines the beam. The Mach–Zehnder interferometer is a modification of this. Using beam splitters for beam division and recombination, the separation between the interfering beams is made large. It is extensively used in hydrodynamic measurements.

8.6 Point-diffraction interferometer

Generation of an aberration-free reference wavefront without using high quality optics is achieved in a point-diffraction interferometer (Smartt and Strong, 1972). A pinhole on a transparent substrate is located slightly off-axis at the focal plane of the lens system under test. Diffracted wavefronts from the pinhole produce reference wavefronts. An absorption coating is placed over the pinhole substrate to enable matching of the intensities of the interfering beams.

8.7 Scatter fringe interferometer

Large concave mirrors, as used in astronomical telescopes, are effectively tested in a scatter fringe interferometer. A scatter plate placed perpendicular to the optical axis passing through the center of curvature of test surface splits a converging beam into two components. The specularly transmitted beam, after reflection through a small segment of the mirror, serves as the reference beam, while the scattered component fills the whole test surface and is the test beam. A semireflecting mirror is interposed to superimpose both the beams onto another scatter plate (identical to the first one). The instrument is comparatively less sensitive to vibrations since the interfering beams are closely confined to each other.

8.8 Shearing interferometer

In a shearing interferometer, the test beam is split into two beams, which are compared against each other, thereby doing away with any reference beam, as is needed in comparison-type interferometers. When superimposed, the beams are laterally, rotationally, or radially sheared relative to each other so that different parts of the two identical wavefronts interfere. There exist several techniques for achieving the desired shear. Figure 7 gives simple examples. By employing birefringent crystal elements, such as the Wollaston prism or Savart polariscope, shear is readily obtained. This type is generally called a *polarizing interferometer*.

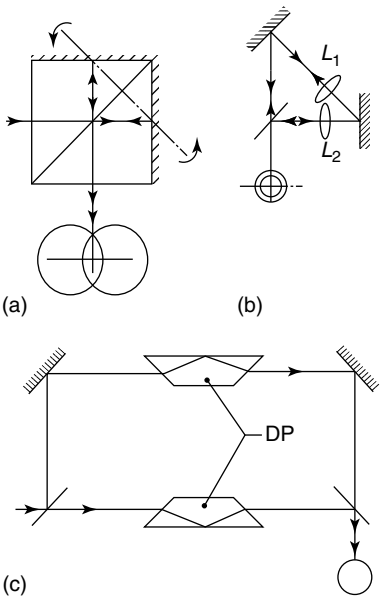


Figure 7. Shearing techniques: (a) lateral shear; (b) radial shear; (c) rotational shear produced by counterrotating Dove prisms (DP).

A shearing interferometer is less sensitive to vibrations (relative to Twyman–Green interferometer) and is suitable for testing a lens, wedge angle, homogeneity of transparent samples, optical flatness, and the optical transfer function.

8.9 Multiple-beam interferometer

If the test surface and the inner surface of the master plate of a Fizeau interferometer are partially reflecting, each ray splits into several components. Consequently, a number of beams of varying path length take part in forming a multiple-beam interference pattern, where high finesse fringes are formed. High precision measurement of thin film thickness, surface contour, and separation of neighboring wavelengths are performed in a multiple-beam interferometer.

9 OPTICAL TRANSFER FUNCTION (OTF)

Various criteria exist to express the performance characteristics of a lens system. Measurement of the individual aberrations gives clues to the designer in respect of its

defects. A standard interferogram depicts the phase error across the pupil. The resolving power test indicates the final resolution limit attainable by the system. However, none of these attributes yields a completely objective assessment of the quality and image-forming capabilities of the optical system under test.

The OTF has been used as a merit function that, to a large extent, serves this need, even though it must be appreciated that this technique also is not the final solution to all lens evaluation problems.

9.1 OTF representation

OTF is basically a resolving power test, but unlike the classical resolution test, it expresses the image contrast at different spatial frequencies of a sinusoidal object. It is, in this respect, similar to the time frequency response tests associated with electrical circuits (Coltman (1954); Schade (1964)).

For elucidation of the OTF concept, consider two bright object points separated by a dark space. Theoretically, the image of this object assembly should be two bright points separated by a dark region. In such a situation, the object is said to be well resolved, and both the object and the image possess unity contrast. However, since any real system is afflicted with aberrations and associated diffraction effects, in actual practice, some light would encroach into the dark space, with consequent reduction in image contrast. As the two object points are brought closer, the contrast decreases until it becomes zero; this is taken as the resolution limit.

According to Fourier mathematics, any periodic function can be expressed in the form of a series containing sinusoidal terms.

Upon combining the above two concepts, the OTF can be taken as a functional representation of the image contrast and location corresponding to various spacings of sinusoidally varying objects.

The OTF and the point spread function are interrelated according to Fourier transformation rules, each being transformable into the other, the point spread function being the representation of intensity variations in the image of a point object. Applying Fourier transform rules, it can further be shown that the OTF is the autocorrelation of the pupil function, where the pupil function is the intensity distribution of the lens system at the exit pupil and autocorrelation is a standard mathematical operation (Francon, 1963; O'Neill, 1963).

Mathematically, the OTF can be expressed in the form

$$O(\xi_x, \eta_y) = M(\xi_x, \eta_y) \ \exp\left[-ik\phi(\xi_x, \eta_y)\right] \qquad (9)$$

where $O(\xi_x, \eta_y)$ is the OTF; ξ_x, η_y are frequency coordinates in the image plane; $M(\xi_x, \eta_y)$ is the modulation transfer function (MTF) and $\phi(\xi_x, \eta_y)$ is the phase transfer function (PTF).

The OTF comprises both the MTF and the PTF. The significance of the phase term is that the image does not remain located at the ideal conjugate point but is displaced by the amount given by the PTF.

The above equation further reveals that for complete evaluation of system performance, both the MTF and PTF are required. However, it has been seen that for a moderate range of object frequencies, the MTF gives sufficient information of practical interest, and is therefore the quantity that is most commonly evaluated.

9.2 OTF evaluation – scanning methods

This method (Murata, 1966) involves scanning the image intensity distribution of a known object, which may be a sinusoidal grating of variable frequency, square-wave grating, single slit, two square-wave gratings rotating in opposite direction (producing Moiré patterns of varying frequency), or of any other arbitrary shape.

The basic experimental setup is given in Figure 8. When the object O is composed of sinusoidally varying intensity gratings of different frequencies, the response of P remains proportional to the corresponding intensity in the image. The image scan therefore directly provides $I_{\max}$ and $I_{\min}$, and hence the contrast. In the case of square-wave gratings, an intermediate electronic subsystem is introduced, which allows only the fundamental of each frequency to be passed, thus indirectly achieving generation of sine-wave targets.

Since, for good performance, the object contrast at all frequencies must be identical, the actual realization of variable frequency sinusoidal or square-wave gratings is in itself a cumbersome task. Furthermore, in both cases, since the object frequency range is limited with regard to generation of higher frequencies, a relay lens located between O and L must be employed. This introduces its own defects into the final result, and hence, it should have much better corrections compared to those of the test lens.

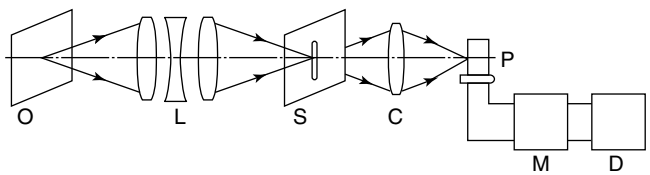


Figure 8. Scanning method of determining OTF: O, object; L, test lens; S, scanning slit; C, condenser; P, photomultiplier tube; M and D, detection and display.

Table 2. Comparison of test objects in OTF evaluation.

Object shape	Advantages	Disadvantages
Sine-wave grating	Direct reading	Difficult to make; limited range of spatial frequencies; intermediate relay lens needed for obtaining higher frequencies
Square-wave grating	Direct reading	Same as in (1). Additional electronic subsystem needed for filtering fundamental of each frequency
Two counter rotating gratings	Rapid direct display	Same as in (2)
Pinhole or single slit	Easy to make; wider range of spatial frequencies	Low light efficiency; computer needed for data conversion; correction factor needed for finite width
Knife-edge	Same as in (4). Also, better light efficiency compared to pinhole or single slit	Computer needed for data conversion; stray light affects the OTF measurement to the largest degree; very low signal to noise ratio at high spatial frequencies

The difficulties are met to the same extent by choosing objects of simple geometry, like a single slit, a pinhole, or a knife-edge, but these have other disadvantages.

Table 2 summarizes the relative advantages and disadvantages of commonly used test objects.

9.3 OTF evaluation – interferometric methods

The desired autocorrelation (or self-convolution) is achieved in lateral shearing interferometers; this provides the basis of an analog technique of OTF measurement (Hopkins, 1955). It is known that if the path difference between the two laterally sheared beams is changed, the total light flux across both of the sheared beams varies sinusoidally. Further, the shear magnitude represents the spatial frequency. Hence, by variation of the shear and measurement of the corresponding contrast, the OTF can be evaluated.

Polarizing type, lateral shear interferometers for OTF measurement have also been developed (Francon, 1966).

9.4 OTF applications

OTF analyzers have been applied with practical advantage in many interesting situations. Examples are as a go/no-go gauge for final checking of optical elements, giving the cumulative effect of residual aberrations, diffraction, and inaccuracies caused during manufacturing processes; providing means of evaluating performance of complex cascaded electro-optical systems; effecting improvement in the quality of optical system at the design stage; and many more.

As an illustration, qualitative representation of MTF curves is made in Figure 9. Curve A is the plot of a diffraction-limited (theoretical best) system; B and C show the effects of adding a central obstruction and of defocusing

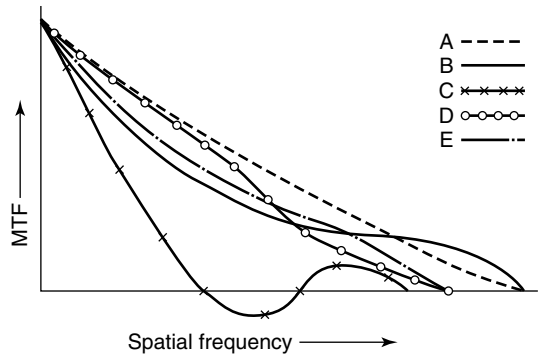


Figure 9. Qualitative representation of MTF curves: A, diffraction-limited system; B, with central obstruction; C, defocusing effect; D, E, different designs of the same optical system.

respectively; and D, E are representations of different practical designs of the same optical system.

It might be noted that while D has better performance at lower spatial frequencies, the quality of E is superior in the high frequency domain. Thus, MTF curves give a reliable quick answer and enable fast selection of a good lens for a particular optical system.

10 HOLOGRAPHY

In the conventional image recording process, the photographic emulsion records the time average square modulus of the complex amplitude function. Hence, the phase information contained in the object wavefront is lost. In 1948, Gabor originated the basic ideas of wavefront reconstruction through which both the phase and amplitude information could be obtained.

It involves two steps. First, a complex interferogram (called a *hologram*), formed by interference between the object and reference wavefronts, is recorded. Second, the hologram is suitably illuminated to reconstruct the original object beam in space.

Holography remained in a dormant stage until the early 1960s when highly coherent (long coherence length) laser sources became available. Considerable improvements in the process were effected. With the laser becoming an integral part of the practice of holography, this subject developed very fast into an active field that has provided new measurement techniques and solutions to a wide range of difficult problems.

10.1 Recording of a hologram

In principle, holography is a kind of lensless photography. The object is illuminated by laser light. The transmitted or scattered object beam is allowed to interfere with a reference beam derived from the same source. A suitably placed photographic plate records the hologram.

Any geometry employed in a comparison-type interferometer, with suitable modification to ensure that the reference beam subtends a certain minimum angle with the object beam at the plane of recording, can be used in making holograms.

The quality of a hologram is influenced by several factors, namely, the various modes present in laser light, diffraction caused by the laser aperture speckles, and stray vibrations. Their effects are minimized to some extent by applying spatial filtering of the incident laser beam using a pinhole placed at the focus of a microscope objective; by properly attenuating the reference wavefront for intensity matching; by mounting the experimental setup on a good vibration isolation table; and by employing high-speed, high-resolution photographic emulsions (nonconventional media are also used) for recording of the hologram.

When the interfering beams fall on the same side of the recording emulsion, a transmission hologram is obtained. To record a reflection hologram, they are introduced from opposite sides of the emulsion, enabling the interference to take place within the depth of the photosensitive layer.

10.2 Wavefront reconstruction

When a transmitting hologram is illuminated by a reference wavefront (spatially filtered as before), the object wavefront and its complex conjugate are reconstructed in space together with two more beams of light travelling near the transparency axis. The latter two beams consist of the directly transmitted reference beam and a beam whose intensity is proportional to the object intensity; these are of little consequence. The object wavefront produces a virtual image of the original object behind the hologram (as seen through it), while the complex conjugate component forms a real image in front. Angular separation between them, and from the axially traveling beams, depends upon the angle between the object and reference beams during recording. The reconstructing beams need not be identical to the recording reference beam. For example, use of a longer wavelength and more divergence give rise to image magnification.

In the case of a reflection hologram, reconstructed wavefronts are obtained on reflection under white light illumination. Holographic images are three dimensional in nature, retaining the parallax effect.

10.3 Applications of holography

The basic concepts of holography were initially introduced to achieve very high magnification in microscopy (Gabor, 1948). Using a shorter wavelength, for example, an electron or X-ray beam, for recording, carrying out reconstruction with a visible radiation wavelength, provides a very high order of magnification in addition to large depth of focus.

Holograms, because of the three-dimensional imaging properties, are used for display and demonstration purposes. Three-dimensional motion and television pictures by holographic principle are potential applications, though yet to be fully developed.

Holographic interferometry utilizes the interference patterns(s) formed by making multiple exposures on the same photographic plate, corresponding to different stages of the object wavefront in a dynamic situation. Vibration analyses have thus been possible by making a continuous exposure of the vibrating object. Instead of photographic plates, the interference patterns can be recorded in a computer, and online monitoring of the vibrating object is now possible. Phase change in the surrounding medium due to the passage of a high velocity missile, thermal gradients of objects, hot spots in electronic integrated circuits, mechanical strain, and particle size distribution are some of the applications of this technique.

The holographic technique has also been applied in the fabrication of diffraction gratings and simulated optical elements and in data storage and information processing.

Standard holographic technique is very sensitive to environmental disturbance, making the measurement difficult. A modified technique called *shearography* (Hung and Durelli, 1979) is robust and free from interference from the surrounding. This technique compares the deformation of one point B in the object with respect to another undeformed point A in the same object (see Figure 10). A coherent optical beam from the light source is split into two beams, one hits the object at B and the other at A. We assume the surface of the object is reflective. The reflected beams

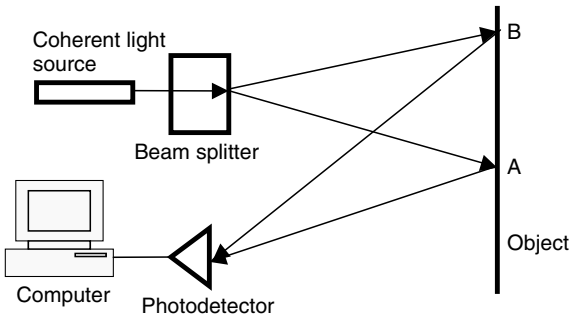


Figure 10. Shearographic setup for monitoring defects.

from A and B fall onto the photodetector, which is a square law detector. Hence, interference fringes are obtained and recorded in the computer. Now, if B undergoes a deformation while A remains unchanged, a new set of interference fringes is obtained and recorded in the computer. The computer then subtracts the two sets of fringes. The result is a set of new fringes due to the deformation at B. Instead of using the computer for signal processing, a photographic film can be used. In this case, the film has to undergo a double exposure before development.

REFERENCES

Ahmad, A. (1996) *Handbook of Optomechanical Engineering*, CRC Press, Baco Raton, FL.

Buchdahl, H.A. (1954) *Optical Aberration Coefficients*, Oxford University Press, London.

Clark, A.D. (1973) *Zoom Lenses*, Adam Hilger, Bristol.

Coltman, J.W. (1954) The Specifications of Imaging Properties by Response to a Sinewave Input. *Journal of the Optical Society of America*, **44**, 468.

Conrady, A.E. (1957 and 1960) *Applied Optics and Optical Design*, Parts I, II, Dover, New York.

Cox, A. (1964) *A System of Optical Design*, Focal Press.

Dyson, J. (1970) *Interferometry as a Measuring Tool*, Machinery Publishing Company, Brighton.

Francon, M. (1963) *Modern Applications of Physical Optics*, Wiley, Chichester.

Francon, M. (1966) *Optical Interferometry*, Academic Press, New York.

Gabor, D. (1948) A New Microscope Principle. *Nature*, **161**, 777.

Hartfield, E. and Thompson, B.J. (1978) Optical modulators, in *Handbook of Optics*, (eds W.G., Driscoll and W. Vaughan) McGraw-Hill, New York.

Hobbs, P.C.D. (2000) *Building Electro-Optical Systems: Making It All Work*, Wiley-Interscience.

Hopkins, H.H. (1950) *Wave Theory of Aberrations*, Oxford University Press, London.

Hopkins, H.H. (1955) Interferometric Methods for the Study of Diffraction Images. *Optica Acta*, **2**, 23.

Horne, D.F. (1974) *Dividing, Ruling and Mask-making*, Adam Hilger, Bristol.

Hung, Y.Y. and Durelli, A.J. (1979) Simultaneous Measurement of Three Displacement Derivatives using a Multiple Image-shearing Interferometric Camera, *Journal of Strain Analysis*, **14**(3), 81–88.

Jacobs, D.H. (1943) *Fundamentals of Optical Engineering*, McGraw-Hill, New York.

Jain, P.K. and Sydenham, P.H. (1980) Radial Metrological Pattern Generating Engine. *Journal of Physics E: Scientific Instruments*, **13**, 461–6.

Levi, L. (1968) *Applied Optics*, Vol. 1, Wiley, New York.

Macleod, H.A. (1969) *Thin Film Optical Filters*, American Elsevier Publishing Company, New York.

Murata, K. (1966) Instruments for the measuring of optical transfer functions, in *Progress in Optics*, Vol. V, (ed E., Wolf), North-Holland, Amsterdam.

O'Neill, E.L. (1963) *Introduction to Statistical Optics*, Addison-Wesley, Reading, MA.

Prasad, J., Mitra, G. and Jain, P.K. (1975) Aberration of a System of Arbitrarily Inclined Planar Surfaces Placed in Non-collimated Light Beam. *Nouvelle Revue d'Optique*, **6**, 345.

Prasad, J. and Singh, R. (1970) Tolerances in the Manufacture of Precision Optical Components. *Research and Industry*, **15**, 15.

Schade, O.H. (1964) An Evaluation of Photographic Image Quality and Resolving Power. *Journal of the SMPTE – Society of Motion Picture and Television Engineers*, **73**, 81.

Smartt, R.N. and Strong, J. (1972) Point-diffraction Interferometer. *Journal of the Optical Society of America*, **62**, 737.

Smith, W.J. (1966) *Modern Optical Engineering*, McGraw-Hill, New York.

Smith, W.J. (1978) Image Formation: Geometrical and Physical Optics, in *Handbook of Optics*, (eds W.G., Driscoll and W., Vaughan), McGraw-Hill, New York.

Smith, W.J. (2000) *Modern Optical Engineering*, McGraw-Hill Professional, New York.

Sydenham, P.H. (1969) Position-sensitive Photocells and Their Application to Static and Dynamic Dimensional Metrology. *Optica Acta*, **16**, 377–89.

Weber, M.J. (2002) *Handbook of Optical Materials*, CRC Press, Baco Raton, FL.

This article is based in part on a chapter by J. Prasad and G. Mitra originally published in Handbook of Measurement Science Vol 2, P.H. Sydenham, Wiley, 1983.

91: Light Sources and Detectors

Miroslaw Jonasz
MJC Optical Technology, Beaconsfield, Quebec, Canada

1 Light Sources 593
2 Light-emitting Diodes (LEDs) 595
3 Lasers 595
4 Detectors 597
5 Thermal Detectors 597
6 Photon Detectors 599
Related Articles 602
References 602
Further Reading 602

1 LIGHT SOURCES

Light sources may be selected on the basis of several criteria, including:

- power
- spectral range
- wavelength tunability
- duration of emission (continuous, also referred to as CW for *continuous wave* or pulsed)
- coherence
- focusability.

Two groups of light source types have evolved commercially

- low-coherence broad-spectrum sources (incandescent filament lamps, arc lamps, pulsed arc lamps, infrared incandescent sources, light-emitting diodes, i.e. LEDs)
- high-coherence, nearly monochromatic sources (lasers).

Key parameters of light sources are summarized in Table 1. Irradiance spectra of representative light sources are shown in Figure 1.

1.1 Incandescent filament lamps

Incandescent filament lamps generate electromagnetic radiation by resistively heating a filament, made usually of tungsten, to about 3000 K. The filament temperature, and, thus, the maximum of the spectral output (Table 1) depend strongly on the operating voltage, typically in a range of few volts to a few hundred volts. Various configurations of the filament, usually coiled-coil, and the envelope are widely available, including designs featuring integrated ellipsoidal dichroic reflectors, transparent in the infrared (IR), that increase the visible output power.

Spectral output of these lamps (Figure 1) resembles, but is lower than, thermal (blackbody) radiation spectrum with about 90% of the spectral output in the infrared. The spectral output of the filament is modified by the transmission spectrum of the envelope (glass or quartz). Being roughly omnidirectional, extended-area light sources, these lamps make it difficult to collimate or focus light output, for example, launch light into an optical fiber.

Incandescent tungsten lamps are available as vacuum lamps, inert gas filled lamps, and halogen lamps, in the order of the increasing filament temperature. In halogen lamps, a small quantity of iodine or bromine compounds added to the inert gas fill used maintains the lamp output at close to 100% throughout the lamp life, provided the lamps are operated at their nominal voltage to maintain the envelope temperature required for the halogen cycle.

Owing to a relatively large thermal mass of the filament, incandescent lamps are used generally as continuous, steady-output light sources. For high-stability applications, stabilized DC power supplies are required.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

Table 1. Key characteristics of electromagnetic radiation sources (CW = continuous output).

Approximate spectral range	Approximate maximum output wavelength	Approximate power range	Approximate pulse length range	Light source type
150 nm–2.5 μm[a,b]	500 nm	1–300 kW 0.1–10 J[g]	1 ns–1 ms	Gas arc lamp[c]
150 nm–20 μm[e]	Varies[h]	0.1 mW–10 kW 1 pJ–20 J[g]	10 fs–10 ms	Laser[d]
250 nm–2.5 μm[b]	0.8–1.2 μm[f]	1–10 kW	–	Incandescent filament lamp[c]
460–950 nm	Varies	0.1–1 mW	1 ns–CW	LED[c]
1–30 μm[a]	Varies	1–100 W	–	Infrared light sources[c]
0.1–100 μm[a]	Varies	0.1–1000 W	–	Blackbody[c]

[a]Broad, thermal spectrum.
[b]Sharp spectral peaks at wavelengths dependent on the fill gas.
[c]Incoherent light source, difficult to focus and collimate.
[d]Coherent light source, easy to focus and collimate.
[e]Nearly monochromatic output at one or more wavelengths within this range.
[f]Depends on the filament temperature and lamp type (halogen, gas filled, or vacuum).
[g]Pulsed output.
[h]May be wavelength-tunable.

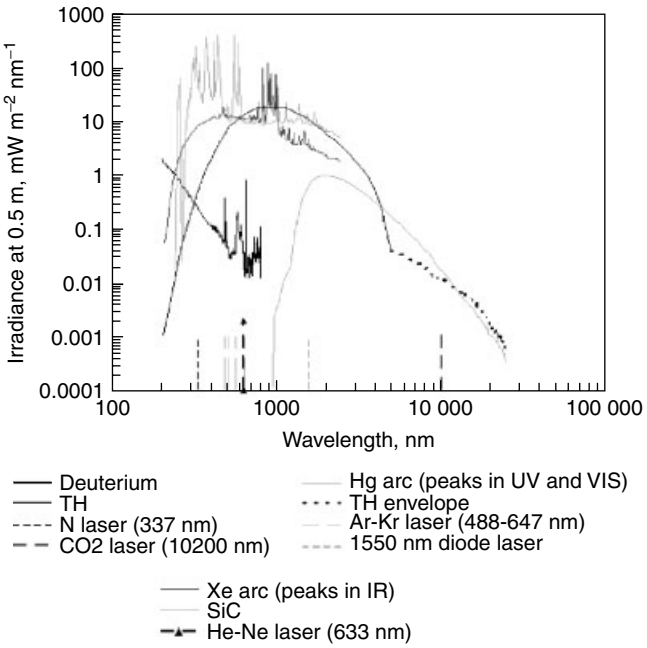


Figure 1. Spectra of representative light sources. Thermal (broadband) light sources spectra are based on the data from a catalog of Oriel Instruments (now Spectra Physics, www.spectra-physics.com): 150-W Xe arc lamp, 200-W Hg arc lamp, 30-W deuterium lamp, 100-W tungsten halogen (TH) lamp (dashed line above about 5000 nm represents the output of the lamp envelope), and 50-W silicon carbide envelope-less infrared element. Wavelength ranges of the lamps' spectra are limited by the envelopes – in the case of a quartz envelope to a range of about 200 to between 2500 and 3500 nm. Spectral lines of few major fixed-wavelength laser types are indicated symbolically with short dashed vertical lines. The vertical range of these lines does not represent the magnitudes of the irradiances that are many orders or magnitude higher than the maximum of the graph scale. For example, a typical 1-mW HeNe laser ($\lambda = 633$ nm, $\Delta\lambda = 0.0002$ nm, beam diameter 1 mm^2) produces irradiance on the order of 5×10^9 mW m^{-2} nm^{-1}.

1.2 Arc lamps

1.2.1 CW arc lamps

In arc lamps, optical radiation is generated by an electric discharge (as an arc) through ionized gas.

The spectral output of an arc lamp (Figure 1) consists of a broad, blackbody-type output and sharp peaks at wavelengths that depend on the gas fill. Gases used in these lamps are typically deuterium, mercury vapor (Hg), xenon (Xe), and a Hg–Xe mixture. The deuterium lamp outputs a smooth spectrum usable from about 160 (envelope limited) to about 400 nm with spectral peaks in an essentially unusable, owing to low spectral power density, visible spectral range. The Hg and Xe lamps (envelope-limited useful range of 250 to 2500 nm) have major peaks at 365.0, 404.6, 434.8, and 546.1 nm (Hg) and between 750 and 1000 nm (Xe). Owing to a high gas pressure in these lamps (of the order of 10 to 100 atm at the operating temperature), the spectral peaks are significantly collision broadened (on the order of several nanometers).

The generally small arc (on the order of 0.1 to 1 mm) allows relatively efficient collimation or focusing of the lamp radiation. Compact Xe arc lamps containing an integrated sealed-in paraboloidal reflector that produces a 'collimated' beam with a moderate divergence are commercially available (CeraLux™, Luxtel, www.luxtel.net). The arc position in an arc lamp varies in time (arc wander) and may cause variations in the output power, especially when this output is collimated or focused.

1.2.2 Xe flash (pulsed) lamps

These arc lamps range from small flashlamps for instruments, to large flashlamps such as those used as pump

sources in lasers. The flash (of the order of microseconds long) is produced by discharging a capacitor. Flash repetition rates are on the order of 100 pulses/s.

1.3 Infrared incandescent sources

The radiating elements of these sources are made of high-temperature-resistant materials such as silicon carbide (Globar), and a sintered mixture of oxides of zirconium, yttrium, thorium, or erbium (Nernst glower). These elements, with sizes on the order of several millimeters, are heated resistively to a temperature on the order of 1000 to 2000 K. The radiating elements are operated in the air with no envelope, limiting the spectral ranges of these sources (Table 1).

Incandescent IR sources have spectral outputs (Figure 1) similar to that of the blackbody, but their wavelength-dependent emissivity is less than unity, typically on the order of 0.9. Blackbody sources operable at temperatures ranging from 400 to 3000 K are brighter (emissivity >0.99), but an order of magnitude more expensive alternatives to these IR emitters.

2 LIGHT-EMITTING DIODES (LEDs)

The light-emitting diodes are relatively narrowband (20–50 nm), low-voltage (2 V), and low-power sources (Table 1) available in a wide spectral range spanning the visible and near IR. White-light LEDs that use a phosphor to convert the base emission wavelength range into a broad range perceived as white light are also available. LEDs convert electric current ($\sim$20 mA) to light by recombining charge carriers in semiconductors. The semiconductors used include SiC, GaN, ZnSe, GaInN (blue), GaP (green), GaP:N (yellow), and AlInGaP, GaAsP (red).

LEDs are available in a wide range of packages, ranging from plastic encapsulation to TO-type metal transistor cans with flat or lensed windows. The emitting area is on the order of a fraction of 1 mm in size. Plastic and glass lensed packages provide rough collimation of the output, but a better result in collimating or focusing a LED output can frequently be obtained with flat-window packages that generally provide a nearly Lambertian spatial radiation pattern.

3 LASERS

The laser (for example, Hecht, 1992) consists of an optical resonator cavity, created by two opposing, high-reflectance mirror surfaces (generally slightly concave), and gain medium inside that cavity. The cavity supports a set of optical radiation modes (resonances) each embodying a discrete wavelength and irradiance pattern in a plane perpendicular to the optical axis of the cavity. The gain medium is selected so that its absorption band overlaps with the wavelength of at least one of the cavity modes and that the radiative energy transition is the most probable transition corresponding to that absorption band. In lasers that employ gain media with wide spectra, prisms or gratings are used to select a subrange of the gain medium spectrum.

By 'pumping' the gain medium with an electric discharge, current, or light, one excites electrons of that medium to an energy state. Electrons falling (initially spontaneously) from that state to a lower one emit light of various radiation modes. Competition, via stimulated emission, of modes supported by the cavity for energy supplied to the gain medium quickly leaves a few dominant modes inside the cavity. The simplest and commonly provided mode, TEM_{00}, has an axially symmetric Gaussian irradiance profile that is frequently associated with the laser beam in general. The process of generation of light by a laser is more akin to that used in an electronic frequency generator than to optical noise generation characteristic of a thermal light source.

The cavity is deliberately made slightly leaky ($<1\%$ loss) in order to let laser radiation out. In pulsed lasers that use Q-switching technique for generation of high-power moderately short pulses (on the order of nanoseconds long), the cavity quality (Q-factor) is spoiled for a sufficient time to build a significant population of electrons in an excited state and then it is temporarily restored to generate a short pulse of radiation. Trains of much shorter pulses (several femtoseconds) can be obtained by mode locking, that is, introducing a fixed phase relationship between the cavity modes. The pulse train is a result of interference between the locked modes.

The cooperation of the resonant cavity properties, gain medium absorption profile, and the stimulated emission gives the laser beam its directionality, high monochromaticity, and coherence. Stabilized frequency lasers have line widths $<1 \times 10^{-11}$ nm (3 MHz), and coherence lengths on the order of 1 km. Key characteristics of representative laser types are summarized in Table 2. Wavelengths of major laser types are shown in Figure 1.

3.1 Gas lasers

These lasers span a wide wavelength range from UV (excimer, for example, XeF, XeCl; nitrogen), through visible (argon, argon/krypton, copper vapor, helium–neon, helium–cadmium) to far IR (CO_2). The gain medium is gas excited by an electric discharge. The gas lasers provide good-quality low-divergence beams with the diameters on the order of 1 mm.

Table 2. Basic data on major laser types listed according to the increasing first wavelength of the spectral range. CW denotes continuous output. This table is mainly based on the *Laser FocusWorld* 2003 Buyer's Guide. Additional data obtained from specifications of selected manufacturers have been also used. Users of this table should exercise caution when applying the data given.

Wavelengths (nm)[m]	Laser type	Power (W) or pulse energy (J)	Pulse repetition rate ($1\,s^{-1}$)[l]	Pulse width (s)	Beam size[a] (mm)	Beam divergence (mrad)
154–532[f], (1064), 2950	Nd:YAG[c] (pulsed)	0.001–180 J	$1–50 \times 10^3$	$20 \times 10^{-12}–10 \times 10^{-3}$	1–14	0.5–130
157, 193, 248, 308, 351	Excimer[b] (pulsed)	0.001–0.2 W	$20–2 \times 10^3$	$0.01–30 \times 10^{-9}$	2.5×22	0.3–4
190–18000[h,k]	Tunable solid state (pulsed)	$0.1 \times 10^{-9}–100$ J	$0–90 \times 10^6$	$17 \times 10^{-15}–100 \times 10^{-9}$	3–10[g]	0.5–1.5[g]
197–4500[h]	Tunable dye (pulsed)[g]	$30 \times 10^{-6}–0.2$ J	$1–10 \times 10^3$	$5 \times 10^{-12}–20 \times 10^{-9}$	$1–3 \times 6$	5
202–1090[i] (514)	Argon[b] (CW)	$2 \times 10^{-3}–25$ W	–	–	0.6–3.5	0.05–3
266–532[f] (1064), 1320	Nd:YAG[c] (CW)	$0.01–5.4 \times 10^3$ W	–	–	0.7–6	1–200
(337)	Nitrogen[b,g] (pulsed)	$0.1 \times 10^{-3}–0.5 \times 10^{-3}$ J	1–100	$\sim 1 \times 10^{-9}$	1×2	2×3
375–760[h]	Alexandrite[c,g] (CW)	0.01–0.15 W	–	–	1	0.1
375–2900[h] (500–1550)[h]	Semiconductor diode (CW)	$0.1 \times 10^{-3}–4 \times 10^{3}$[e] W	–	–	–	$1–17 \times 1–17$[j,d], 87×870[d]
380–1935[h] (600–1550)[h]	Semiconductor diode (pulsed)	$5 \times 10^{-12}–10^{2}$[e] J	$100–100 \times 10^6$	$15 \times 10^{-12}–0.8 \times 10^{-3}$	–	0.3[j], 87×785[d]
543, (632.8), 1152, 3392	HeNe[b]	$0.1 \times 10^{-3}–35 \times 10^{-3}$ W	–	–	0.5–2.7	0.7–4.5
690–2000[h]	Ti: sapphire[c,g] (CW)	0.05–3.5 W	–	–	0.6	1.7
694	Ruby[c,g] (pulsed)	0.08–20 J	0.005–0.1	$20 \times 10^{-9}–2.5 \times 10^{-3}$	7–25	0.2–1
3000–25 000	Lead salt diode[g] (CW)	0.1×10^{-3} W	–	–	–	870
5200–6500[h]	CO[b,g] (CW)	$10–1 \times 10^3$ W	–	–	5–16	4
9000–11 400 (10 600)	CO_2[b] (CW)	$0.04–12 \times 10^3$ W	–	–	0.5–22	0.5–9
9000–11 400 (10 600)	CO_2[b] (pulsed)	0.5–60 J	$0–130 \times 10^3$	0.3 to over 10^{-3}	$1.8–30 \times 30$	1–10

[a] If one value is given, it is the TEM_{00} (Gaussian) beam diameter.
[b] Gas laser.
[c] Solid-state laser.
[d] In two perpendicular dimensions.
[e] The upper end of the range refers to the output of laser stacks.
[f] Through harmonic frequency conversion.
[g] Few data.
[h] Continuously tunable, may be tunable in a subrange of the stated range.
[i] Line selection.
[j] Diode lasers equipped with collimating optics.
[k] Line widths from several GHz down to 1 kHz.
[l] A repetition rate of 0 denotes a single-shot mode.
[m] Dominant wavelengths in parentheses.

3.2 Semiconductor (diode) lasers

These lasers (for example, Camparo, 1985) are similar to LEDs in that they use a recombination of charges in semiconductors to generate emission of light. Owing to a very small emitting area of unequal sizes (on the order of 10 μm) in two perpendicular directions, these lasers typically emit a divergent astigmatic light beam and require complicated optical systems to either collimate or focus it. In fiber-optics applications, the small emitting area can be simply butt-coupled to a fiber.

Semiconductor diode lasers are also available in vertical emitting cavity (VCSEL) configuration that emit low-divergence, circular beams. Although these lasers are more difficult structures to grow on a semiconductor wafer, they do not need to be cleaved to produce the resonant cavity and, thus, can be fabricated solely by using standard semiconductor processing techniques (for example, Lang, 2002).

3.3 Dye lasers

Dye lasers use as their gain media various dyes (including coumarin, fluorescein, and rhodamine) covering a wavelength range from about 190 to 4500 nm. These lasers are pumped with either flashlamps or other lasers, for example, nitrogen lasers. The wide absorption bands of the dyes, as compared, for example, to narrow absorption spectra of gases, support the generation of very short light pulses (several femtoseconds long).

3.4 Solid-state lasers

3.4.1 Fixed-wavelength lasers

Solid-state lasers (SSLs) utilize a solid gain medium, permitting high-quality beams, and various pumping mechanisms. The first solid-state laser used the ruby crystal as the gain medium and a flashlamp as the pump. SSLs are also pumped by other lasers and by semiconductor diode lasers (diode-pumped solid-state lasers – DPSSL). In addition to ruby, the gain media include neodymium-doped (Nd:YAG, Nd:YLF, Nd:YVO_4) and holmium-doped (Ho:YAG, Ho:YLF) crystals and glasses, alexandrite and titanium-doped sapphire (Ti:sapphire). As in dye lasers, the wide absorption bands of the solid gain media support the generation of extremely short light pulses.

These lasers emit in the red and near IR range. Radiation in the near-UV and blue/green region of the visible spectrum can be generated by the harmonic up-conversion (for example, 532 nm in an ND:YAG laser).

3.4.2 Tunable lasers

Tunable solid-state lasers, available mainly as pulsed lasers, operate by using principles similar to those employed in dye lasers except that the gain medium is solid state. In particular, a tunable titanium-doped-sapphire (Ti:sapphire – the broadest range of any tunable laser: 660–1180 nm) laser has become a widely used replacement for tunable dye lasers. However, the Ti:sapphire laser has the noise lower by a factor of 10 as compared to the dye laser and also has a much narrower line width (down to 1 kHz). As with other SSLs, up-conversion is used to provide tunable light in the UV and visible range.

4 DETECTORS

Developments in electro-optical detector technology (for example, Donati, 2000) combined with advantages of direct electronic signal processing have long displaced the human eye and photographic film as detectors of light in all but marginal applications. Two basic groups of electro-optical detectors are discussed here: thermal and photon detectors.

An electro-optical detector of light is selected on the basis of one or more of the characteristics listed in Table 3. Some detector types may have additional selection criteria, for example, after-pulsing for photomultipliers.

The sensitivity is frequently the dominant selection criterion. Hence, the ranges of light power that can be measured with the various detector types are listed in Table 4 and shown in Figure 2 for major detector types.

5 THERMAL DETECTORS

In a thermal detector, the active element is heated via absorption of light. This heating induces electrically sensed changes in that element's properties. Thermal detectors have a flat spectral response in a broad spectral range, determined by the absorption spectrum of the sensing element material or its absorptive coating that converts the energy of the incident radiation to heat. The detector spectral range is limited/modified by the detector window/filter. A weak thermal link to a heat sink allows the detector to correctly respond to the incident radiation power. Thermal detectors are generally less expensive than photon detectors, and do not require cooling when operated in the infrared. However, thermal detectors are typically slower and have a lower sensitivity than photon detectors.

The basic types of thermal detectors are:

- thermocouples
- thermopiles

Table 3. Key characteristics of radiation detectors.

Name and typical symbol	Typical units	Definition
Responsivity, R	A/W, V/W, V/J[a]	Ratio of the detector output to the incident light power dose
Quantum efficiency, QE[b]	Nondimensional	Number of electrons generated by the detector in response to an input of one photon of radiation
Dark current, I_D, or dark voltage, V_D	A, V	The baseline output of the detector in the absence of incident light
Response time, t_R	s	Time interval required for a detector to generate an output commensurate with the input light power (dose) when that detector is illuminated by a step increase in the light power (rise time) or when the illumination ceases abruptly (fall time)
Linearity range	W	A range of the input light power within which the detector output is linearly proportional to the input power
Noise equivalent power, NEP	W, W/$\sqrt{\text{Hz}}$	Light power that generates a detector output equal to the RMS detector noise[c]

[a]For detectors sensitive to the radiation dose.
[b]Applicable to photon detectors only.
[c]With the detector noise being a function of the measurement (electrical) frequency bandwidth, NEP is usually given at a frequency of 1 kHz and for a 1-Hz bandwidth. An amplifier, if needed, may significantly increase the NEP of a complete detection system.

Table 4. The first order of magnitude incident radiation power ranges that can be measured by the various detector types. The minimum power is the detector noise equivalent power at the wavelength of the maximum responsivity. PMT = photomultiplier, APD = avalanche (semiconductor) photodiode.

Approximate power range[h], W		Detector type	Approximate spectral range[b]
Minimum	Maximum		
10^{-19} [a]	10^{-12}	PMT: photon counting[e]	100–900 nm
10^{-19} [a]	10^{-12}	APD: photon counting[e]	200–2500 nm
10^{-15}	10^{-9} [c]	PMT: analog detection[e]	100–900 nm
10^{-13}	10^{-6}	APD: analog detection[e]	200–2200 nm
10^{-12}	10^{-1}	Semiconductor photodiode[e]	0.1 nm–5 μm
10^{-12}	1	Semiconductor photoconductor[e]	400 nm–150 μm
10^{-9}	10^{-3}	Thermopiles[f]	200 nm–200 μm
10^{-9}	10^{-3}	Pyroelectric[f]	200 nm–1 mm
10^{-8} [d]	1	Bolometer[f]	0.1 nm[g]–5 mm

[a]This power corresponds to ~1 photon/s in the visible and near infrared.
[b]Various detectors of the stated type (except thermal detectors) may have their maximum sensitivity located at different wavelengths in this range and be sensitive each in a subrange of this range.
[c]A much higher maximum power (about 7 mW) was measured with short (~1 μs) light pulses (Lasher and Redhead, 1963) at the cost of high mortality of the PMTs.
[d]down to 10^{-17} in the far infrared at a temperature close to 0 K.
[e]Photon detector.
[f]Thermal detector.
[g]Can be used as a photon counter/spectrometer for X rays.
[h]The dynamic range of a detector may be smaller than the min–max range.

- pyroelectric detectors
- bolometers.

Golay cells (pneumatic detectors) were the mainstay thermal detectors in the past but have been all but displaced by other thermal detector types.

5.1 Thermocouples and thermopiles

Thermocouples and thermopiles (a series of thermocouples) utilize the temperature dependence of voltage generated by a junction of dissimilar metals, usually bismuth and antimony. The thermocouple consists of two such junctions connected in series in reverse in order to cancel the effect of the ambient temperature – *see* **Article 196, Thermocouple Temperature Sensors, Volume 3**. One of these junctions is coated with a thin light-absorbing layer. The other is shielded from the incident light, that is, remains at the ambient temperature. The voltage generated by a thermocouple is low (in the mV range); therefore, several thermocouples are typically connected in series (thermopiles).

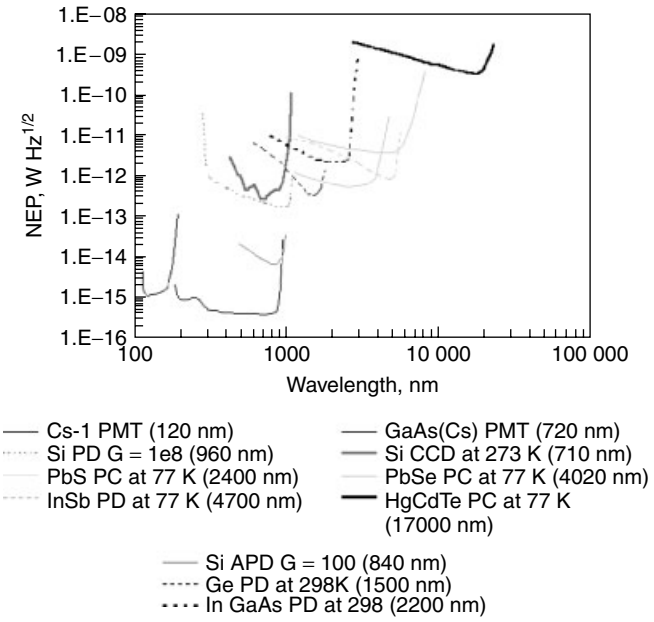


Figure 2. Sample spectra of the noise equivalent power (NEP) for major types of photon detectors (the wavelength of the minimum NEP is given in parentheses in the curve label). They indicate approximately the minimum power that can be measured by a given detector type at a signal-to-noise ratio of 1. The NEP of a specific detector of that type may vary, depending on the detector area and temperature. Abbreviations: APD = avalanche photodiode, LAAPD = large-area avalanche photodiode, G = gain, PC = photoconductor, PD = photodiode, PMT = photomultiplier. PC and PD detectors require external signal amplifiers that typically significantly increase the total NEP of the detection system. Long-term signal integration (on the order of seconds) may reduce the total NEP in such a case. The field of view of the far IR detectors is limited to 60°. The NEP is generally specified at a modulation frequency of the optical input of about 1 kHz, with a bandwidth of 1 Hz. However, an external amplifier with a high gain required for the PDs may significantly limit the maximum modulation frequency. Detector areas are on the order of 0.01 cm^2, except those of the PMTs (~0.5 cm^2) and the LAAPD (~0.1 cm^2). Pyroelectric, thermopile, and bolometer detectors (not included) have a flat spectral response with NEP ~10^{-9} to 10^{-11} (at several Hz) typically from UV to far IR. This spectral range is limited/modified by the window/filter transmission range. The graph is based on the data of the following manufacturers: Si PD and the PMTs – Hamamatsu (1994), www.hamamatsu.com, Si APD – Perkin Elmer, www.perkinelmer.com, Si LAAPD – Advanced Photonics, www.advancedphotonix.com, PbS and PbSe PC – New England Photoconductor, www.nepcorp.com, Ge, InGaAs, InSb PDs, and HgCdTe PC – Judson Technologies, www.judtech.com.

5.2 Pyroelectric detectors

Pyroelectric detectors use materials, typically lithium tantalate, that maintain electrical polarization even when no voltage is applied to the detector. When an active element of a pyroelectric detector is heated via absorption of light, it expands. This changes the polarization of this element and causes current to flow in a circuit connected to it. Pyroelectric detectors thus respond to changes in the incident light power, that is, to chopped or otherwise modulated radiation power.

5.3 Bolometers

Bolometers utilize changes in the conductivity of a resistor to sense its temperature that is affected by absorption of radiation. Room-temperature bolometer arrays allow inexpensive imaging of IR radiation, as compared to other IR imaging arrays.

When operated at room temperature, these devices are relatively poor detectors. Chopping, or modulating the incident radiation power, can be used to reduce the effect of slow-changing ambient temperature. Room-temperature bolometer detectors may also utilize two identical bolometers (one exposed to the incident radiation, the other shielded from it) in two arms of an electrical bridge. This arrangement cancels the effect of the ambient temperature on the detector output. However, the full potential of these devices is realized by operating them at temperatures close to absolute zero, which reduces their inherent thermal noise. Cryogenically cooled bolometers operating in the far IR are by several orders of magnitude more sensitive than other detectors of electromagnetic radiation in that spectral range.

6 PHOTON DETECTORS

Photon detectors convert the energy of incident light photons to electricity directly. These detectors work in relatively narrow spectral ranges, with a distinct long-wavelength cutoff, but can be extremely sensitive, up to the detection of single photons. Biological detectors of light are nearly as sensitive. The human eye can detect a burst of several photons and the rat eye can reliably detect single photons (Hagins, Penn and Yoshikami, 1970).

6.1 Photoemissive detectors

Photomultipliers (PMTs; BI, 1980; Hamamatsu, 1994) are by far the most commonly used detectors in this class, followed by image intensifiers.

6.1.1 Photomultipliers

A classical photomultiplier is a vacuum tube with a photosensitive cathode (Table 5) and on the order of 10 dynodes

Table 5. Basic characteristics of typical photocathodes.

Photocathode type	Photosensitive material	Spectral range (nm)	Wavelength of maximum responsivity (nm)	Responsivity (mA/W)	Maximum quantum efficiency
S-1	Ag–O–Cs	400–1100	350–800	3.5	0.012
S-4	Cs–Sb	160–550	300	50	0.21
S-11	Cs–Sb	290–600	450	55	0.15
S-20	Na–K–Cs–Sb	160–720	400	50	0.15
–	GaAs (Cs)	160–850	400	37	0.22

that collect and multiply the photocurrent generated at the cathode. A high voltage of the order of 1000 V is distributed between the electrodes of a PMT by a voltage divider network. A photon striking the photocathode ejects an electron with the quantum efficiency (QE) of less than one-fourth. That electron is accelerated by the potential differences between the cathode and the following electrode and – upon impacting that electrode – ejects several next-generation electrons. This electron-multiplication process continues for each following dynode until the anode, where electrons are collected, and results in a total electron gain on the order of 10^6 to 10^8. A high gain and a low dark current on the order of 1 photoelectron/s (and less when cooled) enable the detection of single photons of the incident light. Classical PMTs are available generally in two forms:

- side-on (with a reflective photocathode)
- end-on (with a semitransparent photocathode).

The side-on form is typically a faster detector (rise time on the order of a fraction of 1 ns), due to a more compact electron multiplier design. Supercompact multianode PMTs that enable a low-resolution imaging of the incident irradiance distribution ($\sim 10 \times 10$ 'pixels') are also available. The sensitivity spectra of PMTs span the near-UV to near-IR wavelength range (Figure 2).

A radical variance of the vacuum-tube design is a multichannel-plate PMT (MCP PMT), where the multiplier section is a glass plate, about 0.5-mm thick, with closely packed narrow (on the order of 10 μm) cylindrical channels, each acting as a set of dynodes. The MCP offers the most compact PMT design allowing rise times on the order of 100 ps.

6.1.2 Image intensifiers

Image intensifiers (IMIs) use various forms of imaging electron multipliers, typically the MCP, to map the distribution of photoelectrons exiting the photocathode to that at the anode (for example, Johnson and Owen, 1993). The anode is a phosphor screen (metallized to conduct electric current) that emits visible light when impacted by electrons with energy sufficient to penetrate the screen metallization. With the MCP multiplier, the imaging of the input electron irradiance distribution is ensured by restricting each electron cascade to its own microchannel and by placing the photocathode, MPC, and the anode-phosphor screen close to each other (<1 mm). This creates compact devices with imaging resolutions of the order of 50 line pairs per mm. Photon gains up to the order of 10^9 are possible with devices that employ stacks of two or three MCPs.

The IMI is by itself an optical-to-optical converter and requires an array detector such as a CCD (discussed in Section 6.2) to convert an optical image to electronically processable image. Phosphor-less versions of an image-intensified CCD are also available. In these imaging detectors, the phosphor screen is replaced with a CCD embedded inside the vacuum tube.

6.2 Solid-state detectors

The sensitivity spectra of solid-state detectors span a wide wavelength range from the soft X rays to far IR. Spectra of major types of solid-state detectors are shown in Figure 2.

6.2.1 Photovoltaic detectors and detector arrays

In a photovoltaic semiconductor detector, the semiconductor junction separates charge carriers generated by absorption of photons of the incident light and produces current in an electric circuit connected to the detector. Photovoltaic detectors can be used in the low-speed short-circuit photovoltaic mode (no dark current and the related noise) and in the high-speed photoconductive mode with moderate bias current and resulting higher noise. A wide variety of semiconductor photovoltaic detectors are available commercially, including position-sensitive detectors and imaging (array) detectors. These detectors are typically made of Si, GaAs, Ge, InGaAs, InAs, and InSb semiconductors that absorb light in subranges of a wide range between X rays and far IR (Table 6). Detectors specialized for either low-noise, low-speed operation or high-speed operation (>100 GHz) are available. Semiconductor detectors have QE on the order of unity, much higher than that of PMTs,

Table 6. Basic characteristics of typical photovoltaic semiconductor detectors at room temperature, unless indicated otherwise.

Photosensitive material	Spectral range (μm)	Wavelength(s) of maximum responsivity (μm)	Maximum responsivity, A/W	Maximum quantum efficiency
Si	0.2–1.2	0.8–0.95	0.7	0.9
GaAs	0.3–0.75	0.6–0.72	0.4	0.7
Ge	0.8–1.8	1.55	0.9	0.7
InGaAs	0.7–1.7	1.6	1	0.8
	0.7–2.6	2.3[a]	1.1	0.6
InAs	1–3.1	3[b]	2	0.8
InSb	1–5.5	5.2[b]	2	0.5

[a] At 253 K.
[b] At 77 K (liquid nitrogen).

that is relatively insensitive to temperature changes, except near the long-wavelength cutoff (Figure 2).

Photovoltaic detectors are available both as:

- unity gain devices (photodiodes, photodiode arrays, CCDs, CMOS arrays)
- detectors with internal amplification: avalanche photodiodes (APDs) and APD arrays.

In the latter detectors, the cascade multiplication of photoelectrons is induced by applying a high voltage on the order of several hundred to several thousand volts. In the lower part of the high-voltage range, the number of electrons produced is proportional to the applied voltage, as it is in the PMT. In the high part of the voltage range, the cascade multiplication becomes free running. This latter mode of operation (Geiger mode), is used for single-photon counting. In that mode, the electron cascade is quenched to enable photon counting at a rate on the order of up to 10^6 photons/s (active quenching) limited by dead time and APD heating (for example, Dravins, Faria and Nilsson, 2000).

Semiconductor array detectors are the mainstream of imaging detectors. Arrays with pixel sizes on the order of $10 \times 10\,\mu m$ are available in the photodiode array, CCD, and CMOS formats. The photodiode arrays have individually addressable elements and offer high quantum efficiency but suffer from a substantial read-out noise (for example, Johnson, Edwards and Mendenhall, 1994). CCD arrays have a much lower readout noise but also a lower quantum efficiency. In a CCD array, the photogenerated charges, instead of being in real time removed from the sensitive volume as it is done in the photodiode, are stored in a potential well until they are transported out with an extremely high efficiency, along a column or row of pixels that perform the role of an analog shift register.

Charge integration and transfer are achieved at the rate of an external clock signal by varying potentials of an array of electrodes and semiconductor gates located at the CCD chip surface. It is these surface structures that reduce the quantum efficiency of the CCD array as compared to that of the photodiodes.

Back-illuminated operation of the CCDs compensates somewhat for this reduction. CCDs tend to suffer from 'blooming' caused by spilling charge from an overfilled pixel to the surrounding pixels (Janesick *et al*., 1987).

Recently, electron-multiplying CCDs (EMCCDs) have been commercialized (e2v technologies, www.e2v.com). These CCDs utilize, in the readout section, an electron-multiplication technique implemented in the APDs. This allows high gain (single-photon counting) with the low-noise characteristic of the CCDs.

The CMOS imaging arrays utilize silicon as photosensitive material and have a similar spectral response to that of silicon CCDs. However, the CMOS arrays use a different approach to sensing the incident light power and also different manufacturing techniques. In a pixel of a CMOS array, a photodiode typically discharges a storage capacitor, thus preventing blooming. The capacitor charge can be directly read out via an integrated amplifier.

This allows addressing selected sections of the array and enables a high-speed readout. The CMOS arrays employ technology used for microprocessors. This permits integration of preprocessing electronics, such as an analog-to-digital converter, on the array chip. However, additional semiconductor/metal structures implementing these functions take up a significant portion of the area available for light collection. Complexity of these structures also increases the manufacturing defect rate. Such disadvantages have generally limited applications of CMOS arrays to low-cost and low-quality consumer applications (Taylor, 1998).

6.2.2 Photoconductive devices

Photogeneration of charges in semiconductors significantly affects its conductivity. This is utilized in junctionless semiconductor detectors, photoconductors, for the detection of electromagnetic radiation. Typical semiconductors used

Table 7. Basic characteristics of typical photoconductive semiconductor detectors at room temperature, unless indicated otherwise.

Photosensitive material	Spectral range (max sensitivity) (μm)	Maximum responsivity (V/W)
CdS	0.4–0.75 (0.56)	– [b]
CdSe	0.6–0.9 (0.71)	– [b]
PbS	1–2.7 (2.2)	6×10^5
	1–4.2 (3.1)[a]	9×10^5 [a]
PbSe	1–5 (4)	7500
	1–6 (5)[a]	55 000[a]
HgCdTe	1–4[a] – 24[a]	10^5 [a]–200[a]

[a]At 77 K (liquid nitrogen).
[b]These detectors are used mainly as light switches, because of the sensitivity of their response to their light-exposure history.

in these photoconductors are CdS and CdSe (visible), PbS and PbSe (near IR), and HgCdTe (IR) (Table 7). Note that the semiconductor conductivity is also a significant function of temperature, so that the detector temperature must be stabilized.

The literature on optical engineering usually covers sources and detectors. A useful general book is OSA (1993). A useful reference work on infrared engineering is Hudson, 1969.

RELATED ARTICLES

Article 89, Optical Materials, Volume 2; **Article 90, Optical Elements, Volume 2**; **Article 92, Optical Measuring Instruments, Volume 2**; **Article 93, Testing Optical and Other Radiation Systems, Volume 2**.

REFERENCES

BI. (1980) *Photomultiplier Handbook*, Burle Industries, Inc, Lancaster.

Camparo, D.C. (1985) The Diode Laser in Atomic Physics. *Contemporary Physics*, **26**, 443–477.

Donati, S. (2000) *Photodetectors: Devices, Circuits and Applications*, Prentice Hall, NJ.

Dravins, D., Faria, D. and Nilsson, B. (2000) Avalanche Diodes as Photon-counting Detectors in Astronomical Photometry, in *Optical and IR Telescope Instrumentation and detectors*, (eds I. Masanori and A.F. Morwood); *SPIE Proceedings*, **4008** 298–307.

Hagins, W.A., Penn, R.D. and Yoshikami, S. (1970) Dark Current and Photocurrent in Retinal Rods. *Biophysical Journal*, **10**, 380–411.

Hamamatsu. (1994) *Photomultiplier Tube: Principle to Applications*, Hamamatsu Photonics KK, Hamamatsu, Japan.

Hecht, J. (1992) *The Laser Guidebook*, McGraw-Hill, New York.

Hudson, R.D. (1969) *Infrared System Engineering*, Wiley Inter-Science, New York.

Janesick, J.R., Elliot, T., Collins, S., Blouke, M.M. and Freeman, J. (1987) Scientific Charge-coupled Devices. *Optical Engineering*, **26**, 692–714.

Johnson, J.B., Edwards, G. and Mendenhall, M. (1994) Low-cost, High-performance Array Detector for Spectroscopy Based on a Charge-coupled Photodiode. *Review of Scientific Instruments*, **85**, 1782–1783.

Lang, R.J. (2002) Semiconductor Lasers: An Overview of Commercial Devices, in *Photonics Handbook*, Laurin Publishing, Pittsfield, MA, (pp. H215–H218).

Lasher, D.R. and Redhead, D.L. (1963) High-current Operation of RCA Type 2059 Photomultiplier Tubes. *Review of Scientific Instruments*, **34**, 115–116.

OSA. (1993) *The Handbook of Optics*, McGraw-Hill, New York.

Taylor, S.A. (1998) *CCD and CMOS Imaging Array Technologies: Technology Review*, Technical Report EPC-1998-106, Xerox Corporation, Cambridge, UK, (p. 15).

FURTHER READING

Johnson, C.B. and Owen, L.D. (1993) Image Intensified Electronic Imaging, in *The Handbook of Optics*, McGraw-Hill, New York, (pp. 21.1–21.32).

This article is based in part on a chapter by J. Prasad and G. Mitra originally published in Handbook of Measurement Science Vol 2, P.H. Sydenham, Wiley, 1983.

92: Optical Measuring Instruments

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Refractometers 603
2 Curvature Measurement 603
3 Autocollimators 604
4 Focometers 604
5 Microscopes 604
6 Telescopes 604
7 Projectors 606
8 Spectrophotometers 606
9 Opto-medical Instruments 607
10 Laser-based Instruments and Systems 607
Related Articles 609
Further Reading 609

1 REFRACTOMETERS

This article shows how the building blocks of optical systems – materials, **Article 89, Optical Materials, Volume 2**; elements, **Article 90, Optical Elements, Volume 2**; and sources and detectors, **Article 91, Light Sources and Detectors, Volume 2** – are assembled to form whole optical systems. Testing of optical systems is found in **Article 93, Testing Optical and Other Radiation Systems, Volume 2**.

Refractometers are employed for refractive index and Abbe value measurement. Depending upon the precision needed, several instruments are available commercially.

Abbe and Pulfrich refractometers are the most commonly used instruments for precise measurement. The design of both these instruments is based upon the total internal reflection principle.

An Abbe refractometer consists of two $90^\circ - 60^\circ - 30^\circ$ prisms, a pair of Amici prisms, and a telescope. One right-angled prism is used for illumination, while the other serves as the test specimen holder, which is placed over it with a liquid of high refractive index at the contact area. The Amici prisms are rotated in opposite directions to compensate for the dispersion introduced by the test specimen and stage prism. The telescope can be swiveled along an arc to pick up the total internally reflected light and its position is read on a scale directly calibrated in terms of refractive index. The instrument also incorporates the provision for measuring the refractive index of liquid samples.

In the Pulfrich refractometer, the emergent angle I' of the grazing incident rays, as measured by the telescope, is used for the determination of the refractive index N of the test samples, according to the equation $N = (n^2 - \sin^2 I')^{1/2}$, where n is the refractive index of the supporting glass block.

2 CURVATURE MEASUREMENT

Measurement of radius of curvature r of optical surfaces is conventionally made employing mechanical or optical spherometers.

In mechanical spherometers, the sag of the test surface is obtained by locating the correct contact point with the help of a plunger. This quantity in conjunction with other known parameters enables calculation of r.

The optical spherometer is based on the autocollimation principle. The instrument is first focused on the test surface and then at its center of curvature. The distance between the two image positions determines r.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

The well-known Foucault knife-edge test method is usually employed for testing long radius of curvature, concave surfaces, and assessing sphericity errors at different zones.

3 AUTOCOLLIMATORS

Parallelism of a glass plate, prism and wedge angles, and surface flatness can be efficiently tested using an autocollimator. It is also frequently used for alignment purposes.

When the surface under test is not exactly normal to the incident rays, the reflected image is transversely displaced. This image displacement is measured (manually and photoelectrically) precisely, yielding test surface inclination.

In the angle Dekkor method, the test specimen is placed horizontally upon a flat base, while the autocollimator can be swung in a vertical plane. A precision goniometer is essentially a spectrometer with an in-built autocollimating telescope.

4 FOCOMETERS

The two standard techniques conventionally used for focal length measurement of an optical system are the focal collimator and nodal slide methods. For better precision, an autocollimating microscope can be used.

A wide variety of optical instruments are available for various measurements in the industry. Brief details of some common types of instruments are now discussed.

5 MICROSCOPES

A microscope consists essentially of a condenser lens to illuminate the object, an objective lens for image formation, and an eyepiece optical system for viewing. Opaque objects are illuminated from the top, either by a beam splitter placed below the objective or by fixing miniature lamps by the side of the objective. An industrial version of a stereozoom microscope is shown in Figure 1.

A tool-room microscope eyepiece can be fitted with a special graticule pattern. Microthreads, tool shape, and the like are checked by matching their images with respect to a corresponding ideal pattern contained in the graticule.

Dimension-measuring traveling microscopes are mounted on a high-precision translating rail. The travel distance is read from a scale using an optical vernier.

In the case of metallurgical specimens (where it is important to study fine scratches in polished surfaces, faults, and inclusions, which, due to glare, are usually not visible under normal illumination), a special objective, carrying a beam splitter at its back, is used.



Figure 1. Stereozoom microscope. (Copyright © Bausch & Lomb.)

When the object has only variation in phase or refractive index and no differential absorption, a phase-contrast microscope is used. With the help of a phase-shifting device, the variations in the object are converted into intensity variations in the image. On phase retardation, a denser medium appears brighter and vice versa in the case of phase advancement.

In fluorescence microscopy, the specimen (usually biological) is dyed with an appropriate dye and observed, using filters, under ultraviolet light.

Study of the crystals is done on a polarizing microscope. Plane polarized light, after passing through the specimen, changes its state of polarization and the output beam is studied by an analyzer.

Interference objectives are basically Twyman–Green types of microinterferometers, which are coupled with a microscope to measure the microsurface structure of polished reflecting surfaces.

6 TELESCOPES

Basically, a telescope has an objective and an eyepiece. Astronomical telescopes of the reflecting type contain at least a large primary mirror and a small secondary mirror. Terrestrial telescopes are often of the refracting type fitted with additional inverting optics for image erection.

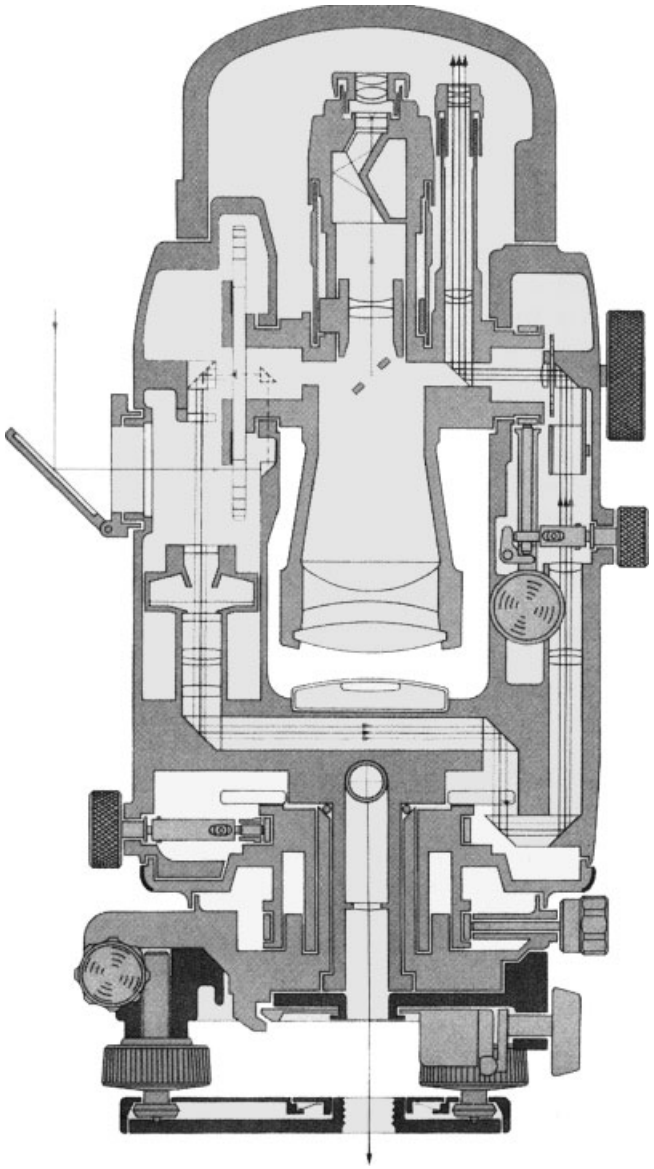


Figure 2. Micrometer theodolite. (Copyright © Wild Heerbrugg Ltd.)

The precision theodolite and sighting level are another important class of telescopic instrument. They are extensively used in land surveying, geodesy, and tacheometry. Figures 2 and 3 illustrate, as examples, the optical configurations of standard versions of the micrometer theodolite and universal automatic level.

As an example of an optical system, consider the technical data of a Wild micrometer theodolite shown in Figure 2. Telescope magnification with a standard eyepiece is 30×. The field of view at 1000 m is 27 m with a shortest focusing distance of 1.7 m. Bubble sensitivity per 2 mm run, for the circular level is 8 min of arc and for the plate level 30 s of arc. Liquid compensator setting accuracy is ±1 s of arc,

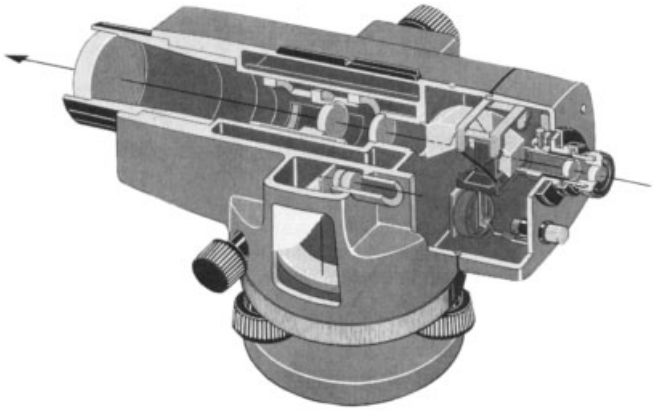


Figure 3. Universal automatic level. (Copyright © Wild Heerbrugg Ltd.)

and the working range is ±2 min of arc. Graduation intervals of the horizontal and vertical glass circles is 1° of arc with direct reading on the micrometer to 6 s of arc and by estimation, 3 s of arc.

In the case of the Wild universal automatic level shown in Figure 3, the compensator setting accuracy is enhanced to ±0.3 s of arc and working range to ±15 min of arc. For precise leveling, a parallel-plate micrometer, reading at 0.1 mm direct and 0.01 mm by estimation, is attached to the basic instrument.

Endoscopes are often employed in industry for observations inside high-temperature enclosures or other inaccessible areas. Figure 4 shows a commercial model of a portable endoscope for photographic and visual examination of the inside walls of operating furnaces and boilers. Photo-visual endoscopes suitable for use in very high pressure (25 kg cm^{-2}) chambers with corrosive atmospheres are



Figure 4. Portable endoscope. (Copyright © S.R. Clave.)

also available. A wide range of forms for medical use is covered in Section 9.

7 PROJECTORS

Projection equipment consists essentially of a high-power lamp, a suitably designed condenser lens system and a well-corrected projection objective lens. The condenser fully illuminates the entrance pupil of the objective, which forms a magnified image on a distant screen. The diascope is used for the projection of transparencies (the overhead projector) and the episcope for opaque objects.

Profile projectors are made epidiascopic in design so that both opaque and transparent objects may be imaged. A set of projection objectives, of different magnifications, is mounted on a revolving turret. The test specimen is mounted on a glass platform provided with X–Y translational motion. The image is received on the rear of a ground glass screen. Reflecting mirrors are used to fold the optical path to make the instrument more compact.

Optical micrometric readers are incorporated on several machine tools; these help in accurate positioning of the workpiece. An optical dividing head is used for precise angular and displacement measurements.

8 SPECTROPHOTOMETERS

Spectrophotometers are used for a variety of applications in industry, including routine batch composition analyses and pollution control.

The principal component of the instrument is the monochromator, an optical source, which provides a selectable continuous range of wavelengths in the specified spectral region.

A commercial model of the instrument together with the schematic layout of the optical system of the monochromator used is given in Figure 5(a) and (b). The specifications, as given by the manufacturer, are wavelength range 0.35 to 0.80 μm; source, a tungsten lamp; dispersing element, a holographic concave grating with 1200 grooves/mm; and wavelength separation accuracy better than 0.02 μm. The detector is a silicon photodiode.

A colorimeter basically employs optical filters instead of the dispersing element.

A large number of other instruments based on optical principles are used for quality control and other purposes in industry. A selection of examples, which have not been dealt with already, includes the infrared-based instruments such as the IR (infrared) spectrophotometer, IR gas analyzers, and noncontact-type online IR moisture analyzer, which are based upon the principle that depending upon its composition, each substance has its own characteristic IR absorption band(s), which serves as a fingerprint for its identification and quantitative evaluation.

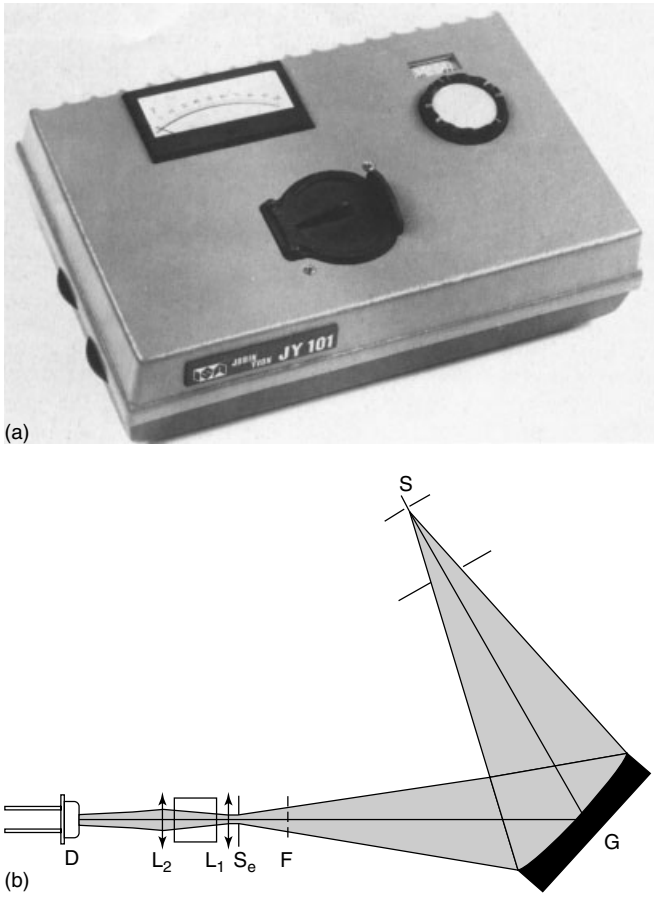


Figure 5. (a) Spectrophotometer; (b) schematic layout of monochromator comprising: S, source; G, holographic grating; F, removable filters; S_e, exit slit; L_1, L_2, lens; and D, detector. (Copyright © Jobin Yvon.)

Optical pyrometers are used for the temperature measurement of industrial furnaces wherein the flame luminosity is compared with that of an incandescent filament under varying electric power input conditions until complete photometrical matching is obtained.

The saccharimeter is used to determine sucrose concentration in the sugar industry: in these the property of optical rotation is used.

Photometers are used for determining the light output of various sources and illumination measurements.

Moiré fringe methods use two sets of ruled lines overlaid at a slight angle to form fringes that move transversely to the relative motion of the grating and its index. Movement of a fringe at a fixed location is counted with a detector driving a high-speed digital electronic counter. They are used extensively by the machine-tool industry for precise determination and control of linear and radial movements,

and in many aspects of gauge-room inspection. Lastly, in these examples, liquid and gas chromatographs are used for rapid estimation of sample purity and for conducting chemical analyses.

9 OPTO-MEDICAL INSTRUMENTS

For diagnostic purposes, observation of different places inside the human body is made by the endoscopic class of optical instruments. The cystoscope, laparoscope, arthroscope, bronchoscope, gastroscope, colonoscope, auriscope are some of the instruments belonging to this family. These vary mainly in their length, diameter, and other configuration parameters, to suit the particular application.

An endoscope is essentially a tele-microscopic system with a large number of intermediate image transfer relay lenses or coherent optical fiber bundles. Usually a low-wattage miniature lamp, fitted at the viewing side, provides illumination. Use of a higher-wattage light source, photography, and electro-optical sensor arrays is feasible with optical fiber–based endoscopes.

Ophthalmic instruments are designed to examine and measure various parameters of the human eye. Retinal observation is done by an ophthalmoscope. It has a miniature low-wattage light source, which is focused by a condenser onto a tiny mirror and the light reflected from the mirror illuminates the retina. During observation, any of the number of small lenses of different dioptric power, mounted on a rotatable disc, can be brought into the return light path. A lens introduced in the correct focus position gives the refraction error of the eye. A retinoscope measures the retinal curvature and is based on the Foucault knife-edge test principle.

Retinal photography is taken by a fundus camera. It focuses the filament of a lamp onto the pupil, while the eye, to be examined, looks at a small steel ball placed suitably in the light path. The same objective, used for illumination, focuses the image of the retina onto a photographic plate. After initial adjustment is made, a photograph or photo array image is taken under electronic flash illumination.

A cold light source, which is basically a noncoherent fiber bundle transmitting only visible light (IR content is filtered out by thin film device) is often used in retinal surgery.

An ophthalmometer is used to measure corneal curvature; the size of the reflected image from the cornea is measured by a variable image-doubling technique.

An Hbmeter (Figure 6) determines the hemoglobin concentration in blood. The instrument compares the absorption of light by hemoglobin in a carefully defined depth of hemolyzed blood with a standard glass wedge of similar absorption characteristics.

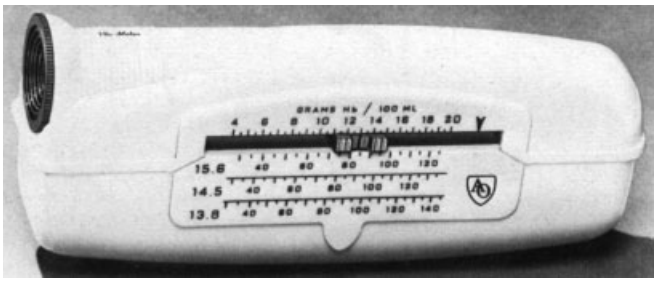


Figure 6. Hemoglobinometer. (Copyright © American Optical.)

Other opto-medical instruments not covered above are the ear microscope, slit lamp, trial sets and refractors, optometer, keratoscope, projection perimeter, Nagel anomaloscope, photocoagulators, focimeter, noncontact tonometer, IR body temperature ear probe, and colony counters.

10 LASER-BASED INSTRUMENTS AND SYSTEMS

The highly coherent, high-power, well-collimated beams obtainable from lasers have rendered them extremely useful in different sophisticated instruments. Both accuracy and range of many conventional systems are enhanced using the laser radiation source. Shop floor, as well as field, alignment problems from micrometers to kilometers can easily be accomplished precisely with a laser beam.

As an example, the schematic diagram of an early military tank, laser-based, range-finder system is shown in Figure 7. The laser transmitter–receiver system and optical sight are housed in the tank. The line of sight in elevation is aligned to the axis of the gun by means of a precision parallel linkage and in azimuth by coincident turret mounting. Bore sighting is achieved using controls on the unit with either a muzzle boresight or muzzle reference system. Gunlaying is achieved through a ballistic graticule.

A laser speckle interferometer is used to measure small displacement, vibration, deformation, turbulence, contour generation of three-dimensional objects, and for thermal mapping; it is analogous to holographic interferometry. It is based on the principle that scattered laser light from a grainy surface (grain size larger than the wavelength of light employed) gives rise to an interference pattern (called *speckle*). Overlapping of two speckle patterns produces Moiré fringes.

Using a laser source and flexible optical fiber as a probe, it has been found possible to measure vibrations inside a closed chamber or of other places where normal optical techniques may not be suitable. The fiber bundle carries the object beam forming one arm of a

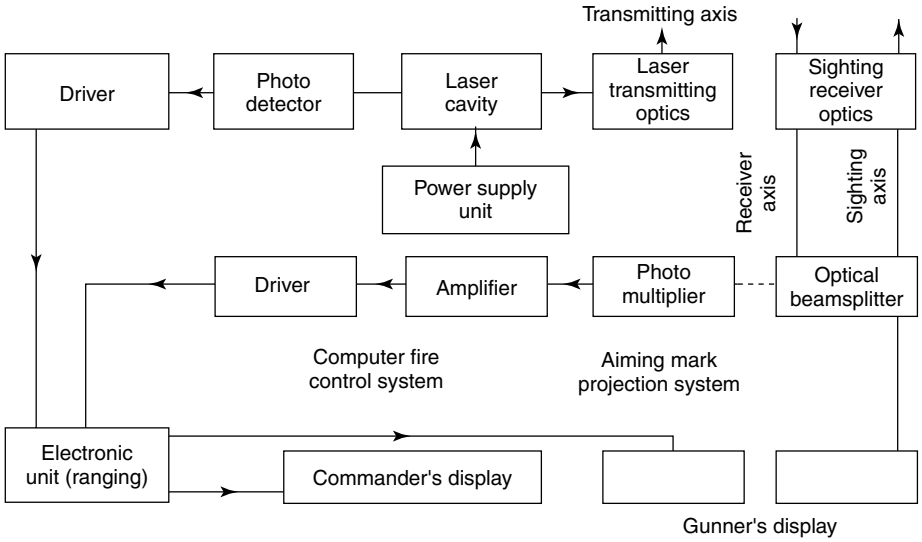


Figure 7. Schematic diagram of tank laser range-finder system. (Copyright © Barr & Stroud.)

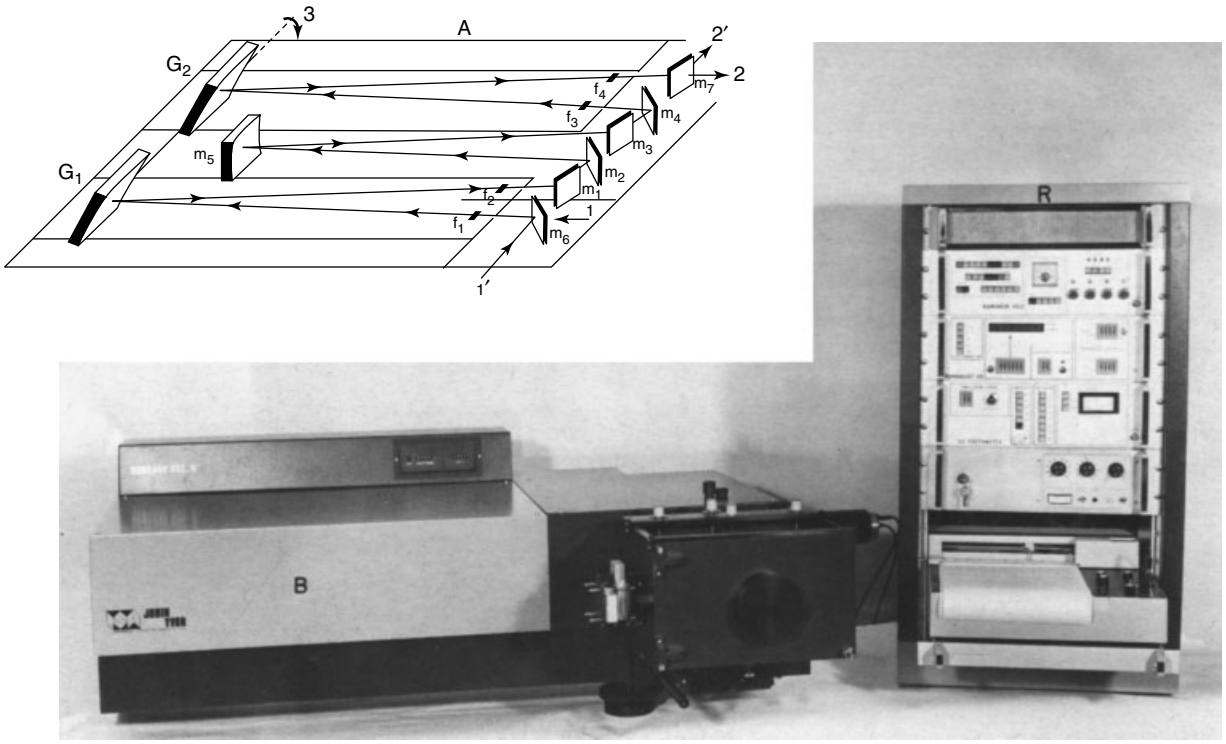


Figure 8. Laser Raman spectrophotometer: A, optical diagram; B, basic unit; R, detection system; 1, entrance beam path; 1′, alternate entrance beam path when m_6 is present; f_1, f_2, f_3, and f_4, stepping-motor controlled slits; m_6, kinematically mounted plane mirror, which permits use of 1′ as entrance path; 2′, alternate exit beam path when m_7 is present; 3, grating holders and shaft; G_1, G_2, holographic gratings; m_1, m_2, m_3, m_4, m_5, plane mirrors, and concave mirror for imaging exit slit of first monochromator f_2 on entrance slit f_3 of second monochromator; and m_7, kinematically mounted plane mirror to permit use of 2′ as exit path. (Copyright © Jobin Yvon.)

Twyman–Green-type interferometric setup. The laser interferometer has also been used in the measurement of refractive index variation, plasma density, and fluid velocity with the laser velocimeter.

A laser Raman spectrophotometer is based on the principle that a change in radiation frequency occurs due to molecular scattering. This has become a very powerful tool for studying vibrational and rotational energy levels of a

test sample. A commercial model of a laser Raman spectrophotometer, together with its optical diagram, is shown in Figure 8. The instrument employs two concave, aberration-corrected, holographic gratings with 2000 grooves/mm and offers a resolution better than $0.5\,\mathrm{cm}^{-1}$ at $0.5145\,\mu\mathrm{m}$.

RELATED ARTICLES

Article 89, Optical Materials, Volume 2; **Article 90, Optical Elements, Volume 2**; **Article 91, Light Sources and Detectors, Volume 2**; **Article 93, Testing Optical and Other Radiation Systems, Volume 2**.

FURTHER READING

Ahmad, A. (1996) *Handbook of Optomechanical Engineering*, CRC Press, Boca Raton, FL.

Hobbs, P.C.D. (2000) *Building Electro-Optical Systems: Making It All Work*, Wiley-Interscience, New York.

Smith, W.J. (2000) *Modern Optical Engineering*, McGraw-Hill Professional, New York.

This article is based in part on a chapter by J. Prasad and G. Mitra originally published in Handbook of Measurement Science Vol 2, P.H. Sydenham, Wiley, 1983.

93: Testing Optical and Other Radiation Systems

Alan J. Cormier
Sira Technology Ltd, Chislehurst, UK

1 Basic Parameters of Optical Instrumentation 610
2 Principles Used in Testing of Optical Instrumentation 610
3 Equipment Used in Testing of Optical Instrumentation 611
4 Limitations to Testing of Optical Instrumentation 613
5 List of Global Optical Calibration Facilities 614
References 615
Further Reading 615

1 BASIC PARAMETERS OF OPTICAL INSTRUMENTATION

In general, optical instrumentation may be classified as either imaging (e.g. camera lens) or afocal (e.g. telescope), where the term 'optical' strictly implies operation within the visible band (∼400 nm to ∼700 nm) although the ultraviolet and infrared spectral bands are also included here.

There are a large number of basic design parameters associated with an optical instrument, such as focal length, numerical aperture, magnification, and distortion; however, here we consider more specifically wavefront and image quality.

1.1 Wavefront quality

Imaging systems: Light from an object point, emerging from the exit pupil of an imaging optical system, will ideally be in the form of a spherical wavefront capable of focusing to an image point the size of which is limited only by diffraction effects. When this spot size is achieved, the system is said to be diffraction limited. Real optical systems generally have aberrations that degrade the wavefront leading to larger spot sizes and lowered resolution.

Afocal systems: For an afocal system (not image point forming), a plane output wavefront is implied. In the case of visual systems, however, a departure from this condition may be used to correct refractive error for the users.

1.2 Overall image quality

In addition to wavefront quality, this depends upon factors such as system transmission losses, chromatic aberration, image distortion, and the effects of stray light. Wavefront vignetting due to internal aperture sizes and positions will also affect final image quality (by modifying the numerical aperture) as will the method of detecting the image. In visual systems, the eye may limit image quality; with conventional camera systems, film grain size and the processing method may be the limiting factor. For digital cameras using CCDs (Charge Coupled Devices), resolution is limited by the pixel size.

2 PRINCIPLES USED IN TESTING OF OPTICAL INSTRUMENTATION

2.1 Interference

Wavefront amplitude and phase: Optical radiation propagates in the form of a wave. A converging wavefront must

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

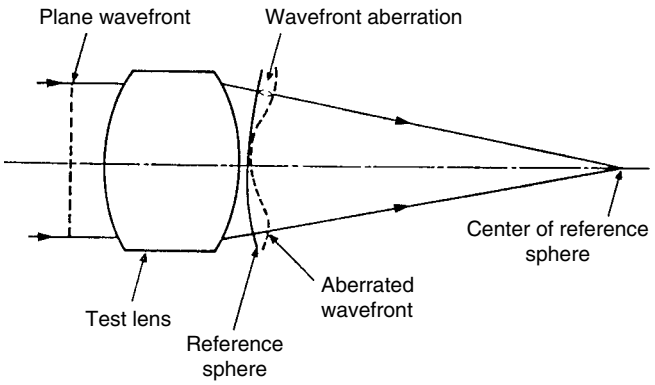


Figure 1. Wavefront aberration related to the reference sphere. (*Source*: Reproduced by permission of Sira and IoP Publishing.)

be perfectly spherical if the minimum focal spot size is to be achieved. By comparing a real wavefront with a perfect reference wavefront, we can specify the wavefront error, Figure 1.

Two-beam interference: If both reference and test wavefronts are 'coherent', then interference fringes will occur. Maximum fringe intensity occurs where the local phase difference in radians is given by,

$$\Delta\phi = n2\pi \text{ radians}(n = 0, 1, 2\ldots) \quad (1)$$

2.2 Diffraction

The diffraction effect: When a beam of light touches an aperture boundary, the beam is spread slightly (diffracted). For a circular aperture, this spreading causes the beam diameter to increase. The amount of spreading is related inversely to the wavelength of light used.

Airy disc and point spread function (PSF): In the case in which diffraction effects were absent, a perfect spherical wavefront could produce an infinitely small focal spot. Diffraction effects, however, limit the minimum possible spot diameter to the Airy disc (a fact made use of in Knife-edge and Star Tests). This diameter increases when optical aberrations are present. The distribution of intensity in the image plane is represented by the PSF.

Line spread function (LSF) and edge spread function (ESF): By scanning a narrow slit across the PSF we obtain the LSF. For practical purposes, the LSF is usually obtained by scanning a slit aperture across the image of a slit source (or alternatively by imaging the slit source onto a 2D array detector). This function is useful when we wish to assess image quality using an OTF Optical Transfer Function bench. An associated function, the ESF, can also be used to determine OTF. This function is derived by recording the energy passed by an edge boundary moved across the image plane (Williams, 1999).

2.3 Optical transfer function

We can consider image quality measurement in terms of both 'spatial' (e.g. millimeters) and 'spatial frequency' (e.g. cycles per mm) units. These two domains are related through a Fourier Transform. By performing a Fourier Transform of the (spatial) image, LSF, we obtain the OTF having both amplitude modulation (MTF or Modulation Transfer Function) and phase (PTF or Phase Transfer Function) terms. MTF is of particular significance, giving information on image contrast (Williams, 1999).

3 EQUIPMENT USED IN TESTING OF OPTICAL INSTRUMENTATION

3.1 Radiation test sources

Incandescent sources: One of the most useful is the quartz tungsten halogen lamp where a tungsten filament is enclosed in a quartz envelope (*see also* **Article 91, Light Sources and Detectors, Volume 2**). The halogen gas and quartz envelope allow the lamp to be run at high temperatures, thus shifting the continuous spectral curve to shorter wavelengths (down to ∼350 nm).

Gas discharge sources: At low pressures, gases such as Na and Ne, and Hg vapor, will radiate just at fixed resonant wavelengths. Such spectral line lamps are useful for wavelength calibration purposes. As pressure is increased, so will light intensity; however, the lines will gradually broaden and be superimposed on a spectral continuum. Mercury lamps are widely used, with the 546.1 nm spectral line being a common optical standard.

Arc lamps: Two closely spaced electrodes located within a quartz envelope filled typically with high-pressure xenon gas or mercury vapor. Very bright concentrated sources are produced with resonant lines superimposed on a thermal radiation continuum. Compared with tungsten halogen sources they produce more ultraviolet light and much higher brightness in the visible band.

Lasers and Light emitting diodes (LEDs): A laser source has the special property of 'coherence', which is

put to good use in the interferometer. Laser sources also have intrinsically very high brightness and narrow bandwidths. For optical testing purposes, the He–Ne (633 nm) laser is frequently used; others are available such as the CO_2 laser for testing in the infrared (10.6 μm). The LED is similar to the laser except that it does not have an optical cavity; consequently, the spectral bandwidth is broader and the coherence lower.

Blackbody infrared source: An ideal blackbody source completely absorbs all radiant energy striking it and, therefore, appears perfectly black at all wavelengths. A perfect blackbody source emits radiation with an emissivity of 1 (see Burnay, Williams and Jones, 1988). Real sources, using controlled high temperature cavities, can emit wavelengths between 1 μm and >25 μm. The shorter wavelengths require higher temperatures. When used with a pyroelectric detector, the radiation source must be used with an optical chopper.

3.2 Test targets

USAF 1951 target: A significant number of targets formed from opaque line patterns deposited on substrates such as fused silica are available. These are used to assess resolution limit directly. One of the most common is the USAF 1951 target, shown in Figure 2. Use of this and other similar targets is considered in the Military Standardization Handbook (MIL-HDBK-141, 1962).

Slit and bar targets: Formed either by deposition techniques, or machined into a solid opaque substrate. In the latter case, substrate transmission is not an issue. These targets are frequently used to produce an LSF from which the system OTF can be calculated. Bar targets are generally multiple, wide slits with equal mark-space ratio.

3.3 Integrating sphere

Used to produce an output beam that is insensitive to any spatial, angular, or polarization changes occurring in the input beam. They do this by collecting input light through one aperture, which, after multiple reflections and scattering at a high reflectivity surface, eventually emerges through a second aperture at right angles to the first (see Figure 3). Integrating spheres are used to overcome problems such as spatial and angular response variations exhibited by many detectors and the effects of input beam position and profile changes.

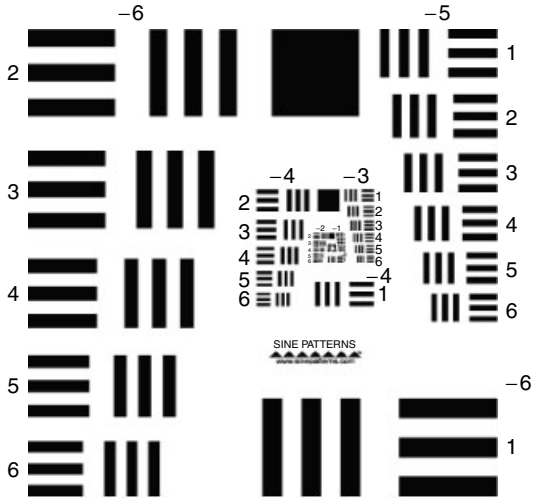


Figure 2. USAF 1951 target. Public domain. (*Source*: Copyright Sine Patterns LLC.)

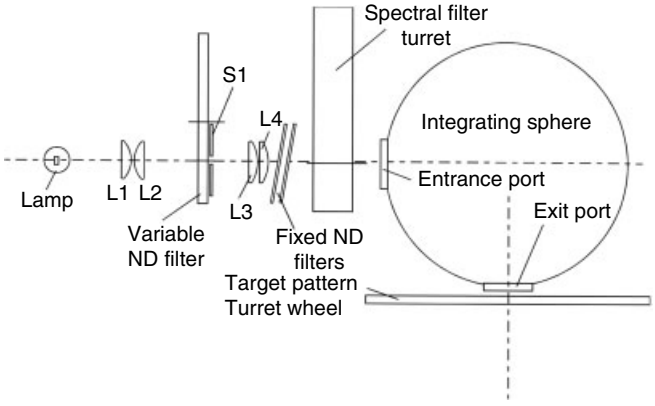


Figure 3. The integrating sphere as part of Sira's helmet mounted display test facility. (*Source*: Reproduced by permission of Sira and IoP Publishing.)

3.4 Knife-edge and star tests

Knife-edge or Foucault test: As shown in Figure 4, a sharp edge is positioned close to a beam focus, produced by a lens or mirror, with the eye immediately behind and focused at the exit pupil position. By moving the edge across the beam, sections of the exit pupil are seen to go dark in correspondence with the blocked rays. From the pupil patterns seen, experienced observers can deduce the aberrations present (Malacara, 1978).

Star test: For this test, the optical system is used with a point object (similar to a star) to produce a geometrical image that is smaller than the Airy disc. By using an eyepiece to examine the image, the experienced observer can

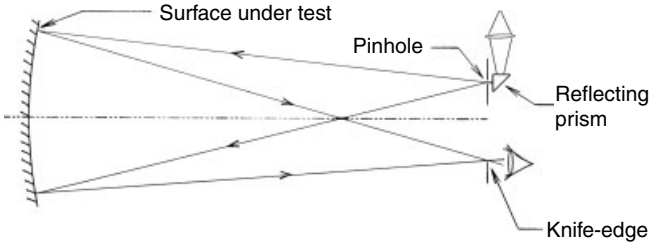


Figure 4. Knife-edge test configuration. (*Source*: Reproduced by permission of Sira and IoP Publishing.)

again learn a great deal about the aberrations present that affect the shape of the diffraction pattern (Malacara, 1978).

3.5 Interferometer

A number of different types of interferometer exist; however, in each case, coherent beams derived from the same source are made to interfere in the optical domain.

Perhaps the most common type is the He–Ne Fizeau. The laser beam passes into an air gap formed between two surfaces one of which is a plane (flat) reference surface. The other surface is the plane test surface. An interference fringe pattern is produced, which maps the errors on the test surface. Spherical test surfaces may also be examined when the interferometer is fitted with a special lens called a *transmission sphere* (see Figure 5). In this case, the last concave surface acts as the reference surface.

Because two-beam interference effects repeat for every 2π phase difference, the interferogram will be ambiguous outside this range. Phase shifting of the reference beam is commonly used to overcome this problem (Zygo Corp, 1998).

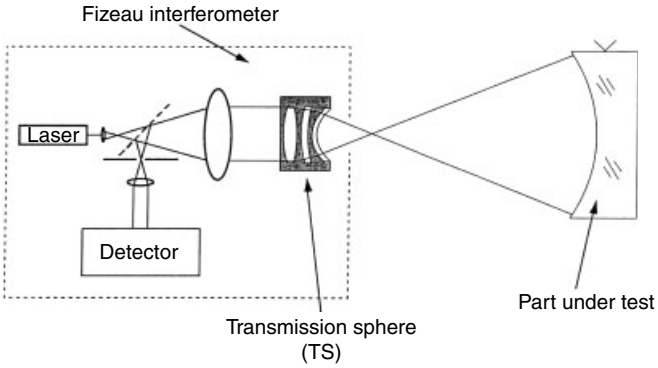


Figure 5. Fizeau interferometer test configuration using a transmission sphere. (*Source*: Reproduced by permission of Sira and IoP Publishing.)

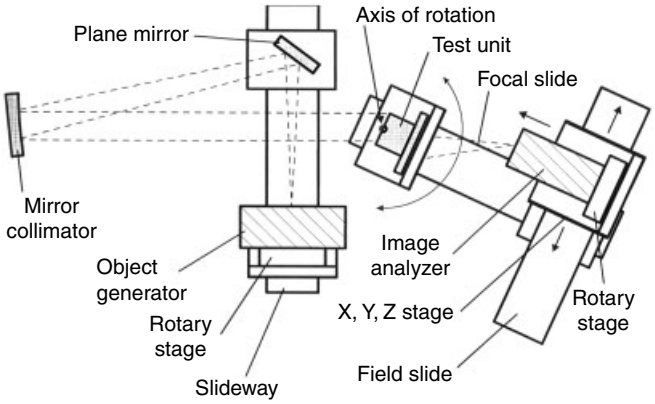


Figure 6. Typical OTF test bench configuration. (*Source*: Reproduced by permission of Sira and IoP Publishing.)

3.6 Camera/OTF bench

The camera bench is an important piece of basic test equipment and is used in the measurement of fundamental system parameters, such as focal length, magnification, and distortion. When used with a single element detector behind a slit aperture and a (scanned) slit source, an OTF bench is formed, Figure 6. Because the detector axis direction usually remains fixed, test limitations can occur at larger field angles in cases in which the image slit cannot be placed close to the detector surface. This is due to the signal falling outside the detector boundary. OTF benches often use an area array detector in place of the originally used image slit, which overcomes this problem (Williams, 1999).

4 LIMITATIONS TO TESTING OF OPTICAL INSTRUMENTATION

Limitations are broadly related to

1. accuracy of the measurement equipment,
2. suitability of the measurement configuration, and
3. environmental effects.

In some cases, the limitation may be absolute, in other cases the limitation may be reduced by careful recalibration of measurement equipment or changing the measurement configuration. Any tests made must take into account the instrument's intended mode of use. This is particularly true in the case of visual instrumentation where the users perception of image quality is a key factor (Haig and Burton, 1987).

In the case of an imaging system, such as a film camera, a common test would be to photograph a resolution test chart;

however, the results are generally subjective. It is therefore usually necessary to test the lens objectively. When performing OTF tests, measured image quality is limited through aberrations introduced by the optical components used to provide the input beam (although it is usually easy to make them insignificant). Wherever possible, additional components should not be used to relay the test image to the detector. Where this is necessary (for example, when testing afocal systems), the relay optics should be diffraction limited.

The use of interferometers, in conjunction with phase-shifting techniques, has the potential for high measurement accuracy. However, poor agreement between different instruments and operators has been a source of concern (Briers, 1999). Special techniques are also required when using high aperture (e.g. NA 0.5) transmission spheres (Selberg, 1990).

For visual system tests, it is usual to perform photometric measurements where an observer's response is standardized through a combination of detector and matching filters. While the overall detector spectral response achieved may appear close to the ideal form, small localized variations can result in significant measurement errors (Ohno and Thompson, 1998). These errors may be increased through effects such as detector ageing and filter degradation, uncorrected thermal dependence, and stray light. Periodic recalibration of such detection equipment is crucial if the highest measurement accuracy is to be achieved.

Testing of thermal (or heat) imaging systems operating in the mid- and longwave infrared presents additional problems. In particular, the optical elements and support structures are themselves a source of radiation that can lead to reduced image contrast. A common practice is to chop the infrared source, which can then be separated electronically from the DC background – *see* **Article 184, Photon Counting, Volume 3**. This technique is effective when performing OTF measurements; however, limitations will occur when accurate radiometric measurements are required, unless the temperature and emissivity of the chopper blades are known. The limitation may be reduced by using a chopper with mirror blades, which allow a calibrated blackbody to be imaged during the obscuration period.

Precise knowledge of target and background plate temperature and emissivity is also required when undertaking other forms of differential temperature imagery (e.g. Minimum Detectable Temperature Difference, or MDTD, and Minimum Resolvable Temperature Difference or MRTD) as thermal imagers respond to radiance and not temperature differences. Further consideration is given to this aspect in (Burnay, Williams and Jones, 1988)).

In the above text, we have seen that thermal and stray light effects can act to limit the accuracy of test measurements. Other environmental effects, such as vibration, shock, and humidity, can similarly limit performance. Vibration, for example, can affect significantly the accuracy of interferometric measurements. This is particularly the case when using phase-shifting hardware.

A very wide range of optical instrumentation exists, from complete imaging systems to instruments that are themselves used in the testing of other optical instruments. As a consequence, a wide range of test requirements will also arise. By understanding the physical limitations involved and adopting environmental control procedures and regular equipment recalibration, an acceptable result can usually be achieved.

5 LIST OF GLOBAL OPTICAL CALIBRATION FACILITIES

The foremost optical metrology laboratory in the United Kingdom is the National Physical Laboratory (NPL):

National Physical Laboratory (NPL)
Queens Road
Teddington, Middlesex TW 11 OLW
Phone: (44 181) 977 3222
Fax: (44 181) 943 6458
E-mail: enquiry@npl.co.uk

Other key laboratories include

France: Bureau National de Metrologie
1, rue Gaston Boissier
75 015 Paris – France
Phone: (33 1) 56 56 88 63
E-mail: francoise.le-frious@bnm.fr

Germany: Physikalisch-Technische Bundesanstalt (PTB)
Bundesallee 100
Postfach 3345
D-3300 Braunschweig
Phone: (49 531) 592–3005
Fax: (49 531) 592–4006

Italy: Istituto Elettrotecnico Nazionale
Dept. of Photometry
Strada delle cacce 91
10135 TORINO, Italy
Phone: (39 11) 348 8933
Fax: (39 11) 650 7611

United States of America:
National Institute of Standards and Technology (NIST)
Optical Technology Division

NIST, 100 Bureau Drive, Stop 3460, Gaithersburg, MD 20899-3460
E-mail: inquiries@nist.gov
Phone: (301) 975-NIST (6478) or TTY (301) 975–8295

Canada: National Research Council Canada
Institute for National Measurement Standards
Montreal Road, M-36
Ottawa, Ontario K1A 0R6
Phone: (613) 993–9326
Fax: (613) 952–5113
E-mail: gary.hysert@nrc.ca

Other countries:

A global list of metrology laboratories may be found on the US National Institute of Standards and Metrology (NIST) Web site at http://www.nist.gov/oiaa/nat_pg2.htm

REFERENCES

Briers, J.D. (1999) Interferometric Optical Testing: Considerations for the Proposed New International Standard, *SPIE: Optical Fabrication and Testing*, Vol. 3739, SPIE (pp. 496–507).

Burnay, S.G., Williams, T.L. and Jones, C.H. (1988) *Applications of Thermal Imaging*, IOP Publishing.

Haig, N.D. and Burton, G.J. (1987) Effects of Wavefront Aberration on Visual Instrument Performance and a Consequential Test Technique. *Applied Optics*, **26**(3), 492–500.

Malacara, D (ed.) (1978) *Optical Shop Testing*, John Wiley & Sons, New York.

MIL-HDBK-141, ed. (1962) *Military Standardization Handbook – Optical Design*, Department of Defense, Defense Supply Agency, Washington, DC.

Ohno, Y. and Thompson, A.E. (1998), Photometry – The CIE V(λ) Function and What Can be Learned from Photometry, *International Symposium on Measurements of Optical Radiation Hazards*, 1–3 September, NIST, Gaithersburg, MD.

Selberg, L.A. (1990) Interferometry Accuracy and Precision, *SPIE Optical Fabrication and Testing*, Vol. 1400, SPIE (pp. 24–32).

Williams, T.L. (1999) *The Optical Transfer Function of Imaging Systems*, IOP Publishing.

Zygo Corp (1998) *Application Note – Fringe Analysis versus Phase Measuring Interferometry*, Zygo Corporation, Web site http://www.zygo.com.

FURTHER READING

Labsphere (2002) *A Guide to Integrating Sphere Radiometry and Photometry*, Labsphere Inc, Web site http://www.labsphere.com.

94: Human Factors Engineering

Nicholas I. Beagley
DSTO, Canberra, ACT, Australia

1 Ergonomics and Human Factors Engineering 619
2 Task Capture and Design 620
3 Characterizing the User Population 621
4 Establishing Context 621
5 Training Specification 622
6 Modeling and Prediction 622
7 Prototyping and Testing 622
References 623

1 ERGONOMICS AND HUMAN FACTORS ENGINEERING

Translated literally from the Greek words *ergon* (work) and *nomos* (science/laws), Ergonomics means the 'Science of work'.

The study of work has been traced back to a publication in 1713 by Bernardino Ramazinni on work related injuries.

Before World War II, the focus on workplace studies was on efficiency and productivity through time and motion studies by Taylor (1911) and Gilbreth (1911). Concentration on the design of military systems prompted a shift in focus to human performance, behavior, and safety.

The term *Ergonomics* was proposed by Prof. Hugh Murrell in 1949 and first applied to this field through the naming of the UK's human research focused Ergonomics society that year. It is this broader description of the field for which the term Ergonomics has been adopted. As the study of the human in the context of their working environment expanded so other terms such as the more general Human Factors (HF) and the more design specific Human Factors Engineering (HFE) have increased in use – for more on HFE *see* **Article 97, Design Methodology, Volume 2**.

Despite hot debate about the appropriate use of terminology, the terms Ergonomics and Human Factors Engineering are frequently used interchangeably. There are numerous definitions that seek to encapsulate all aspects of the discipline, each serving a purpose in setting out the scope for the discipline. In trying to cover this range, they tend to be relatively long and wordy. An alternative method of communicating the scope and purpose of this area is to convey the underlying philosophy of the discipline. This is put succinctly in the statement that Ergonomics is about 'equipping the man, *not* manning the equipment'.

This conveys the idea that in order to get the most from the total system, designs should be tailored to the capabilities of the human rather than relying on people to adapt to unsuited designs.

The practice of Ergonomics/HFE draws on a number of broad activities that combine to improve our understanding of human/system performance and ensure that this knowledge is available to positively influence system design in the goal of achieving more efficient and safer systems.

Ergonomics is underpinned by a knowledge base informed by research. Fundamental research in areas such as vision, memory, and motor skills provides data that can often be generalized across a wide range of contexts.

At a higher level of human function where basic functions, such as the senses and cognition, combine, for example, human behavior, the influence of functional combinations and context increases.

The consequence is a reduction in our ability to apply the findings of generic research for the purpose of prediction.

Given the limitations on prediction, final confirmation of design suitability is normally based on testing

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

within a close representation of the system's intended context.

To effectively influence design, it is important to package the accumulation of research knowledge and experience in a way that is accessible to multidisciplinary human factors and engineering design communities.

As with all areas of science, the published research literature is the main repository for knowledge. The level of domain knowledge required to retrieve, review, and apply this knowledge limits its widespread applicability. Countering this problem, a large number of documents have been developed to bring together principles of design, guides to best practice and agreed standards. Key documents in this area include Def Stan 00–25, UKMOD (2000); Mil Std 1472F, USDOD (1999); and ISO 9241, ISO (1993).

In the absence of an all-encompassing deterministic predictive model of human/system performance, it is necessary to study the human/system in context.

A wide range of tools and techniques has emerged from the human factors community over the years to support this activity. These tools include techniques for measurement, analysis, and for influencing the design cycle. Tools such as task analysis, that is, the systematic decomposition of an activity to support understanding and redesign, are a fundamental activity for most ergonomists seeking to understand the goals and associated activities carried out in pursuit of a system objective.

Human/system performance measurements, using a range of objective and subjective measurement tools and techniques within designs sensitive to the human/system variable, are a vital part of design testing.

At the top-level process, methods such as the Human Factors Integration (HFI) Program (UKMOD, 2000; UKMOD, 2002) – see other documents on http://www.ams.mod.uk/ams/content/docs/hfiweb/hfihome.htm, and the US Manpower Integration (MANPRINT) Program (see documentation on http://www.manprint.army.mil/manprint/guides/mpref-guidebook.asp), are followed to ensure consistency of analysis and to provide an audit trail of the human science support to the design process.

2 TASK CAPTURE AND DESIGN

Ergonomics at its most fundamental remains the science of work. It is not the study of the human in isolation, nor is it the study of the world around the person, although it draws on each of these disciplines. At its core, ergonomics is concerned with the activities of people, the design of the systems that define these activities, and an understanding of how the capabilities of the human and the demands of the environment impact on the way people interact with the world around them.

Given the central role of the activity of the human to the practice of Human Factors, it is not surprising that the documentation and analysis of these activities lies at the heart of many human factors studies.

The process of task description is often one of the starting points of any human factors analysis. The capture of a baseline description of a person's activities within the context of a system's function is an essential part of assessing opportunities for, and the achievement of, progress through design and development. It is not possible to demonstrate that a design change has improved a situation without establishing what went before.

Task description of an existing process can be a useful baselining diagnostic tool for revealing parts of an activity that are inefficient or error prone. At this point, there is a blurring between the terms *task description* and *task analysis*.

The widespread application of task analysis within the fields of ergonomics and HFE became popular in the 1950s. It has since been modified and tailored to a variety of objectives.

Task analysis has a key role in process and product design. This concept is akin to a ConOps – *see* **Article 67, Principle of Concept of Operations (ConOps), Volume 2**. It can be used to aid the allocation of function, product testing, and for the evaluation of workload. Accidents and assessments of the risk of accidents can benefit from the use of task analysis. A more complete description of task analysis techniques is available in Kirwan and Ainsworth (1992).

Task analysis methods are often distinguished by their method of subtask separation and organization. A widely applied approach to task analysis is called hierarchical task analysis (HTA). It starts with a high-level objective oriented description of the activity and proceeds to decompose the activity into goal focused tasks and subtasks.

The approach to HTA descriptions can be freely adapted to capture specific information relevant to the objective of the analysis such as information flows, or associated technologies. Frequently, the sequence and logic underlying the conduct of a task sequence is represented within the HTA box diagram.

A baseline task description of an existing process, when analyzed systematically, can reveal opportunities to rationalize the number, or order, of steps to gain time savings or reduce errors.

The form and content of the task description should be based on the objectives of the method. If the issue of concern is the mental workload placed on the operator, the task description should capture and represent the information flows over time.

This information, when applied to user walk throughs using representative scenarios designed to test the limits of the system, can provide a means of assessing operator loading and the associated potential for failure.

An alternative form of task description and analysis is the operational sequence diagram (OSD). A representation of the task sequences and the logical links between tasks can be a useful tool for examining the implications of changes to a system on its performance. This approach to modeling forms the basis of computer-based task-modeling tools such MicroSaint and the Integrated Performance Modeling Environment (IPME).

Occasionally, it can be useful to consider the linkages between subcomponents of a system. An example might be the analysis of communications between different people within a control room. A diagrammatic representation of the control room overlaid with lines depicting the distribution of communications amongst people within the control room can be a useful tool for examining the location of workstations and the information necessary to meet the control room's operational role. This method of system description is called *link analysis*.

3 CHARACTERIZING THE USER POPULATION

If a design is to take account of the characteristics of its intended user population, the key characteristics relevant to a system's emerging concept must be captured as a reference for the designers and decision makers.

If assumptions made about the intended population, such as the ability to recruit and retain highly skilled engineers as users, are unfounded, then the procuring organization could be left with an expensive system that cannot be operated due to insufficient numbers of capable operators. Within the UK's military procurement process the tool for characterizing the user is called the Target Audience Description (TAD).

The TAD is a document that is created early in the design concept phase of a procurement program. The TAD has become a standard tool whose structure prompts the gathering of information likely to be relevant to the procurement program.

The document starts with a general description of the current group who would be intended to use a system. A reasonable assumption for the manning of a system is that the future users will have the same level of skills and capabilities as the existing population.

There is a danger in specifying a highly technologically sophisticated replacement for an existing low tech system when the new system would require a range of skills, knowledge, and experience considered to be beyond the reasonable training expectations of the current population.

If the requirement for skills and training is noted at an early stage, the parallel processes of training and recruitment can be put in place over the timescale for system procurement.

It should, however, be recognized that in an increasingly technological world, the number of trades that will be competing for the pool of highly skilled individuals could also be expected to increase.

For the purposes of ergonomic design, the parameters of the population, including anthropometric, that is, body size, dimensions, should be captured.

The population ranges normally captured in databases and tables and formulae include stature, weight, strength, functional reach, leg length, sitting height, and hip breadth.

A wider list of anthropometric measures should be provided when the design has a large human/equipment integration component.

System designs are normally required to safely and comfortably accommodate potential users within the 5th to 95th percentile, that is, the percentage of the target population that can fit, reach, and operate the system.

In addition, cognitive attributes of the target population have become increasingly relevant for systems designers. This may be specified by selection tests applied to people within the target group. Selection criteria may provide an additional source of information regarding the group's normal spread of knowledge, experience, and skills as relevant to the design under consideration.

The TAD is simply a method of formalizing and reporting the capture and consideration of user attributes. Small-scale design programs may not justify the development of a full TAD. However, the process of capturing and considering these attributes at an appropriate level of detail has a place in all human system procurement and development activities.

4 ESTABLISHING CONTEXT

The study of human centered systems is inextricably linked with the context in which they are placed. It is on this premise that the ecological approach to human factors research is based. The ecological approach starts from the position that there is a reciprocal coupling between the person and their environment. It takes the view that the environment is both an opportunity for human action (or affordance) and a constraint. For example, a maze affords a path for progress while at the same time constraining the path's direction. From this perspective, it is clearly of particular importance to analyze it.

An analysis of the environment should seek to identify those aspects of the environment that exert an influence over human behavior and contrast them against those behaviors that are specific to the individual.

5 TRAINING SPECIFICATION

As the systems we expect people to use become increasingly complex, there is a need to pay special attention to the demands this places on providing an effective mechanism for teaching the knowledge and skills necessary to support these interactions.

Training is a necessary overhead to effective system performance. This training must be considered alongside the more traditional cost estimation parameters of purchase and maintenance. Training programs should be designed to ensure that the necessary skills are transferred from the training environment to the operational environment. It is through this demonstration of transfer of training that the training program can be formally evaluated.

The normal approach to demonstrating a transfer of training is to compare the performance of a trained group against an untrained control group to demonstrate a statistically significant difference between the performances of the two groups. Under some circumstances, it is possible for a training program to have a negative transfer. An example might be the training in a flight simulation with a slightly different interface style to the real aircraft. Errors might arise because of the interface differences that have the effect of surprising the operator when they encounter the different real system.

Training systems cannot be solely assessed on their ability to support a transfer of appropriate skills to the real world. While real flight might offer a better transfer of training than the use of flight simulation, the cost of this approach can, in some circumstances, be prohibitive.

Training through simulation has the added advantages of graded, and total, control over the scenario and the ability to automatically gather performance measures with which to set up trainee development.

6 MODELING AND PREDICTION

The complexity of human interaction within the real world reduces our ability to predict outcomes. This leaves the practice of Human Factors with no alternative other than to conduct user testing in representative environments.

The time and logistical costs of running a user experiment is compounded by the time load on the subjects. Any opportunity to achieve the design benefits of user testing through performance prediction and modeling will have obvious savings in costs and timescales.

Human modeling and performance prediction is already used as a tool to support the design and analysis of human systems. There is a range of commercially available models that can be used to predict human performance in a number of ways. Stochastic, that is, based on probability, task models such as MicroSaint™ and The Integrated Performance Modelling Environment™ provide the researcher with the ability to specify task sequences and their associated procedural logic and performance parameters. By defining a concept task in this form, it is possible to make predictions about the likely performance of the task in terms of defined parameters such as completion time. The addition of other parameters such as information flow or physical load can be used to reveal peaks in workload that could have consequences for task performance.

There are several commercially available body modeling tools on the market including JACK™, Safework™, and Sammie™. These tools allow the user to specify the key anthropometric dimensions of virtual mannequins.

Most of these tools allow the user to import virtual Computer Aided Design (CAD) models of concept equipments for evaluation within the virtual environment. To these are added the mannequin with the needed anthropometric data; the two then can be made to interact.

These techniques have demonstrated benefits for the evaluation of issues of fit, reach, and vision. Some include basic modeling of the physiological stresses that might accompany varying posture and loads, although the true complexity for the human musculoskeletal system prevents these models from offering validated predictions of human biodynamics, that is, the study of dynamic forces on the body.

7 PROTOTYPING AND TESTING

The objective of HFE is often to draw conclusions about the merits of concept systems. The difficulty associated with this form of analysis is the fact that these systems do not yet actually exist. Human Factors has developed a wide range of tools and techniques focused on performance measurements that support the comparative evaluation of concept human/systems.

When a system component exists only as a concept, the researcher must be inventive in testing the projected capabilities. The normal approach to this is prototyping, that is, the creation of a model or simulation of a product.

If the availability of a new technology is anticipated within a few years, there may be a requirement to examine the potential issues that may arise in its usage or the implications for the normal approach to associated tasks.

If an early prototype of the technology is available, this can form the basis of product testing to examine these issues. In many instances, the early prototypes of new technologies are extremely limited in their functionality or are unduly bulky. In these cases, there is a danger that the generic capability illustrated by the prototype technology will be missed by the user who becomes distracted by the limitations of the prototype.

In cases in which the prototype technology is too immature for useful user testing, it may be possible to simulate the performance of the system. An example of simulating an immature technology is the use of a researcher to take the place of a proposed artificial intelligence system intended to aid the user in a task. The subject of the evaluation is not made aware of the method used to provide this functionality thus allowing the researchers to examine issues such as modes of interaction and user confidence in the system. This approach allows conclusions to be drawn that will contribute to design decisions that will have to be taken when the prototype technology reaches maturity.

REFERENCES

Gilbreth, F.B. (1911) *Motion Study*, Van Nostrand, Princeton, NJ.

ISO (1993) *Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs)*, European Standard 9241 (Parts 1 to 17), International Standards Organisation, Geneva.

Kirwan, B. and Ainsworth, L.K. (eds) (1992) *A Guide to Task Analysis*, Taylor and Francis, London.

Taylor, F.W. (1911) *The Principles of Scientific Management*, Harper & Bros, New York (Reprinted 1967 by W.W. Norton & Co.).

UKMOD (2000) *Human Factors for Designers of Military Equipment (Parts 1 to 14)*. Defence Standard 00–25, UK Ministry of Defence.

UKMOD (2002) *A Practical Guide to Human Factors Integration: A Guide for Integrated Project Teams and Capability groups*, Version 0.9.3, 19 July 2000, UK Ministry of Defence.

USDOD (1999) *Department of Defence Design Criteria Standard Human Engineering*, Military Standard 1472F, US Department of Defense.

95: Human–Machine Interface

Nicholas I. Beagley
DSTO, Canberra, ACT, Australia

1 Comparing Humans and Machines 624
2 The Interface 625
3 The Visual Interface 625
4 The Auditory Interface 626
5 The Other Senses 626
6 Control Devices 626
References 627

1 COMPARING HUMANS AND MACHINES

This article continues the development of the topic of Human Factors begun in **Article 94, Human Factors Engineering, Volume 2**. There has been an uneasy alliance between people and machines over the course of history. The use of simple tools is cited as one of the first steps in the civilization of the human species.

Subsequent developments of agriculture and weapons are the basis on which civilization in its current forms evolved (Diamond, 1999).

With each technological advance, people have been able to achieve goals in increasingly efficient and effective ways. This is not to say that technology is consistent in delivering advantages. However, the underlying trend is a world in which the traditional labors of the human race have been reduced or replaced by the application of technology by those able to afford to develop and acquire them.

Our understanding of human performance is based on fundamentals revealed through human physiology, for example, Paul Fitts established a law for the speed and accuracy of motor movements, such as, hand control, through laboratory testing (Fitts, 1954).

A wide body of research reveals many parameters of human performance in the areas of speed, strength, and endurance as well as sensitivity to stimulus and the capacity for mental storage, retrieval, and processing.

In most cases, these parameters vary between individuals across the population and with the variations in both the task to be completed and the demands of the environment.

Despite the appeal of a reductionist model of human performance, the ability to accurately predict human performance within a conceptual system remains elusive.

The relationship between man and the engineering design of machines has been discussed in Sydenham (2003).

While the pursuit of this ideal continues, Human Factors (HF) practitioners are left to apply a combination of modeling and human testing within an iterative design cycle to achieve the objectives of true human/system integration.

Fitts contributed further to the toolset of the HF practitioner with his list comparing the performance of humans with machines (Fitts, 1951).

The original Fitts list consists of two columns, comparing the strengths of the person versus strengths of the machine. As technology continues its relentless advance, the ability to establish clear distinctions between the capabilities of people and machines is diminishing.

Opportunities exist to overcome human weaknesses such as physical and mental fatigue with automated solutions capable of quickly and accurately completing repetitive tasks.

Advances in computer technology have allowed machines to outperform people in areas originally considered to be the domain of the human. As a result, the

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

original Fitts list serves little purpose beyond illustrating the rapid progress of technology.

That said, the underlying concept holds true. There remains value in considering the allocation of function between the different components of a system, including the person. An example of a modified Fitts list is shown in Table 1.

2 THE INTERFACE

People interface with the world around them by physical actions directed by the nervous system based on sensory inputs. Our ability to sense our posture in relation to the earth's gravitational force and coordinate our muscular response accordingly is the basis by which we are able to stand erect. The rapid feedback loop, from our sense of vision, balance, touch, hearing and so on, allows us to walk, run, and jump in a dynamic environment. The evolution of human capabilities is based on the demands of survival in the natural world. It is against the framework of these capabilities that we must consider and tailor the interfaces of the modern world.

Many of these fundamental parameters of human capability have been well established.

Wavelengths of energy visible to the human eye are defined (400 to 700 nm).

Frequency and intensity of sound waves comfortably heard by the human ear have been established (20 to 20 000 Hz with a sensitivity ratio of 1 : 1 billion).

Resolution of tactile sensation that can be felt over different areas of the human body has been mapped out (2 to 3 mm at the finger).

Across this range of sensory capabilities some modes are better suited to communicating certain information than others, depending on the context of the task and the environment.

Variation in sensory capabilities across a user population places further limits on the method of information presentation.

Table 1. Modified Fitts list.

Human	Machine
Inductive reasoning	Deductive reasoning
Pattern detection	Parallel processing
Improvisation	Processing capacity
Task shedding when overloaded	Accuracy of calculation
Relational memory retrieval	Speed of response
Integration of multiple senses	Absence of fatigue or boredom
Adaptive learning and skill	Supra-human sensory capabilities
Mobility	Data storage and search

Many issues face the interface designer. The task is to clearly and unambiguously present the person with the necessary information without overloading them and to make it as easy as possible for them to implement the actions they decide upon in a timely and error-free manner.

Measuring systems are there to interface the use to the state of selected physical variables; human-factors engineering is therefore part and parcel of measuring instrumentation design.

3 THE VISUAL INTERFACE

Human vision is the primary source of input for most human tasks. The eye shares many similarities with the camera focusing light with a lens on the retina.

The retina is made up of approximately 100 000 low light sensitive rods and 60 000 color sensitive cones that provide daylight vision, which in combination offer a range of light sensitivity from 1 billion to 1.

Photoreceptors are not evenly distributed on the retina. In the region of the fovea, the high concentration of photoreceptors provides a resolving power of 1 min of arc as compared to several minutes of arc in the periphery of the retina. As a consequence of this distribution, visual attention is based on head and eye movement focusing the area of interest on the fovea with the peripheral vision providing an additional cue to visual attention through the detection of movement. Movement of the eye is finely controlled and can range from slow to extremely fast (over 400 degrees per second).

It is important to recognize the capabilities and limitations of the human visual system when designing interfaces to take advantage of this primary sense.

A key factor to recognize in the design of visual interfaces is the singular attention of the human. People can look quickly between different locations but can only focus on one point at any one time.

Efficient design of displays requires a consideration of the task they are supporting and the load the design places on the operator. If there are a number of display elements that need to be monitored at the same time, they should be colocated in a logical fashion to reduce the time taken to search between locations.

A good example is the design of the head up display for fast jet aircraft cockpits. The display places the key information required by the pilot to operate the aircraft and its supporting systems, as an overlay on his view of the world. This allows the pilot to spend more time 'head out', observing the world through which they are flying thus raising the opportunities for developing situational awareness.

Color should be used with care. Although color can help highlight important features of a display, or indicate the

Table 2. Color associations.

Color	Conventions	Other associations
Red	Danger Emergency Failure	Stop Fire Hot
Blue	Mandatory	Water Sky Off Cold
Green	Safe	Safe passage, Exit Continue
Yellow and black	Caution	Ambient Delay Warm

grouping of common elements, its overuse can have the opposite effect, distracting the user instead.

Inappropriate color combination can reduce the contrast between different elements of a display disguising information and reducing the effectiveness. It should be noted that people link colors with particular attributes or task specific associations. In addition, it should be noted that individual variations amongst the population, for example, color-blindness, and changes to the ambient light levels can dramatically alter the perception of a color scene.

Table 2 includes a list of common associations for colors and their combinations.

4 THE AUDITORY INTERFACE

Hearing is the next most important source of information input after vision for most tasks. Hearing and speech is a highly evolved method of communication.

When we talk to another person, we are using an interface in which participants sense and interpret the speech and expressions of the other person, forming a response and then responding through the formation of speech or some other action.

Humans learn to speak through a process of training, trial and error, when growing up. It is a highly effective and adaptable form of communication containing subtle nuances that convey information at a number of levels that are received and recognized in a way that is currently impossible by machines.

When communicating with a person who does not share your language, it is often possible to adapt your technique using gestures and a few shared words to convey meaning, but the speed and quality of the information is degraded.

At a lower sound level, the introduction of noise may mask some of the information leading to misinterpretation. In these circumstances, we tend to use visual cues such as the movement of the other person's mouth as well as an anticipation of what might be said, based on the subject of the discussion, to fill in gaps in the information masked by the noise.

Hearing can be a useful alternative to vision for the communication of information, particularly when we do not wish to be distracted from a visual task. For example, it is possible to conduct a simple visual search task while maintaining a conversation.

It can also be a useful medium when the display of visual information would be difficult for example, at night. Auditory warnings are used extensively to attract our attention away from a visual task. The characteristics of the sound can be tailored to convey the importance of the message thus increasing the amount of information a relatively simple warning sound conveys.

5 THE OTHER SENSES

While vision and hearing are the main sources of information to the human user of a system, we are sensitive to many other attributes of our environment.

Walking along the street draws on visual input to judge where to place our foot when taking the next step. Auditory information warns us if there is traffic approaching from behind. Our vestibular system continually updates our orientation without which balance would be impossible. Proprioceptors in our feet tell us when our feet are in contact with the ground. Our linked senses of smell and taste might play a part in directing our destination in search of somewhere to have lunch.

The ability of the human to filter and integrate the vast amounts of information continually received from our broad range of sensory capabilities and our ability to adapt our processing of this information in a changing environment sets us apart from machines.

6 CONTROL DEVICES

The human ability to sense our environment is just one part of the story when considering the human, or the Man Machine Interface (MMI).

As mentioned above, the information we receive is processed at several levels in our central nervous system from reflex responses up through lower brain control to higher cortical control. Having decided on a response to the information gathered by our senses and their relation to other memory and cognitive processes we are able to influence our environment. Our methods for influencing the environment are based on bodily control through the

movement of our limbs and our ability to communicate using speech.

From the most basic tools, such as a rock used to open a nutshell, through to a pilot's yoke linked to a sophisticated fly-by-wire jet aircraft, we convert physical movement, through gripping objects, into a desired outcome. Our ability to tailor these tools to our requirements is central to our ever-advancing technologies.

The shaping of the earliest flint tools to the hand size of its intended user illustrates our early skills in tailoring to our requirements. In the case of aircraft controls, the need to tailor the design to the capabilities of potential pilots continues to hold true with the need to avoid the exclusion of users on the grounds of hand size or grip strength, and so on.

The usability of controls extends beyond the person's ability to physically grip or manipulate the control device. Our ability to use controls is based on an expectation of their function. We are able to use a hammer because we have past experience of hammers and their use. The hammer itself provides us with cues to its function with an obvious area to grip and a flattened metal area for hitting the nail.

More advanced tools/machines rely on the same components of knowledge and cueing to convey their function to the human operator. For complex systems such as an aircraft flight deck, this places a requirement for consistency in design and operation without which pilots experienced in flying planes with different controls will require retraining and, even once trained, may make errors owing to a tendency to revert to the operational expectations of other systems.

This expectation of how something operates can be seen on a wider scale with the existence of population stereotypes such as association with twisting a tap head clockwise to turn it off. When designing a system that may be used by a multinational population, these stereotypes may change with a consequent risk of error.

Where clear conventions exist, such as turning a wheel clockwise to move a vehicle to the right, they should be adopted.

Where there is some doubt over the natural physical relationship between the cause and effect, for example, using a horizontally mounted joystick to move a cursor up or down a screen, the stereotype of the population should be explored.

As the design of interfaces shifts from physical design toward one of human computer interaction design, this tailoring of the controls has developed a new emphasis on information flow and the associated mental workload. The basic objective of ensuring that the capabilities and limitations of the potential user population are respected continues to hold true across this range of technologies from the very simplest through to future technologies yet to be conceived.

This article has provided an introduction to the Man Machine Interface. The topic is extended by explaining the six human factor domains in **Article 96, The Domains of Human Factors Integration, Volume 2**; and **Article 97, Design Methodology, Volume 2** covering how to go about designing for Human Factors.

REFERENCES

Diamond, J. (1999) *Guns, Germs, and Steel; The Fates of Human Societies*, W.W. Norton & Company.

Fitts, P.M. (1951) *Human Engineering for an Effective Air Navigation and Traffic Control System*, National Research Council, WA.

Fitts, P. (1954) The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement. *Journal of Experimental Psychology*, **47**, 381–391.

Sydenham, P.H. (2003) *Systems Approach to Engineering Design*, Artech House.

96: The Domains of Human Factors Integration

Nicholas I. Beagley
DSTO, Canberra, ACT, Australia

1 Human Factors Integration and MANPRINT 628
2 Manpower 629
3 Personnel 629
4 Human Factors Engineering 630
5 Training 630
6 Safety 631
7 Hazards 631
References 631

1 HUMAN FACTORS INTEGRATION AND MANPRINT

Large-scale acquisition, for example, military procurement, is normally approached at a high level with the aim of achieving a general capability. The lower level objectives and the functionality required from any system designed to fulfill this capability are often unclear at the outset.

Procurement programs generally set out a vision of the capability and then set about establishing, in ever-increasing detail, the system design solutions necessary to achieve the vision – *see* **Article 63, Outline of Systems Thinking, Volume 2** and other articles within the section *Engineering a Measuring System.*

Commonly adopted procurement processes share similarities in their approach, stepping through defined sequences of specification, design, refinement, acceptance, delivery, through life support, and finally disposal.

The process is normally interspersed with milestones by which the progress of the program is assessed and approval to continue is attained – *see* **Article 66, Phases of System Life Cycle, Volume 2** and **Article 69, Requirements Allocation, Volume 2**.

This process is well suited to the specification and testing of tightly defined engineering solutions designed to meet a clear functional requirement – *see* **Article 69, Requirements Allocation, Volume 2**.

However, where human performance contributes to total system effectiveness, the users' capabilities and limitations must be taken into account. At the same time, the added complexity of human involvement reduces the predictability of design alternatives taken within the development process, thus limiting the degree to which the requirements can be tightly specified or met.

Despite these difficulties, accounting for the human component must not be overlooked. To the contrary, the need was recognized for a more effective way of inserting Human Factors Analysis into the defence procurement lifecycle. From this requirement, emerged the Manpower Integration (MANPRINT) program, developed by the US Air Force in support of their system acquisition program, MANPRINT (2003).

MANPRINT sets out the necessary Human Factors inputs to the procurement process in the form of prescribed analysis and reporting for each stage of a system's procurement. Further guidance was provided through its positive direction to design teams to consider defined areas of Human Factors assessment, called *domains*. These are considered here.

UKMOD adapted the MANPRINT framework to map onto their own updated procurement process (UK MOD 2000). The UK's Human Factors Integration (HFI) program defines the Human Factors specific roles and responsibilities of the Capability Working Groups and Integrated Project

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

Teams prescribed by the overall procurement process thus giving HF a formal place in the development cycle.

Central to the HFI process is the activity of Early Human Factors Analysis, in which concepts are reviewed within the domains of HFI. This is started at the earliest opportunity to avoid where possible costly changes to the final design due to the discovery of incompatibilities with the human user. The specified activity of Early Human Factors Analysis compiles a log of the issues identified within each domain, the plan for each issue's resolution, and the process toward an acceptable design.

As defense equipment includes numerous measuring systems that have to perform in demanding conditions, MANPRINT methodology will be found useful for measuring systems design of all kinds.

2 MANPOWER

The Manpower domain of HFI is concerned with the number of people required to operate and maintain a system.

This consideration goes beyond the issue of the person sitting at the controls of a piece of machinery, extending to the people required to train that person, maintain the system, relieve that person when they become fatigued, and so on.

When considering large system structures involving many people, such as the manning of a ship, the logistical issues of manpower can become critical to the sustainability of the system. For example, the total cost of a naval ship owes far more to the manning and sustainment cost than it does to the hardware purchase costs.

This situation is commonly met in many of today's systems. A design or operational approach that efficiently uses the capabilities of the people to reduce the total number of people required to operate a large system can represent a significant financial saving.

Reducing the number of people available for operating a system is likely to raise the responsibilities and associated workload placed on those remaining. The sustainability of this approach must be considered across a range of scenarios. For example, it may be possible to conduct the normal operations of the ship under a peaceful scenario with a given staffing level. If, however, a wartime scenario places raised demands on the crew and their ability to consistently perform at the required level cannot be sustained, this situation must be accounted for.

Important implications can arise if the design demands a higher entry requirement for the people chosen to operate the ship than those currently employed in similar roles. If this is the case, then the availability of these people in the workforce and the ability to motivate them into working on the future ship will be important assumptions that must be planned for in the full life considerations of setting up and maintaining a capability. Similarly so with complex measuring systems such as are found in aircraft, transport systems, industrial plants, and medical health systems.

Where a system relies on a particular set of skills at different levels within the system, its design should consider the program's phasing demands to supply the right people with the right level of skills at the right time along with a sustainable succession plan over the lifetime of the system.

3 PERSONNEL

The Personnel domain of HFI is concerned with the characteristics of the people in relation to the requirements of the total system, in pursuit of the optimal system performance.

People's characteristics vary considerably; it is necessary to recognize their capabilities and limitations in the context of the system's demands on the person. It is also important from the point of selection and training and their associated costs to recognize when and where, in which subsystem, a system design is likely to demand particular aptitudes or knowledge.

In some cases, for instance in aircraft and race car cockpits, the system places physical constraints on the individual such as body size, physical strength, or a resistance to motion sickness.

Whilst training and experience can help develop candidates to meet a system's requirements in terms of physical strength, endurance, and motor skills, some attributes such as anthropometry are fixed.

Some trades require certain physical criteria for selection designed to ensure that the people selected are capable of meeting the requirements of the systems they support. For example, selection into the army demands a level of physical fitness that allows the designer of supporting systems to make certain assumptions about the target population's physical abilities.

The mental aptitude of the user group can greatly decide their ability to develop the cognitive skills necessary for contributing to the performance of a system. For example, market traders must demonstrate a capacity for extracting relevant information from a multitude of market information sources and apply this information, along with relevant knowledge and experience, to make informed decisions on the purchase and sale of assets. The personal attributes necessary to be successful in this occupation are not common across the whole population.

The design of any system should clearly establish the assumptions about the personnel it is making with regard

to the skills, limitations, and expectations of its intended user population; this will help ensure a significant mismatch does not arise. For large-scale system acquisitions, this is normally captured in a document called a *Target Audience Description* (TAD).

4 HUMAN FACTORS ENGINEERING

The Human Factors Engineering (HFE) domain of HFI is concerned with the physical and operational integration of the user with a system's process and associated technologies. These considerations extend to all aspects of human/system design and include guidance documentation, and a range of tools and techniques designed to ensure the full integration of the system's characteristics with the capabilities and limitations of the human operator, the demands of the task, and the attributes of the environment.

HFE within the HFI process is normally concerned with the specification of mandatory criteria such as are given in Def Stan 00–25 (A UK MOD Defence Standard) and Mil Std 1472F (A US DOD Standard) against which designs will be judged. These not only prescribe documents but also offer guidance for specifying both the known user specific–design constraints and the appropriate methodologies for ensuring that user requirements are established and accounted for throughout the design process.

An increasing concern of HFE is in the area of Human Computer Interaction (HCI). An analysis of the underlying task is frequently the starting point from which concept designs are created and iterated on the basis of subjective user feedback, workload assessments, and objective measures of human performance.

The traditional area of ergonomics also begins with the need to understand the task and the characteristics of the user to ensure an appropriate fit. Human body modeling, prototype development, and user-based testing and evaluation, all contribute to the activity of ensuring the fit between the human and the system. The knowledge, experience, and skills, both physical and cognitive, must also be determined to inform the design process. All of these processes have a large supporting literature – Chapanis (1996), Jacko and Sears (2002), Sanders and McCormick (1993), Tilley (2001), Wickens, Gordon and Liu (1997).

Human Factors analysis techniques must be sensitive to variations in the user population and the different contexts of operations. Designs must take into account different task demands dictated by possible scenarios, in which the system might be required to function.

The impact of different environments on the user population's ability to use the system in the intended manner needs consideration to allow appropriate changes to be made to the design or the intended operational procedure.

5 TRAINING

The Training Domain of HFI is concerned with the knowledge and skills required for operating and maintaining the system. The importance of training increases with the complexity of a system and with the negative consequences of system failure. Some instrument systems are easy to learn to operate but some require specific training to be used correctly. For example, using a simple digital voltmeter should not tax any competent technical person but setting up an automated test equipment console is another matter.

Consideration of the training requirements of a system should produce recommendations regarding what should be taught and the method that should be adopted for teaching.

There exist many approaches to training and many different scales of training appropriate to the system under consideration. Training for a simple system such as a common consumer product, for example, a television set, might simply take the form of a reference leaflet. People with experience of similar systems might not require the leaflet to access the regular features of the television, using it only to explain unfamiliar features.

Complex systems, for example, large software applications, such as setting up an exhaustive in-flight test plan for a new car, railway, or aircraft, are often accompanied by a training course that leads potential users through the available functionality and provides worked examples to ensure an understanding of the concepts on which the software's functionality is based.

Techniques, such as computer-based training and self-instructional packages, have increased in popularity given their ability to support training at a pace determined by the trainee without the overheads of a human trainer.

Computer-based training tool development should be based on a careful analysis of the knowledge and skills required by the trainees, and so are not suited to all training demands for a given system.

The use of simulation as a training tool has expanded the range of tasks that can be trained for using a computer. A current limitation to computer-based training is that it is only able to provide limited feedback. Humans still remain the most effective trainers across a wide range of contexts, offering insights based on a deep base of knowledge and experience combined with the power of inductive reasoning and wisdom to appropriately guide the trainees.

The normal form of training analysis and reporting applied to large-scale procurements is called the *training needs analysis* (TNA). Development of a TNA is conducted early in the design and development process setting out anticipated training system requirements such as the required facilities, the necessary fidelity, the time requirement for trainees, and so on.

6 SAFETY

An important responsibility of the system designer is to safeguard the well being of people using or coming into contact with the system. The System Safety domain of HFI is concerned with the risks to human health that might arise from system failure or use in an unintended manner.

Consideration of system safety is in reality an analysis of risk. This requires consideration of all aspects of system design and operation, in which the demands of the task place the operator, or bystanders, under a risk of injury.

This analysis must include opportunities for error in the operation of the system.

In each case, a standard risk assessment is normally undertaken, in which the likelihood of the risk occurring and the consequence for human safety are rated.

Issues revealed by this process can have additional implications for system performance that may not represent a significant risk to user safety, but may still result in the failure of a system to meet its objectives.

The output of a safety assessment and its risk analysis is usually a hazard log that characterizes the risk and draws on relevant expertise to establish measures that should be taken to mitigate the risk, for example, the mitigation of the risk of sailors falling from their yachts in high seas is usually a combination of harnessing, lifejackets, and location beacons.

In some cases, a conflict will arise between the proposed safety measure and the performance of the system. On rare occasions, users may be faced with a potential compromise between the risk of injury against the performance of the system.

When failure of the system has significant implications for the welfare of the individual, for example, in some military systems, there may be circumstances in which the individual may elect to bypass a safeguard in the interest of system performance. In such cases, it is important for designs and procedures to be developed that ensure safe use under normal circumstance and a reasonable balance of risks in life-threatening situations. Within the guides of relevant legislation, the design should allow the user to make these decisions when necessary. Safety in systems is also covered in **Article 99, Safety Organization, Volume 2**; **Article 100, Safety Analysis Methods, Volume 2**; and **Article 101, Assessing and Demonstrating Safety, Volume 2**.

7 HAZARDS

Whereas the Safety domain was concerned with risks related to the use and abuse of a system, Health Hazards domain of HFI is concerned with the conditions of operation inherent in the system and their impact on the health and safety of the human operator.

At the heart of hazard analysis is the consideration of the system's operational characteristics and the environment in which it operates.

Factors such as noise and vibration can have health implications for its operators and those around them.

Noise dose limits are set in legislation on the basis of dose limits worked out as a combination of intensity and frequency range. Vibration levels can have significant implications for the health of the individual with the link between prolonged doses of mid to high frequency vibrations to back pain arising from musculoskeletal injury. The combination of vibration, repetition, and wrist deflection when using power tools, has been linked to high incidence of repetitive strain injuries (RSI) such as vibration white finger and carpal tunnel syndrome.

Many mechanical systems include an inherent risk to the operator. Many systems contain toxic substances or produce toxic fumes from which the operator must be protected.

Similarly, moving parts present a risk to the operator with opportunities for cutting, trapping, and crushing limbs and digits. Well-established methods of guarding the operation of mechanical systems should be in place and should be tested against a range of potential operational scenarios including the risk of circumvention by the operators themselves.

Inherent risk to the vision of the operator can arise. Activities that require long periods of visual attention should be broken up into shorter periods to reduce fatigue. Some systems such as the use of high-magnification optics require safeguards ranging from training and procedures through to reactive optical filters that prevent operator blindness caused by pointing the device at a high luminance source such as the sun or a laser.

Extremes in temperature are a hazard to the operator whether they are a result of the external environment given the scenario or because of the heating or cooling of a component.

Where a risk is identified, consideration should be given to guarding against the risk, including an assessment of whether the system's operational procedure raises or lowers the individual's exposure to the risk.

REFERENCES

Chapanis, A. (1996) *Human Factors in Systems Engineering*, John Wiley.

Jacko, J.A. and Sears, A. (2002) *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications (Human Factors and Ergonomics)*, Lawrence Erlbaum Associates.

MANPRINT (2003) MANPRINT documentation is only available as a web document on http://www.manprint.army.mil/manprint/guides/mp-ref-guidebook.asp.

Sanders, M. and McCormick, E.J. (1993) *Human Factors In Engineering and Design*, McGraw-Hill Science/Engineering/Math.

Tilley, A.R. (2001) *The Measure of Man and Woman: Human Factors in Design*, John Wiley.

UKMOD (2000) *A Practical Guide to Human Factors Interaction: A Guide for Integrated Project Teams and Capability Groups*, Version 0.9.3, 19 July, Ministry of Defence, UK, also available on http://www.ams.mod.uk/ams/content/docs/hfiweb/hfihome.htm.

Wickens, C.D., Gordon, S.E. and Liu, Y. (1997) *An Introduction to Human Factors Engineering*, Addison-Wesley.

97: Design Methodology

Nicholas I. Beagley
DSTO, Canberra, ACT, Australia

1 The Objective of Design 633
2 Requirements Specification 634
3 The Design Life Cycle 634
4 Accounting for User Variation 634
5 Design Standards and Guidelines 635
6 Concept Modeling 635
7 Product Evaluation 636
Further Reading 636

1 THE OBJECTIVE OF DESIGN

The underlying objective of the design process is to provide a new or enhanced capability. Design can be either evolutionary or revolutionary in nature.

An evolutionary design takes discrete steps forward from an existing design's form and functionality.

The alternative is a revolutionary design in the capability that is normally associated with the availability of a new technological development that will support a new capability or an ability to do something in a new and improved way.

One difficulty associated with evolutionary design is that the expectations of the user are firmly aligned to the current product. This can have the effect of coloring their opinion of prototype systems that operate in a slightly different way.

Revolutionary design faces a different problem in that with no model to relate to, the potential users of a revolutionary system may have difficulty in grasping the role and function of the new system sufficiently to provide useful feedback for the system's design and evaluation.

In either case, the role of the user, as the foremost expert in their requirements, is vital in guiding the design process. The challenge facing the human factors practitioner is to ensure that users are provided with appropriate levels of information and experience of emerging concepts, and that any bias within the population are recognized and balanced where appropriate.

When designing a system, many attributes must be considered on the way to meeting its associated capability. Designs are a trade-off of attributes such as functionality, ease of use, cost, maintainability, training demands, and so on. For example, if you are designing a consumer product such as a kettle, the product's overriding objective is to appeal to the consumer market and thus achieve sales. The attributes necessary to achieve this go beyond the technology required for boiling water – to the cost of production, ease of use, and aesthetics of the design. Other issues such as brand styling and reputation may place additional demands on a design team.

Any design can be judged on the basis of its comparative achievement in providing a defined capability, that is, it must be fit for the purpose. It must not only provide the necessary functionality but also take account of the users' capabilities, the demands of the task and the restrictions of the environment.

As a consequence, designs should be considered in the broad context of their operating environment. Setting up the boundary limits of a project are covered in **Article 68, Setting the System Boundaries, Volume 2**. If a design does not account for the user's requirements, it will at best be inefficient and at worst be rejected outright by its intended user. Phases of the life cycle of a development are covered in **Article 66, Phases of System Life Cycle, Volume 2**. More on requirements generation

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

is found in **Article 69, Requirements Allocation, Volume 2**.

2 REQUIREMENTS SPECIFICATION

Specification of large-scale systems is frequently conducted as a two-stage process starting with a capability focused user requirement that is then refined to become a more detailed, functionality focused, systems requirement.

This is essentially driven by engineering-based, quality control processes. Guidance for the input of Human Factors into this process has been developed in the form of the UK's Human Factors Integration (HFI) Process and the US's MANPRINT program. In these programs, human analyses and their deliverables are mapped onto a defined acquisition process, in which the various customer and supplier responsibilities are prescribed, and the milestones for acceptance, delivery, and payment are set out within the framework of the design and development process.

The user requirements document normally sets out the characteristics of the current capability, its users and the procedures for use whilst describing the enhanced concept, its associated functions, and the context of operation.

Also defined is an initial assessment of the constraints, such as user recruitment, that might be placed on the future system and the measures of effectiveness, (MoE) by which a system could be evaluated. MoEs should be sensitive to all aspects of the concept capability and are particularly important for the evaluation of design options. Example measures include subjective measures such as user acceptance and objective measures expressed in terms of system performance and usage errors.

For large scale systems, the user specific information should include an overview of the likely capabilities of the target population, available numbers, existing training demands, career, retention issues, and so on.

The system requirement document (SRD) is normally a more tightly defined description of the required system and its components. It is derived through a process of user based–concept refinement using prototypes developed in response to the user requirements document (URD). This document sets out clearly defined functions and their associated performance criteria.

Human specific performance criteria should relate to issues of safety, operability, reliability, maintainability, training, and so on, and might include such performance standards and metrics as completion times, error rates, lifting limits, workload limits, and so on.

It will specify the necessary standards to adopt and should include appropriate Human Factors standards for design such as the UK's Def Stan 00-25 or the US's Mil Std 1472F.

3 THE DESIGN LIFE CYCLE

The design life cycle is a process of concept refinement to the point of production of a solution that can be implemented. The design of a new car or a new training program each start with a concept that brings together particular attributes perceived to present the intended product with a niche in the marketplace.

The design life cycle should extract and question the assumptions surrounding these perceived qualities whilst seeking to embody them in the final design.

The scale of the design process should be tailored to an appropriate level given the available resources and the consequence of design failure.

Safety critical systems rely on stringent safeguards that prolong the design cycle whilst some low cost commercial products can afford to fail in the market place as long as lessons are learned that improve the chances of success for subsequent products.

Design should be considered to be a competitive process. There are costs associated with the design process in time and money as well as disruption to the intended user of the new design, particularly when it is intended to replace an established system. The value of the proposed design enhancement should always be compared to a baseline taken from what has gone before. Where available, a comparison should be made with alternative designs.

If the life cycle process of a design raises the costs beyond the resources of the potential purchaser, or delays the delivery to a point that it becomes obsolete before delivery, the life cycle process will have failed.

There has been increasing pressure to reduce the time and costs associated with the design life cycle. This, in parallel with recognition of a user's inability to completely specify their requirements in the early stages of the design process, has led to a shift toward a process of concept prototyping and design evolution through user testing and virtual prototyping.

Within this process, there remains a need to establish clearly defined criteria for acceptance with which to evaluate the progress of a design and to manage the contractual obligations between the customer and the supplier.

At the same time, there has been recognition of the risk of failure associated with design and the need for a closer partnership between the user/specifier and the supplier/designer in moving toward a product that achieves the best balance between the inevitable compromises of design.

4 ACCOUNTING FOR USER VARIATION

Human variation must be recognized when designing for a heterogeneous population. Each person is an individual

in the whole population and it is important to recognize and design for the range of capabilities in a targeted user population.

Variation in size, weight, strength, sight, hearing, memory, experience, skills, and so on, can each influence our ability to use an interface or play a part as a component of a wider system.

From the perspective of design, there are four main approaches for accounting for human variation. These are product tailoring, self-adjustment, multiple versions, and one size fits all.

The tailoring of a suit to our body shape has similarities with bespoke software and product design offering confidence in a product's fitness for purpose, at a price.

The approach of allowing self-adjustment is highly suitable for many products. Car and office chairs must accommodate a wide range of the population and must therefore include the capability for adjustment, without which large proportions of the community would find the product unusable. Many of the mainstream software applications allow a degree of user adjustment to suit their working practices. It should be noted that too much adjustment capability can bring with it costs in adjustment time or even weaknesses in the structural integrity of a product.

The approach of producing multiple versions is familiar to us all through the purchase of off-the-shelf clothing. This term off-the-shelf has been adopted more widely to describe products that have already been developed and marketed. The two usages bear similarities in their associated cost savings.

The final approach of one-size fits all is appropriate for a subset of products depending on their intended purpose. The term is frequently applied to T-shirts that will look like a tent on some people and stretch to fit others. This design compromise frequently offers considerable savings at the cost of poor fits for proportions of the potential user population.

5 DESIGN STANDARDS AND GUIDELINES

The normal repository for the knowledge gathered through research on human attributes is the published literature and teaching textbooks; it is widely dispersed. The broad application of human sciences findings to system design and their implications for system success has led to the need for design standards and guidelines that are accessible to the design community.

These documents often have the dual role of offering advice to a designer as they seek to develop a product that integrates with the human, and at the same time setting out the expectations of the procurer by defining acceptable levels of performance of a system.

This is particularly important for areas such as human health and safety, in which design standards play an essential role in safeguarding future users of a system.

Several organizations develop standards for engineering design, including the International Standards Organization (ISO), the UK's Defence Standardization Organization (Dstan) and the US Department of Defence (USDOD).

Key Human Factors standards from these organizations include ISO 9241, Def Stan 00-25 and Mil Std 1472F. Whilst some are created for defense application, they are also useful to all human factors uses. Standards from the government military and some standards organizations are freely available on their Web sites. The ISO standards are available in some libraries, or can be purchased on-line.

Standards such as the UK's Def Stand 00-25 cover a broad range of human, task, and environmental attributes and their design implications.

Def Stan 00-25 assists the designer by defining known parameters of the military user population it serves providing the designer with reference tables for operator size, strength, sensory capabilities, and so on. It also provides guidance on appropriate strategies for design and testing to ensure that the government's requirements for integration of the human users are met.

6 CONCEPT MODELING

It can be quite difficult to tightly define specifications for a concept early in the design cycle. Without a product or system on which to base their assessment, users and designers alike encounter problems in specifying and refining their concepts.

A common approach to making these concepts more concrete as a tool for evaluation or design refinement is through modeling.

A model of a system can take many forms. The form chosen should reflect the purpose of the intended model. A simple paper-based prototype of a series of screen interface designs can be a low cost, quick to develop, and flexible tool for communicating a concept and seeking user feedback. High-end prototyping applications can be used to simulate the design and intended functionality of a proposed system. This might be a logical progression from the use of paper-based prototypes to refine a concept. The degree to which these methods are used depends on the complexity of the system being designed and the consequences of inappropriate design features remaining in the final design.

Given the costs associated with altering physical designs, there is an obvious emphasis on prototyping to reveal

design issues prior to the final production run. Physical prototypes can take many forms from cardboard mock-ups through to machined pieces. Computers are playing an increasing role in the modeling process with computer aided design techniques used to refine design concepts and to provide visualizations. Human body modeling tools offer the designer the ability to examine issues of population fit and reach at the very earliest stages of concept formulation, helping them to weigh up alternative design strategies and avoid the cost of large changes that would have been associated with discovering these problems further along the design path. More recently, the use of stereo lithography has increased to provide physical outputs to computer designs providing further realism to the process of evaluation.

7 PRODUCT EVALUATION

The evaluation of designs can be conducted at several levels. Evaluations might be an assessment of the design's fitness for an intended purpose or it might seek to compare two existing designs to determine which is the best, or the least bad, design for the intended purpose.

In the process of either evaluation, it is possible to reveal attributes of a design that are particularly suitable or unsuitable thus providing a third option by which desirable components of competing systems are brought together, or unsuitable components are enhanced in the process of developing a new and improved system.

It should be recognized when evaluating a design's fitness for a purpose, that it may be that the design was developed to meet a slightly different requirement of an alternative user group.

Variations in user, task, or environmental context can have important implications for the success of a design. Whilst there is much that can be done to assess the attributes of a potential system through basic analysis, the findings cannot assure one that it will provide absolute assurance of suitability.

Prior to final acceptance a design should be tested in its context, or perhaps within a close representation of that context, to reveal issues of suitability.

For example, the down selection between a number of new cars might be made on the basis of their promotional literature but the final selection before purchase normally involves a test drive of the remaining options. If you were buying a fleet of cars, there may be value in arranging for a longer term loan of a vehicle to consider other issues of suitability within your context of intended operation.

The process of comparative evaluation is similar with the availability of many tests possible for comparing competing products against a priority list of desirable features and negative attributes to be avoided.

There exist a number of tools and techniques that can be used to help in the development of a priority list of attributes in situations in which different users can be expected to hold different opinions. Approaches include workshops, ballots, and a tool called the *repertory grid*, in which attributes are compared individually and analyzed to produce a ranked list.

With all approaches, there is a need to ensure that users are given sufficient access to the designs under consideration of representative circumstances. Where required, training should be provided and the implications of skill development with training over time accounted for in the analysis and the interpretation of the user's subjective responses to a system.

FURTHER READING

DOD-HDBK-763, *Human Engineering Procedures Guide*.

ISO 9241, *Ergonomic Requirements for Office Work with Visual Display Terminals*, Parts 1 to 17.

MIL-STD-1472E, Department of Defense, *Design Criteria Standard: Human Engineering*.

MIL-H-46855B, Department of Defense, *Handbook: Human Engineering Guidelines for Military Systems, Equipment, and Facilities*.

MIL-HDBK-759C, Department of Defense, *Handbook for Human Engineering Design Guidelines*.

MIL-STD-1801, *User/Computer Interface*, 29 May 1987.

Ministry of Defence, Defence Standard 00-25, *Human Factor for Designers of Equipment*, Parts 1 to 13.

NASA-STD-3000, *Man-Systems Integration Standards (MSIS).*

98: Reliability and Maintainability

Patrick D.T. O'Connor
Consultant, UK

1 Concepts of Reliability and Maintainability 639
2 Causes of Failure 640
3 Design for R&M 641
4 Testing for R&M 642
5 Reliability Data Analysis 642
6 Related Concepts 644
Related Articles 644
Reference 644
Further Reading 644

1 CONCEPTS OF RELIABILITY AND MAINTAINABILITY

1.1 Reliability

Reliability is the ability of an item or system to keep operating, or to be available for operation, over a period of time without failing. Reliability can be defined as a mathematical probability of not failing or of being available for use.

Durability is the ability of an item to withstand wear out mechanisms such as fatigue, wear, and so on. Durability is usually expressed as a minimum time before the occurrence of wear out failures.

Reliability is also expressed as the mean time to failure (MTTF) (for an item that can only fail once, like a light bulb, microprocessor or interplanetary spacecraft), or the mean time between failures (MTBF), for an item that can fail and be repaired a number of times, like a car or an electronic assembly.

The MTBF measure implies that the rate at which failures occur is, on average, constant over the time considered, and then MTBF is equal to the inverse of the failure rate.

In the case of components that can only fail once, the expression *hazard rate* is used. It is often expressed as failures per million hours (fpmh). Failure rate and hazard rate are conventionally denoted as λ.

In the case of a constant failure/hazard rate, the reliability (R = probability of no failures over a time interval t) is

$$R = (1 - \text{probability of failure}) = \text{EXP}[-\lambda t] \qquad (1)$$

If the hazard/failure rate is not constant, other relationships apply. The reliability formula most often used is the Weibull distribution:

$$R = \text{EXP}\left[-\frac{(t-\gamma)}{\eta}\right]^{\beta} \qquad (2)$$

where γ is the failure-free life (durability), η is the location parameter, and β is the shape (or slope) parameter.

If $\beta = 1$, then the constant hazard/failure rate is represented, and η is the MTTF/MTBF. If $\beta > 1$, it represents a hazard/failure rate that is increasing over time (wear out).

If $\beta < 1$, it represents a hazard/failure rate that is decreasing over time (infant mortality).

Figure 1 shows the three basic patterns of failure that typically occur. This is often called the *bathtub curve*. As shown, it applies best for electronic systems but not so well for mixed ones with mechanical components.

Reliability statistics and methods are fully described in O'Connor (2002).

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

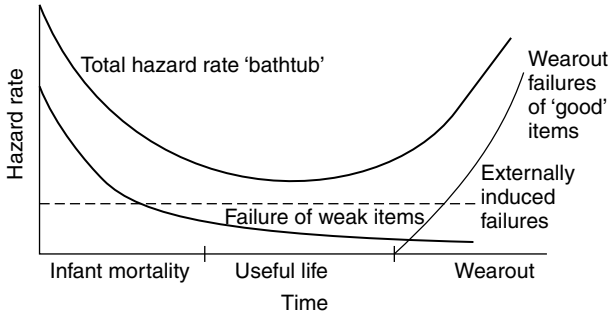


Figure 1. Patterns of failure: the bathtub curve.

1.2 Maintainability and availability

Maintainability is the property of an item to be returned to a working condition after failure (corrective maintenance, CM), or to be kept in a working condition by performing preventive (or planned) maintenance (PM). CM is expressed as the mean time to repair (MTTR). PM can be expressed as total time (e.g. man-hours) over a period.

Availability is the probability that an item or system will be available for use at a particular time, or the ratio of available time to total time. In the steady state, assuming a constant hazard/failure rate and no PM:

$$\text{Availability(A)} = \frac{\text{MTBF}}{(\text{MTBF} + \text{MTTR})} \qquad (3)$$

2 CAUSES OF FAILURE

There are many reasons and causes why a product might fail. The main causes of failures are the following:

1. The design might be inherently incapable. The list of possible reasons is endless, and every design problem presents the potential for errors, omissions, and oversights. The more complex the design or difficult the problems to be overcome, the greater is this potential.
2. The item might be overstressed in some way. If the stress applied exceeds the strength, then failure will occur.
3. Failures might be caused by variation. The actual strength values of any population of components will vary: there will be some that are relatively strong, others that are relatively weak, but most will be of nearly average strength. Also, the stresses applied will be variable. Figure 2 shows this general stress–strength situation. Failure will not occur if the applied stress does not exceed the strength. However, if there is an overlap (interference) between the distributions of stress and strength, and a stress value in the high tail of the stress distribution is applied to an item in the weak tail of the strength distribution so that there is overlap, then failure will occur. Obviously, it becomes more difficult to design against failure, or to predict reliability, in this kind of situation. We can no longer rely on simple, deterministic, values of load and strength. We now need to know how these values vary, particularly in the high tail of the stress distribution and in the low tail of the strength distribution. Values near the average, or in the opposite tails, are of little interest to engineers when considering this kind of failure, during design and test.
4. Failures can be caused by wear out. Figure 3 shows how time-dependent strength variation affects reliability. The level of uncertainty in predicting reliability increases considerably over time.
5. Failures can be caused by other time-dependent mechanisms. Battery run-down, material creep, and progressive drift of electronic component parameter values are examples of such mechanisms.

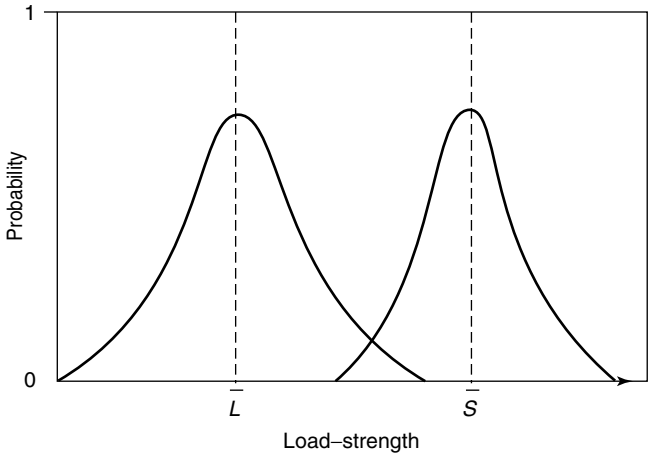


Figure 2. Stress and strength variation.

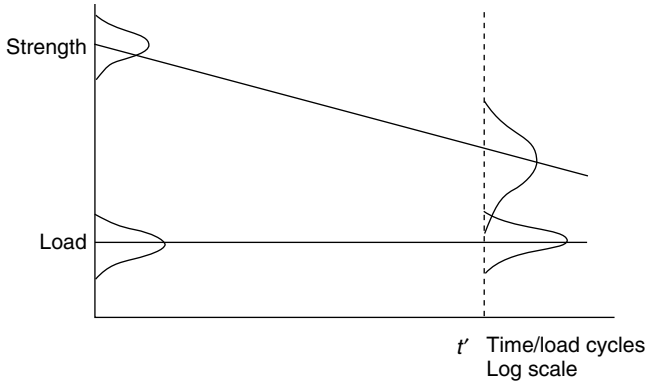


Figure 3. Time-dependent strength variation.

6. There are many other potential causes of failure. Gears might be noisy, oil seals might leak, display screens might flicker, operating instructions might be wrong or ambiguous, electronic systems might suffer from electromagnetic interference, and so on.

The important point to appreciate is that failures are by no means always clear cut and unambiguous. They are often open to interpretation and argument. They also differ in terms of importance (cost, safety, other effects). Therefore, we must be careful not to apply conventional scientific, deterministic thinking to the interpretation of failure data and the estimation of reliability.

For example, a mere count of total reported failures of a product is seldom useful or revealing. It tells us nothing about causes or consequences, and therefore nothing about how to improve the situation. This contrasts with a statement of a physical attribute such as weight or power consumption, which is usually unambiguous and complete.

3 DESIGN FOR R&M

Manual and computer-driven analyses are usually cheaper than testing, and testing also adds to project timescales. Therefore, designs should be analyzed as far as is practicable, (*see* **Article 71, Modeling Methodology, Volume 2**) to obviate or to reduce the need for testing.

Analysis can also show up which aspects of the design might need to be further investigated or confirmed by testing, by highlighting aspects that present risks or uncertainty.

Reliability, durability, and performance aspects that can be affected by variation are difficult to analyze effectively, so further testing is nearly always necessary to confirm these aspects.

Designs can be analyzed manually, with mathematics software, or with computer aided engineering (CAE) software. These include computer aided design (CAD), finite element analysis (FEA), electronic design automation (EDA), electromagnetics (EM) analysis, system simulation, and other tools.

The tools enable engineers to create and analyze complex and difficult designs quickly and economically. However, it is not safe to rely on the software to replace the need for testing. All CAE methods involve assumptions and simplifications that can, to varying degrees, generate erroneous or misleading results. No design analysis software can deal with the whole range of possible operating stresses, environments, variations, and degradation mechanisms that can cause failures.

Some methods are used specifically for R&M analysis of new designs. The methods most often used are the following:

Reliability prediction: The expected reliability of the new design is estimated, using data and assumptions regarding application environments, and so on. Reliability data is published for many component types, most notably the US Military Handbook 217 MIL-HDBK-217 (Reliability Prediction for Electronic Systems). However, this approach is nearly always misleading, on several grounds. The most notable are the following:

- Engineering components have no intrinsic, quantifiable property of failure rate.
- The mathematical models used to represent the effects of application conditions (temperature, etc.) on failure rate are not always consistent with the physical and other causes of failure.
- The data provided in such sources is nearly always of dubious quality and several years out of date in relation to the technology and capability, thus giving pessimistic/conservative predictions in comparison with what can actually be achieved with adequate engineering effort being applied.
- Failures do not generally occur at constant average rates over time. They can be time-dependent, cyclic-stress dependent, and so on.
- Component failures do not necessarily cause system failures, and system failures are not always the result of component failures.

Failure modes and effects (and criticality) analysis: Failure modes and effects analysis (FMEA) or failure modes and effects (and criticality) analysis (FME(C)A) is a method for tabulating all of the components (or functions) within a design, and asking the following questions of each one:

1. How can it fail in the application? (failure mode).
2. How likely is each failure mode? (failure probability or failure rate).
3. What would be the effect of each failure mode? (failure effect).
4. How critical is the effect? (failure effect criticality).
5. What should or could be done about it?

FME(C)A is widely used in many industries, particularly in those for which failures can have serious consequences, such as military, aerospace, transport, medical equipment, and so on.

Some industries have established standardized approaches (the US Military Handbook is MIL-HDBK-1629, and the US Society of Automotive Engineers (SAE) has also produced a guidance document).

Variation analysis: Variations (tolerances, environmental conditions, interactions, etc.) should be analyzed to ensure that they can be controlled and their effects are minimized,

see **Article 55, Uncertainty Determination, Volume 1**; **Article 52, Error Models, Error Budgets and their Calculation, Volume 1**.

Maintainability analysis: Maintenance tasks that might be necessary (failure diagnosis and repair, and PM tasks such as lubrication, cleaning, replenishment, calibration, etc.), must all be analyzed to estimate the effort involved and to ensure that they can be performed correctly by the people likely to be involved.

Conclusions: Since the analytical methods indicate the aspects of designs that are critical or might present risks, they provide an excellent basis for optimizing development and manufacturing tests. Therefore, the results of the analyses should be used to help plan and prioritize the tests, and the engineers involved should be part of the test team.

R&M design analysis methods are described in O'Connor (2002). Information on the most popular software for design analysis, with Internet contact information, is provided on the homepage of O'Connor (2002) *see also* http://www.pat-oconnor.co.uk/practicalreliability.htm.

4 TESTING FOR R&M

4.1 Reliability testing

Generally, functional and compliance testing of a new design needs to be performed only once, or maybe a small number of times or on a small number of items.

However, reliability/durability/variability testing usually involves multiple tests and test items, because of the wider uncertainty associated with these aspects. Another important difference is that we do not deliberately plan to cause failures in the first category, but we should do so in the second, using the methods of accelerated testing, as explained in **Article 214, Accelerated Testing, Volume 3**.

4.2 Statistical design of experiments DOE

In a statistically designed experiment, a range of tests is performed in which every variable is set at its expected high and low values (and possibly also at intermediate values), and the effects are recorded for each set of values. The results are analyzed using the technique of analysis of variance (ANOVA), which determines the magnitude of the effect of each variable, and of the interactions between variables.

Taguchi developed a framework for statistical DOE adapted to the particular requirements of engineering design, with the objective of making the design robust in relation to expected variations. The design process is explicitly considered to include the design of the production methods and their control.

Statistical experimental methods of optimization in engineering design can be effective and economic. They can provide higher levels of optimization and better understanding of the effects of variables than is possible with purely deterministic approaches, when the effects are difficult to calculate or are caused by interactions.

Table 1. Accelerated test/DoE selection.

Important variables, effects, etc.	DoE/Accel. test
Parameters: electrical, dimensions, etc.	DoE
Effects on measured performance parameters, yields	DoE
Stress: temperature, vibration, etc.	Accel. test
Effects on reliability/durability	Accel. test
Several uncertain variables	DoE
Not enough items available for DoE	Accel. test
Not enough time available for DoE	Accel. test

Note: These are by no means clear-cut criteria, and there will often be shades of gray between them. We must decide on the most appropriate method or combination of methods in relation to all of the factors: risks, knowledge, costs, and time.

4.3 Accelerated test or DoE?

Accelerated test and DOE are complementary approaches in development testing for reliability. Table 1 (from O'Connor (2002)) gives some guidance on which approach to select for particular situations.

4.4 Maintainability test

Testing for maintainability involves the timed performance of specified tasks. The maintainability parameters can be estimated using weighted averages, based upon the expected rates of occurrence of the tasks. See O'Connor (2002).

5 RELIABILITY DATA ANALYSIS

5.1 Reliability demonstration

Reliability demonstration testing (RDT) is based upon the method of probability ratio sequential testing (PRST).

The equipment being evaluated is operated for a period of time expressed as multiples of the specified MTBF, and as failures occur they are plotted against the operating time. Testing is continued until the staircase plot of failures versus time crosses a decision line. The decision lines are

constructed from the test criteria that have been decided beforehand. These are as follows:

- The upper test MTBF, θ_0, the level considered to be acceptable.
- The lower test MTBF, θ_1, the agreed minimum value to be demonstrated.
- The design ratio, $d = \theta_0/\theta_1$.
- The decision risks, expressed as percentages. The producer's risk, α, is the probability that equipment with an actual MTBF that is higher than required to pass, will by chance fail. The supplier's risk, β, is the opposite.

The tests are required to be performed using operating and environmental conditions that represent expected in-service conditions.

The method is the basis for US MIL-HDBK-781 and other standards, as described in O'Connor (2002). Figure 4 shows an example of a PRST plot.

Care should be exercised when considering its use. It should not be used during development testing for the following reasons:

- It implies that all failures are of equal importance, so that simple summation can provide the main criterion.
- It implies that all failures occur at constant average rates in time, in particular, that there are no significant wear out failure modes which could cause increasing numbers of failures later.
- It ignores action taken to correct causes of failures that have occurred.
- It ignores failures that have NOT occurred, but that might occur in future.
- By being based upon simulated instead of accelerated stress conditions, it does not effectively or economically stimulate failures to identify opportunities for improvement, as described in **Article 214, Accelerated Testing, Volume 3**.
- It generates an attitude that discourages the occurrence of failures, since the objective is to pass the test, rather than to generate improvement opportunities. The parties concerned argue about aspects such as whether or not particular failures should be counted and statistical interpretations, instead of concentrating on the engineering.

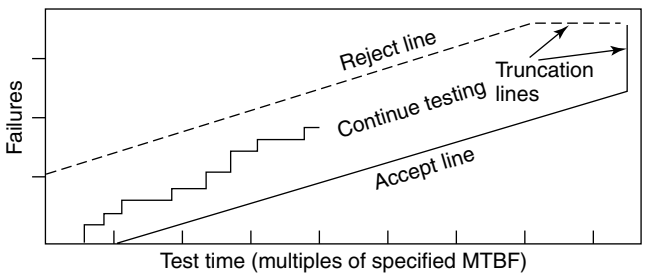


Figure 4. Probability ratio sequential test (PRST).

5.2 Reliability growth monitoring, RGM

Reliability growth monitoring (RGM) is the term used to cover a range of statistical methods that have been developed for comparing the MTBF measured during tests or in service with targets or requirements. The best known is the Duane method. The cumulative MTBF, measured at intervals, is plotted and compared with the target or required value, using logarithmic scales. The slope gives an indication of the rate of MTBF growth, and extrapolations indicate when the target or requirement might be achieved. Figure 5 shows an example.

These methods are subject to most of the same criticisms that apply to reliability demonstration tests. RGM should never be used for monitoring reliability growth during development testing, since the objective in this phase should be to generate failures as quickly as possible so that real reliability growth can be achieved by identifying and implementing improvements. It can be used as part of an in-service monitoring process, provided that it is supplemented by other methods that can help to identify failure causes and priorities.

5.3 Trend analysis

Times to failure of items on test or in service can be analyzed using a range of methods, to determine the underlying trends. The most useful of these are

- probability plots and hazard plots,
- time series analysis,

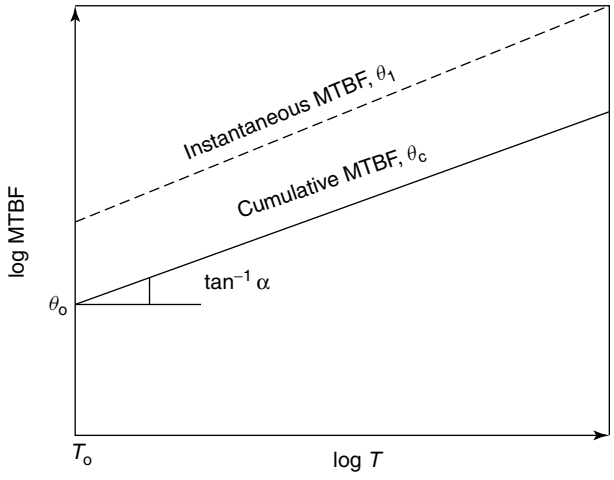


Figure 5. Duane reliability growth plot.

- proportional hazards analysis,
- M(t) method.

These methods are described in more detail in O'Connor (2002). Computer software is available for performing the analysis methods described. Software is listed in the book's homepage – http://www.pat-oconnor.co.uk/practicalreliability.htm.

6 RELATED CONCEPTS

6.1 Manufacturing quality

Manufacturing quality is just as important a driver of reliability as is design. Therefore, the reliability effort must include this aspect, see O'Connor (2002).

6.2 Safety

Safety and reliability are closely linked. The analysis and test methods used are mostly the same. Therefore safety and reliability effort and management should be integrated. *See* **Article 99, Safety Organization, Volume 2**; **Article 100, Safety Analysis Methods, Volume 2**; **Article 101, Assessing and Demonstrating Safety, Volume 2** and O'Connor (2002).

RELATED ARTICLES

Article 71, Modeling Methodology, Volume 2; **Article 99, Safety Organization, Volume 2**; **Article 214, Accelerated Testing, Volume 3**.

REFERENCE

O'Connor, P.D.T. (2002) *Practical Reliability Engineering*, Wiley, Chichester.

FURTHER READING

O'Connor, P.D.T. (2001) *Test Engineering*, Wiley, Chichester.

99: Safety Organization

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Definitions of Safety 645
2 Concepts of Safety and its Application to Measurement Systems 646
3 The Safety Conscious Culture 647
4 Responsibility for a Safe Design 647
5 Planning for Safety 648
6 Management of Safety in a Change Environment 649
7 Communicating and Coordinating Safety Issues 650
Related Articles 650
References 650
Further Reading 650

1 DEFINITIONS OF SAFETY

Safety is simply defined as 'Freedom from acceptable risk' (Railtrack, 2000).

This is a high level intention that needs elaboration.

What the concept means to an organization that designs and/or operates a system is better understood from a typical Health, Safety, and Environment Policy statement, the one chosen here being that of Rolls-Royce (2003), the designer and maker of aircraft engines, nuclear reactors, and other highly sophisticated engineering equipment that need the highest levels of safety in their manufacture and operation:

> 'The Rolls-Royce Group will make proper provision for the health, safety and welfare of its people, visitors and contractors and those in the community who may be affected by its activities. It will care for the environment through a commitment to good environmental practices. Continual improvement will be achieved by implementing the Group's HS&E Management System and related standards, which will include the setting of objectives and targets. As a minimum, the standards will meet relevant legal, industrial, and other requirements. ... etc.'

The statement continues, over a page, to address key aspects including how it will do this, who is responsible, the role of employees, its internal audit and reporting, training commitment, and, at the end explains how each person involved is responsible for implementing HS&E.

Safety statements cannot be formally precise in their expression; the concept of safety is necessarily subjective.

Greater appreciation is seen by considering the consequences of failure of a system that leads to death, injury to persons, and damage to property.

Safety cannot be designed into a system by dictating specific solutions but has to be developed by effective risk estimation and consequent risk reduction through good design.

Of the many risk factors that can be attributed to a system, it is those that cause catastrophic failure that are most closely associated with getting the safety aspects right.

The design and operation of measuring systems need safety consideration just as much as the larger systems aspects do.

Issues affecting safety are not proportionate to the cost or size of a component or subsystem. A critical element of any size and cost, such as a small hole in a test valve, could be the cause of a catastrophic failure, as it was for an early submarine that sank on its maiden voyage losing all those aboard her when the hole, being filled solidly with paint, did not show that the torpedo tube was open to the sea when the test valve was opened.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

Measuring systems are too often seen as a minor part of the larger systems in which they are placed. They are, however, as critical as other subsystems in themselves, but further, they can be part of a critical safety assessment system such as a smoke alarm in an aircraft engine compartment.

First, to state it as it is. Safety could only be totally assured if absolutely every aspect and future circumstance were known and fully understood in advance. The practical reality is that even the best design effort might still miss making a critical assumption. The safest designs are those that have their safety and operation systematically addressed by very experienced persons, within a continuous improvement regime that learns and upgrades the system as an ongoing process.

2 CONCEPTS OF SAFETY AND ITS APPLICATION TO MEASUREMENT SYSTEMS

The issues involved in safe systems are many. Here are a few of the salient ones:

Possibly the first question to be resolved is where does safety start in a design? Can it be added at some later date when the system and its use are better understood? At what stage of the system development life cycle should it be addressed?

The answer here is categorically as soon as development commences.

It will only be fulfilled well enough if seen as a 'whole of life cycle' issue. It is a matter for all to take most seriously. It cannot all be left to the person designated to take responsibility for setting it in place. The whole organization must take it seriously and develop a culture that makes designers constantly conscious of the need to assure they create safe systems.

Key issues of system safety are

- the definition of safety,
- safety acceptance levels,
- hazards, accidents, and causal factors,
- organizational fundamentals,
- management considerations,
- standards of safety and safety assessment,
- safety planning.

Safety as a concept has many interpretations (Mugridge, 2002) such as

- totally zero rate of dangerous defects (a nice idea, but impractical),
- absence of all danger ('cotton wool' mentality),
- acceptably unsafe (this is the usual principle applied). Benefits outweigh the risks is a criterion; different groups of people have different views on what this is. It is usually enshrined in the law of the country involved. For example, the UK law is the Health and Safety at Work (H&SW) Act 1974.)

A large systems development organization will certainly have its own policies and practices in place for safety aspects of systems under its development. Smaller organizations may not have their own safety policy and may have to rely on national laws for guidance.

What is acceptable depends on whom you ask:

In the United Kingdom, the ALARP (As Low As Reasonably Practicable) principle is often applied. In France, the GAMAB (Globalement Au Moins Aussi Bon) principle based on '... a level of risk globally at least as good as the one offered by any equivalent existing system' is applied.

In Germany, they use the Minimum Endogenous Mortality (MEM) principle in which '... hazards introduced by the system should not significantly increase the mortality rate, due to technological facts, above the MEM'.

In the United States, numerous safety principles statements exist. For example, the Threshold Level Value (TLV) used by the US Government Conference of Hygienists as the 'safe' level states it is the

> '... airborne concentration of a material to which nearly all persons can be exposed day after day without adverse effects'.

Note that all of these statements are subjective – it is not possible to define safety in a strictly formal way.

The public has very strong views on safety risk, but again with much subjectivity being apparent. The public perception of safety risk must be taken into account to suit the area of application of the measuring system.

For example, in rail travel, the general public will tolerate many single fatality accidents much more readily than a single multiple fatality incident. The level of safety acceptable for road travel is way below that for air travel.

A sensor system built to protect a small electric hand tool from overheating can use a more relaxed approach to safety design than a sensor used to detect overheating of an aircraft engine.

It also depends on whose safety is threatened; some groups are more tolerant than others:

- *Armed forces*: becoming more safety conscious for its friendly users but a hard area to make safe under all conditions.
- *Workers in industry*: now much more safety conscious than at prior times.

- *Civilians, for example, commuters*: often seen as secondary to the system they use but are equally in need of safety as are those who build or run it.
- *Children*: sentimental views easily come to the fore more than for the above classes.

It also depends on what type of system application the company design work is directed toward. If it is for electronic games equipment, then it will not need the same level of attention as it does for sensors to be placed to monitor a nuclear reactor vessel.

More detail on safety is provided in Brauer (1994), Hammer (1993), McPartland and McPartland (2002), Roland and Moriarty (1990) and Roughton and Mercurio (2002). Web sites exist where safety information and group activities are available, Safety US (2003).

A useful guide to the issue of safety engineering is Railtrack (2000). Although compiled for the UK railway industry, it is written as a general guide and in a language that is easily understood.

When preparing a measuring system design, however, it is at the least necessary to adhere to the safety standards in force for the project.

With so many variations being in use, it is necessary to adhere to those defined as being needed by the system development in which the measuring systems are placed.

It is also not usually feasible to cater to all standards at once, so the measuring system design built for one application may not be deemed satisfactory for another.

3 THE SAFETY CONSCIOUS CULTURE

'Safety' then, is much more than making sure a design is safe in the engineering sense and that it will be used safely.

Engineering Safety Management (ESM) is a systemic (all encompassing) and systematic (follows rules) process and a state of awareness that collectively leads to a system being sufficiently safe in its design and in its likely applications.

For the safety culture to develop, several important aspects need attention. These are

- understanding of the nature and language of safety;
- clear expression of safety responsibilities;
- existence of a sound safety culture;
- competence in safety issues by staff and users;
- maintaining that culture with all concerned including contractors and users;
- tight management of system changes to ensure safety levels are not compromised as systems are repaired and upgraded (there was the instance of an explosion-proof instrumentation enclosure being left open after some maintenance was done!);
- effective communication and coordination of safety related issues to all who should know;
- ability by those who need to know of how to develop a safety case for a system in terms of requirements, purposes, and detailed engineering;
- understanding and application of relevant standards;
- ability to plan safety into a system.

4 RESPONSIBILITY FOR A SAFE DESIGN

Even though the law will point to one, or a few, persons as being finally responsible, safety is not one person's responsibility alone. To reach good safety levels in the development of a project requires the operation of an effective safety organization. Different types of responsibility exist. They need to be defined, recorded, and formally accepted by signed agreements. Delegation of safety duties can be made – but not the responsibility.

A key organizational need is for some form of authority such as a Board of Safety Management (BSM). This group needs to have excellent communication with all necessary parts of the organization, its contractors, and its system users. After a tragedy occurs is not when such a board should be set up – as is seen in the press so often – but it should be set up as soon as the project begins. If it does not have one, it should immediately be formed.

Responsibilities suggested in Railtrack (2000) are

- setting, maintaining, and monitoring a safety policy;
- ensuring that the ESM is effectively implemented and maintained;
- agreeing to safety classifications for projects;
- endorsing key safety documentation;
- monitoring ESM performed on each project;
- appointing a system of auditors and assessors.

Allocation of safety responsibility is a serious task. It should be done in a clearly traceable manner, which usually implies a top-down military type line-management structure.

Everyone should have access to that structural information in order to know who is responsible for what.

Delegating safety responsibility to outside organizations and individuals can be fraught with weakness if not firmly handled. As people get removed from the safety structure set up, the messages about being safety conscious attenuate with regard to the whole, but can enhance for the individual concerned.

When commissioning systems, all the necessary documentation and data on safety needs to be passed into a sound depository and a clearly capable safety organization must exist for that system.

More detail is to be found in Railtrack (2000) and for the various safety standards – *see* **Article 100, Safety Analysis Methods, Volume 2**.

5 PLANNING FOR SAFETY

Some of the task of planning will already be apparent. As the Systems Engineering life cycle is traversed – *see* **Article 66, Phases of System Life Cycle, Volume 2**, a safety system must be running alongside as a whole life activity.

It is rarely possible to economically bootstrap a project with safety systems, that option getting increasingly harder as the life cycle stages move on to production. At the production stage, the cost of redesign and rework just gets too great to bear.

Figure 1 shows a suggested framework of plans and their timings (Railtrack, 2000).

For a new development, or to upgrade, the final Safety Plan usually has to pass through a Preliminary Safety Plan stage for, at the Concept Stage, **Article 67, Principle of Concept of Operations (ConOps), Volume 2**, things are still far too fluid to start to dictate issues of detail. However, the major parts of the safety plan are formed as the project gets better defined. Safety issues can have a great impact on engineering design.

It is recommended that once the Preliminary Safety Plan is ready, it should be submitted to the safety authority that would have been set up by this time.

The next major activity is to create a Hazard Log. This log records details of expected hazards on the basis of relevant past experience for a newly started development and as they arise in an ongoing one. Not only is this a source design data but it is also an important part of the necessary burden of proof that will invariably be needed to defend any legal actions that may result.

With this log in place, it is possible to conduct various forms of safety analysis – *see* **Article 100, Safety Analysis Methods, Volume 2** and **Article 101, Assessing and Demonstrating Safety, Volume 2**. This stage is often called the *Hazard Identification and Analysis* (HIA) stage; it permits the all-important Risk Analysis to be done.

Risk analysis allows Safety Requirements to be prepared as part of the overall system requirements engineering task.

The required full Safety Plan can then be prepared and authorized to be placed alongside the System Engineering Plan, and the Test and Evaluation Plan.

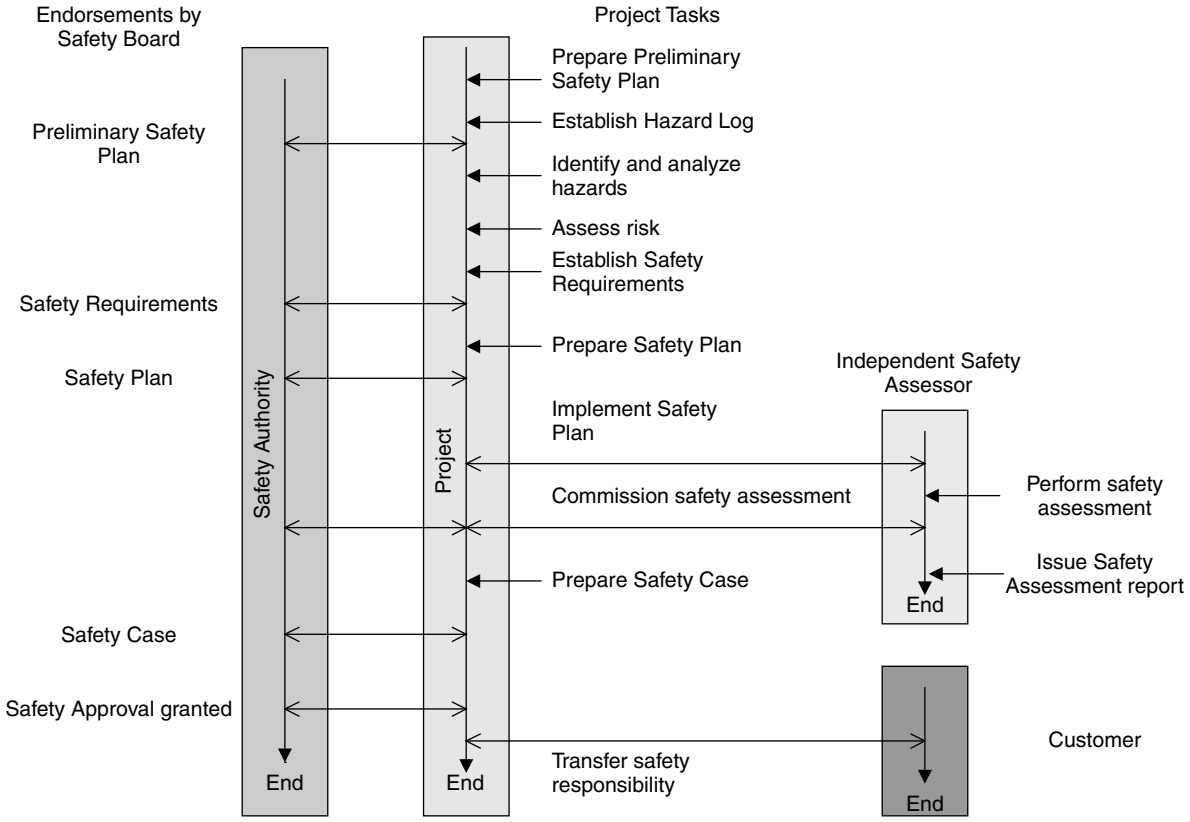


Figure 1. Suggested framework of safety activities. (Reproduced from Railtrack (2000) *Yellow Book – 3*, Railtrack PLC, Euston Square, London, distributed by Praxis Critical Systems Ltd, 20 Manvers Street, Bath BA1 1PX, UK, pp. 1–4, Fig 1-1 of *Yellow Book*.)

Once all this has been done, it provides the substance needed for a Safety Case to be prepared, which, after receiving endorsement, allows Safety Approval to be granted.

This may seem to be a large overhead on a project. Its extent is tailored to suit the type and size of project. The driving force for doing it well – and beyond just professional responsibility – is the fear of legal repercussions. A catastrophic failure can lead to very large legal settlements and perhaps to imprisonment. Even a low level breach of safety could lead to product recall for a simple modification, a process that has the potential to bankrupt a company due to the cost of the recall and fixes, without the cost of legal action.

6 MANAGEMENT OF SAFETY IN A CHANGE ENVIRONMENT

Safety is not an issue to be addressed at the start of a development and then ignored.

Each time a design change is made, its impact on safety needs to be investigated to see if the level of safety reached has been compromised. This is not always easy to establish.

Each engineering change has to be assessed via the change management process adopted for the development. That adding safety features will change the level of safety is obvious; but so also can removing them.

A Change Management Committee, or the like, is needed. They review proposals for change that is put to them. No change is permitted without correct approval.

All changes are accounted for within the configuration management process; each has to have its impact on safety assessed. This process needs to record

- identification of versions,
- history and status of versions,
- parts forming items,
- relationships between items.

Figure 2 is a change proposal management process suggested in Sage and Rouse (1999). The Engineering Change Proposal (ECP) is created and subsequently put to the Change Control Board (CCB) for approval before it can be put to use.

Software tools are available to assist the task of configuration management, but its use will need safety approval in safety critical systems developments.

Lack of sufficient safety is an insidious parameter for it is usually not one event in isolation that leads to failure but a complex chain of repercussions that lead, often unexpectedly, to failure. Reasons become blatantly obvious only after the event!

Somewhere, in one or more places in a chain, unpredicted events arise. Once they have been seen – usually too late at the first occurrence – they can be avoided. Not knowing about them leaves the situation waiting for that particular

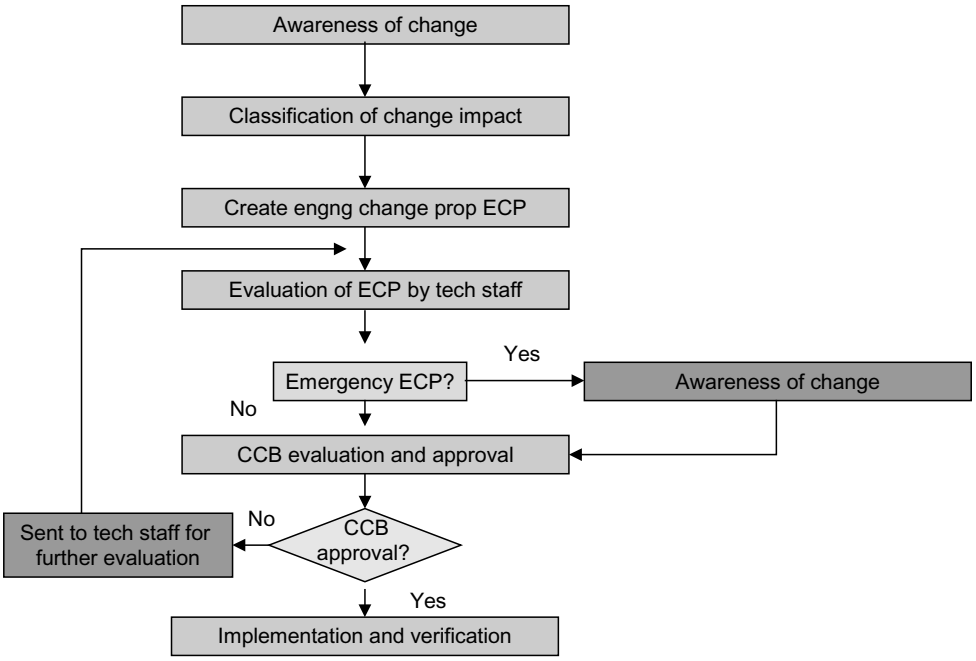


Figure 2. Change proposal management process. (Reproduced from *Handbook of Systems Engineering and Management*, Sage and Rouse, Copyright © (1999), Reprinted with permission of John Wiley & Sons, Inc.)

set of circumstances to arise. Because something has never happened before, does not demonstrate it never will.

The software component of safety critical systems is particularly risky and has to have far more effort expended on its development than do run-of-the-mill applications. In these cases, even the tools used to check the software have to be understood to the point at which their standard of operation will not compromise safety levels by, for instance, not finding potential faults. For this reason, safety critical system developers will often build their own safety examination tools.

The safest system could well be the one that is not used! It is not possible to build a useful system that is perfectly safe. The cost to provide it would not be practical. Such a system would be so overburdened with safety features as to be not workable. Safety then is a gradually improved facet of a systems development.

Swapping of safety cases and data is recommended as that increases the pool of event situations; but organizations are not always willing to do this because of legal and media ramifications of things they may wish to keep low key.

7 COMMUNICATING AND COORDINATING SAFETY ISSUES

Given that a project development often involves many people who would normally never meet for discussion of common problems, it is necessary to ensure that issues that arise, no matter what the source, are communicated to whoever needs to know.

This can be a difficult issue to address, as people are often loath to be critical of the work of others, or to offer too much about aspects of their own work that could be construed as a weakness on their part. Good, effective communication is essential between all involved.

Usually, the project will operate under many mandatory issues. If so, one will assuredly be that all persons have the obligation to report safety matters to a central place for action. All parties involved, including contractors, are subject to this ruling. An example is found in the UK Engineering Council 'Code of Professional Practice on Risk Issues'. In essence, it states the following:

- Communicate up and down the chain of responsibility.
- Seek to raise awareness of potential hazards and risks with colleagues.
- Make clear limitations and implications of risks.
- Discuss near misses and incidents with peers and take action on findings.

Commence communication verbally and look for a sound, simple solution. Many issues would have been resolved by this stage. Use all forms of communication, especially those with formal records once the issue is seen to need more central action. Formal records – the auditable paper trail – tend to be heeded more than discussion does.

It is to be expected that organizations will naturally set up liaison groups between their respective safety boards.

As safety issues are risks, the use of the risk control register should be the database and control mechanism for safety.

Formal committee type units usually have associated with them long delays in getting issues resolved. As with any committee of importance, there needs to be an executive group who can act rapidly when safety issues need urgent attention.

RELATED ARTICLES

Article 63, Outline of Systems Thinking, Volume 2; **Article 98, Reliability and Maintainability, Volume 2**; **Article 100, Safety Analysis Methods, Volume 2**; **Article 101, Assessing and Demonstrating Safety, Volume 2**; **Article 102, Introduction to the Legal Process, Volume 2**.

REFERENCES

Brauer, R.L. (1994) *Safety and Health for Engineers*, John Wiley, New York.

Hammer, W. (1993) *Product Safety Management and Engineering*, American Society of Safety Engineers.

McPartland, J.F. and McPartland, B.J. (2002) *McGraw-Hill's National Electrical Code® Handbook*, McGraw-Hill/TAB Electronics,.

Mugridge, C. (2002) *Notes on Safety for a Short Course*, Private Communication, Witchwood Engineering, UK.

Railtrack (2000) *Yellow Book – 3*, Railtrack PLC, Euston Square, London, distributed by Praxis Critical Systems Ltd, 20 Manvers Street, Bath BA1 1PX, UK.

Roland, H.E. and Moriarty, B. (1990) *System Safety Engineering and Management*, Wiley-Interscience, New York.

Roughton, J.E. and Mercurio, M.J. (2002) *Developing an Effective Safety Culture: A Leadership Approach*, Butterworth-Heinemann.

Sage, A.G. and William B.R. (eds) (1999) *Handbook of Systems Engineering and Management*, Wiley, New York.

FURTHER READING

Safety US (2003) http://www.quality.org/config/cm-guide.html Safety Groups in US.

100: Safety Analysis Methods

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Some Terms and Definitions Commonly Used in Safety Design and Evaluation 651
2 Standards and Other Information on Safety 651
3 Overview of Safety Assessment Methods 652
4 Work Safety Assessment (WSA) 653
5 Failure Mode and Effects Analysis (FMEA) 653
6 Failure Mode and Effects and Criticality Analysis (FMECA) 654
7 Fault Tree Analysis (FTA) 655
8 Hazard and Operability Analysis (HAZOP) 655
Related Articles 657
References 657

1 SOME TERMS AND DEFINITIONS COMMONLY USED IN SAFETY DESIGN AND EVALUATION

Many terms will be encountered in the safety aspects of designing and operating a system. No single definitive set is used; several exist. Where a safety standard is used for a project, the terms used there should be adhered to. Table 1 provides a composite set for use where they are not defined elsewhere for a project.

2 STANDARDS AND OTHER INFORMATION ON SAFETY

It is essential that terms, methods, and reporting be done consistently throughout a project. Consistency can be established by the use of a suitable safety standard that is appropriate for the project.

Some key safety standards are

- DEFSTAN 00-56
- IEC 61508
- CENELEC EN50126

All of the above are useful, but can be a little difficult to come to grips with. For overall guidance, each has its own kind of explanations. The defense safety standards are written in a directive tone with numerous updates and revisions to be taken into account as they are used; these are not always easy to read and maybe difficult to appreciate.

Several texts on safety are available – see Brauer (1994), Railtrack (2000), Roland and Moriarty (1990) and Roughton and Mercurio (2002).

Information given in texts should not be applied to the letter where definite standards are clearly defined to be used in a development. The texts should be viewed as useful reference works and are essential for building up one's personal safety attitude and design experience. The difficulty is that whereas they might maintain adherence to one standard, they, more often than not, pick up terms and ideas from several sources and can also use the author's own terminology from time to time.

One useful, Web accessed, general introduction and guide that was compiled recently for use by a very wide range of civilian users – as are found in a large railway system, is Railtrack (2003) – known as the '*Yellow Book*'. Its features are as follows:

- Prepared only as a guide to concepts and methods.
- Developed for use within the breadth of activities of the railway sector.
- Principles are applicable to all sectors.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

Table 1. Composite set of terms and definitions used in safety.

Term or acronym	Meaning
Accident	Unintended event that results in harm
Accident trigger	Event that causes a hazard leading to an accident
ALARP	As Low As Reasonably Practicable
BSM	Board of Safety Management
Causal factor	State of system that is systematic, for which a cause is clearly defined – once established
CCB	Change Control Board
CENELEC EN50126	Electrical Industry Standard on safety
CMC	Change Management Committee
DEFSTAN 00–56	UK Ministry of Defence. Standard on safety
ECM	Engineering Change Proposal
ESM	Engineering Safety Management
ETA	Event Tree Analysis
Failure	Deviation from intended performance, usually in a catastrophic manner
FTA	Fault Tree Analysis
FMEA	Failure Mode and Effects Analysis
FMECA	Failure Mode and Effects and Criticality Analysis
GAMAB	Globalement Au Moins Aussi Bon
H&SW	Health and Safety at Work Act of UK
Hazard log	Record of hazards and potential accidents identified from analysis of safety
HAZOP	Hazard and Operability Analysis
HIA	Hazard Identification and Analysis
HS&E	Health, Safety, and Environment
IEC 61508	International Electrotechnical Commission safety standard
Incident	Event that could have safety significance
Individual risk	Risk experienced by the individual. Usually expressed as probability of fatality per unit time
Intolerable risk	Risk that is totally unacceptable
MEM	Minimum Endogenous Mortality
MIL-STD-882C/D	US Defense Department Standard on safety
PSP	Preliminary Safety Plan
Risk analysis	Analysis of system risk, by various methods
Safety	Freedom from unacceptable risk
Safety Approval	Process through which the Safety Board grants approval to make changes
Safety Audit	Activity to verify that project is following the Safety Plan and maybe the veracity of that plan
Safety Case	Formal presentation aiming to provide assurance system will be safe and fit for purpose
Safety certificate	Formal statement authorizing the system for use
Safety integrity level	SIL. Discrete level, usually from 5, for specifying safety integrity zero being the lowest level where safety is not relied upon
Safety Plan	Documented plan that details how the system will be made sufficiently safe
Safety Requirements	Requirements needed to make a system safe under the definitions that apply
TLV	Threshold Level Value
Value of Preventing a Fatality	VPF. Defined monetary figure used to indicate reasonable expenditure for preventing a single fatality
WSA	Work Safety Assessment

- Developed by a committee of Railtrack® employees and supported by the committee as versions with updates provided on the Web site, Railtrack (2003).

It is important to recognize that it is all too easy to spread unsafe practices by sloppy use of terms due to the semantic nature of the necessarily subjective statements and varied use of terms.

3 OVERVIEW OF SAFETY ASSESSMENT METHODS

Failure of a system can cause death, injury, and property damage to a product or system. Examples are loss to an individual and the more geared-up situations where, say, a ship and its passengers are affected. It can also cause destruction of facilities and the environment.

Safety standards are essential as they contribute to safety management by providing generic means to design safe systems: they do not specify design solutions.

Methods deployed must be acceptable to all involved and cannot be imposed without user ownership across the system.

To prevent or minimize impact of human error, make use of key mechanisms like

- adoption of fail safe, or not fail at all, operations, activities, and events;
- redundant system units;
- fault-tolerant (error tolerant) design of hardware and computing;
- preventing unsafe actions by the way the system is designed and operated;
- providing assistance with making key decisions on a safe basis;
- using comprehensive integration and testing procedures.

Some key characteristics of failures are

- failure of individual small components and subassemblies are relatively easily controlled with low risk – that is, high confidence, by allocation of more resource;
- many disastrous failures have been the result of a series of apparently unimportant events that line up to create a pathway through the whole system, which adds up to major failure – like it did in the Challenger disaster and some Mars explorer missions;
- Reason's 'Swiss cheese' model points out that the 'holes in the cheese need to line up for both good and bad actions to result'.

Sources of safety requirements can be from a Regulatory Authority or be derived from assessment and analysis.

Types of safety requirement are many such as

- failure probability target – random or systematic;
- hazard removal requirement;
- functional safety requirement;
- process requirement.

Safety requirements must be

- auditable
- testable
- reasonable
- unambiguous
- uniquely identified.

Safety can be regarded as fitting into two groups of interest – safe technical behavior and safe human operation.

Several well-developed and documented methods exist for establishing safety risk. All are similar in their methodology. Whichever method is adopted, it must be used under advice to prevent shortcomings arising from differences in use of terms, process steps, and documentation.

There is a particularly strong need for a tightly specified, hardware and software codesign: the simplest of changes to software can negate prior safety analyses so regression for safety assessment can be very expensive to carry out.

Several methods of safety analysis are now introduced to set the scene. For application, these short overviews will need expansion using the appropriate standards in force for a project and the texts referred to above.

4 WORK SAFETY ASSESSMENT (WSA)

Work Safety Analysis (WSA) is a simple means for systematically identifying hazards and potential corrective measures, Sage and Rouse (1999). It is used mainly for physical operational work by people, but its relatively basic methodology has a place in safety assessments where deep detail is not required.

This is a hierarchical task analysis in which the hazards for each step are identified; once these are established, their likely causal effects are identified.

It is predominantly based on the judgment of an expert(s), plus considerable relevant data from past studies. It is often found that a lay person, one unfamiliar with the system, can often 'see' additional hazards that experts do not!

Likely issues are best handled using a checklist to ensure things already known from prior histories are not overlooked. Checklists for use by safety assessors, auditors, and designers are provided in Railtrack (2000, 2003).

Once established, each hazard is classified in terms of its relative likelihood of occurrence on a 0 (hazard eliminated) to 5 (very serious) scale and its severity of consequence.

Relative risk is then calculated as the product of probability and consequence.

Once the risk table has been reduced by calculation it will become clear where corrective measures can be taken to reduce the salient hazards to acceptable levels.

WSA is extended in different ways, as follows.

5 FAILURE MODE AND EFFECTS ANALYSIS (FMEA)

Failure Mode and Effects Analysis (FMEA) is another commonly used method. It has the same basic steps as WSA but is made more objective by use of the methods of reliability analysis. It too is used to establish potential systems failures that are then analyzed to establish the safety regime and make improvements.

This method was developed from hardware reliability engineering and thus there exist many similarities between safety and reliability analysis; this is to be expected for both are for investigating the risk of failure.

It tabulates the percentages of failures by the mode of failure and then categorizes these as critical or noncritical. It leads to *event trees*.

As with most safety assessment methods, it is most effective if used early in a development to flush out likely failures at the time when they can be avoided at lesser cost by design or operational correction.

An FMEA study is also of value to setting up the logistics of through life maintenance as it determines what needs to be, or not to be, done when servicing the system.

Carrying out an FMEA study can be done on either hardware, or a functional basis. The hardware approach is only practical once sufficient design drawings and analyses have become available – that is, in the later development life cycle stages. It can be carried out with various levels of system subassembly.

In the functional approach, the system is studied using the system functional decomposition breakdown levels. It starts studies of likely failure at the top level, working down to the level of detail needed. It is used to precede a hardware study.

It is suggested, extending the list given in Sage and Rouse (1990), that an FMEA be carried out with the following steps:

- Use the functional or hardware approach according to the stages of development available.
- Define the blocks using a clear numbering system such as that commonly used in functional flow diagrams (FFD) used in system functional decomposition.
- Locate the system requirements documents.
- Define and record failure modes and mechanisms using a checklist that defines them along with the relevant environmental parameters to be applied.
- Work on all potential failures found suggesting corrective actions that avoid them or sufficiently attenuate them
- List any special conditions of use that are essential for safe operation.

Traceability is essential for all of the above process documentation for any open link may leave issues unaddressed or see rekeyed starts using different semantic appreciations.

6 FAILURE MODE AND EFFECTS AND CRITICALITY ANALYSIS (FMECA)

Failure Mode, Effects, and Criticality Analysis (FMECA) is a similar process to the FMEA in that it is for identifying and investigating the safety and failure spots of potential systems designs.

The impact of unsafe situations will vary from low to extreme. FMECA extends FMEA to take in the criticality factors.

It is mainly applicable in the concept and preliminary design stages of the development life cycle for preventing unsafe events but it can also be used to study the safety of existing designs.

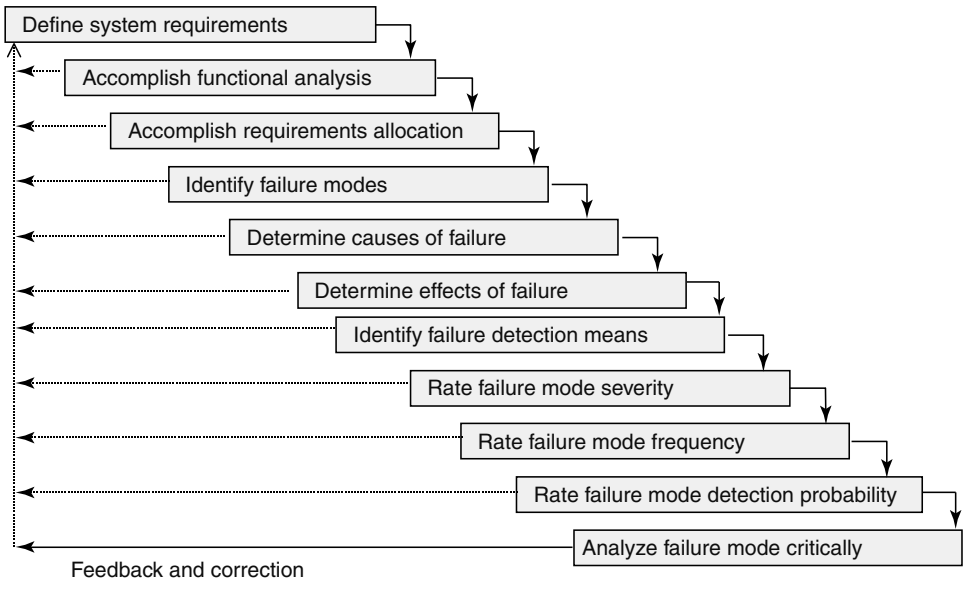


Figure 1. Steps of a typical FMECA analysis. Reproduced by permission of Pearson.

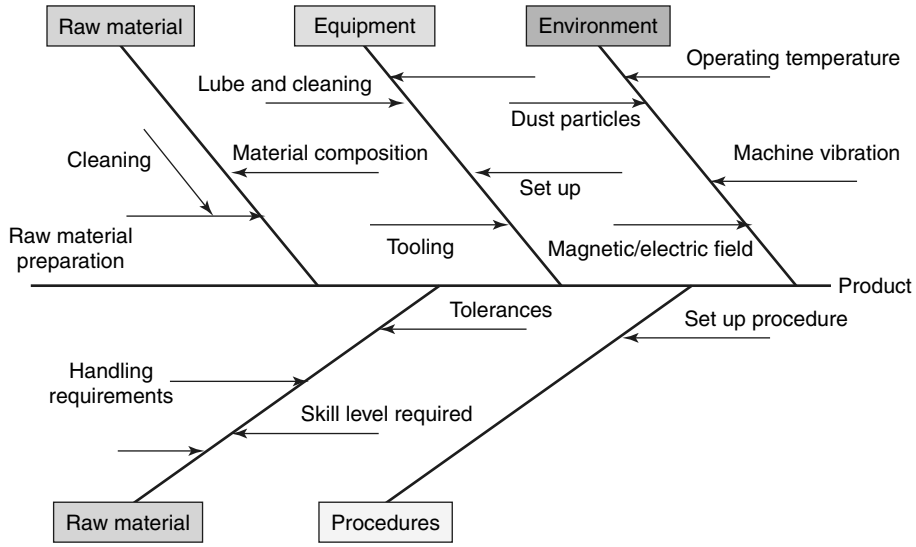


Figure 2. Fish bone diagram example: identifying faults in a manufacturing process. Reproduced by permission of Pearson.

The steps of a typical FMECA study are shown in Figure 1. More detail of each step is available applied to an example in Blanchard and Fabryky (1998).

An essential difference between FMECA and FMEA studies is that the former allocates tighter understanding of the failure data using statistical methods. This allows a more discerning study of the criticalness to be produced where many parameters are involved.

7 FAULT TREE ANALYSIS (FTA)

Fault Tree Analysis (FTA) studies failure, and thus safety, using deductive use of a graphical enumeration and analyses of the different ways in which a fault can arise. Its methodology can be applied in the above methods to tease out potential safety problems. It is a method for charting the hazards in a hierarchical manner.

Its process is similar to the widely applied decision tree method in that it is a top down procedure that identifies how issues can be broken down. Here it is used to locate causal reasons for failure.

As causes of potential failure are determined they branch out with probabilities of each arising being allocated at each branch point. Once the full tree has been identified, each of the many paths is calculated back to the start to identify each risk, thus revealing the most unsafe sets of circumstances. High-risk paths are rectified by redesign to obtain a more suitable risk distribution across the whole project.

In analysis of the safety-regime application of the decision tree concept, use is often made of the Ishikawa cause and effect or the fish bone diagram. The one used in the Blanchard and Fabryky (1998) example is given in Figure 2.

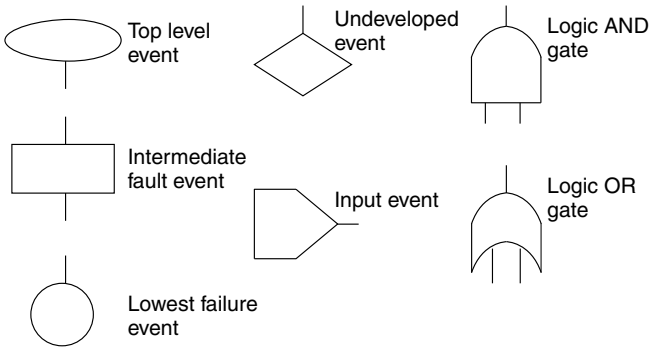


Figure 3. Some symbols used in fault tree construction.

Fault trees are constructed using commonly agreed standard symbols – see Figure 3 for the basic ones used. A more complete list, with explanations of each, is available on http://www.relexsoftware.com/reliability/brief/fta_brief2.asp.

An example Fault tree for events that can lead to failure in an elevator, drawn with the Reflex Fault Tree™ tool, is given in Figure 4.

8 HAZARD AND OPERABILITY ANALYSIS (HAZOP)

Hazard and Operability Analysis (HAZOP) was developed in the 1960s for the UK chemical industry. It was quoted as being a

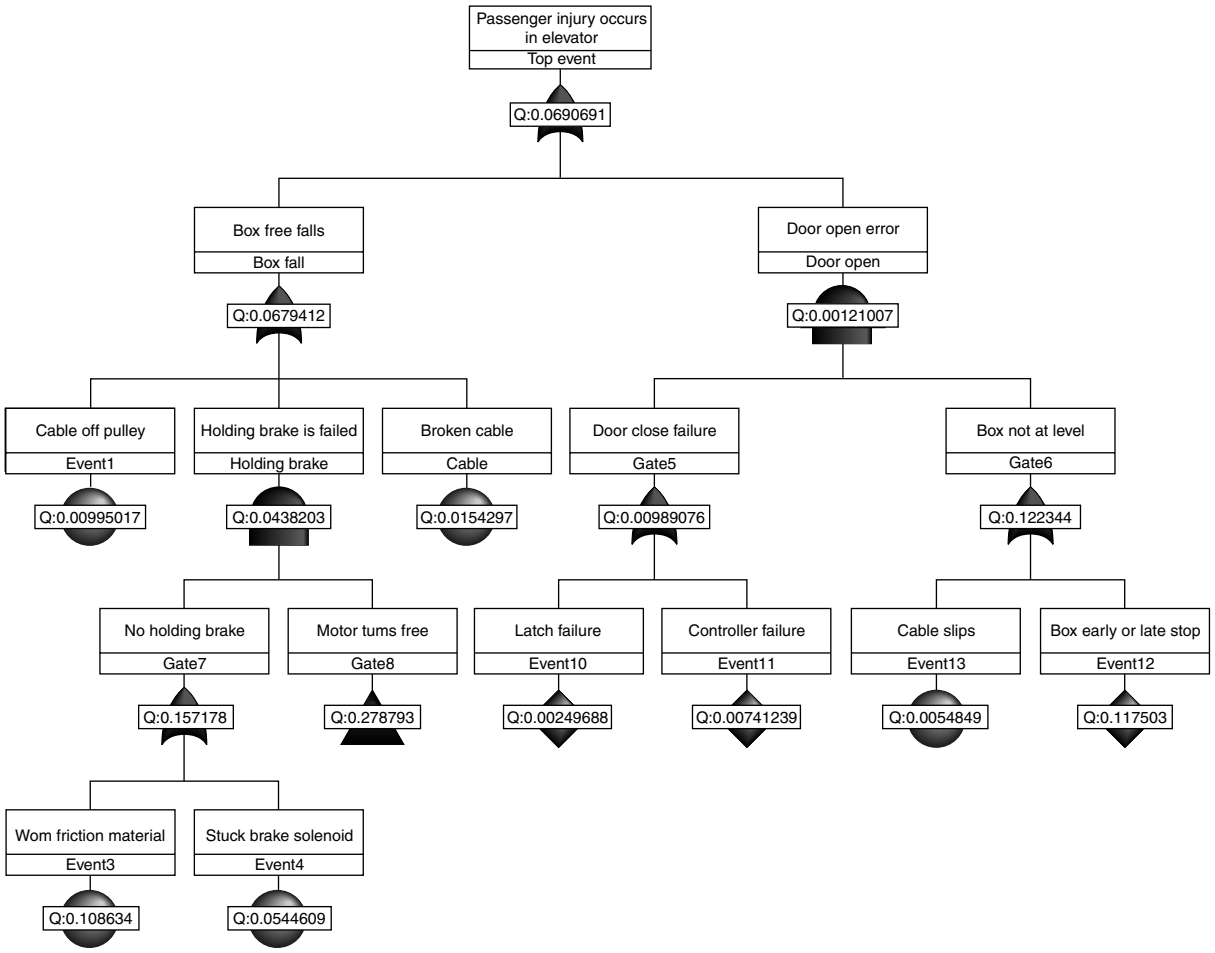


Figure 4. Example fault tree created with the Reflex Fault Tree™ tool.

> 'Qualitative technique whose purpose is to identify all deviations from the design's expected operation and all hazards associated with these deviations'

It is a systematic, creative examination of a design that is best performed by a multidisciplinary team who can bring a wide variety of viewpoints to the study.

It has been widely adapted for use in other sectors.

In use, each system component is inspected in turn keeping in mind the design intention – what is this component supposed to do?

Standardized words are used to minimize the semantic nature of normal language in this use. A selection from the thesaurus is as follows:

- NO/NOT: No part of the intention is achieved.
- MORE: Some quantitative increase over what was intended.
- LESS: Some quantitative decrease compared to what was intended.
- AS WELL AS: Some qualitative increase over what was intended.
- PART OF: Some qualitative decrease compared to what was intended.
- REVERSE: Logical opposite of intention.
- OTHER THAN: Something completely different.
- EARLY: Something happens before it is expected, relative to clock time.
- LATE: Something happens after it is expected, relative to clock time.
- BEFORE: Something happens before it is expected, relating to order or sequence.
- AFTER: Something happens before it is expected, relating to order or sequence.

The first step, identifying the system elements, for a simple example HAZOP study of a two component fuel mixer (Mugridge, 2003) is given in Figure 5.

From each such study is derived plausible (that is, may well be true but not necessarily so until investigated) deviations from the design intention. Avoid prejudgment of any issues before they are analyzed; include those that

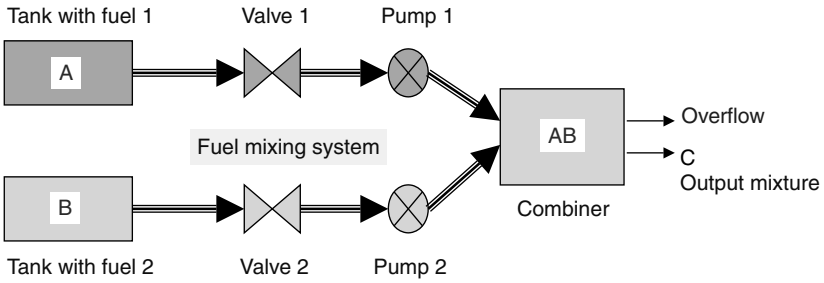


Figure 5. First step in an HAZOP study – the system block diagram.

Guide Word	Deviation	Cause	Effect
NO	No flow of A	Pump failure	Explosion
MORE	Excess flow	Valve failure	Poor quality
LESS	Reduced flow	Pump/valve failure	Explosion
AS WELL	Contamination with air	Line rupture	Fire
REVERSE	Pump reverse flow	Maintenance error	Explosion
OTHER THAN	A solidifies	Low temp	Explosion

Figure 6. HAZOP table for fuel mixer system.

seem wildly unlikely. Work on the basis that all elements are unsafe until proven to the contrary.

Figure 6 shows part of an analysis table formed from study of the system given in Figure 5.

It can now be seen that while each method has the same basic features, they variously incorporate features of their own. Each of the above analyses methods finds use according to the Standard requirements in force and the experience of the analyst. It is essential to use only one methodology for all related parts of a system; mixing methods can lead to confusion, the enemy of reliable safety studies. Safety is built up over years and can be destroyed in moments by simple design or operational changes.

RELATED ARTICLES

Article 63, Outline of Systems Thinking, Volume 2; **Article 64, Executing A Measuring System Design, Volume 2**; **Article 99, Safety Organization, Volume 2**; **Article 101, Assessing and Demonstrating Safety, Volume 2**.

REFERENCES

Blanchard, B.S. and Fabryky, W.J (1998) *Systems Engineering and Analysis*, Prentice Hall, Englewood Cliffs, NJ.

Brauer, R.L. (1994) *Safety and Health for Engineers*, John Wiley, New York.

Mugridge, C. (2003) *Teaching Notes for Safety Course*, Witchwood Engineering, UK.

Railtrack (2000) *Yellow Book – 3*, Railtrack PLC, Euston Square, London, distributed by Praxis Critical Systems Ltd, 20 Manvers Street, Bath BA1 1PX, UK.

Railtrack (2003) Engineering Safety Management Guide, http://www.yellowbook-rail.org.uk/site/the_yellow_book/the_yellow_book.html.

Roland, H.E. and Moriarty, B. (1990) *System Safety Engineering and Management*, Wiley-Interscience, New York.

Roughton, J.E. and Mercurio, M.J. (2002) *Developing an Effective Safety Culture: A Leadership Approach*, Butterworth-Heinemann.

Sage, A.G. and Rouse, W.B. (eds) (1999) *Handbook of Systems Engineering and Management*, Wiley, New York.

101: Assessing and Demonstrating Safety

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Safety as a Whole of Life Cycle Factor 658
2 The Seven Stage Process 658
3 Stage 1 – Hazard Identification 659
4 Stages 2 and 3 – Causal and Consequences Analyses 659
5 Stage 4 – Loss Analysis 661
6 Stage 5 – Options Analysis 661
7 Stage 6 – Impact Analysis 663
8 Stage 7 – Demonstration of ALARP 663
9 Safety Process Reviews 664
Related Articles 665
References 665

1 SAFETY AS A WHOLE OF LIFE CYCLE FACTOR

Safety must be treated as a whole of system life cycle issue. Each design activity, when adding its own contribution, must carry out safety assessments as well as is possible at the time.

Overarching the individual design contributions, to ensure that they integrate and do not degrade the overall system safety, will be the duty of the Safety Board with its safety policies and practices.

Applying a well executed, top-down, System Engineering, SE planning and project management, and working within a design change management process will help in reducing rework at a later stage.

The difficulty, however, is that safety is an ongoing maturing process. A system that has been 'safe' for years can have its integrity reduced by subtle changes in the chain of circumstances in application, by new requirements of regulators, and because of the occurrence of a previously unseen event that attracts attention. All of these can require extensive system redesign.

The detailing designer should work within existing directives aiming to second-guess failure mechanisms that are then mitigated by the use of appropriate design strategies.

2 THE SEVEN STAGE PROCESS

The 'Yellow Book' of Railtrack (2000) suggests that a seven stage process be used to demonstrate that the system risks are as low as reasonably practicable (ALARP). (While ALARP is used here, this basis of safety is not the only one in use – *see* **Article 99, Safety Organization, Volume 2**.) Notes of a short course on safety (Mugridge, 2003) are summarized here to introduce the methodology for assessing the risk level of a system.

The steps of the Seven Stage process are shown in Figure 1. The sequence is as follows:

Stage 1. Identify the hazards.
Stage 2. Carry out an analysis of the causes of each hazard.
Stage 3. Analyze the consequences of the hazards. Stages 2 and 3 can be done in either order.
Stage 4. Assess the loss in terms of harm to people.
Stage 5. Consider the options for risk reduction.
Stage 6. Assess the likely cost benefits of each hazard reduction measure.

Then, after systems adjustment to reduce the risks:

Stage 7. Demonstrate that the system risks are as low as reasonably practical.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

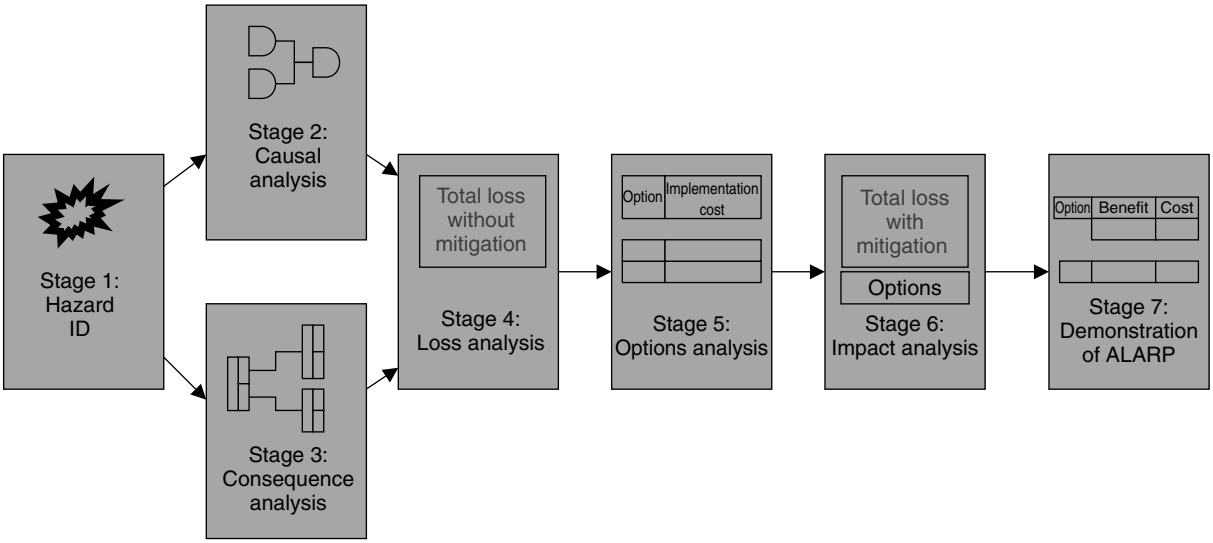


Figure 1. Seven stage process for demonstrating ALARP.

Carrying out this process needs speed, but with adequate care. If the process takes too long, the system may change significantly, making the analyses following the data used to perform the task no longer useful.

Each step is now covered in turn.

3 STAGE 1 – HAZARD IDENTIFICATION

The purpose of this stage is to identify any potential hazards associated with the system and the way in which it is to be used.

For best results, the task should not be guided by intuition to come up with likely hazards, but should be approached in a systematic and comprehensive (that is, systemic) manner. Ideas and inspiration can be found from a variety of sources such as

- expert knowledge,
- layperson experiences,
- previous experience for the same kind of circumstances, and from the prior history of the project (if recorded – all too often this has not been done!)
- new analysis of the current situation.

Some of the various techniques that can be deployed to make use of the above sources are as follows:

- Creative thinking by individuals including a wide cross section of users.
- Brainstorming that teases out ideas as members of the team spark ideas from each other's suggestions. They should make use of charts to record ideas trees and other connections between the different ideas.
- Hazard and operability (HAZOP) studies – *see* **Article 100, Safety Analysis Methods, Volume 2**.
- Empirical approaches, while not well ordered can reveal unexpected issues.
- Checklists from other projects, and the organization's safety documentation.
- Structured Walk-throughs of each likely, plausible situation.
- (FMEA) Failure Mode and Effects Analysis – *see* **Article 100, Safety Analysis Methods, Volume 2**.
- Task Analysis.

More details of these methods are available in Railtrack (2000, 2003). These methods take time to do and need care to identify issues. Bias and preconceived ideas of improvements should not be allowed to prevail as potential risks are uncovered. This is not the stage to decide whether their level of importance is such that they are not worth including. The chain effect must be remembered for even small, apparently innocent, events that might lead to unacceptable risk.

4 STAGES 2 AND 3 – CAUSAL AND CONSEQUENCES ANALYSES

The purpose of causal analysis is to determine credible combinations or sequences of causal factors that can lead to hazards.

It is carried out using a diagrammatic, functional, hierarchical representation as its main schema. A fault tree is generated from a study of drawings, component lists, and operational instructions – see Fault Tree Analysis

(FTA) and Failure Modes and Effects Analysis (FMEA) in **Article 100, Safety Analysis Methods, Volume 2**. The probability of a condition arising is best stated as the quantitative probability estimate.

FTA starts with an undesired top event being identified. Logic gates are used to form a logic network showing how the various fault-inducing conditions are related. An example is given in **Article 100, Safety Analysis Methods, Volume 2**.

Working downwards, one identifies the next immediate causes, eventually stopping when basic events are reached. Probabilities are then inserted for the basic events from an FMEA – *see* **Article 100, Safety Analysis Methods, Volume 2**.

FTA is usually carried out using a software package. The calculations can be done manually, but they are tedious. Tools must be used with great care in safety-critical situations, for a tool may have an inbuilt fault error of its own in the way it computes the calculations or in the assumptions it uses. Super safe systems usually need to have their own safety tools developed under the control of the safety process in order to keep control over all the factors of a safety assessment.

Figure 2 shows how an undesired top event (e.g. a train collides with another train) can cause a top event. In this case, an OR gate is appropriate – for one train OR the other can cause the collision.

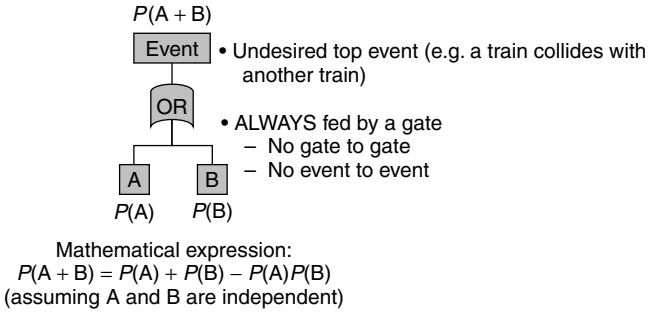


Figure 2. OR gate condition for causing an unwanted event.

Some rules of fault trees are that

- a top event is ALWAYS fed by a gate,
- no gate-to-gate states can occur,
- no event-to-event links can exist.

As well as either train causing the unwanted event, it is also feasible that both trains can cause the unwanted event together. In this case, one train AND the other is a condition to be investigated. Figure 3 shows this condition.

Stage 3 is a Consequence Analysis. The purpose here is to determine what are the intermediate conditions and the final consequences resulting from the occurrence of the identified hazards.

It is carried out using cause consequence diagrams (CCD) and Event Trees. An example of a CCD is given in Figure 4. A fire starts with ignition. If the fire protection system is operating as planned, the sprinkler systems operate and the fire is put out. If it does not operate, but the fire alarm sounds, then the severity of damage caused by the fire is

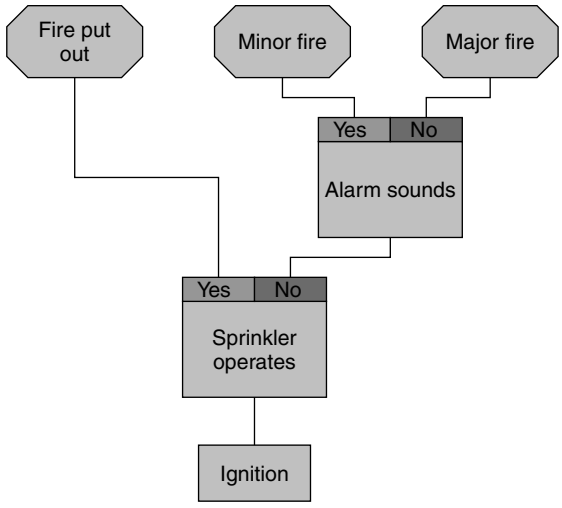


Figure 4. Simple example of a cause consequence diagram (CCD).

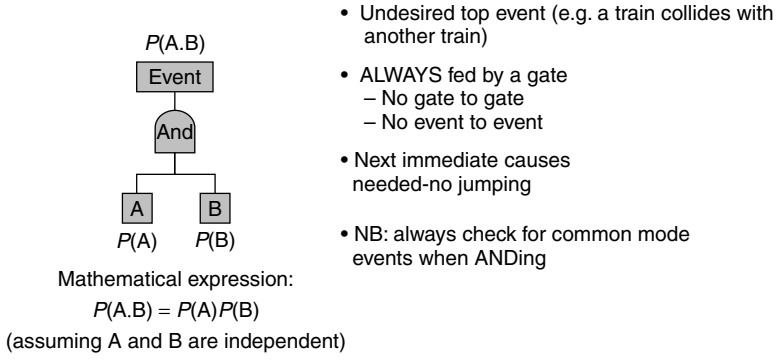


Figure 3. AND gate condition for causing an unwanted event.

lessened. Good design for least risk sets up situations that have less risk due to alternative protective states being in place.

System safety is aided by the existence of safety barriers. These are design situations that will assist in preventing the escalation of a safety risk – a barrier mechanism will have a barrier strength associated with it. Assessment of this value is needed in quantitative mathematical terms.

It is important to recognize that safety rectification temporarily put into place can lead to reduced safety because of its use. A study of such regression steps is needed as much as the main lines of expected risk.

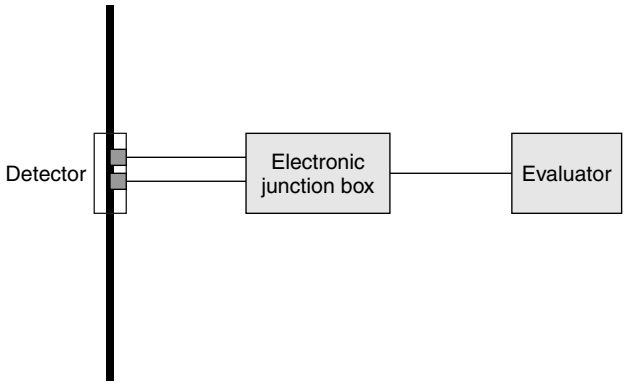


Figure 5. Schematic of a train axle detector.

5 STAGE 4 – LOSS ANALYSIS

The purpose of carrying out a Loss Analysis is to determine the loss associated with each consequence identified during Consequence Analysis. This study is done to see what the current situation is before rectification is implemented. It is not the loss of the improved system.

A loss here is harm to people, ranging from slight effects to fatalities. The principles used can, however, be adapted to take in other losses such as environmental and commercial but that is more for studies of risk assessment in general.

A Loss Analysis needs to estimate the extent of harm, damage, or disruption that could potentially arise from the identified consequences. Harm is initially expressed in qualitative terms for loss assessment and starts from guided estimates, rather than from exact measures.

For example, if a measuring system fails and causes its aircraft to crash, it is not possible to be exact about the harm to life that might arise. Guidelines commonly accepted are that 1 fatality = 10 minor injuries, and 1 major injury = 20 minor injuries. These subjective conversions allow the various levels of harm to be reduced to the common unit of fatalities per annum.

Once the subjective figures are decided, the next step is to convert them into objective values to ensure that any further work with them has minimal added subjectivity. This results in tabulation of the nature of the loss against the cost associated with each consequence.

The next step involved is arriving at the value of preventing a fatality (VPF), a value put on a human life. This varies considerably; in the rail sector, the VPF ranges over £1 to 3 million for a life, or its equivalent.

A higher figure is needed for multifatality accidents for the public is disproportionably averse to this! A weighting coefficient is used – called the *Factor of Gross Disproportion*.

Other sectors are often less open about this value. A road accident death is far less than the above, but any long-term harm to a surviving person can cost more depending on individual legal judgments.

The VPF is, therefore, project specific. Advice is needed as to what values to use; it will come from the legal fraternity. Note that these figures are used only for supporting ALARP decisions – and they are not a guide to legal settlements.

To illustrate some of these principles, consider the simple example of an axle detector set up on the side of a railway track, see Figure 5, to warn a railway gang working on the track of an approaching train.

The CCD for this element of a railway system is developed in Figure 6. The various situations are self-evident from inspection of the tree. The failure F for the base fault is decided along with the probabilities for dividing risk at each branching point. The latter values are used to decide the failure values at the end of each branch of the tree.

With the CCD now set up, it is possible to carry out the calculations needed to convert the data for the various severities of collision into monetary values. This is done in Figure 7 where the total cost is shown for major, minor, and zero collisions.

Those values can then be used – Figure 8 – to calculate the total annualized cost of the systems as it is before any improvements are carried out.

More details for conducting a Loss Analysis are available in Railtrack (2000, 2003).

6 STAGE 5 – OPTIONS ANALYSIS

The purpose of the Options Analysis is to ensure that all reasonable measures for reduction of losses have been considered. There are two ways to consider this – reduce the rate of occurrence of the hazard, or limit its consequences.

For each system hazard, it is necessary to identify physically possible and known mitigation strategies and then to calculate the cost of each option. Only direct costs are

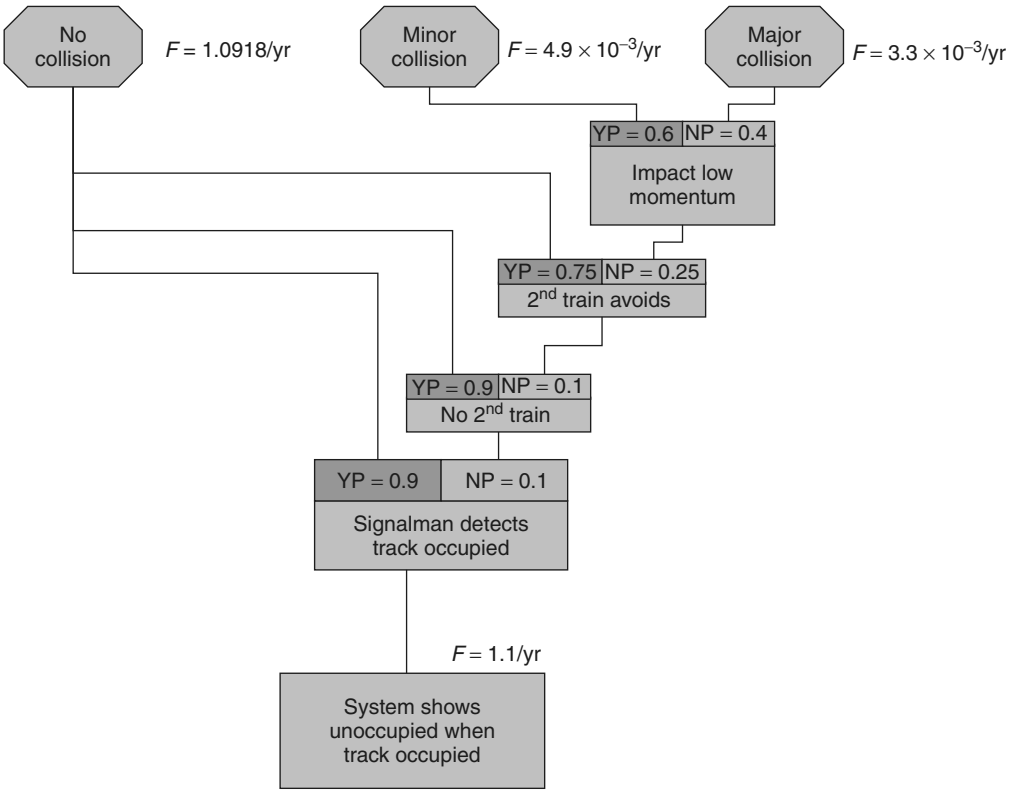


Figure 6. Cause consequence diagram (CCD) for axle detector system.

Consequence	Losses	Cost £m
Major collision	30 fatalities 50 major injuries 200 minor injuries = 36 equivalent fatalities	115.9
Minor collision	2 major injuries 6 minor injuries = 0.23 equivalent fatalities	0.3
No collision (emergency application of brakes)	2 minor injuries = 0.01 equivalent fatalities	0.01

Figure 7. Cost of the consequences of collision for the axle detector system.

included here – indirect benefits arising due to mitigation are costed into the safety assessment at a later stage.

It is done as a risk reduction exercise using the same kind of discovery methods as are listed for Stage 1. Indeed, some ideas may well already be on record as earlier stages were undertaken.

Consider an Options Analysis for the train axle detector given earlier.

Suppose two risk reduction options have been identified:

1. Rework the evaluator component's software

   - Total cost £2 million
   - Expected life 20 years

   – Annualized cost: £0.1 million

Consequence	Calculation	Cost £m
Major collision	115.9×0.0033	0.38
Minor collision	0.3×0.0049	0.01
No collision	0.01×1.0918	0.01
Total annualized loss		0.40

Figure 8. Annualized loss of the consequences of collision for the axle detector system.

2. Complete system redesign

   - Total cost £5 million
   - Expected life 20 years

   – Annualized cost: £0.5 million

7 STAGE 6 – IMPACT ANALYSIS

The purpose of an Impact Analysis is to assess the effects of implementing each of the risk reduction options. Its approach is to estimate the net cost or benefit of each option.

For each option (separately), revisit the previous stages of the analysis, inserting revised figures. The safety value of a change is calculated from the differences revealed. In this revision process, carry out each modification on the same qualitative or quantitative basis that was used originally; to change this, will not allow the baseline to be properly compared with the revision state.

Considering the two options identified for the axle counter and placing in some realistic data leads us to the following:

1. Rework evaluator software

   - Reduces failure rate by a factor of 3
   - Reduces annualized loss from £0.4 million to £0.13 million

   – Reduction in risk = £0.27 million per annum

2. Complete system redesign

   - Reduces failure rate by a factor of 10
   - Reduces annualized loss from £0.4 million to £0.04 million

   – Reduction in risk = £0.36 million per annum

Impact analysis is continued as iterative action loops until all variables under control have settled to give the safest situation for allowable factors.

From these outcomes, selection decisions need to be made using background factors existing at the time – political, financial, public pressure, and so on.

8 STAGE 7 – DEMONSTRATION OF ALARP

ALARP is a well-documented legal principle used in some health and safety activities – to see more on this topic, search the Internet under ALARP. It is used here to show the last stage of a system safety analysis. It is not, however, the only one required to be used, so always checkout the appropriate safety standard relevant to the project – *see* **Article 100, Safety Analysis Methods, Volume 2**.

Two things must be shown in order to justify that the acceptance of identified risks are ALARP:

- The overall risk, and each individual risk is below the upper limit of the tolerability that has been decided for the project and its circumstances.
- All practical and cost-effective risk reduction measures have been implemented. This is largely a scientific and engineering exercise but interpretations of what constituted such parameters as 'all practical and so on' need to be done against a background of appropriate legal advice.

Figure 9 shows the zones of the ALARP principle.

Figure 10 shows how to tabulate the cost/benefits resulting from the Loss and Impact Analyses.

Setting limits of acceptability are usually within the prerogative of the appropriate Board of Safety, or are dictated by national or international law. The task of safety analysis is to produce a Safety Case that demonstrates adherence being as low as it is reasonably practical – ALARP.

Bodies issuing the safety compliance certificate can inspect the report of the 7 Stage Safety Assessment process to ascertain that a sound and thorough study has been made.

Legal litigation is heavily based on the principle of reasonableness. This means that best practice needs to be reached in safety design because the knowledge of what can be done and how it can be achieved is available, making it reasonable to expect it to be used. This basis can, however, be slow to encourage new technology to be used, so some legal systems now include judgments that need a designer to apply current technology that would make the difference as long as that technology is reasonably available; it is not

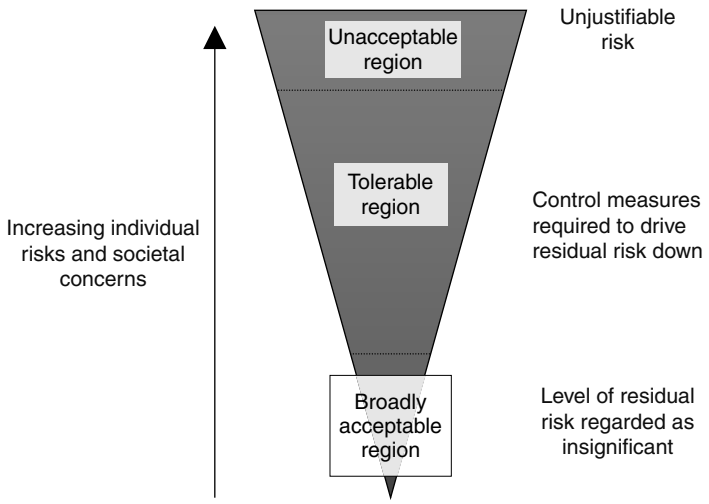


Figure 9. Zones of the ALARP principle.

	Cost/benefit calculations	Loss per year (£m)	Annual benefit over current design	Annual cost of modification (£m)	Option required?
Loss analysis	Current design	0.4	n/a	n/a	n/a
Impact analysis	Option 1 (rework evaluator s/w)	0.13	0.27	0.1	Yes
	Option 2 (complete redesign)	0.04	0.36	0.5	No

Figure 10. Table demonstrating the costs and benefits of a change to the system.

possible under those rules to retain old systems that could be made safer.

9 SAFETY PROCESS REVIEWS

It is necessary to give time to internally review the safety process put into place.

Ask these questions of that process:

- Is the hazard identification up to date?
- Have all potential accidents been assessed?
- Have all relevant probabilities been checked?
- Is probability assessment conservative?
- Has risk tolerability been assessed?
- Have the risks reduced, the system still being ALARP?

Beware of

- common cause failures,
- excessive optimism regarding systematic failures,
- the use of unstated assumptions and unexpected events – we cannot totally, definitively, predict all likely failure mechanisms! Their discovery is an ongoing process.

An independent professional review is essential to make sure internal bias does creep into assessments. The purpose of reviews is to

- provide assurance that the work that has been carried out has been (or is being) done in an adequately safe manner (i.e. in accordance with the agreed Safety Plan);
- form a judgment as to whether the risk associated with the system has been (or will be) reduced to an acceptable level.

Reviews are carried out as a series of appropriately scaled and scheduled audits and assessments that review both process and product.

External audits and assessments are also needed. These are carried out on the safety process by suitable third party authorities.

Safety Audits focus on the Engineering Safety Management (ESM) processes. Are the processes adequate? Are they being adhered to adequately?

Safety Assessments focus on the project's product; they evaluate the risk. Is the system safe? Is it fit for the purpose? Does it meet its safety claims?

In practice, there is some overlap between these two kinds of activity.

Audits and assessments must be conducted in an independent manner. The level of independence will depend upon the system involved. Several systems that are under the care of an organization may warrant different levels of independence.

These are driven by the system's defined Safety Integrity Level (SIL). Other factors, however, that may need to be considered are

- degree of developer's previous experience with a similar design,
- degree of complexity,
- degree of novelty,
- novelty of design,
- novelty of technology.

The timing of audits will generally start some way through a development when it is late enough for something to be there to be audited, but early enough to allow improvement to the processes while error correction costs are low.

Extra audits are carried out after transfer of safety responsibilities to the user agency and during all of the operational life. The system may be very well maintained and be kept within original specification, but the working environment, and political and public scene factors may have changed to degrade the ALARP levels of understanding in the originally certified system.

Assessments are done upon completion of the detailed engineering design phase. If done earlier, sufficient details will not be available. Prior checks will be carried out to uncover any obvious safety defect, but there they are more a matter of requirements and configuration control.

They will also be made immediately prior to safety case submission to ensure that all is in order right up to the minute.

The audit/assessment process starts off with the formation of a suitable plan. This includes

- familiarization with the project,
- review of previous audit or assessment,
- requirements, activities, timescales,
- project dependencies,
- audit/assessment team,
- delivery of documentation.

The plan is then carried out and the results reported to the agency calling for the assessment or audit.

Useful checklists of parameters to address in assessments and audits are given in Railtrack (2000, 2003).

RELATED ARTICLES

Article 63, Outline of Systems Thinking, Volume 2; **Article 64, Executing A Measuring System Design, Volume 2**; **Article 99, Safety Organization, Volume 2**; **Article 100, Safety Analysis Methods, Volume 2**.

REFERENCES

Mugridge, C. (2003) *Teaching Notes for Safety Course*, Witchwood Engineering, UK.

Railtrack (2000) *Yellow Book – 3*, Railtrack PLC, Euston Square, London, distributed by Praxis Critical Systems Ltd, 20 Manvers Street, Bath BA1 1PX, UK.

Railtrack (2003) Engineering Safety Management Guide, http://www.yellowbook-rail.org.uk/site/the_yellow_book/the_yellow_book.html.

102: Introduction to the Legal Process

Christopher Sweet
Minter Ellison, Adelaide, South Australia, Australia

1 Introduction to Legal Issues 666
2 Definitions 666
3 Legal Systems in Force 667
4 Application of the Principles and Process of the Law 667
5 Difficulties in Application of the Law 668
6 The Litigation Process 669
7 Commencing Litigation 669
8 Evidence and Documentation 669
9 Reducing Liability 670
Further Reading 670

1 INTRODUCTION TO LEGAL ISSUES

The sorts of legal issues that arise in measurement systems can be many and varied. This is because there are different systems of law in place, and different legal issues can occur at different stages of the design, implementation, and application times of the life cycle of a measuring system. For example, there might be specific legislative requirements such as those imposed by quality standards that have to be met in the design and production of a particular component of a measurement tool. Then, during the operational phase in which the measuring system is used, there may be other standards such as environmental standards that the designer of the measuring system needs to have in mind in setting the parameters for the measurement system. The sorts of legal issues that arise will depend on the nature of the system or product being developed and on the laws of the countries in which the instrument or system is to be developed and used.

The instrument or system designer will need to be aware of the law that will impact on the design and test processes such as the preparation of contracts and patents and, then, the legal context in which the system or product is to be tested and subsequently used.

Designers and test engineers should carefully consider how the law might impact on each step in the design, development, and test process, whether it be the contractual arrangements, protection of intellectual property, or liability, and seek appropriate advice to reduce the risk of adverse legal consequences, whether these be losses due to inadequate contracts or liability due to a failure in the system. This article covers a broad introduction to some of the legal concepts and issues that might arise. A case study about legal liability is given in **Article 103, Legal Liability Issues for Designers – A Case Study, Volume 2**.

A brief outline of some of the basic concepts follows.

2 DEFINITIONS

In most countries, the main sources of law are legislation and case law.

2.1 Case law

This is law made by the courts. The process of courts making law through judicial officers (e.g. magistrates or judges) depends on a number of factors:

- Level of the decision-making court in the hierarchy of courts. Most countries have a hierarchy of courts, and it usually follows that the higher the court in the hierarchy,

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

the more binding or persuasive its decision will be on the courts below.
- Nature of the cases or disputes coming before the courts, that is, the fact situations that allow the courts to adjudicate and make law.
- Case law from other countries, which may influence courts or be considered binding on courts in another country.
- As cases come before the courts for adjudication, a body of law is developed by judges applying judicial decisions from other cases to new fact situations. Many court proceedings are recorded in official court reports and so become available to judges and the public for future reference.

2.2 Legislation

Legislation is law made by a parliament. Legislation sets out the law relating to its subject matter in one document. Amendments can then subsequently be made by further amending legislation.

Legislation is produced following a lengthy process of debate in parliament and public consultation. The primary document is termed an *Act of Parliament*, which is usually required to be passed through parliament.

Some acts are termed *Consolidating Acts* because they consolidate legislation on a particular subject. Acts that amend legislation are usually called *amending acts*.

Some Acts adopt international conventions into the law of a particular country. An example is the Warsaw Convention, which concerns aviation liability and which has been introduced into the laws of many countries through aviation liability legislation.

Acts also allow for regulations to be made under the Act, which give effect to certain administrative matters. An example would be regulations that contain published documentary and physical SI standards. Another might be a law to protect the environment, that is passed in an act of parliament, such as 'The Environmental Protection Act'. The specific standards for emission of certain industrial gases would then be set out in regulations made under that act.

Both case law and legislation comprise the law in most countries, and both will need to be taken into consideration by all designers and testing processes at various stages.

3 LEGAL SYSTEMS IN FORCE

Many countries have different systems of law, which govern the way the law regulates processes. For instance, many countries have a system of common law such as that in England. By common law, we refer to a system of case law developed by the courts in which a complex set of precedents are gradually developed by judges considering cases coming before the courts.

Most countries that were previously in the British Commonwealth have a system based on the common law. The common law is law made by judges when they adjudicate disputes between parties. The authority to make common law is derived from the authority exercised by English judges over many centuries. A few examples of common-law countries are Great Britain, Australia, many of the states of America, and Canada. In these countries, judges have retained a significant direct role in creating law rather than just interpreting law made by legislation.

Many European countries have a different system of 'civil law', which is based on systematic legislation extending over more areas of law rather than case law being made by the judiciary. The law is 'codified' in detailed legislation and interpreted or applied by the courts.

These are just some of the systems of law in force throughout the world. It is important to interpret a situation according to the particular system in place in each country in which a system is developed or used. This also means where the system is to be used rather than where it is designed and made.

4 APPLICATION OF THE PRINCIPLES AND PROCESS OF THE LAW

In practice, the law is applied in various ways.

During the initial contract phase, this can be protecting intellectual property, compliance with standards, or dealing with a liability claim.

In the case of complying with standards, which is one of the most common applications of the law, that is, via legislation, those responsible for compliance will normally familiarize themselves with the standards relevant to their system design and incorporate the standards into the design as part of the design process. This will involve researching or obtaining advice on the standards that apply to the particular design. Most final contracts will dictate the standards to use.

An initial nondisclosure agreement is usually signed by the would-be contractor and the customer. Another preparatory legal mechanism used at the first encounter stage is the use of a Memorandum of Understanding (MOU), other names used being 'a letter of agreement', 'heads of discussion', and more.

In the case of protecting the intellectual property of a system, applications can be made for patents or by dealing with the rights of the parties in a contract. All these

processes involve the application of the law and the legal rights of the parties.

Another common application of the law is in the adjudication of liability disputes between individuals or corporations. As mentioned earlier, this is the way in which case law is developed and applied by courts in factual situations.

This is more difficult to research because case law is developing everyday on different subject matter and not all cases are officially reported. While the most significant decisions of the highest courts are reported, those decisions of lower courts may not be reported. Other sources of information on court decisions are textbooks and journals and Web services.

All jurisdictions have a hierarchy of courts. The criteria for each level of the hierarchy will determine the court in which a particular dispute is adjudicated. In most jurisdictions, the level of the court is set by criteria such as the type of crime or the monetary value of the compensation being sought. For instance, most jurisdictions have a supreme or high court at the top of the hierarchy, a district or county court at the middle level, and a local or magistrate's court at the lower level. With all the various different levels in the court system, cases on the same sort of subject matter may be subject to adjudication in different courts for the same case. You might find that the legal liability for a person injured in the same accident is being determined in different courts because each person is seeking different amounts of compensation.

5 DIFFICULTIES IN APPLICATION OF THE LAW

The common law probably presents a more difficult concept than legislation. Legislation can be obtained from most government printers and is usually identified by subject matter. Because legislation is constantly undergoing change by parliament, subsequent amendments also need to be checked. Helpfully though, many government printers will produce a consolidation from time to time, which brings together the principal act and subsequent amendments. There are also many annotated versions of legislation produced by academics and law firms.

The common law, on the other hand, requires constant research for new decisions by the courts. Judicial consideration of previous cases and facts can sometimes produce only a minor variation to the law but one that can mean the difference between liability and no liability for a particular act or omission. One of the most relevant common-law concepts for professionals is the law of torts (or wrongs). This aspect concerns civil liability for damage caused to another. For those providing professional services, the tort of negligence is probably the most frequently used common-law action.

Explaining the law of torts and the concept of negligence easily needs a large text book to introduce its many facets. In very simple terms, it involves the concept of owing a duty of care to another, the standard of that duty of care, breaching the duty, and causing damage to the other party. Unlike a situation in which you may have a contract with another party, under which you have certain remedies against that other contracting party for a breach of the contract, the duty of care in the law of torts does not depend on a contract existing between the parties. For instance, if an engineer contracts with an aircraft manufacturer to design an aircraft and if the aircraft is faulty, the manufacturer has contractual rights against the engineer. If the aircraft crashes and injures passengers, the passengers may have a claim against the engineer, under the tort of negligence.

Laws also vary from country to country and, therefore, so do standards. Those designing measurement systems for use in different countries need to be aware of the differing standards between countries. A measurement system used in one country may be subject to different thresholds of liability in another.

Fortunately, with the development of more international standards and more uniformity of standards for products between countries, it is becoming easier for a designer and manufacturer to produce a product, which will ultimately comply with the standards in a wider variety of countries. While complying is more straightforward with legislation, it is not so with the common law.

In the same way as mandatory products standards vary from country to country, the threshold for tortious liability at common law will also vary even between those countries with the same basic system of common law. For instance, the standard of care required by a design engineer in one common-law country such as the United Kingdom may be more or less stringent than in another common-law country such as Australia even where the second country's common-law system is largely based on that of the first.

Because the standard of care expected by any common-law system will be subject to change at any time by the courts, there is a much greater imposition on any designer or manufacturer, particularly one offering a professional service in the nature of expert advice, to be conscious of the relevant standard in the jurisdiction in which they operate and the likely standard in any other jurisdiction in which the expert advice is likely to be relied upon. Fortunately, many common-law countries are now incorporating the concepts of standard of care into legislation to try to provide more certainty and consistency into these concepts.

Where legislation exists, it will generally override the common law on the same subject matter. However, legislation is still subject to interpretation by the courts, so expert advice from those legally trained should be considered where there is doubt about which law or standard applies.

6 THE LITIGATION PROCESS

The litigation process is complex. Litigation involves the process by which a series of events (facts) become a dispute between two parties and ends up in court. The term often used is *suing* another party.

An event that starts the litigation process is usually one where something goes wrong, which leads to injury to a person, damage to property, or a loss to business caused by failure. For instance, it may be a failure that leads to interruption of a business activity. There are any numbers of failures that can cause such an event.

The party (person or corporation) suffering the loss can make a claim for compensation against the other. If a contract exists between the parties, the usual claim is for breach of contract. If no contract exists, the claim can be in tort if such a right is recognized by law.

The process of litigation often starts with a letter of demand written between the parties to try to resolve the problem or to negotiate a settlement for money. A lawyer may be consulted at this stage or the parties may chose to try to reach a resolution between themselves. They may have consulted a lawyer to find out their rights under the law or the strength of their respective positions. In more serious events, most corporations involve lawyers from the time of the event.

If a letter of demand is not successful in providing resolution, lawyers are often called in to advise the parties on their rights. These rights may include commencing litigation in the courts. There are other alternatives such as further negotiation, mediation, and arbitration. Each has pros and cons in terms of cost and effectiveness.

7 COMMENCING LITIGATION

Litigation is usually commenced with a formal document called a *writ* or *summons*. This document is filed with the court and served on the other party being sued. The formal court process is called an *action*. The receipt of a writ or summons may be the first notice a party has of a claim or action against them.

Once the litigation process has commenced, the suing party may then file a more detailed statement of its claim against the other, called a *statement of claim* or *particulars of claim*.

The defending party is then usually required by the court rules to file a responding document called a *defense*. If no defense is filed, the court rules usually provide for the suing party to enter judgment by default. It is, therefore, extremely important that a party being sued responds to the service of a letter of demand or writ. Usually, a lawyer is consulted when either of these documents is received. The statement of claim and defense are called the *pleadings*. They define the issues of fact and law in dispute between the parties and are usually drafted by lawyers.

The type of claim and the monetary value of the claim determine the forum in which the court action is commenced. The process of commencing the claim is the same for both individual and corporation.

The action then proceeds through the court list toward a trial. Certain steps need to be taken by the parties before trial. These are

- finalization of the pleadings;
- disclosure of all relevant documents to each other;
- formal pretrial conference to facilitate settlement of the dispute.

At any stage, the possibility for a negotiated settlement or other form of resolution such as mediation can be explored. In general, 90% of all litigated disputes settle before going to trial. This is because the outcome of a trial is considered to be uncertain and many parties prefer to reach a compromise between them.

At all times during the process, parties (clients) are expected to cooperate with the court and their lawyers by providing frank and timely instructions and by making evidence available. Evidence can be in any form such as products, components, data, information either written or recorded in any other form, or oral testimony.

8 EVIDENCE AND DOCUMENTATION

In any situation in which legal issues are concerned, but especially in a litigated case, evidence is crucial to the way the case is conducted.

Because a party must prove its case to a court, it is by the evidence called at a trial that this is achieved. Whether the evidence is the oral testimony of a witness, documents or data, or expert evidence, the case will only be as good as the evidence available to prove the case.

It is, therefore, important that good quality and complete information of the various steps in the design and implementation of measuring systems is recorded and retained. Claims may not arise for many years after the product or system is released into the public environment, and disputes may not reach trial for many years. Memories as to

what took place will fade and the information recorded by the individuals involved will often be the only reminder of what took place. There are legislated time limits imposed on claims for damages and under the law of contract in most jurisdictions, but these can still be many years and extensions to these are possible in some cases.

Information and data that might be needed to trace a process for those involved should therefore always be stored in a secure environment for long preservation. Legal advice can be obtained before information is generated to ensure it is likely to be suitable. Legal advice is needed before destruction as to the relevant time limits that might apply for claims.

Good preservation of sound and relevant information is one of the best risk-management procedures for litigation.

The important issue for each designer to keep in mind is to inform themselves of the relevant legislative standards and the relevant common-law duties that they are required to comply with.

In most common-law jurisdictions, for instance, the standard of care likely to be used to judge a professional service, such as might be provided with expert advice in the design of a measuring system, is that of a test of reasonableness. While it is often difficult to identify the precise standard as to reasonableness, the best guide is usually what is considered to be appropriate industry practice by one's peers. A rider to this is that today what is reasonable to use because of technological change also needs to be considered. All this becomes quite difficult when the technology is considered to be at the very cutting edge of industry practice and standard or where, for instance, research into a completely new system is being carried out.

9 REDUCING LIABILITY

In order to reduce the risk of legal liability arising during the design and development of a system, and to best assist in defending a claim if it does arise, an organization or individual should

- be familiar with the applicable law and standards that apply and that will apply to the system, test environment, and the final system or product;
- maintain and preserve clear and comprehensive records throughout the entire design and test process, whether it is written notes, data, or hardware;
- seek appropriate expert advice to establish a sound and comprehensive contractual basis for any activity to protect intellectual rights and to deal with possible consequences such as by maintaining adequate insurance to cover all potential liabilities;
- seek timely expert legal advice in responding to all major incidents and threats of legal action such as a letter of demand or writ.

Following these basic principles will not, of course, guarantee a litigation-free project, but it should help reduce the numbers of adverse situations and assist in dealing with such a situation, if it arises, in the best and most efficient way. By being prepared and being able to demonstrate, a case can be well defended and will deter claims being made.

Some useful reference texts for further material on the legal processes are given.

FURTHER READING

Cooke, J.R. (2001) *Architects, Engineers and the Law*, 3rd edn, The Federation Press, New South Wales.

Kutz, M. (1998) *Mechanical Engineers' Handbook*, Wiley, New York.

Many other more detailed law texts are available on specific legal concepts and topics.

103: Legal Liability Issues for Designers – A Case Study

Christopher Sweet
Minter Ellison, Adelaide, South Australia, Australia

1 Introduction 671
2 Facts and Background of Role Study 671
3 Legal Consequences in General 672
4 Liability for the Deaths of the Test Crew and Damage to Property 672
5 Public Inquiries 673
6 Criminal Investigations 673
7 Product Liability and Recall 673
8 Legal Process 675
9 Legal Risk Mitigation 675
Reference 676

1 INTRODUCTION

A general introduction to the legal aspects of engineering applications is presented in **Article 102, Introduction to the Legal Process, Volume 2**. It is shown in that article that there are many legal situations to guard against and that expert legal assistance is often essential throughout a project.

This case study extends that discussion by illustrating some of the legal liability issues that can arise in the perspective of an application. It is intended to highlight issues, not to constitute legal advice: it is important to note that legal advice must be taken on a case-by-case basis using experts who can chart the way through the law and its processes.

In particular, this study demonstrates how a system designed for measurement during the design phase of a sophisticated product, such as a commercial aircraft, can fail, leading to legal liability being attached to several parties including, perhaps, the measuring system designer. The liability might be for damage to person, property, or for loss of business.

It is quite easy to see how an omission, however minor, during the test and evaluation phases of a part of the design of an instrument or system can lead to catastrophic consequences for the testing body, the manufacturer, and any third parties injured by the omission.

The legal concepts in the example can be readily applied to systems designed for measuring or testing or those that might be used during the operation and performance of a system or product. Equally, the concepts can be applied to the provision of advice as well as the provision of services that are based on software or data.

Legal risk is not totally avoidable but it can be mitigated by taking appropriate actions from the start of a project. This is covered in Section 9.

2 FACTS AND BACKGROUND OF ROLE STUDY

The case study example involves a fictitious university research unit, which has tendered for a contract concerned with testing the accuracy of a navigational instrument to be installed in the cockpit panel of a commercial aircraft. The navigational instrument incorporates the use of a novel method for processing the output signals from a conventional directional gyroscope, which indicates the direction of flight to the flight crew.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

The university unit decides to tender as a joint venture with a private technology company. In particular, the university brings to the contract the expertise of one of its senior researchers on the behavior of measuring instruments under extreme temperatures. Temperature is a critical feature in the proper operation of the new instrument.

A contract is drawn up between the university and private company for each to provide certain aspects of the tender services. The tender is successful and provides a significant grant to the research budget of the university department.

This university has never really been overly concerned about its exposure for liability on these sorts of projects and, therefore, decides that its normal insurance arrangements are adequate for the joint venture.

The contract goes ahead without event until there is a change in personnel within the university department. The senior researcher with the expertise in temperature effects has a major disagreement with the head of the department and resigns on short notice. His area is quite specialized, and the university finds it hard to replace his expertise. The head of department assures the private company that they can continue the contract with other researchers and so it continues.

The testing process is completed and the instrument is rated to perform satisfactorily at its declared operating temperatures. However, due to a miscalculation by one of the researchers during testing, it is actually out of specification and indicates incorrectly when subject to very low temperature.

The instrument is fitted to a prototype aircraft by the manufacturer for final operational test and evaluation. On its first test flight, the directional instrument fails to function accurately at high altitude (due to the low temperature there) and the prototype aircraft crashes, resulting in the loss of the test crew.

3 LEGAL CONSEQUENCES IN GENERAL

A number of significant legal consequences can flow from this scenario. The key issues are

- liability for the deaths
- liability for damage to property, that is, loss of the aircraft and damage to any buildings it destroyed while crashing
- public inquiry to ascertain what went wrong
- criminal investigation
- product recall if the component has already been placed into service.

These are addressed below as separate issues but they usually will be interdependent in reality.

These are each likely to be very expensive and time-consuming exercises. Their impact on the providing parties and their affiliations can be very harsh, leading to loss of employment, criminal penalties, corporate collapse, payment of compensation, and loss of goodwill in the market place, to name a few.

4 LIABILITY FOR THE DEATHS OF THE TEST CREW AND DAMAGE TO PROPERTY

There is likely to be a claim by the dependants of the test crew killed during the test flight. This claim may be brought under one or more of the following

- aviation law
- at common law
- under fatal accident legislation
- under employee/employer liability.

In each case, there is likely to be a complicated legal analysis of the various competing claims.

Any claim brought by the dependants is most likely to be against the manufacturer of the prototype, but it may also be contemplated against the contractors such as the university and private technology company or the flight licensing authority. To make matters more complex, those influencing the legal forces at play can be constantly changing.

Even if the claim, or claims, is brought against the manufacturer and not the university unit/private company, the manufacturer or their insurer will almost certainly seek to recover its losses from the university/private company. These losses may include compensation for the loss of the prototype hull and/or for loss of profits.

The contracts for the project may (or may not) have included detailed clauses providing for indemnities between the parties in the event of certain types of claims. This may further complicate (or possibly simplify) matters as to which party is ultimately liable.

The same could be said in the case of claims by any person injured on the ground or any property of such persons damaged by the falling wreckage.

If the various companies involved have insurers, each is likely to become involved with their own team of lawyers in addition to the lawyers involved for the parties, adding further complexity and cost to the resolution of any claims arising from the crash.

In all the legal scenarios above, the manufacturer, university, or private company could be exposed to considerable loss. A potential lack of adequate insurance could also expose the university to severe financial embarrassment,

depending on the level of insurance arrangements for the university, under normal circumstances.

5 PUBLIC INQUIRIES

In many countries, the crash of an aircraft will involve an official investigation by the local air safety bureau. Human fatalities in such circumstances will usually also result in an investigation by the local coroner.

In each case, an investigation can take months or years to complete, will cause much anxiety for all those involved (particularly those who will appear as witnesses), and will potentially expose the organizations to adverse publicity during the investigations, not to mention the close scrutiny each organization will undergo during any legal process.

During the safety bureau investigation, documents and design material will be taken and fully considered by a whole series of experts. Those personnel involved in all stages of the design and test process including the university team, both past and present, will be interviewed a number of times for witness statements. For instance, it is possible that each could be interviewed up to four to five times by the crash investigators, coroner's investigators, police, and lawyers in connection with each legal part of the case. Each legal inquiry may result in significant consequences for each of the parties involved.

The coronial investigation may result in findings, which lead to criminal investigations or recommendations, which affect the licence of the manufacturer or the way in which the aviation research industry is regulated.

The air safety bureau investigation and its findings may also lead to further investigations by the police or to recommendations for regulation of the industry.

All of these actions will produce adverse publicity for all those involved but particularly for the commercial reputation of the aircraft manufacturer. In this example, there are likely to be at least two unavoidable public hearings irrespective of the compensation claims. Prior to the hearings and for the hearing itself, staff time and traveling costs will be needed to prepare maternal and to attend the inquiries when called.

6 CRIMINAL INVESTIGATIONS

This is a third category of possible legal process that may affect this example. In disastrous events such as plane crashes, train collisions, or building collapses where lives are lost, there is an increasing trend for those in positions of responsibility to be held publicly accountable for their actions or omissions.

Historically, the corporate veil provided a shield for individuals against criminal prosecution for reckless or irresponsible acts that led to the loss of life. In many countries, that situation has been changing over recent years. Courts are now more prepared to look into the inner workings of a company to find those persons whose key actions may have led directly or indirectly to the event.

So, in this example, an investigation might reveal that the head of the university department misled the manufacturer as to their ability to continue the project and consider that his or her actions led directly to the failure of the instrument.

Also, one of the directors of the manufacturer may have known or suspected that the university did not have the competence to satisfactorily complete the project but may have remained silent for the sake of the project being completed.

In all cases, prosecutions for these sorts of disasters can carry significant penalties for companies and the individuals. Companies can be fined; individuals can be imprisoned or fined. The more severe the disaster, the more severe the penalties for those found to be responsible.

7 PRODUCT LIABILITY AND RECALL

In addition to the above legal issues, another costly allied liability situation, which could occur if any component has already been distributed in another product or in addition to the above types of liability is product liability. For a component that has already found its way into the market place in a product, the need to carry out a product recall may result. This is usually done to correct a defect in a design or its manufacture or to call back products with subsequent refund of purchase price and possibly compensation.

When these involve large numbers of a product item, they are commonly seen advertised in newspapers for such things as defects in automobiles, consumer goods, and foodstuffs. Where it is not possible to accurately identify all customers, these notices are directed at the purchaser or user.

In the notice issued for the defect, the serial numbers and product names of the defective lot are identified. The defect is explained in lay terms and the recommended immediate action to be taken is stated. Persons who believe they have a defective item are asked to make contact to arrange for withdrawal and refund, replacement, or rectification, free of charge.

Other less publicly noticeable statements are sent to wholesalers, franchisers, distributors, and retailers.

In the case of a defect in a sophisticated aircraft and where there are a low number of early models, the manufacturer will be aware of all sales and can make rapid contact with the buyers, and a public notice will not be needed.

Where it is not economic to rectify the item, the maker will usually need to recover it, compensating the buyer.

In some cases, it will be economic and prudent to make a modification to each item via a rework activity through provision of parts to be replaced or added by the owner, or as a factory operation by the supplier.

Recalls can be surprisingly expensive exercises. The overall cost might well not only take an individual project into a loss balance sheet but also place the organization into liquidation.

One way to demonstrate what is involved is by listing the factors involved for the example given, along with likely costs; refer to Table 1.

In this case study, assume that the maker had also delivered the same incorrectly calibrated prototype gyro and its data processor unit to an aircraft instrument systems' supplier. Also assume that no unit, apart from the one in the crashed aircraft, had yet been sold for deployment.

For safety reasons, the maker needs to immediately notify the supplier to not sell or use any of the units and to retrieve them for design alteration or destruction.

Many of the costs of a recall are for tangible items. Some items, however, are less easy to identify in clear-cut monetary terms, examples being the value of loss of goodwill and the impact of disturbing normal design and manufacturing operations.

Assume that the defect in this case study was shown to be quite simple, being that an electrical resistor involved in the calibration could marginally overheat, changing its resistance, which in turn could alter the gain of a critical signal-processing preamplifier.

Table 1. Costs associated with recall to repair or withdraw a defective low volume item.

Item	Time estimate (h)	Cost ($)	No. of units	Recall to fix option	Refund and withdraw product option
Designing a change fix					
Engineer labor (20 h @ $75/h^{-1})	20	75	All units	1500	
Drawings, part lists, work instructions	20	80	All units	1600	
Press release to lessen impact					
Media labor costs	8	80		640	640
Newspapers/radios costs				5000	5000
Freight and handling					
Freight of units back from customer		100	10	1000	1000
Tracking costs for 20 days	100	2	10	2000	2000
Outward dispatch labor	1	60	10	600	
Freight of units to customer		200	10	2000	
Compensation paid for inventory changes – not for any liabilities				2000	10 000
Media relations costs				5000	5000
Execution of repair					
Disassemble labor	10	80	10	8000	
Repair assemble	10	80	10	8000	
Test and recertification	40	80	10	32 000	
Other costs					
Loss of other production due to diverted staff – assessed as $1000 per recalled item repaired		1000	10	10 000	
Replacement loan gyros		500	10	5000	
Legal advice on change made				5000	
Total cost estimate of recall and repair of defective instruments				89 340	
Total cost to withdraw and refund for items					23 640
Minimum cost for option chosen				89 340	23 640
Sales income for 10 units			5000	50 000	
Overall income (loss in this case) for option chosen				($39 340)	($23 640)

Note: The assumption is that 10 instruments have been sold and all need to be rectified. They were sold for $5000 each, with a profit margin of 20% of sales. The problem was soon found to be quite simple, being that an electrical resistor's power-carrying capacity was inadequate and could overheat, changing the gain of a critical signal-processing amplifier.

This problem could have arisen from such reasons as

- supply of the wrong part by errors of specification or purchasing, or selection;
- ventilation around the resistor had been reduced from the design value to save cost and size;
- another component generating local heat was placed next to this critical resistor;
- the whole unit was placed in an environment hotter or colder beyond specification.

From the details of Table 1, several points are clear:

- The costs of the 'recall to fix' and the alternative 'refund and withdraw' options are both major – but one has to be chosen. The cost of recall cannot be avoided.
- The 'recall to fix' option is marginally the dearest to implement, but it will possibly be seen to be necessary for business to continue and for more sales to be made.
- The 'refund and withdraw' option is the least costly option, but it loses not only goodwill but also considerable income from sales of the product.
- The cost of either option may well exceed the costs of development, influencing the future of the project.

These costs are apart from any legal liability arising from the aircraft crash but would involve a commercial decision taken after the legal consequences have been considered. It is often the case that commercial decisions are taken in conjunction with advice from lawyers. Clearly, design errors like this need to be controlled early to reduce legal risk.

8 LEGAL PROCESS

In all of the legal consequences mentioned above, the process will begin with the personnel being contacted for interviews. Either the air safety investigators or the police or the coroner's investigators will be the first to investigate. They will request documents and other materials such as data, software, and hardware for preservation and for the purposes of their inquiries. Equipment and files may be impounded.

One would also hope that the companies have contacted their own lawyers and insurers at the earliest opportunity to assist with responding to the investigations.

The quality of the documentation made during the various processes used in design and testing becomes critical while responding to any allegations. All the documentation and other materials have the potential to become evidence during the investigation. Notes made by each of the personnel during the project will become their aide memoires when called upon to give statements. Human memory is notoriously inaccurate after time has elapsed and the notes made may be the difference between accurately recalling ones actions or not. That is why the concept of proper record keeping is reinforced constantly by the legal profession. When a person gives evidence to a court, their credibility to recall events is often determined by whether they have properly recorded their past actions.

9 LEGAL RISK MITIGATION

The above examples show how legal liability will affect all concerned if a design causes harm in some way in its normal or abused use. Designers need to be able to defend any design that has been practiced to 'best practice' standards.

Some indicators on how to limit that risk as design proceeds are now presented. It would be ideal to have a lawyer in the design team to give advice on every decision made but that is often not practical. Designers must exercise judgment and make use of regular legal reviews of the project.

The main line of defense will usually be the ability to demonstrate 'sufficient standard of care' or that 'due diligence has been practiced'. Where a situation cannot be formally described in complete reductionist terms – most of those that give rise to litigation – the case will be argued on the basis of 'reasonableness', a highly subjective cognitive concept that lawyers are experienced in arguing one way or the other.

Preparation for legal defense starts when the project commences. Following practices of reasonable design and keeping a suitable document trail are absolutely necessary. Without these, it would be difficult to provide the evidence needed to withstand scrutiny during any of the legal processes discussed above.

If adequate records are not made at the time, it may well be impossible to later find the documented evidence needed to support witnesses or the defense case. It is acknowledged that detailed documentation will impose significant additional overhead on a project and record information that may never be called upon. Taking such precautions is, however, much like taking out additional insurance; one hopes it is never needed but it is a comforting thing to have in place when needed.

The design team is not expected to provide for every possible circumstance in a design, as that would suppress creation and cost-effectiveness. The team should allow for that which is reasonable. Allowance has to be made for all situations that can reasonably be expected to arise.

Some key points to assist in the better defense of a legal liability suit are:

- description of, and adherence to, a formalized design process by project personnel;
- peer design review and approval of project design information and documentation;
- management of system safety by a formally constituted Safety Board, with a sound and documented safety and design change process;
- reference and adherence to Standards and Regulatory requirements for the appropriate aspects of the design;
- appropriate training for designers, with records of attendance and achievement in learning;
- findings of independent quality audits conducted during the execution of project activities that confirm that adequate duty of care has been exercised;
- availability of the above evidence that is suitably framed to suit legal requirements;
- maintenance of records of appropriate calibrations and certifications;
- availability of information to show that current practice used in the project is in accordance with other similar designs;
- well-run risk and safety assessment process that has been in place.

Adequate security of records and equipments is also necessary to support the veracity of any case made. Relevant details must be secured and staff access controlled.

Just how far these various issues should be addressed will depend on factors such as

- nature of the project (such as defense, government, sea, air, land, utilities, commercial, personal, health, etc.)
- kind of system delivered (benign, hazardous, domestic, factory)
- environmental impact, guidance being found in an environmental impact statement (EIS) for the project, if one exists
- locality and extent of application (office, city, country, international)
- level of perceived public acceptance (automobile safety vs aircraft safety, natural environment, low or heightened public interest, child or adult user)
- lifetime of system use (days to decades)
- business issues (profits made, risks to be accepted)
- potential population affected by the system.

Clearly, legal advice is crucial when considering procedures and processes to minimize the risk of legal consequences for any project. Lawyers can anticipate the consequences and advise on how to protect the project from such risk.

A well-prepared set of documentation and a sound best practice process form a sound basis for fending off claims; either an adverse consequence will not occur or, if it does, a claimant will find it harder to argue that standards have been breached, leading to liability. Sound processes will also make it much easier for a legal team to assess whether a claim will succeed and so assist them to properly manage the consequences and reduce public and commercial damage.

A corollary is that a legally sound process being in place is highly likely to avoid litigation in the first place because the delivered system is less likely to incur difficulties because of its greater integrity.

Further information on this topic is available in Sydenham (2004). An extensive table of issues to address at each phase of the project life cycle is presented there.

REFERENCE

Sydenham, P.H. (2004) *Systems Approach to Engineering Design*, Artech House, Norwood, MA.

104: Models of the Measurement Process

Luca P. Mari
Università Cattaneo, Castellanza (VA), Italy

1 Measurement as a Comparison Model 681
2 The Output/Input Black Box Model 682
3 Set-theoretical Model 683
4 Generalized Model 683
Related Articles 684
References 684
Further Reading 684

1 MEASUREMENT AS A COMPARISON MODEL

Measurement is an operation of data acquisition and presentation, aimed at expressing in symbolic form the information empirically obtained on a system about a quantity, the measurand (we accept the common ambiguities of calling 'measurand' both the system under measurement and the measured quantity, and the latter in both its general and specific forms, for example, length and length of a given object in a given time).

Peculiar to measurement is the requirement of being objective and intersubjective, where objectivity implies that measurement results convey information only related to the system under measurement and not its environment, and intersubjectivity requires that measurement results convey the same information to different subjects. As such, these properties appear an ideal target, justifying the efforts to constantly enhance measurement devices and procedures.

To achieve an acceptable degree of objectivity and intersubjectivity, measuring systems are adopted, which include selective and repeatable sensors and traceable standards. Indeed:

- Although human beings are able to directly sense a fair amount of quantities and are well trained to express in linguistic form their perception (e.g. 'it is rather cold', 'this is heavier than that'), their statements are affected by subjectivity, that is, they report information on both the sensed system and the perceiver state; to avoid the influence of the latter, and thus to enhance the objectivity of the operation, the measurand is transduced by a sensing system whose output ideally depends only on the measurand and is unaffected by influence quantities and internal imperfections.
- While related to the measurand, the quantity provided by sensors still depends on their specific behavior; as a consequence, distinct sensors, even if perfectly repeatable, produce different outputs from the same input; furthermore, in many cases the sensor output quantity, appropriate for signal conditioning and for driving presentation devices, is not dimensionally homogeneous to the measurand. The sensor output must then be dealt with as an instrument reading, not as a measurand value. To make the information obtained by the measurement intersubjective, a common reference must be adopted so that measurand values are expressed in comparison to such a standard. Therefore, the possibility of tracing the readings to the agreed standard is critical, a condition operatively ensured by instrument calibration.

The requirement of empirical comparison to traceable standards is so fundamental that it can be assumed as distinctive of measurement; generic scale-preserving evaluations can be formalized as homomorphisms from empirical to symbolic relational systems, as shown in Figure 1 (*see also* **Article 8, Formal Theory of Measurement, Volume 1**). In the case of measurement, such mappings are

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

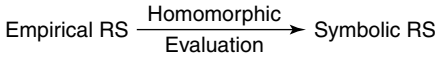


Figure 1. A generic scale-preserving evaluation.

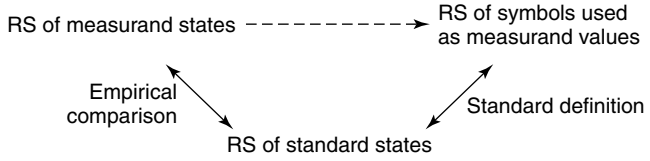


Figure 2. Measurement as a scale-preserving evaluation obtained by the comparison to a standard.

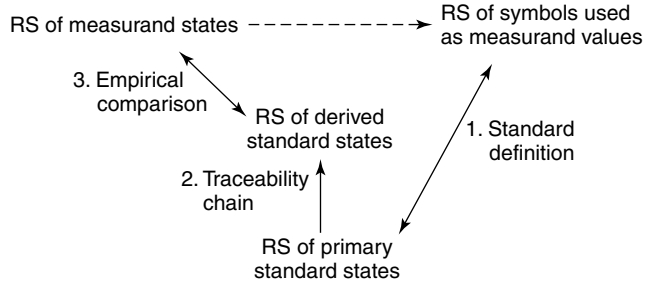


Figure 3. Measurement as a scale-preserving evaluation obtained by the comparison to a standard derived by a primary standard.

not direct but mediated by the comparison to standards, as shown in Figure 2. Finally, the situation in which primary standards are not directly available can be schematized as in Figure 3. In such a case, operations 1 and 2 are usually carried out before measurement: nevertheless, measurement cannot be completed without them and therefore such operations play an essential role in the definition of measurement itself.

As a consequence, measurement results must state a measurand value with reference to the adopted standard, usually expressed in the form of a measurement unit (*see also* **Article 43, Units, Volume 1**).

2 THE OUTPUT/INPUT BLACK BOX MODEL

It is a well-known fact that different methods of measurement exist, each of them corresponding to a specific technique to perform the comparison between the measurand and the standard (see Figure 4). While some methods require the synchronous presence of the measurand and the standard (e.g. following the paradigm of the two-arm balance provided with a set of standard weights: a direct comparison), many others are based on the usage of devices acting as serializers of the comparison, so that a

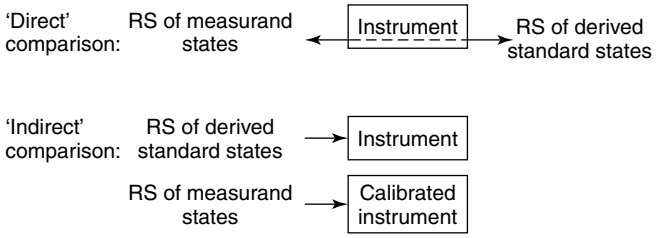


Figure 4. The different usages of the measuring systems as comparators.

measurement involves (at least) two interactions: standard-instrument and measurand-instrument.

In its interaction with the measurand, the instrument generates an output; a general problem of measurement can then be stated as follows: *from the output of the measuring instrument ('the reading') its input (the state of the system under measurement and its environment) must be reconstructed, and from this state a measurand value must be inferred*.

To cope with this input-from-output inference problem, two basic strategies can, in principle, be followed:

- The analytical model of the measuring system behavior is identified and the obtained characteristic function is inverted, so that from the output readings the input signals are computed. Because of its complexity, this approach is seldom adopted.
- The system is regarded as a black box and only its input–output behavior is taken into account: the instrument operates by interacting with a set of (known) standard states and the corresponding output readings are recorded; by a suitable interpolation, this collection of couples becomes the so-called *calibration curve*, shown in Figure 5, which can be thought of as a mapping from measurand values to instrument readings. This function is then inverted so that each instrument reading can be associated with a measurand value.

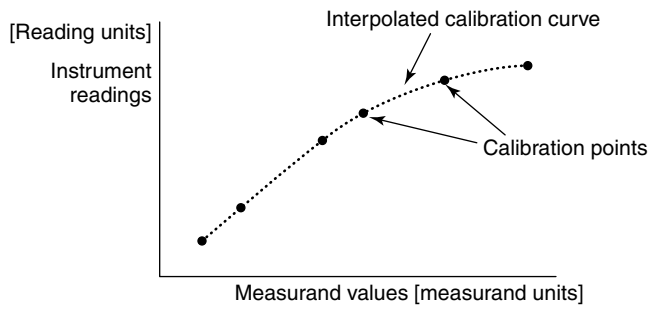


Figure 5. A diagram with the example of a curve generated by calibration.

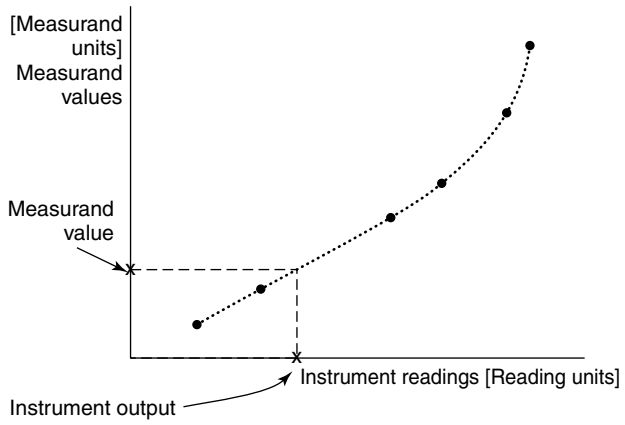


Figure 6. The example of an inverted calibration diagram, for usage in measurement.

The interactions standard-instrument and measurand-instrument have therefore a complementary function: while the former is aimed at creating a calibration diagram, the latter uses the inverted diagram, shown in Figure 6, to find the measurand value that corresponds to the obtained reading.

To enhance the user-friendliness of the measuring systems, it is customary to set up their presentation component so that the data they display are expressed directly in measurand units, that is, the calibration diagram is embedded into the systems. While measurement always requires calibration information, in these cases one can specifically speak of calibrated instruments.

3 SET-THEORETICAL MODEL

The sensor behavior, therefore critical for both calibration and measurement, is usually expressed as a characteristic function formalizing the input–output conversion performed by the sensor itself.

The sensor input, a couple $(x, \overline{w})$ where $x = x(t) \in X$ is the measurand and $\overline{w} = \langle w_1, \ldots, w_n \rangle = \langle w_1(t), \ldots, w_n(t) \rangle \in \overline{W}$ is a collection of further quantities influencing the sensor behavior, is transformed to its output $y \in Y$. Therefore, the sensor characteristic function:

$$f : X \times \overline{W} \times T \to Y \tag{1}$$

takes the measurand $x(t)$, the influence quantities $\overline{w}(t)$ and the current time t, included to take into account possible time-dependent effects, and associate them with the output signal $y(t) = f(x(t), \overline{w}(t), t)$ to which both the measurand ('the signal') and the influence quantities ('the noise') contribute.

This simple formalization allows us to introduce some basic parameters describing the static behavior of a sensor:

- *Sensitivity*: ideally, $x_1 \neq x_2$ implies $f(x_1, \overline{w}, t) \neq f(x_2, \overline{w}, t)$, that is, distinct measurand values always produce distinct outputs; the ratio $\Delta y / \Delta x$ expresses the aptitude of the sensor to reproduce measurand variations to output values.
- *Selectivity*: ideally $f(x, \overline{w}_1, t) = f(x, \overline{w}_2, t)$ even if $\overline{w}_1 \neq \overline{w}_2$, that is, the sensor output is not affected by the variations of influence quantities; the less is the variability of y due to $\overline{w}$, the better is the sensor (therefore, selectivity corresponds to nonsensitivity to influence quantities: the relative contribution of the measurand to the output can be formalized as a signal-to-noise ratio).
- *Repeatability and stability*: ideally $f(x, \overline{w}, t_1) = f(x, \overline{w}, t_2)$ even if $t_1 \neq t_2$, that is, the sensor output is not affected by short-term (fluctuations) and long-term (aging) time effects; the less is the variability of y due to t the better is the sensor (a stable sensor does not require frequent recalibrations).
- *Linearity*: ideally $y = ax + b$ (where a and b are given coefficients, possibly with $b = 0$), that is, f is a straight line, the better the actual sensor behavior is approximated by this equation, the better the sensor is usually considered to be (a linear, zero-crossing sensor is calibrated in a single operation, aimed at determining the slope a).

In addition to these static parameters, the dynamic behavior of the sensor is synthesized by parameters such as its frequency response (*see also* **Article 29, Relationship Between Signals in the Time and Frequency Domain, Volume 1**; **Article 27, Signals in the Frequency Domain, Volume 1**; and **Article 36, Systems in the Frequency Domain, Volume 1**).

The technical specifications for sensors usually include some quantitative evaluation for these parameters in the nominal conditions of usage, expressed by the allowed ranges of measurand and influence quantities.

4 GENERALIZED MODEL

The inference process that leads to the evaluation and the expression of a measurand value is always only plausible in its results, and in general nothing can be inferred with certainty about the measurand value. The causes of this lack of certainty are various, and in particular the following:

- The model of the measurement system has not identified all of the relevant influence quantities, and any one of them may have a significant variability, such that the

environmental conditions (including human operators) change after the calibration.
- The measuring system is less stable than expected when the calibration procedure was defined, that is, the instrument would require a recalibration before its usage.
- The interpolation shape of the calibration curve does not adequately map the actual instrument behavior (e.g. it is significantly nonlinear where a piecewise linear interpolation was chosen), so that for some instrument reading subsets the instrument is wrongly calibrated – *see* **Article 58, Description of Accuracy, Linearity, and Drift, Volume 1**.

All these cases can be formally characterized by recognizing that the certainty implied in the choice of a single-valued association between instrument readings and measurand values is not adequate. In the interaction with the measuring system during calibration, each measurand value generates an instrument reading that should be considered a sample drawn from a whole set of possible readings.

Such variability can be formalized according to a set-theoretical model, so that the information obtained in the calibration is expressed by a calibration strip, in which an interval of possible readings, whose center and width can be considered as the nominal reading and an uncertainty interval respectively, is associated with each measurand value (see Figure 7). (Note the changes of the calibration strip width along the measurand axis, taking into account nonuniformities in the uncertainty evaluation.)

As in the previous (certain, and therefore ideal) case, this diagram is used in its inverted form during measurement: for any given instrument reading, an uncertainty interval of possible measurand values is obtained together with a nominal value (*see also* **Article 54, Explanation of Key Error and Uncertainty Concepts and Terms, Volume 1**).

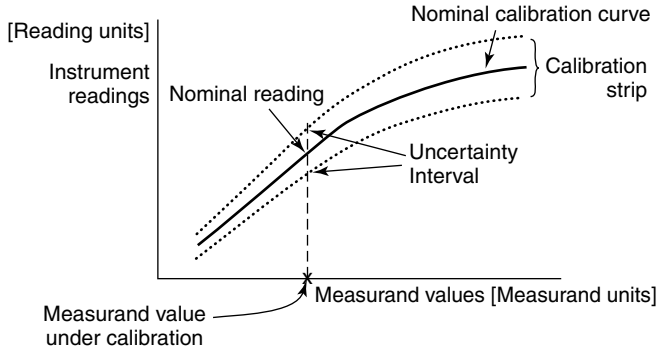


Figure 7. A diagram with the example of a strip generated by a calibration in which uncertainty has been taken into account.

An even more general approach could be adopted by expressing the uncertainty estimation as a standard deviation, and therefore in a probabilistic framework, as recommended by the ISO *Guide to the Expression of Uncertainty in Measurement* (1993) (GUM). The *Guide*, based on a recommendation by the International Committee for Weights and Measures (CIPM, 1981), states that measurement uncertainty can be estimated on the basis of both statistical and nonstatistical methods, and specifies a procedure to combine such components into a combined standard uncertainty. The set-theoretical formalization can then be regarded as a specialization of this framework: if the combined standard uncertainty is multiplied by a coverage factor, then an expanded uncertainty is obtained, which is thought of as the half-width of an uncertainty interval.

The inherent presence of uncertainty justifies the fundamental assumption that the result of a measurement must state not only a (nominal) measurand value but also its uncertainty estimation. Uncertainty in measurement is covered in **Article 54, Explanation of Key Error and Uncertainty Concepts and Terms, Volume 1**.

RELATED ARTICLES

Article 54, Explanation of Key Error and Uncertainty Concepts and Terms, Volume 1.

REFERENCES

CIPM, BIPM (1981) *Procès-Verbaux des Séances du Comité International des Poids et Mesures*, **49**, 8–9, 26, (in French); Giacomo, P. (1982) News from the BIPM, *Metrologia* **18**, 41–44.

ISO (1993) *Guide to the Expression of Uncertainty in Measurement*, International Organization for Standardization, Geneva.

FURTHER READING

Krantz, D., Luce, R., Suppes, P. and Tversky, A. (Vol. 1: 1971, Vol. 2: 1989, Vol. 3: 1990) *Foundations of Measurement*, Academic Press, New York.

See also the volumes of the *Proceedings on the International Workshop Advanced Mathematical Tools in Metrology*, in the *Series on Advances in Mathematics for Applied Sciences*, World Scientific, Singapore, New Jersey, London, Hong Kong (currently published: Vol. 1: 1994, Vol. 2: 1996, Vol. 3: 1997, Vol. 4: 2000, Vol. 5: 2001).

105: Modeling with LabVIEW™

Wiesław Tłaczała
Warsaw University of Technology, Warsaw, Poland

1 General Overview and Nature of the LabVIEW™ Environment 685
2 LabVIEW G-programming Rapid Information 687
3 Data, Wires, and Terminal Types in LabVIEW 688
4 Quick Start with LabVIEW – VI Example 688
5 General Application to Different Classes of Measuring Systems with VIs 689
6 Relevance to Presented Applications 690
7 Communication and Data Transfer Inside and Outside VI 691
8 VI Behavior Customizing 692
9 Modularity and Hierarchy 693
10 Data Flow Principle Presentation 694
References 694
Further Reading 694

1 GENERAL OVERVIEW AND NATURE OF THE LabVIEW™ ENVIRONMENT

LabVIEW (**Lab**oratory **V**irtual **I**nstrument **E**ngineering **W**orkbench) is the trademark of National Instrument Corporation (NI). It is a programming environment including all of the tools necessary for instrument control, data storage, data analysis, and data presentation. It uses graphical programming and a graphical language called *G*. The main goal of LabVIEW is the creation of virtual instrumentation.

1.1 Virtual instruments

Virtual instruments are computer-controlled instrumentation systems. The physical instruments, which exist in the background of the computer (Figure 1), may be controlled using the computer keyboard, mouse, and display. Also, a wide variety of the computer-based applications for process modeling, instrument simulation or animation of the execution of the measuring system are possible using virtual instruments. In this sense, all LabVIEW programs are Virtual Instruments, called *VI*s for short.

1.2 Graphical programming language G – VI components

The graphical programming language G uses icons and wires to create graphical source code. In graphical language, icons are used instead of text-lines, which are used in the sequential text-based programming languages. Graphical code is created on the Block Diagram behind the user interface called the *Front Panel* (Figure 2). The SubVI icon represents a program placed as a subprogram within another VI. During program execution, data are transferred between the Front Panel and the Block Diagram and carried between icons on the block diagram. The VI connector pane gives the possibility of passing data to/from the VI. The path of data flow on the block diagram is determined by wire connections. LabVIEW is a system based on the data flow principle.

1.3 Data sources

Data used by the program can be supplied by controls on the front panel, loaded from files as off-line data, and received

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

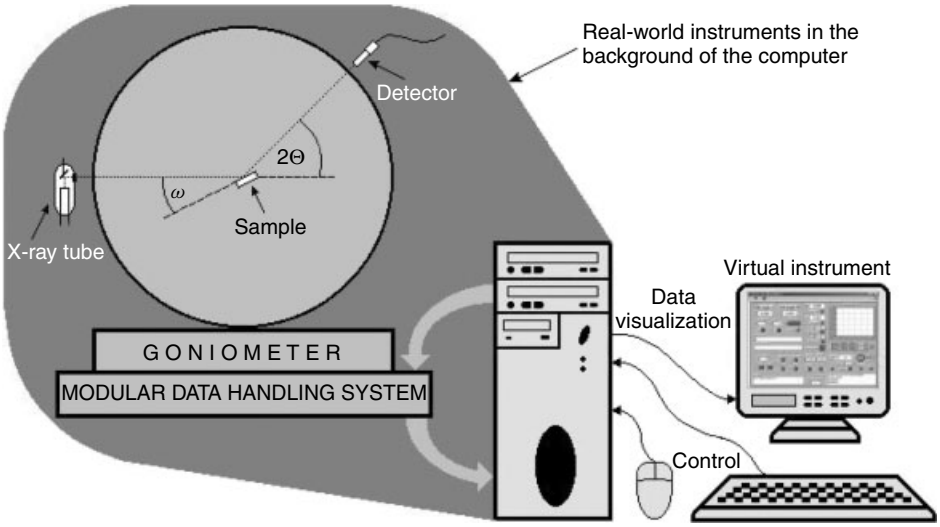


Figure 1. Virtual instrument presentation.

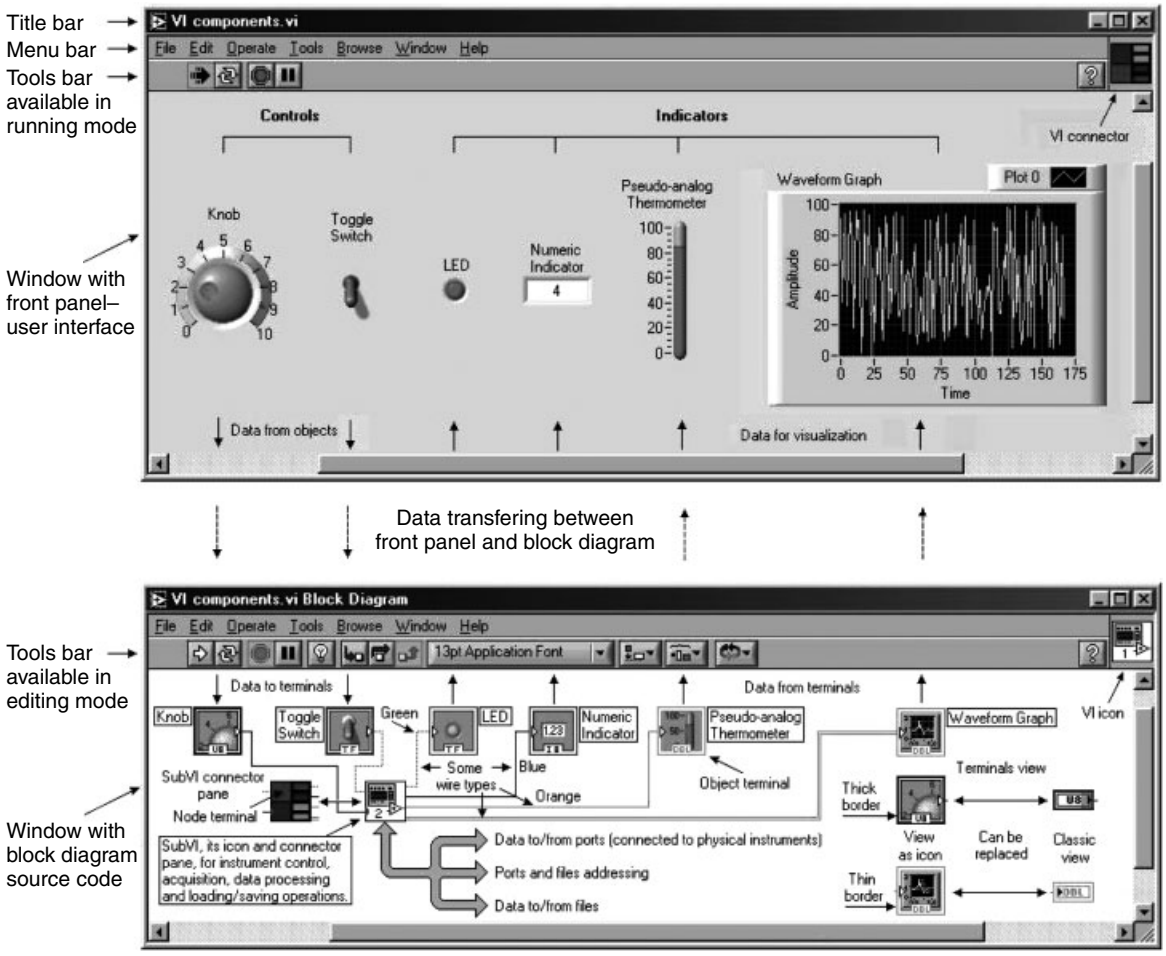


Figure 2. VI components.

through ports from physical instruments as on-line data and also from the Internet.

After modification, data can be displayed by indicators on the front panel, saved to files, or sent back to peripheral physical instruments.

Each control and indicator placed on the front panel is represented by its corresponding terminal, which appears automatically on the block diagram when VI is built.

Data are passed to/from the object on the front panel through the object terminal on the block diagram.

2 LabVIEW G-PROGRAMMING RAPID INFORMATION

After launching the LabVIEW application, the start-up screen appears and a user can select a New VI. A new VI named *Untitled 1(2,3...)* appears as a composition of two windows that comprise a VI, the front panel, and its associated block diagram. At the initial stages of building a new VI, it is helpful to see both windows side by side to determine what can be done from the menu bar.

2.1 Editing tools

Editing tools used to build a VI are the Tool bar and Tools, Controls, and Functions palettes (Figure 3). Palettes can be visible after selecting them from the menu bar.

The Tool bar contains buttons and indicators to build, edit, debug, and run VIs.

The Tools palette is a graphical panel that contains various tools used to create and operate VIs. A tool is a special mouse cursor for performing specific operations such as positioning, wiring, labeling, operating, scrolling, setting breakpoints, probes and shortcut menus, and also for operations usually used in a paint program. The Tools palette is common to both front panel and block diagram operations.

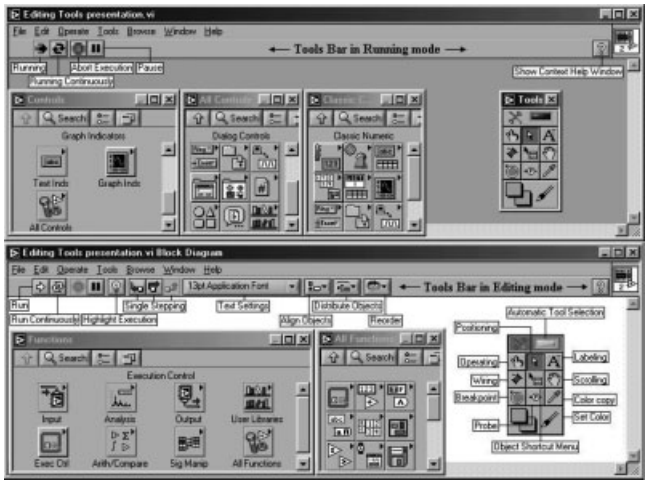


Figure 3. LabVIEW editing tools.

2.2 The front panel

The front panel window contains various types of controls and indicators and models a real-world instrument front panel. Front panel design activity needs to place on it objects used by a user to control VI and to display data produced by the program. Figure 4 shows, as an example, the front panel of the virtual X-rays diffractometer, which gives a user possibility to calibrate the sample and detector positions; select parameters for the scanning; acquire and present data; and save data in the file.

The complete set of front panel objects is available from the Controls palette and the All Controls subpalette.

2.3 The block diagram

The block diagram window contains the program code in a graphical form. It contains terminals of the front panel objects, programming structures, functions, constants, subVIs, and wires.

The complete set of block diagram functions is available from the Functions palette and All Functions subpalette.

Objects can be placed on the front panel/block diagram by dragging and dropping, or copying. Front panel objects and constants on the block diagram can be also created by selecting from the shortcut menus of the objects placed earlier on the block diagram.

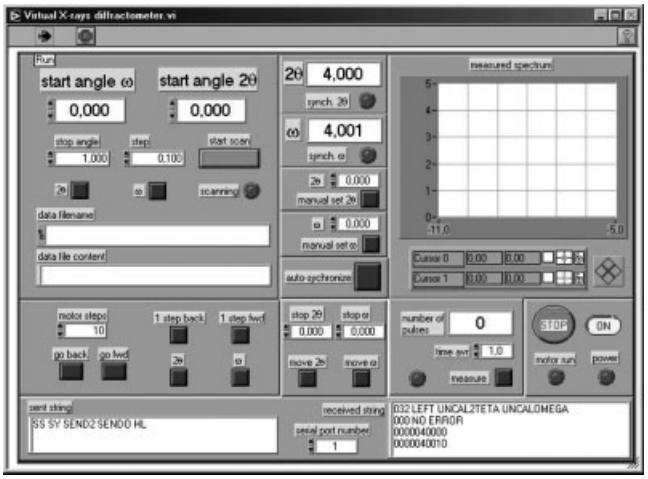


Figure 4. The front panel of the X-rays diffractometer. The application has been developed by M. Sarzyński, a PhD student at the faculty of physics in WUT, under the author's direction.

2.4 Wires and terminals

Wires are used to carry data from one terminal to another. Wires are analogous to variables in traditional programming languages. Terminals are objects or regions on nodes through which data pass.

3 DATA, WIRES, AND TERMINAL TYPES IN LabVIEW

3.1 Data types

Data type is a format for information. Data can be the instrument status information data, reference data representing node properties or status, data obtained as results of measurements, or data used for addressing. Data types used in LabVIEW and acceptable for most VIs and functions are numeric, Boolean, string, waveform, array, cluster, refnum, enumeration, and path.

3.2 Wire types

Wires can be of different types, colors, line styles, and thickness, depending on carried data types. Wire types are unique for each data type. For example, orange color is reserved for floating point numbers, blue for integers, green for Booleans, and pink for strings. Basic wire styles used in block diagram are as follows: thin, solid lines are used to carry scalars; a bit thicker solid lines for 1D arrays; double lines are reserved for 2D arrays of numbers; and thick lines are used to carry 2D arrays of Booleans or strings. A broken wire appears as a dashed black line. To find out why the wire is broken, the Wiring tool is moved over the broken wire to display a tip strip with an error description.

3.3 Terminal types

Control terminals and node outputs terminals have thick borders, while indicator terminals and node inputs terminal borders are thin. Terminal color indicates what data type will be supplied or accepted by this terminal. The programmer can open the Context Help window, move the cursor onto the selected node, and get information about data types accepted or supplied by the terminal. The Context Help window also displays the function and VI parameters, the object description, and the parameter type definitions. It can be opened by selecting the same from the menu bar. A broken wire can be removed by selecting and deleting it.

3.4 Debugging techniques

The broken Run button appears on the tool bar in the case of a broken VI. A broken VI is a VI that cannot compile or run because of programming errors. LabVIEW has many built-in debugging features like Execution Highlighting, Probe and Breakpoint tools, Pause, and Single-Stepping through a VI.

With the Execution Highlighting, an animation of the execution of the VI block diagram can be viewed. Using Probes, the intermediate values created during VI execution can be checked. Single-Stepping through a VI and its subVIs can be used to execute a block diagram node by node. Breakpoints can be set at certain locations of the VI using the Breakpoint tool. The Pause button temporarily stops execution.

The programmer can also open the Error List window by clicking the broken Run button and locate the object reporting error.

4 QUICK START WITH LabVIEW – VI EXAMPLE

4.1 Building VI

As an example, consider the VI created for presentation of the ^{137}Cs spectrum data loaded from the file. For building a user interface as shown in Figure 5, there is need to create a waveform graph, a slide, and path objects. These are created as objects by selecting them on the Controls palette. The Positioning tool is used to rearrange or resize selected objects. Right-clicking the object and selecting Properties from its shortcut menu can open the dialog box to configure the object. The Labeling and Operating tools change the defaults for the object. Data paths between output terminal and input terminals on the block diagram are defined using the Wiring tool.

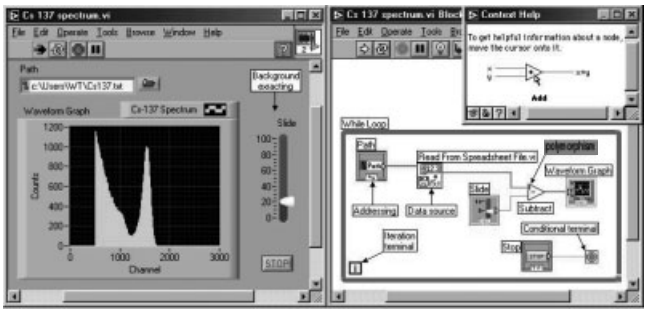


Figure 5. Simple VI example – its front panel and block diagram.

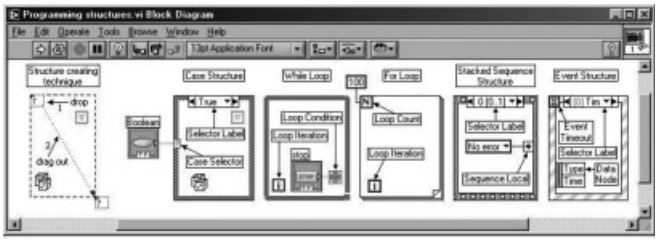


Figure 6. LabVIEW programming structures.

For VI execution control, the While Loop is created by selecting the same on the Functions palette. The While Loop is a structure used to repeat the subdiagram inside it until a conditional terminal receives a particular Boolean value. The iteration terminal provides the current loop iteration count.

4.2 Programming structures

Programming structures, shown in Figure 6, are program control nodes, such as Flat/Stacked Sequence structures, Case structure, Event structure, For Loop, or While Loop. Structures can execute code repeatedly or conditionally. They are not created on the block diagram immediately after being selected. Instead, a small icon representing the structure appears in the block diagram, giving the programmer the opportunity to size and position it to encompass all the objects that need to be placed inside the structure and then execute within it. The structure-creating technique is shown on the left-hand side in Figure 6.

4.3 Running VI

A user can run a VI by clicking its Run button. If a user runs the VI, LabVIEW compiles it and starts the VI execution. While the VI is executing, the Run button changes appearance to indicate that the VI is running.

A user can stop execution by clicking the Abort Execution button or by using the Stop button placed on the front panel. The program will stop immediately when the Abort Execution button is selected, but data can be corrupted this way. The VI will reach the end before stopping when the Stop button is used. This is the recommended stopping mechanism. The continuous Run button causes the VI to run over and over until the Abort Execution button is pushed.

5 GENERAL APPLICATION TO DIFFERENT CLASSES OF MEASURING SYSTEMS WITH VIs

5.1 Where can LabVIEW be used?

Areas of the typical uses of LabVIEW are electrical and mechanical engineering, physics, chemical engineering, biology, biomedical engineering, civil engineering, and many others. For example, scientists at NASA used LabVIEW to analyze and display Mars Pathfinder Sojourner rover engineering data (Bishop, 1999). At the Laboratory of Dielectrics and Magnetics at Warsaw University, X-ray measurements of the layer spacing in the SmA phases were performed using the diffractometer (Figure 4) controlled by the software developed in LabVIEW (Pociecha *et al*., 2002).

5.2 LabVIEW in industry and research

LabVIEW applications are implemented worldwide in many industries and in research for test and measurement applications, process monitoring and control, machine monitoring and control, and design process control. For example, Construzioni Apparecchiature Elettroniche Nucleari (CAEN) S.p.A, offers a VME (Versa Module Eurocard) controller which can be operated from the parallel port of a standard PC using LabVIEW software developed in CERN. Analog Devices (AD) and Crossbow Technology, Inc. used LabVIEW software for the demonstration dual axis, 2 g accelerometer.

5.3 LabVIEW in universities

LabVIEW programs used in university laboratories improve the didactic process. These applications can be used in e-learning and distance learning. With its Remote Panel, the LabVIEW application can be turned into a remote laboratory. With their Web browser, users can open a link to a remote application, and control it using the Remote Panel, and observe how an experiment runs. Such a panel, as an example, is shown in Figure 7. This is the statistical distribution analyzer for remote control of the experiment devoted to the statistical character of the gamma decay investigations (*see also* **Article 106, Virtual Instrumentation in Physics, Volume 2**).

5.4 Communicating with LabVIEW

LabVIEW simplifies the tasks of communicating with GPIB, USB, serial, Ethernet, VXI, and PXI instruments. For

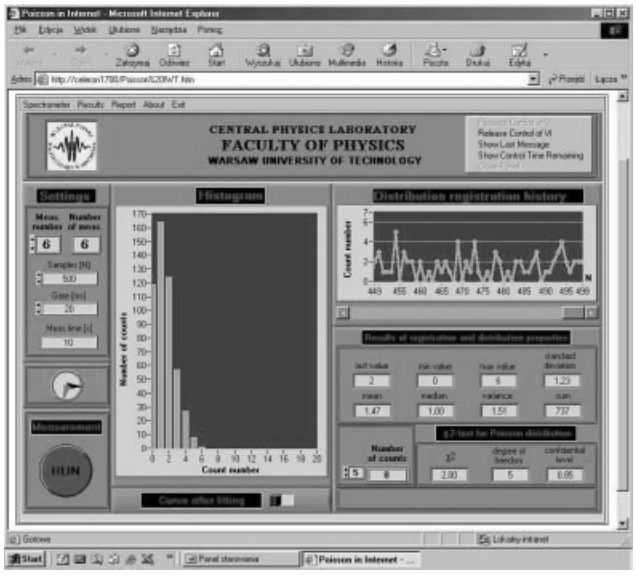


Figure 7. The front panel of the poisson distribution analyzer as a web page.

example, the Metex M-3660D multimeter uses the built-in RS 232 port for serial communication, and with the Pt100 (*see also* **Article 197, Metallic Resistance Temperature Detectors (RTDs), Volume 3**) element, it can be used as a digital thermometer controlled by a computer in thermodynamics experiments. The GPIB (General Purpose Interface Bus) interface is optionally installed in the Hameg HM-507 oscilloscope and can be use to communicate between a PC and the HM-507. It is relatively easy to build applications using LabVIEW to control these real-world instruments.

5.5 Data analysis with LabVIEW

LabVIEW provides more than 400 built-in functions for extracting useful information from acquired data and for signal processing and analysis. Tools available from advanced libraries can be used for solving systems of linear algebraic and ordinary differential equations; for integration and differentiation (with G Math Toolkit); for probability and statistics analysis; for curve fitting; and for signal generation and signal processing with DFT (Discrete Fourier Transform) and FFT (Fast Fourier Transform) analysis (*see also* **Article 143, DFT and FFTs, Volume 3**), windowing and filtering.

6 RELEVANCE TO PRESENTED APPLICATIONS

In Figures 4 and 7 are shown two front panels of the exemplary VIs built to control the different types of experiments. One of them is used to control X-rays diffractometer (Figure 4) and the other to perform experiments devoted to the statistical character of gamma decay investigations (Figure 7).

The presented graphical user interfaces are intuitive in operation and simple to apply. Designing of similar interfaces using a traditional computer language can take many days. With LabVIEW, this can be done in an extremely short time, even with very little experience in G-programming.

LabVIEW terminology and the block diagram representation of VIs (Figure 8) are easy to interpret and natural to use by scientists and engineers. Typical advanced operations usually needed for communication or analysis have been encapsulated in VI libraries for quick application.

The productivity with LabVIEW is significantly higher, and generally, a system's development time is much reduced compared to traditional programming. For example, the LabVIEW flexibility allowed rapid modification of the virtual diffractometer (Figure 4) to then make it applicable (Figure 7) to measurements devoted to the statistical character of gamma decay investigations. The modification can be made easily because both instruments use the same subVIs needed to control gate time, collect data in the time limited by the gate, present acquired data on the XY graphs, and communicate through serial RS 232 interface with peripheral instruments.

LabVIEW also improves the teaching process and the way by which students learn in universities. The applications used during the laboratory work help students to understand better how an experiment is running because VIs can provide a lot of useful information that is quickly available from the menu bar.

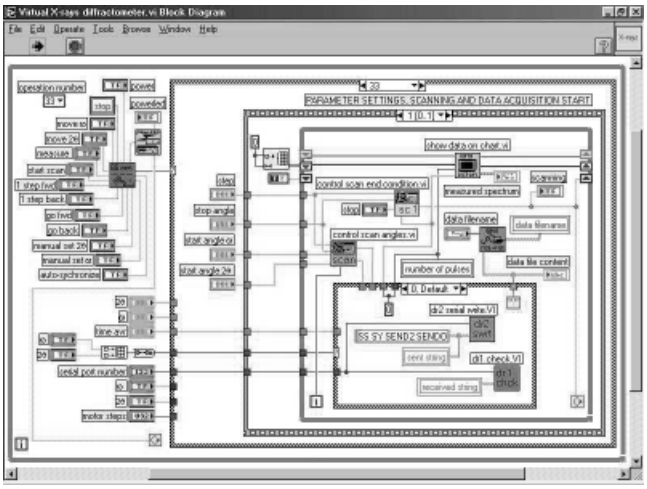


Figure 8. The block diagram of the X-rays diffractometer.

7 COMMUNICATION AND DATA TRANSFER INSIDE AND OUTSIDE VI

7.1 Local variables

A local variable can be used to represent the front panel object in several locations on the block diagram of a single VI. It can also be used to pass data between nodes that are not connected by a wire. It is necessary to use local variables to control several While Loops by using one switch. The local variable can be created from the shortcut menu of the object or its terminal. Figure 9(a) shows how data passes in the VI A using local variables. A local variable always contains the value of its associated terminal.

7.2 Global variables

Global variables can pass data between several VIs. A global variable can be used to represent the front panel object in several locations on the block diagrams of the several VIs. The global variable can be created by dragging the Glob from the All Structures panel and then dropping it on the block diagram. From the shortcut menu of the global variable, the front panel of the special kind of VI has to be opened first and the controls that define the data types accepted by this global variable have to be placed on it. A global variable always contains the value of the selected item placed on the global front panel. Figure 9(b) shows how data passes between the VI B and the VI C using global variables.

7.3 Serial communication

The program developed for communication between serial instruments using Virtual Instrument Software Architecture (VISA) is shown in Figure 9(c). The VISA Configure Serial Port initializes the selected port to the specified settings. VISA Write writes a value in a string to the port. VISA Read reads the number of characters specified by the requested byte count from the port after a timeout period controlled by the Property Node. VISA Close is used to terminate communication.

The sequence of the operations in this program is determined by wiring the output terminals to the input terminals used for addressing in the line of nodes. GPIB (*see also* **Article 170, IEEE-488 Instrumentation Bus, Volume 3**) communication can also be performed using the VISA.

7.4 In/out port communication

Most of the peripheral devices in the PC are controlled using the I/O ports. Figure 9(d) shows the VI built for data loading to counter card using Out Port functions and Sequence structure. The first frame of the Sequence is used

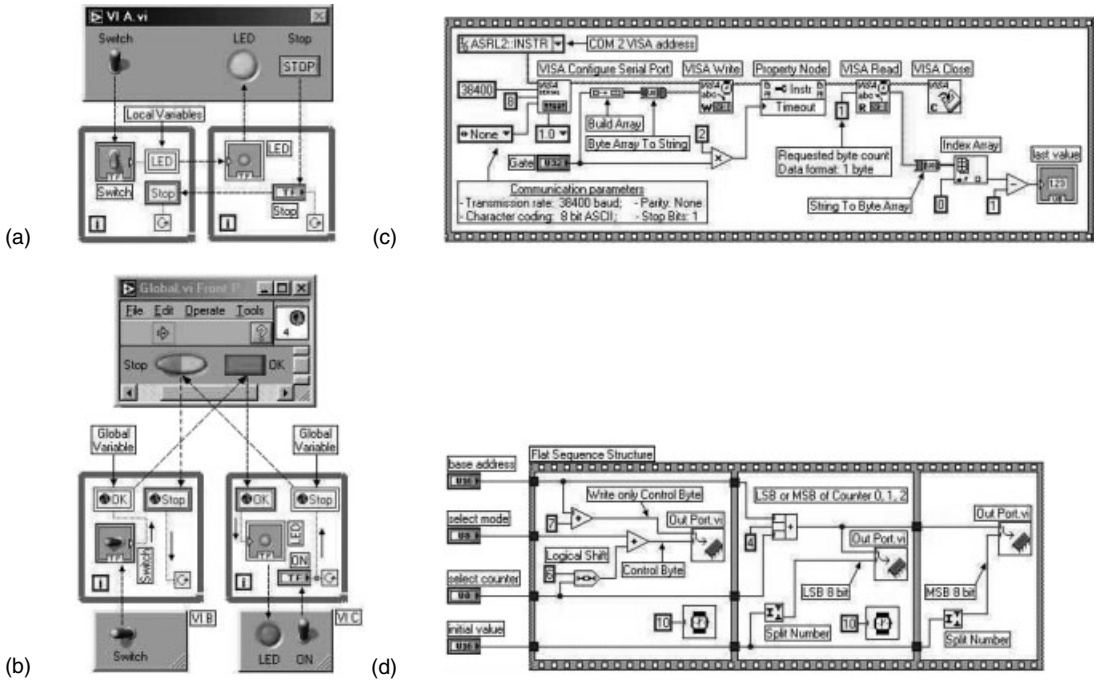


Figure 9. Communication and data transfer inside and outside VI.

for the Control Byte writing to the card through the Out Port before the initial count is written. Then, the initial count in two bytes is loaded to the card, LSB byte followed by the MSB byte.

8 VI BEHAVIOR CUSTOMIZING

After the VI building, a user can customize the VI Menu bar by its editing and the VI behavior using the VI Properties dialog boxes.

8.1 Customizing the menu bar

The custom menu template can be built at the time of editing after opening the Menu Editor window by selecting the Run-Time Menu from the default menu bar. In Figure 10(a), for example, the Menu Editor window for the virtual spectrum analyzer (VSA) is shown (*see also* **Article 106, Virtual Instrumentation in Physics, Volume 2**). On the left-hand side of this window, the menu hierarchy is displayed. The tools needed for the menu hierarchy editing are available from the Menu and Tools bars. After the highlighting of the menu item in the menu hierarchy, the item properties appear on the right-hand side of the window. A user can select the item type, write the item name and tag identifier, and define for it the shortcut key combination. A user item can be enabled/disabled and checked/unchecked by setting the respective attributes. The built menu can be viewed in the Preview area. A tag identifies the custom item in the block diagram from the output of the Get Menu Selection function (Figure 10b). The current VI's Menu bar returns the menu bar refnum of the current VI. The Item Tags are strings and can be used, for example, to select the subVIs or functions placed in the different frames of the Case structure.

The VI documentation, user's manual, and so on, can be selected from the custom menu bar, for example, to instruct a user on how the instruments can be executed and how the measurements can be performed.

8.2 Customizing the VI properties

By popping up the VI icon in the top right of the front panel corner and selecting the VI Properties, a user can open the dialog boxes and turn on/off all the properties available for the current VI.

From the Window Size category, for example, the width and height of the front panel window can be determined. From the Execution category, the dialog box, used to turn on/off the execution options, and the Window Appearance category (Figure 10c), used to turn on/off the window appearance options available for the current VI, can be accessed. The window options apply to VI when in the execution mode.

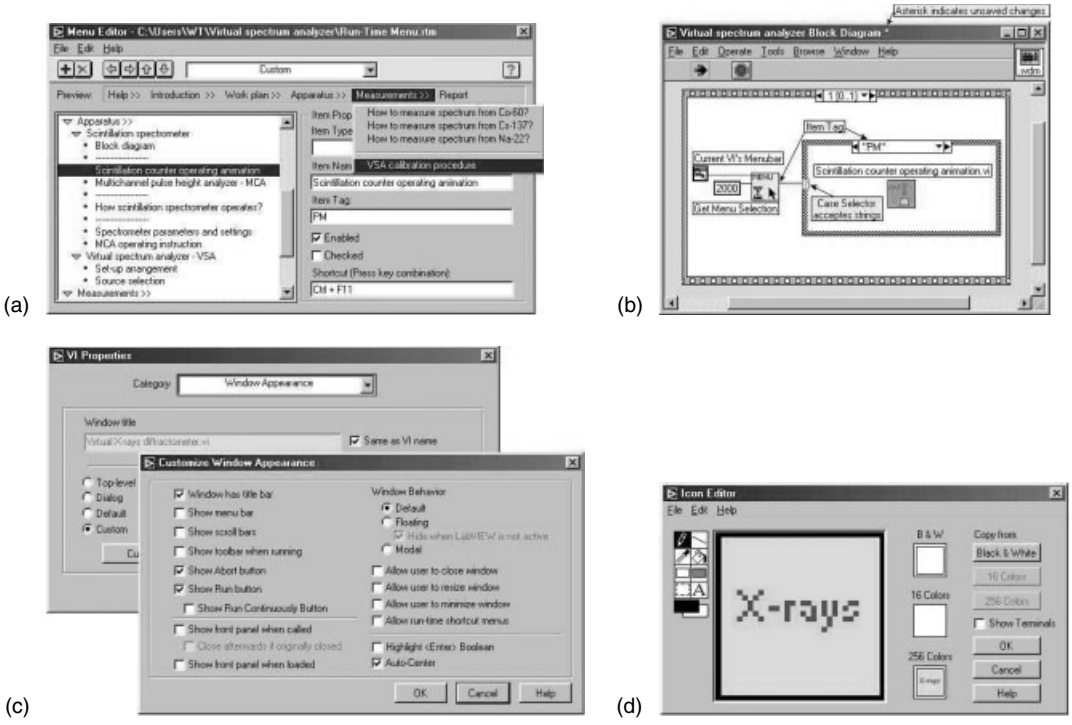


Figure 10. VI customizing windows.

The customizing of VI also needs the icon editing, which can be done in the Icon Editor window (Figure 10d) by accessing the pop up on the VI icon and then selecting the Edit Icon.

9 MODULARITY AND HIERARCHY

VIs are modular in design, so any VI can be used as the top-level program and can run by itself or as a subprogram within other programs. A programmer can divide an application into a set of simple subtasks and then build VIs to accomplish all the subtasks and finally combine those VIs on the top-level block diagram to complete the larger task. The concept of the modular programming, promoted by LabVIEW, gives a user possibility to execute each subprogram by itself, making debugging relatively easy. Furthermore, many low-level subprograms often perform tasks common to different applications and can be used independently by every individual application.

The hierarchy nature is also the significant characteristic of LabVIEW and VIs. From the VI Hierarchy window (Figure 11a), a user can simply find the file path for every subVI, open it selectively by double clicking, and customize individually. There is an unlimited number of layers in the hierarchy structure of every VI. Each subVI can be also opened from the VI Browse available in the menu bar (Figure 11b).

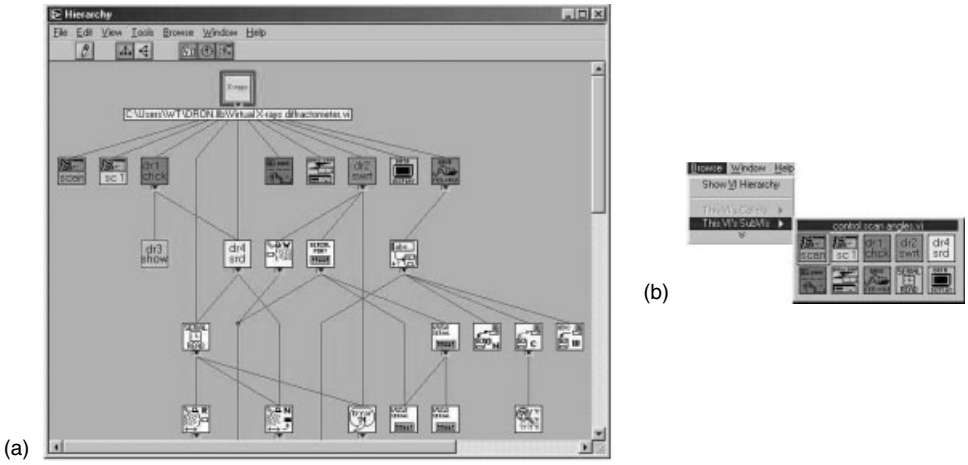


Figure 11. The VI hierarchy window.

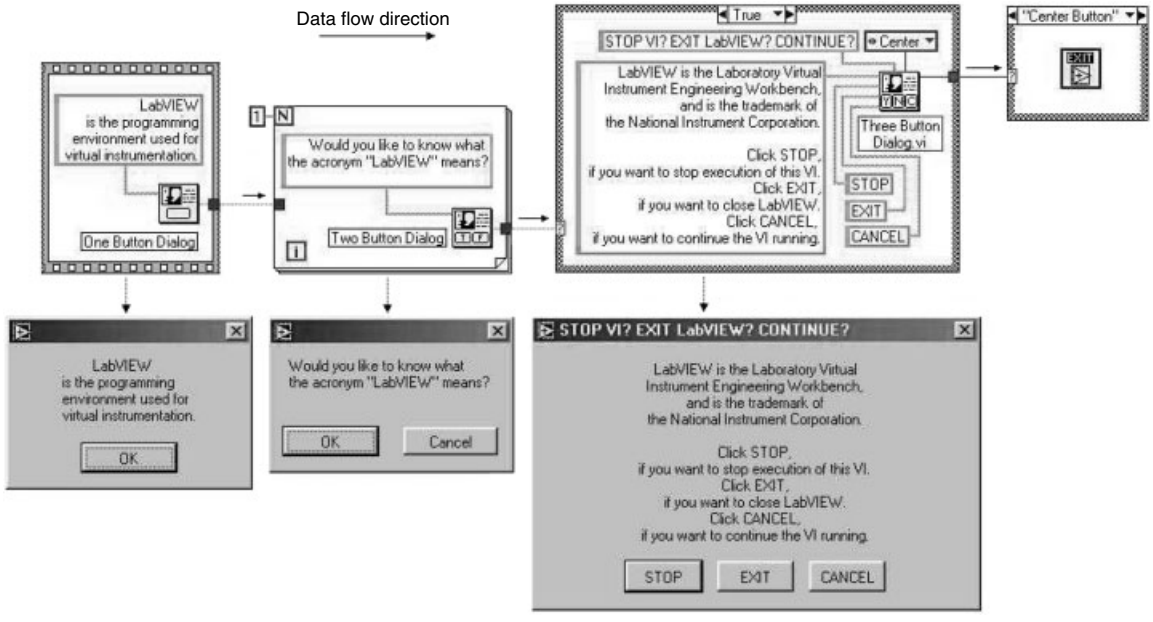


Figure 12. The data flow principle presentation.

10 DATA FLOW PRINCIPLE PRESENTATION

10.1 The Data Flow principle

LabVIEW is a programming environment based on the Data Flow principle. In practice, it means that the nodes execute only when they receive all required input data and produce the output data immediately when they execute. Because the path of the data flow on the block diagram is determined by the wire connection, the wiring can be used to control the VI execution order. Such a solution is applied in the program shown in Figure 12, presenting different types of dialog functions responsible for the messages that appear on the VI front panel in the order determined by wire connections between corresponding structures.

10.2 Application–user communication

For the application–user communication, LabVIEW offers the One/Two/Three Button Dialog functions (Figure 12), which automatically display the messages in the situation where a user needs to be instructed or has to take a decision about what to do next. They are available by selecting the same from the Time & Dialog panel on the Functions palette.

The Button Dialog function displays a dialog box that contains a message addressed to the user and a set of the buttons, one, two, or three, on the VI front panel. The dialog box has a standard form view and stops VI execution until the moment a user pushes selected button. A message should be prepared earlier as a string constant on the block diagram and wired to the corresponding input of the dialog function. Usually, it is placed inside the Case structure and appears when the frame containing it is selected. The button names can be edited. By default, they are Yes, No, or Cancel. The programmer can replace them by wiring the string constants with the needed names to the corresponding dialog function inputs.

The One/Two Button Dialog functions return a value of True or False depending on which button is pushed. This output value can be used to select the following VI behavior.

The Three Button Dialog .vi is a more advanced function and can be edited in a more advanced form. A programmer can use one, two or three buttons for the application–user communication, justify message, place the Window Title on the box, and select the keyboard shortcuts and the window close button, which allow a user to close window without selection. The Yes, No, or Cancel buttons can be edited. The enumerated output of the Three Button Dialog .vi indicates which button was selected.

REFERENCES

Bishop, R.H. (1999) *Learning with LabVIEW*, Addison Wesley Longman, Inc., Menlo Park, CA.

Pociecha, D., Szydłowska, J., Gąsowska, J., Kardaś, D., Mieczkowski, J. and Tłaczała, W. (2002) New Mesogenic Compounds having Fork-like or Cyclic Amide Terminal Groups. *Liquid Crystals*, **29**(5), 663–667.

FURTHER READING

Johnson, G.W. (1997) *LabVIEW Graphical Programming: Practical Applications in Instrumentation and Control*, 2nd edn, McGraw-Hill, New York.

National Instruments (2004) *The Measurement and Automation Catalog 2004*, Austin.

106: Virtual Instrumentation in Physics

Wiesław Tłaczała
Warsaw University of Technology, Warsaw, Poland

1 Nuclear Physics Laboratory for a Student Training with VIs 695
2 Gamma-rays Intensity Analyzer – VI Example 696
3 Data Analysis and Presentation with LabVIEW 696
4 Multiuser Measuring Systems with VIs 697
5 VIs Application to a Teaching Process 698
6 Virtual Spectrum Analyzer (VSA) – VI Example 699
7 Virtual Nuclear Physics Experiment for e-learning 699
Further Reading 701

1 NUCLEAR PHYSICS LABORATORY FOR A STUDENT TRAINING WITH VIs

The Central Physics Laboratory (CPL) is where students of the Warsaw University of Technology study the fundamentals of physics in practice. The Nuclear Physics Laboratory (NPL) is a part of the CPL. The basic equipments used in the NPL are six scintillation spectrometers, for performing basic nuclear physics experiments. Three of them are presented here as examples of the use of LabVIEW™. An introduction to the use of LabVIEW is given in **Article 105, Modeling with LabVIEW™, Volume 2**.

1.1 Determination of the energy of γ-rays

The exercise aims at instructing students about how the scintillation spectrometer works and what typical spectrums of γ-rays look like. Two sources, ^{137}Cs and ^{60}Co, are used to calibrate the spectrometer. The third one, ^{22}Na, is then examined with respect to the energy of γ-photons emitted by it. The front panel of the spectrum analyzer, used in this real experiment to control data receiving, its analysis and presentation, is shown in Figure 1.

1.2 Nuclear decay statistical character

In this exercise, students compare the experimental distributions of emitted particles with theoretical distributions known to be true for a small number of events, which can be described mathematically using the Poisson distribution theory. The front panel of the Poisson distribution analyzer, used in this experiment to control data acquisition, its analysis and presentation, are shown in Figure 2.

1.3 Absorption of γ-rays passing through matter

This exercise helps students to obtain elementary knowledge about the different types of nuclear radiation interactions with matter. The front panel of the Gamma-rays Intensity Analyzer (GIA) used in this experiment to control data acquiring, its analysis and presentation, is shown in Figure 3.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

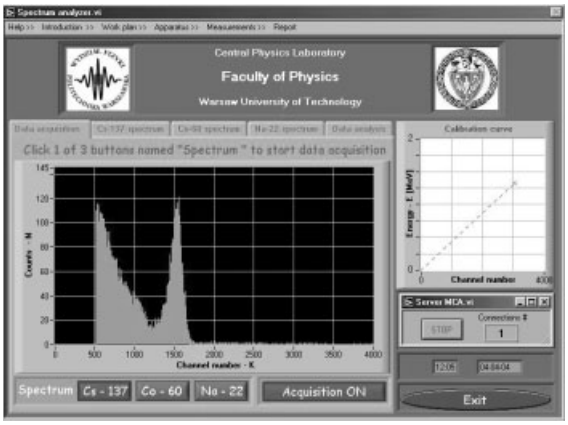


Figure 1. Front panel of the VI spectrum analyzer.

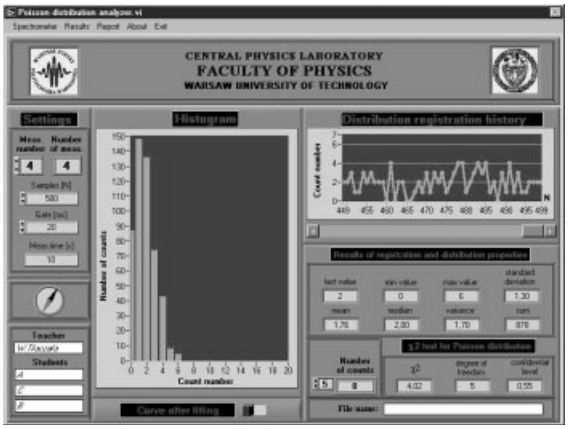


Figure 2. Front panel of the Poisson distribution analyzer.

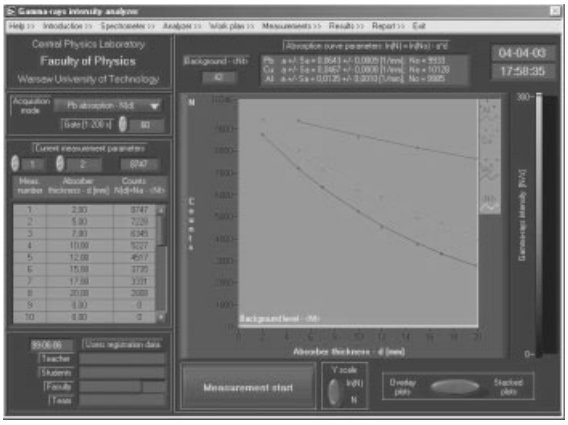


Figure 3. Front panel of the γ-rays intensity analyzer.

2 GAMMA-RAYS INTENSITY ANALYZER – VI EXAMPLE

The absorption coefficient is a quantity, which is characteristic of the absorbing material. A beam of photons passing through matter interacts with it. Students measure the γ-rays with respect to the thickness of Pb, Cu, and Al by using a GIA (Figure 3). The acquired data are then used to examine the attenuation law and calculate the absorption coefficients.

Using the analyzer, a user can obtain the radiation source spectrum, measure the apparatus background or the radiation intensity by selecting the corresponding acquisition mode.

The front panel of the analyzer contains graphical objects used for mode and parameter settings, measurement control, measurement data presentation, and registration data display. The measurement parameters can be set using numeric controls. With the Measurement start button, a user can start/stop measurement. The analyzer takes data from the counter card connected to the scintillation spectrometer. The current Counts number and the average apparatus Background value are displayed in corresponding numeric indicators.

The current gamma intensity value is indicated by the Color Ramp. The set of collected data is visible in the table and presented on the XY graph. The Exponential Fitted Curves are plotted on the same graph. Currently, after every measurement, the absorption coefficients and their errors are computed and presented in the string indicator as the Absorption curve parameters. From customized pull-down menus, the user can open programs, which help perform measurements.

3 DATA ANALYSIS AND PRESENTATION WITH LabVIEW

A beam of photons passed through matter is attenuated in intensity. The fraction of photons surviving a distance, d, is exponential with respect to the thickness. The logarithmic form of the attenuation law is shown in (1)

$$\ln N(d) = \ln N_0 - \mu d \tag{1}$$

where N_0 – the incident intensity; d – thickness of the absorber; μ – absorption coefficient.

Measurement of γ-rays' intensity with respect to the thickness of Pb, Cu, and Al are performed with the GIA. Obtained data are used to examine the attenuation law and calculate the absorption coefficients for different types of absorber using the Least Squares Method (LSM).

The program shown in Figure 4(a) considers the operations used for the curve fitting performed by the GIA.

If the data set is collected as a 2D array, like those presented in Figure 4(b), it is first necessary to transpose it using the Transpose 2D Array function, and then to reshape it into a 1D array with the Reshape Array. Next is the need to extract elements, which represent the data

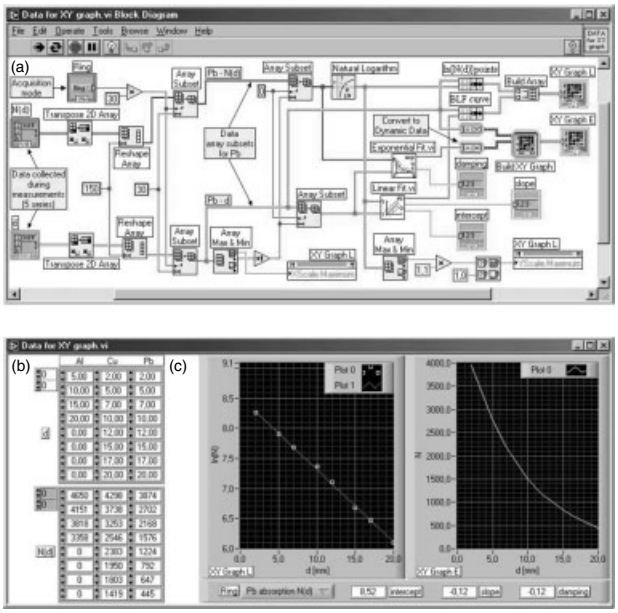


Figure 4. Data analysis and presentation with LabVIEW.

set for a particular curve. For this operation, the Array Subset functions and the Array Min&Max function are used twice. In this way, two arrays are extracted, $[d_i]$ and $[\ln N_i]$, which represent an 8-point data set for γ-rays attenuation for Pb. Obtained this way, the data set is used for the curve fitting with the Exponential Fit. vi. Before fitting with Linear Fit. vi, the data has to be computed with the Natural Logarithm function. After fitting curves to the data, they can be presented on the graphs.

Two methods are used for this purpose. One, the classic method, uses the Bundle functions to prepare two sets of clusters $(d_i, \ln N_i)$, placed on the XY Graph L after building an array with them.

The second method, Express VI method, uses the Build XY Graph subVI, which needs only to convert arrays with the Dynamic Data Converters before wiring them to the corresponding inputs. This method was used for data preparation before placing them on the XY Graph E. The Linear Fit. vi also finds the slope and intercept coefficients, and the Exponential Fit. vi finds the damping coefficient. The slope gives the gamma absorption coefficient. The Array Min&Max and X/YScale. Maximum property nodes are used for X/Y Axis scaling. In Figure 4(c), two XY graphs are shown with data points and fitted curves, and numeric indicators that display the curves coefficients.

4 MULTIUSER MEASURING SYSTEMS WITH VIs

With virtual instruments it is easy to build relatively low cost multiuser measuring systems. This is particularly important when multiple measuring systems are needed in the laboratory for training students. Scintillation spectrometers are the base parts of the instrumentation used for data taking in the NPL.

The analog signal from the detector is shaped by the amplifier and then sent to the pulse height analyzer devices, which are of different types in different experiments.

4.1 Multiuser measuring systems

The multiuser measuring systems used in NPL are presented in Figure 5. In two measuring systems, threshold discriminators are used for pulse height selection. The measuring systems used for the statistical distribution, Figure 5(a), and gamma absorption investigations, Figure 5(c), need data only in the logical form. The discriminators deliver a standard logic signal for every analog signal with its amplitude higher than a given threshold. In each of the presented multiuser system, the discriminator output signals are distributed with fan-out circuits to four computers.

In the experiments shown in Figure 5(a), the 16-bit counters, controlled by the microcontrollers, (*see also* **Article 158, Microcontroller Systems, Volume 3**) are used for data collection. The counter cards with the microcontrollers are serial instruments and communicate with the VI through RS 232 (*see also* **Article 168, RS 232 and EIA/TIA 232 Serial Interface, Volume 3**) port. In this measuring system, one spectrometer delivers signals continuously to four computers, but users perform measurements independently.

In the experiments shown in Figure 5(c), counter cards are used for data collection. Digital I/O ports control the counter cards by using the VI developed for gamma absorption investigations. In this measuring system, one spectrometer supplies signals continuously to four computers, but users are asked to perform measurements synchronously.

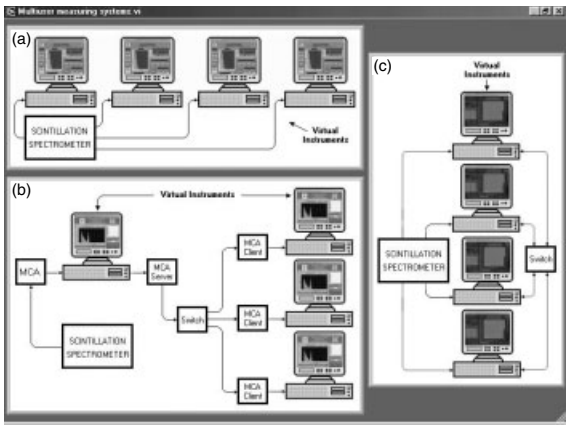


Figure 5. Multiuser measuring systems with VIs.

The local area network (LAN) (*see also* **Article 171, Local Area (LANs) and Wide Area Networks (WANs), Volume 3**) is used for message sending that allows synchronized measurements to be performed with computers working in parallel. In these experiments, users can also obtain a radiation source's spectra. In this case, the threshold discriminator is replaced with a single-channel analyzer.

Figure 5(b) shows the multiuser measuring system using a LAN for data distribution. In this measuring system, only one computer is equipped with the multichannel analyzer (MCA) used to take spectra. Data acquired by the MCA are distributed by the server-client software combination, which is a communication option available in LabVIEW.

Remote computers receive online data as soon as the MCA sends it to the MCA server, which then distributes it to the MCA clients. Data acquisition is synchronized by the MCA, but remote users can still analyze the spectra independently.

5 VIs APPLICATION TO A TEACHING PROCESS

LabVIEW can be used not only for VI building of real-world measurements but also for building applications that simulate real-world instrument operation, useful in the teaching process, for example, teaching the fundamentals of physics or to demonstrate experiments in physics.

5.1 Planck's Law and Wien's Displacement Law

Figure 6 shows how LabVIEW application can be used for demonstration of the Planck's Law and the Wien's Displacement Law. When the temperature of the object is known, the distribution of the radiating energy is given by the law of black body radiation, the Planck's Law, as shown in (2)

$$M(\lambda, T) = \frac{b}{\lambda^5} \frac{1}{e^{hc/\lambda kT} - 1}, \tag{2}$$

where $b = 10^7$ in relative units; λ–wavelength; h–Planck constant; k–Boltzmann constant; c–speed of light in vacuum; T–temperature in Kelvin degrees.

The block diagram, built to place on the graph the radiation emission curves for objects at different temperatures, is shown in Figure 6(b). The relative radiation intensity as a function of the wavelength and temperature is plotted by the implementation of (2), using the Formula Node structure.

The Formula Node is similar to other structures available in LabVIEW (*see also* **Article 105, Modeling with LabVIEW™, Volume 2**), but instead of containing a subdiagram, it contains one or more formula statements delimited by a semicolon. Using the Formula Node, a user can enter algebraic formulas directly into the block diagram. The input and output terminals of the Formula Node can be created by selecting it from the shortcut menu. After creating terminals, a user has to enter variable names into the input/output boxes.

The data for one point is calculated in one iteration of For Loop 1 using the Formula Node. The data set for one curve is collected in the tunnel with an array symbol, during calculations performed in 10 000 iterations. After For Loop 1 completes all iterations, the data collected in the tunnel are provided to the XY graph using the Bundle function, which also determines the scaling factors, such an offset – λo and a multiplier – $\Delta\lambda$. Using For Loop 2, a user can observe one by one the set of 31 curves at the temperatures in the range 200 to 6200 K. Also, it is possible to observe how the peak of the black body spectral distribution shifts to the shorter wavelengths as the temperature increases. This relationship is described by the Wien's Displacement Law,

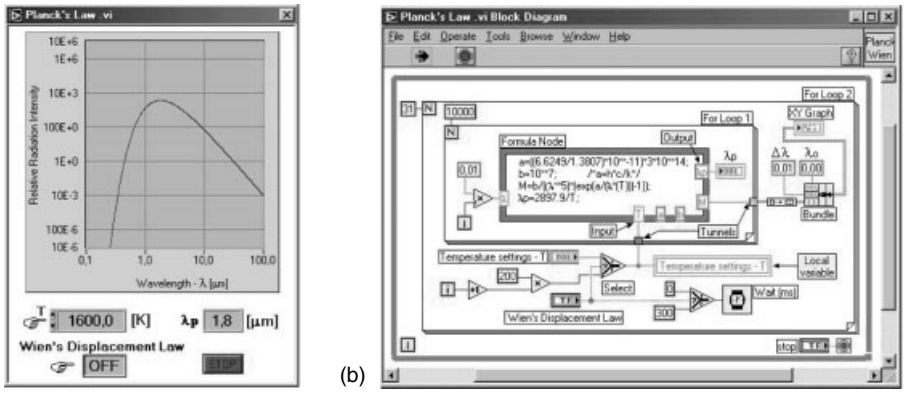


Figure 6. A model of Planck's Law and Wien's Displacement Law.

as shown in (3)

$$\lambda_p T = 2897, 9, \tag{3}$$

where λ_p – wavelength of peak energy in microns; T – temperature in Kelvin degrees.

A user can select a static or dynamic presentation by switching on/off the Wien's Displacement Law button.

Tunnels are terminals on a structure used to get data into the structure and to pass data out of the structure. Tunnels created on the For Loop or While Loop border can be used with the Enabled Indexing for arrays or the Disabled Indexing for scalars.

6 VIRTUAL SPECTRUM ANALYZER (VSA) – VI EXAMPLE

6.1 The virtual spectrum analyzer (VSA)

The Virtual Spectrum Analyzer (VSA) front panel is visible in the background of the window, with the signal converting simulation opened from the VSA menu bar (Figure 7). The buttons, switches, selectors, graphs, and so on are used to select the type of spectrum to acquire, save and load data, calibrate the spectrometer, start and stop data acquisition, present the spectrums, and finally present the calibration curve. The complete VSA front panel is presented in Figure 9.

The VSA not only helps a user to conduct an experiment but also gives the full information concerning it, for example, elementary knowledge about the γ-rays interactions with matter, how to turn on electronics, set correct parameters and run experiments, how to collect and analyze data then verify obtained result, and also how to prepare a laboratory report. This information can be selected from the VSA menu bar. A user can also open animation programs developed for demonstrating how scintillation spectrometer and electronic devices, like a scintillation detector, an amplifier, and the MCA (Figure 7), operate.

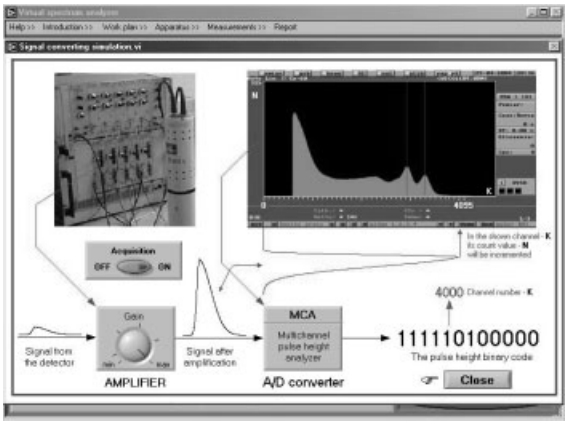


Figure 7. LabVIEW application for the spectrum building animation.

With the VSA a user can virtually measure spectrums of the γ-rays source, calibrate the spectrometer and finally, determine energy of the photons emitted by ^{22}Na. The data collected in the virtual experiment are similar to the data acquired in the real physics laboratory.

6.2 Virtual physics experiments

Several virtual physics experiments were built, using LabVIEW, in the Warsaw University of Technology (WUT) to demonstrate physics phenomena or the methodology of the measurements. There are nuclear physics experiments, thermodynamics experiments based on the Ångström Law, experiments for the Hall effect and Curie temperature investigations, and electric resonance measurements. These experiments can be run through quickly during a lecture or in the 'real' time mode in a virtual physics laboratory for training students. One of these experiments is devoted to determination of γ-rays energy. This experiment, likewise other experiments, has its prototype in the real experiment existing in the CPL (see Figure 1) and simulates it.

6.3 Determination of γ-rays energy – a virtual nuclear physics experiment

The experiment aims at instructing students as to how the basic nuclear apparatus works, how the typical spectrum of γ-rays looks, and how to perform measurements to determine energy of the unknown γ-photons. Two sources, ^{137}Cs and ^{60}Co, are used to obtain calibration curve of the scintillation spectrometer. The third source, ^{22}Na, is then examined with respect to the energy of γ-photons emitted by it.

7 VIRTUAL NUCLEAR PHYSICS EXPERIMENT FOR E-LEARNING

After switching on the apparatus, the parameters setting (Figure 8(a)), and the source selection (Figure 8(b)), a user can start to perform the virtual experiment. The experiment for energy determination needs at first the spectrometer to be calibrated. It can be done to measure the energy spectra of the sources with emissions of the known energies.

7.1 Spectrum acquisition and presentation

The calibrating sources used in the virtual experiment are ^{137}Cs and ^{60}Co. To perform the measurement for obtaining

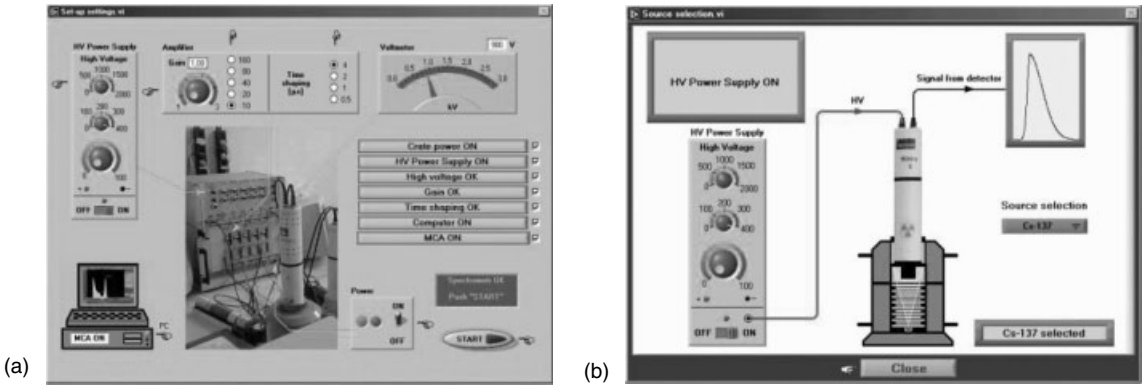


Figure 8. LabVIEW applications for the experimental setup arrangement and the source selection.

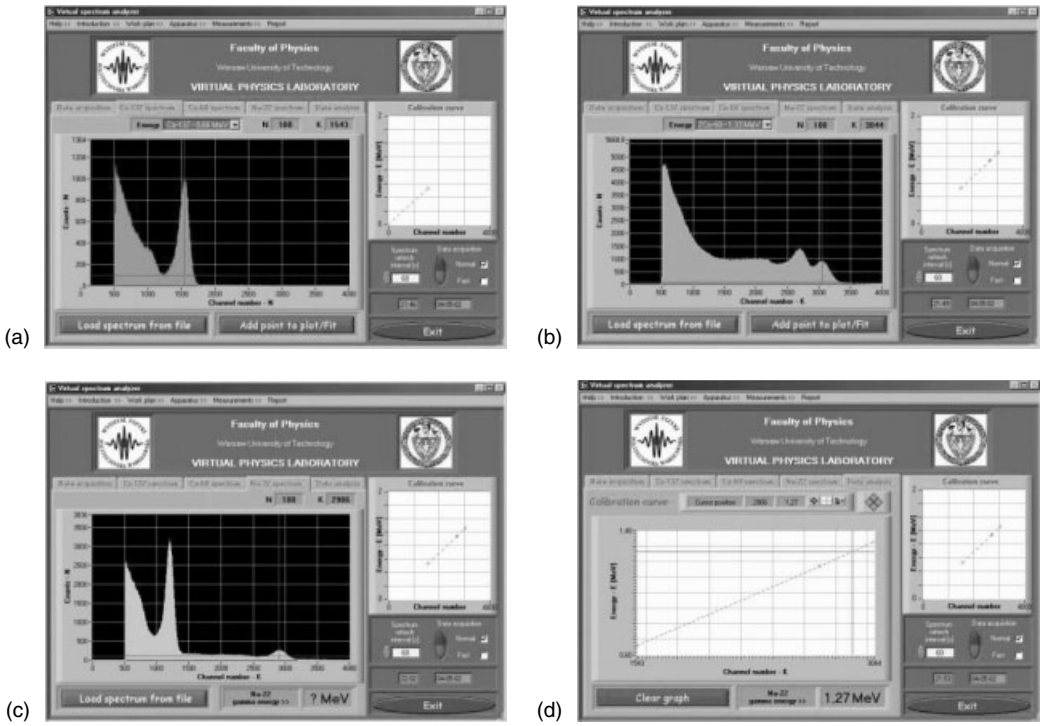


Figure 9. Spectra for the VSA calibration and the ^{22}Na gamma energy determination.

the ^{137}Cs spectrum, a user has to select the Data acquisition mode, the Normal or Fast mode on the VSA front panel, then select the Data acquisition window, and finally push the Cs-137 button. After some time, depending on the acquisition mode to be selected the ^{37}Cs spectrum will be obtained.

7.2 VSA calibrating and data analysis

After finishing data acquisition, performed with two calibrating sources, a user can calibrate the spectrometer by determining the calibration line. For this, a user has to identify the peaks with the aid of the isotope tables, or by selecting the needed values in the Ring control labeled Energy, which is visible on the VSA front panel when the windows with spectrums of ^{137}Cs and ^{60}Co are selected. The spectra obtained during the experiment are collected in Figure 9(a–c).

After selecting the channel number of the characteristic peak on the calibrating source spectrum with the help of the graph cursor, and selecting the corresponding energy value in the Ring control, placed above the graph, the one point can be plotted on the calibration curve graph, using the Add point to plot/Fit button. A correspondence between the MCA channel number and energy can be made this

way. With the ^{137}Cs and ^{60}Co spectrums, three such points are plotted on the graph. The VSA automatically draws the calibration curve by taking the best fitting straight line through the points. The energy scale is indicated on the left-hand side, but the MCA channel numbers are indicated on the X-Axis of the graph shown after opening the Data analysis window.

After taking the ^{22}Na spectrum (Figure 9(c)) and selecting a corresponding peak, the energy value of the γ-rays emitted by ^{22}Na can be read by using the calibration curve available in the Data analysis window.

The data collected in the virtual experiment are not simulated. They are similar to data obtained in a real physics laboratory because they are based on the real measurement results, which have been incorporated in the software when the application was built.

FURTHER READING

Bishop, R.H. (1999) *Learning with LabVIEW*, Addison Wesley Longman, Inc., Menlo Park, CA.

Johnson, G.W. (1997) *LabVIEW Graphical Programming: Practical Applications in Instrumentation and Control*, 2nd edn, McGraw-Hill, New York.

National Instruments (2004) *The Measurement and Automation Catalog 2004*, Austin.

Tłaczała, W. and Zagórski, A. (2003) A virtual nuclear physics laboratory for e-learning, *Advances in Technology-Based Education: Toward a Knowledge-Based Society*, *Proceedings of 2nd International Conference on Multimedia and Information & Communication Technologies in Education*, Badajoz, Spain, December 3–6th.

107: Principles of Sensor Science

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 The Nature and Scope of Sensor Science 707
2 The Central Role of Systems Engineering 708
3 Recursion and Reticulation Principles 709
4 Importance of Classification in Sensor Science 709
5 Sensor Systems as Information Machines for Measurement 710
6 Holistic Design 711
7 Maintenance Engineering 711
8 Principles of Sensor Modeling 712
9 Parametric and Nonparametric Models 713
10 Materials Science for Sensors and Sensing 714
Related Articles 714
References 714

1 THE NATURE AND SCOPE OF SENSOR SCIENCE

Sensors and transducers are manufactured systems (McGhee, Henderson and Sydenham, 1999), which extend the human abilities to sense the two highly relevant problems of measurement described in **Article 5, Humans in the Real World, Volume 1**. Hence, since they are the front-end elements in Instrumentation and Measurement Technology, or I&MT, they are of basic importance in all fields of science, industry, and commerce. I&MT, which may also be referred to as *Information Technology*, or IT, is that field of scientific activity where humans use machines to extend their ability to perform the information handling operations associated with measurement, calculation, communication, and control (**Article 14, Hierarchical Aspects of Measurement Systems, Volume 1**). It is also concerned with the design, analysis and utilization of these machines.

A transducer is a device that transforms the input signal of one energy form into an output of another energy form. It does this in such a way that a prescribed relationship is maintained between the input and output signals. An example is the mercury-in-glass thermometer in which heat energy is converted into mechanical energy, forcing the mercury to move along the glass as it expands or contracts.

To provide a wide view of sensing and transduction, it is important to examine what constitutes the science of sensors. Sensor Science is closely akin to Measurement Science (Bentley, 2004; Finkelstein and Grattan, 1994; Sydenham, 1982, 1983; Sydenham and Thorn, 1992; Sydenham, Hancock and Thorn, 1992).

This article focuses upon sensors, sensing and sensing systems. The specific goal of the article is to analyze the elements of sensor science. An analytical approach leads to proposals that specify the constituent parts defining the ordered study of sensors.

The extensive literature on sensors is the principal tool of this analysis. The article shows that Sensor Science has a sufficiently distinct and noteworthy constitution, which demands both greater emphasis by measurement scientists and more recognition in the wider scientific community.

Calling the discipline under consideration Sensor Science may be controversial. However, there has been notable success in developing specific types of sensors and sensor systems as exemplified by the authoritative 8-volume treatise edited by Göpel, Hesse and Zemel (1989), with its subsequent update (Baltes, Göpel and Hesse, 1996), and the handbook of Fraden (1996). In addition, the excellent texts by Hauptmann (1993), Sinclair (2000) and Usher and Keating (1996), have emphasized the importance of organizing the basic material of sensors. This need is also strongly supported by other books grouped according to

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

the COMETMAN energy form classification due to Stein, which is described in some detail in **Article 109, Structure and Energy in Sensor Systems, Volume 2**. Good examples of books according to this grouping are as follows:

- For Chemical sensors (Göpel *et al*., 1989; Janata, 1989; Taylor, 1996), including gas (Maskell, 1996; Moseley and Tofield, 1987), immunological (Ngo, 1987), pharmaceutical (Cosofret and Buck, 1992) and food industry units (Kress-Rogers, 1993; Karube, 1994);
- For Optical sensors (Wagner, Dendliker and Spenner, 1994; Agousti, 1996; Wolfbeis, 1991; Krohn, 1992; Dadswell and Dean, 1995);
- For Mechanical sensors (Bau, de Rooij and Kloeck, 1994) such as in manufacturing (Soloman, 1994), robotics (Everett, 1995; Ruocco, 1987), robotics and medicine (Webster, 1988) as well as in mechatronics (Bradley *et al*., 1991);
- For Electrical sensors, Chapter 15 in Fraden (1996);
- For Thermal sensors (McGee, 1988; Michalski *et al*., 2001; Nicholas and White, 2001; Scholz and Ricolfi, 1990);
- For Magnetic sensors (Boll and Overshot, 1990; Maenaka, 1994);
- For Acoustic sensors (Crocker, 1998; Lerch, 1994; Lynnworth, 1994);
- For Nuclear sensors, Chapter 14 in Fraden (1996).

In addition to the classification of sensors by the energy form, alternative schemes are grouped by the sensor manufacturing technology. Examples using this approach are solid state and semiconductor sensors (Gardner, 1994; Middelhoek and Audet, 1989; Sze, 1994) and their on-chip compatibility (Wolffenbuttel, 1996) as well as thin film type sensors (Ciureanu and Middelhoek, 1992).

Mention also needs to be made of smart sensors (Barney, 1988; Chapman, 1995; Frank, 1996) as well as the associated conditioning and support circuits (Carr, 1993; De Sa, 1990; Pallas-Areny and Webster, 1991; Regtien, 1992; Van Putten, 1996) – *see* **Article 160, Smart Sensor System Features, Volume 3**.

Complementing the multitude of published books about sensors, there are also a large number of technical articles that have contributed to the overall understanding of sensors. Journals where these articles appear, are *Measurement; Sensors and Actuator – A Physical, Sensors and Actuators: B Chemical; IEEE Trans on Instrumentation; Instrument Science and Technology; and Measurement* and *Sensor Review* to name only five.

Achievements in the area of classification of sensors, which can be traced back to the work of Kurt Lion (1969), have been built upon by Peter Stein (1969) and more recently by many other researchers. Significant progress has been made possible by the cataloging and classification of physical material effects (Ballentyne and Lovett, 1980; Schubert, 1984), by sensor modeling (Finkelstein and Watts, 1983; Neubert, 1973), sensor design (Finkelstein and Finkelstein, 1983; Cook, 1993a,b; Sydenham, 1994a,b; Finkelstein *et al*., 1993), system reliability (Cluley, 1974; Hoyland, 1994; O'Connor, 2002; Ramakumar, 1993; Lyn, 1996) and fault diagnosis (Patton, Frank and Clark, 1989), all of which represent a selection of these contributions.

In spite of this progress, the constituent body of coherent principles of the discipline lack the organization necessary before that group of principles may be referred to as a science.

The purpose of this article is to address this question. By doing so, proposals are made as to what constitutes the science of sensors and giving an approach to its systematic organization is achieved.

As shown above, there is substantial literature dealing with all aspects of sensors, sensing, and sensor systems. Included in this area are those given specific reference above and at different points in this article.

2 THE CENTRAL ROLE OF SYSTEMS ENGINEERING

Many different methods and theories are used in Sensor Science. The bubble and block diagram of Figure 1 is a good way to visualize some of these Sensor Science disciplines. In the search for these contributory disciplines an important key is found in the definition that sensors are *systems*, which refine and extend the human faculties of *sensing* and perception. Because sensors are systems, the science of sensors must be holistic by always using the whole-life-whole-system approach characterizing the methods of Systems Engineering described in **Article 63, Outline of Systems Thinking, Volume 2**; **Article 64, Executing A Measuring System Design, Volume 2**; **Article 66, Phases of System Life Cycle, Volume 2**; **Article 65, Life Cycle Concept, Volume 2**; **Article 67, Principle of Concept of Operations (ConOps), Volume 2**; **Article 68, Setting the System Boundaries, Volume 2**; and **Article 69, Requirements Allocation, Volume 2** (Sandquist, 1985; Weinberg, 1975).

Thus, sensors and sensor systems should be regarded as information machines. This class of machine performs a diversity of sensing functions involving the holistic relations among specific sensor structures within different energy domains for the acquisition, capture, communication, or distribution of information.

The information, which is acquired about the states of equilibrium and motion of solids, liquids, gases, and

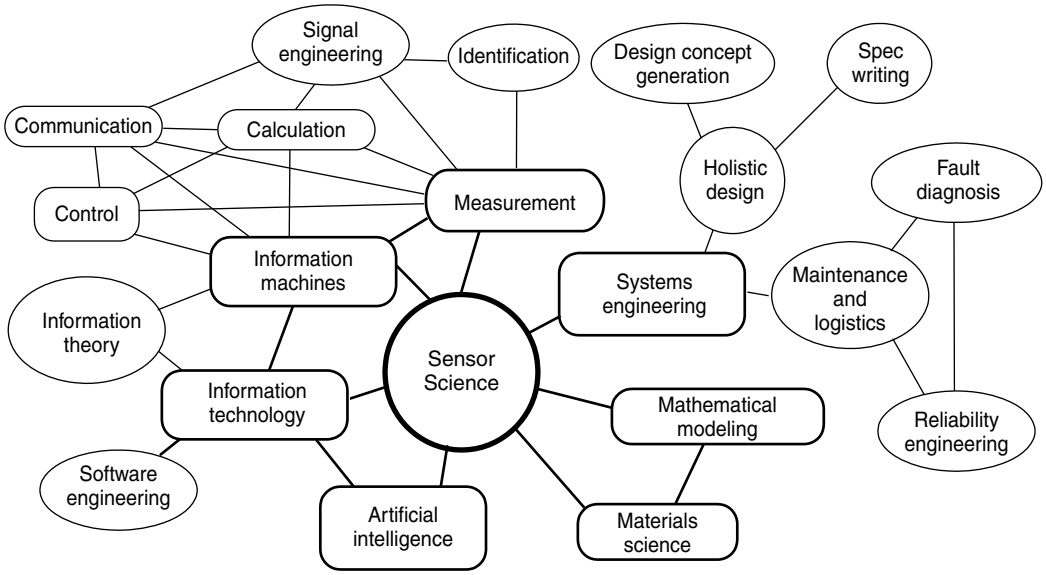


Figure 1. A bubble and block diagram illustrating the constituent disciplines of Sensor Science.

their constituent systems, may be carried or captured in a variety of signal forms. This statement is similar to the definition given in **Article 2, Organization of Instrument Science, Volume 1**. As both function, or 'what they do', and structure, or 'how they do', are specified in these definitions, the importance of considering sensors as embedded within sensor systems is reinforced. In other words, sensors are human-made elements embedded within human–machine systems, which help humans to acquire information, by the process of sensing, and to handle data, by performing information handling operations.

Sensor systems possess physical, also called *organic*, structure. The systems approach to their specification and design is basic to Sensor Science.

3 RECURSION AND RETICULATION PRINCIPLES

Important tools for use in Sensor Science, taken from the Systems Engineering approach, are the techniques of reticulation and recursion. Reticulation is that approach to the analysis of systems, which allows them to be successively subdivided down to the simplest element.

Recursion is that property of a procedure that allows operations to be performed in a nested manner using the same basic method of analysis. By this means, all of the subdivided portions or elements of the sensor system may themselves be individually analyzed using the approach of Systems Engineering. It is the systems aspect of sensors, with its holistic implications, which provides the key to the nature and scope of their science.

4 IMPORTANCE OF CLASSIFICATION IN SENSOR SCIENCE

A science is a systematically organized body of knowledge. Hence, since Sensor Science possesses a corporate body of knowledge with an organized structure, it is as necessary to organize it as it is to organize other scientific fields. The nature and scope of the constitution of Sensor Science may be objectively examined using the methods of taxonomy. When applied in Sensor Science, taxonomy requires the identification of that series that depends upon the increase or decrease of some variable in the scheme of ordering.

A taxonomy of Sensor Science will be erroneous if it is based upon its ends, as this will only lead to a cataloging of sensors. Hence, the only legitimate scheme of sensor classification will ensure that all of the divisions of Sensor Science are always determined by one common principle. Indeed, only by organizing the constitution of the topic on the basis of contributory disciplines can Sensor Science be arranged according to its basic nature and inherent characteristics. This will ensure that its arrangement is only based upon its basic nature and inherent characteristics. It will also make sure that its organization is not based upon either its ends or upon anything lying outside itself. In other words, applying classification in Sensor Systems is not about the sum of their ends but about coordinating the science of sensors so as to give it an organized or systematized structure. The systemic nature of sensors leads to the holistic approach in their classification as shown above.

By considering the nature and scope of the disciplines constituting the taxonomy, analysis, design, and utilization of sensors and sensor systems, this article shows that Sensor

Science is important in its own right. Hence, it should be given the attention it deserves by all of the scientific community, despite the close relationship it shares with other aspects of Measurement Science within IT.

The Four Orders or Problems of Classical Taxonomy are considered in **Article 2, Organization of Instrument Science, Volume 1**. An understanding of Sensor Science requires the application of the principles of classical taxonomy.

Previous work has well established the application of these principles within the context of I&MT. Sensor systems, as the front-end elements in information machines, form one of the two divisions of the machine kingdom. These require an identifiable body of scientific principles, which depend upon the holistic relations among specific but diverse sensors and sensor systems. Classical taxonomy has always been used in the life sciences to bring order into the questions of likeness between different living entities (Daly and Linsley, 1970). It is also important in the earth sciences (Von Engelhardt and Zimmermann; 1988) for the classification of minerals, in technology transfer, and in Knowledge Engineering or KE.

Bringing order to Sensor Science recognizes the universality of the four principal orders or problems of classification **Article 2, Organization of Instrument Science, Volume 1**. The classification task in Sensor Science depends on discriminational criteria using reticulation, which requires the whole system to be 'broken down' by a process of subdivision.

These criteria are based upon two main approaches. In the life sciences, ordering may be built around evolutionary, or phyletic, analysis.

Phenetic analysis, based upon physical appearance or structure, is also widely used in bioscience classification. In the theory for classifying sensors and sensor systems, the same names may be adopted. Hence, the phyletic and phenetic approaches are also appropriate for machine kingdom grouping.

Phyletic analysis in the life sciences is based upon genetic considerations. Functional analysis, which is similar to evolutionary/propagative analysis, concentrates upon the functions performed by physical systems in the machine kingdom. Hence, analyzing the function of a machine may be called *phyletic analysis*. For the same reason, structural analysis, performed using methods analogous to the phenetic criteria of bioscience taxonomy, may also be called *phenetic* in machine kingdom classification. These represent some of the necessary modifications to the theory of bioscience taxonomy before it can be applied to instrumentation.

A taxonomy has three objectives and three functions. These are also relevant to Sensor Science. Thus, the three objectives of sensor classification are

1. the concrete discrimination between different sensors,
2. the consensus regarding standards for the principles of description,
3. the bringing of order or systematization to sensor systems.

Similarly, the three functions of classification should allow

1. the organization of the means of communication and retrieval of the descriptions used for sensors,
2. the acquisition of new information in the extension of sensor descriptions,
3. the highlighting of unifying factors between sensors without diminishing the importance of any existing differences.

The materials of taxonomy in Sensor Science are the diverse types of sensors and their operating principles. Assembling the various sensor types is the main activity of classification in Sensor Science, because it allows the possibility for further study. Assembling sensors into progressively larger groups to obtain a hierarchical ordering by function, structure, and energy form constitutes the final ingredients of discrimination and ordering in Sensor Science.

5 SENSOR SYSTEMS AS INFORMATION MACHINES FOR MEASUREMENT

Sensors are important elements belonging to the class of Information Machines, which is made up of machines for measurement, calculation, communication, and control – *see* **Article 14, Hierarchical Aspects of Measurement Systems, Volume 1**.

Consequently, the underlying theory of information handling is also relevant to Sensor Science. This theory indicates that an economical, possibly minimum, number of alphabet symbols should be coded as messages for carrying or capturing information.

This distinction between information carrying and information capturing is important. While information carrying (i.e. communication) machines perform the primary function of high fidelity information transmission, information capturing (i.e. measuring) machines perform the primary function of high precision information capture. As is stated in **Article 5, Humans in the Real World, Volume 1**, precision measurement may be concerned with either direct or inferential measurement.

Both of these problems of measurement require the use of sensors. In the sensing role, the flow of information and energy, characterizing the states of equilibrium and motion of solids, liquids, and gases, are captured. Identification,

essentially a problem in precision measurement, captures information about those features of a system's ability to store, transform, transmit, or dissipate energy and/or information by a process of interrogation.

This discussion indicates the important contribution made by signal design (Godfrey, 1993; Henderson, McGhee and El Fandi, 1997) and signal processing (Bellanger, 1989; Ludeman, 1987) in Sensor Science. The interrelationships between these contributors to Sensor Science are shown in Figure 1.

6 HOLISTIC DESIGN

Sensing in real environments can only be performed with real structures. These sensor forms and the systems within which they are embedded need to be purposefully designed. It could be argued that, like Measurement Science, Sensor Science should be considered within the context of design rather than solely analysis.

This valid point of view deserves widespread support. However, another point of view concerning design needs to be expressed. It is highly unlikely that all measurement technologists will find themselves involved in the actual design of sensors. It is much more likely that they will be involved in including existing sensors into new designs of sensing systems. Consequently, design would seem to be better addressed within the context of measurement system design.

Effective design methodologies in Sensor Science require a combination of design processes, which involve a number of phases (Finkelstein and Finkelstein, 1983). These include (i) a feasibility assessment, (ii) a specification formulation, which may use KE in the same way as (iii) design concept generation. Further knowledge on how systems engineering impacts engineering detail design is found in Sydenham (2003).

7 MAINTENANCE ENGINEERING

Maintenance engineering, an important branch of all engineering activity, is also essential for the efficient day-to-day activities in Sensor Science where it is now receiving the attention it deserves. Three important aspects of this technology are

- hardware (Cluley, 1974; Hoyland, 1994; Lyn, 1996; O'Connor, 2002; Ramakumar, 1993) and software reliability engineering (Henderson *et al*., 1991; Lyn, 1996);
- fault diagnosis and prediction (Henry and Clarke, 1993; McGhee, Henderson and Baird, 1997; Patton, Frank and Clark, 1989);
- signal processing (Bellanger, 1989; Ludeman, 1987).

When equipment is referred to as *faulty* there is an implication that some abnormality in the operating conditions exists (McGhee, Henderson and Baird, 1997).

A more general definition of faults is that there is a substantial degradation in system performance. This may be due to gradual (graceful degradation) or abrupt changes (catastrophic failure) in the parameters of some system or process parameter or malfunction of equipment causing uncertainties in measured values.

The most common way of implementing such a philosophy is by using some scheme of maintenance or plan for maintenance, which will replace 'worn out' parts before they malfunction.

A second way of preventing malfunction of a process is to provide some monitoring scheme.

Thus, it appears that fault detection, fault location, fault diagnosis and fault correction are the principal functions constituting a monitoring scheme.

In short, such a monitoring scheme may be called *Fault Detection and Isolation*, or (FDI). For fault diagnosis purposes, the system has three subsystems. These consist of the actuators, the main structure of the plant or process and the instruments or information machines making up the technical instrument subsystem. For this reason, the three FDI schemes are

1. Actuator Fault Detection, or AFD, scheme,
2. the Process Fault Detection, or PFD, scheme,
3. Instrument Fault Detection, or IFD, scheme.

Sensor Fault Detection, or SFD, constitutes an important part of IFD. In the case of SFD, the field of both sensor validation (Henry and Clarke, 1993) and *in situ* testing are the focus of some interest (Kerlin, Hashemian and Petersen, 1982; Michalski *et al*., 1993; McGlone, McGhee and Henderson, 1998).

Although many diagnostic systems are in use, they only provide testing for most of the sensing conditioners up to the point where the sensor signal is available. At the moment, it is unusual to find sensor system diagnostics where the sensor is one of the system elements to have both its static and dynamic behavior tested. In the search for more reliable systems, the importance of whole-system-each-element-in-situ-testing will almost certainly lead to more sophisticated fault diagnosis and prediction. Detection and prediction of faults uses similar processes as found in system safety assessment – *see* **Article 99, Safety Organization, Volume 2**; **Article 100, Safety Analysis Methods, Volume 2**; and **Article 101, Assessing and Demonstrating Safety, Volume 2**.

8 PRINCIPLES OF SENSOR MODELING

Since sensing essentially ascribes numbers to physical variables in such a way as to describe both them and the relations between them, it is necessary to develop quantitative methods to describe these numerical assignations. Although sensors must be considered within the wider context of Information Technology, there are aspects of their analysis and design, which require a specialized, unique treatment. An important example illustrating this assertion is the contribution made by the theory of modeling (Abdullah *et al*., 1994).

Sensor models also allow a deeper understanding of sensor behavior to be gained. Sensor models were originally visualized as extensions of those widely used for most other dynamic systems (Finkelstein and Watts, 1983; Neubert, 1973). Although this approach was very useful for the visualization of the state and rate variables, it has some shortcomings in the representation of the states of the local mass element in a material and how they relate to the environmental states. The only approach that accepts this possibility uses mathematical models to describe the flow of energy and information (Kwaaital, 1993; Van Duyn, 1993; Van Duyn and Middelhoek, 1990; Ylilammi, 1989). Developments in this area have shown that sensor modeling should be based upon the three laws of thermodynamics (Van Duyn, 1993).

This indicates that as the acquisition and communication of information are impossible without the flow of energy or matter, considerations of energy form and material type are essential.

The diversity of energy forms in which systems are manifest are described in **Article 109, Structure and Energy in Sensor Systems, Volume 2**. These may be summarized using the COMETMAN acronym. Consequently, the underlying nature of all physical effects in diverse materials, and in whatever energy form, must play an essential role in all aspects of the modeling of sensors and their systems. The concepts of *intensive* and *extensive* state variables may be used to describe the equilibrium relations of the local mass element and the manner in which these are coupled to the environmental conditions respectively. In most cases, this representation is useful for implementing computer-based solutions for the analysis of sensor design candidates. However, it lacks the transparency necessary for deep and simplified qualitative understanding. Nevertheless, the complexity of the model should be of such a level as to be of practical use.

The process of modeling consists of six main steps, which may form a recursive process. When the important objectives of the model are fixed, it is then necessary to gather whatever knowledge is appropriate for the modeling purpose.

In order to simplify the process of modeling, it is necessary to apply methods of abstraction, which have five discriminable levels depending upon the purpose of the abstraction. The three main processes of abstraction allow increasing generality, objective approximation, or reduction, and perspectivization, or partial viewing.

In most cases of sensor modeling, it is necessary to conduct some kind of decomposition so that one large problem to be solved is broken down into a number of smaller more manageable and presumably more easily solved problems.

Formalization is another important aspect of the modeling process where formal models take the form of linguistic groups, pictorial representations, functional representations, or mathematical models. Since all models require validation, this is an important step in all modeling activity.

The modeling process, which may also be summarized as shown in Figure 2, covers the three main problems associated with the

- system, plant or process under measurement,
- goal or aim of the measurement,
- result of the measurement.

The central focus is placed upon the sensor whose dynamic properties are to be modeled and determined. This element, which is characterized by its input and output, may be visualized by humans who represent a valuable source of prior knowledge about the sensor.

This knowledge allows the proposal of appropriate mathematical models as well as the design of appropriate model validation experiments.

Deriving the model of most dynamic sensor systems is intimately connected with the principles of conservation of mass, energy, and momentum in time and space.

Consequently, modeling starts with the equations describing this balance. These relations are typically nonlinear, partial differential equations. Since these equations are difficult to solve, they are not really of much practical use.

Most practical systems are so complex that it is difficult to find an accurate model to describe the behavior of the process by physical modeling.

To give models that are practically useful and still realistic, the experimenter needs to make some well-informed assumptions for the purposes of simplification.

The first of these steps is the routine of linearization. By assuming typical average operating conditions, all of the parameters are represented by nominal values at the operating points of the process. Parameters may be material properties, environmental conditions, and other physical influences. In most industrial systems, it is possible to recognize some geometrical property of its structure. Such a property can be employed usefully by using the ideas

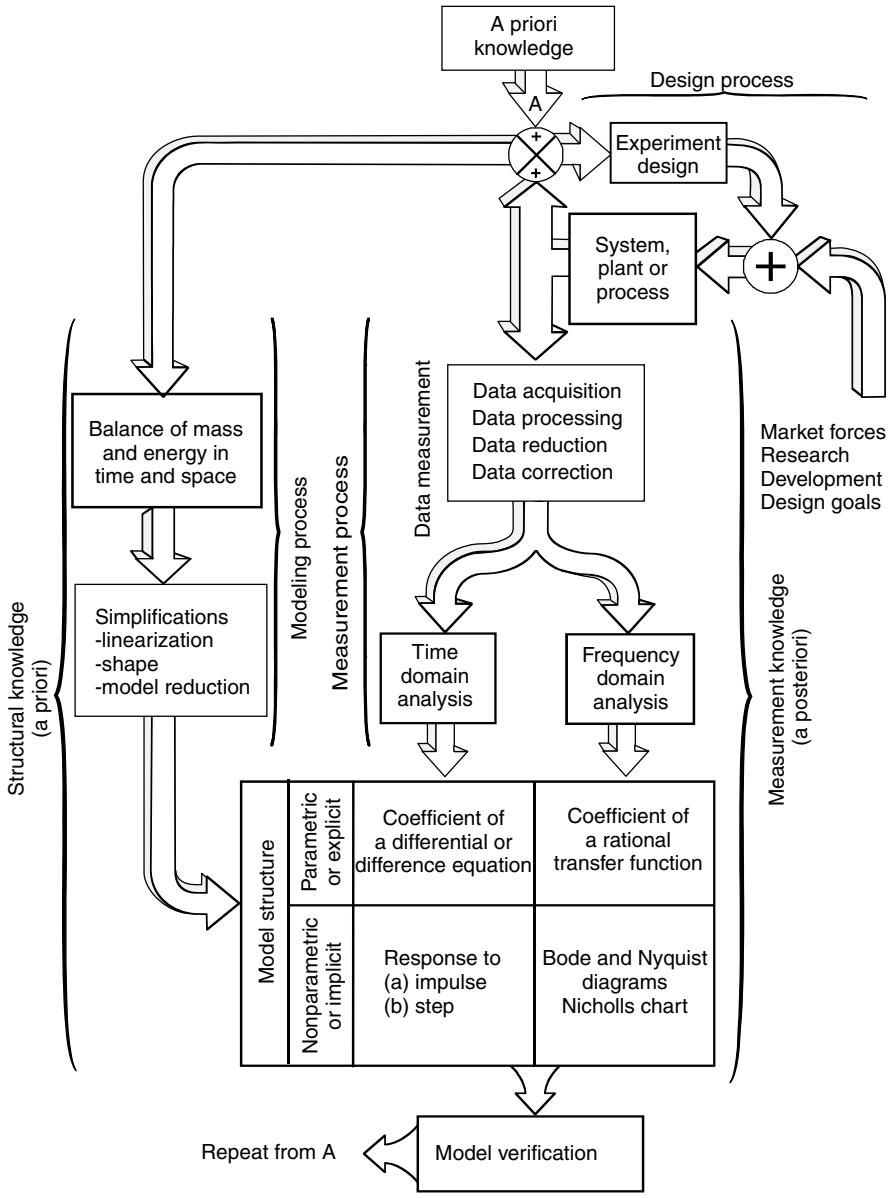


Figure 2. Steps in the modeling problem.

of shape simplification. Examples of this occur when there may be cylindrical symmetry in the real physical object.

Having linearized the model and introduced geometrical features in model simplification, the resulting mathematical model may still be difficult to understand and use. In such a case, the experimenter may be forced to apply a process of *model reduction*. Model reduction usually replaces linear partial differential equations with constant coefficient ordinary differential equations. McGhee, Henderson and Sydenham (1999) apply these techniques to the modeling of a resistance thermometer in its sheath. Modeling in measurement systems pursuits is found in many places – *see* **Article 71, Modeling Methodology, Volume 2**; **Article 114, Models of the Sensor Interface, Volume 2**; and **Article 105, Modeling with LabVIEW™, Volume 2**.

9 PARAMETRIC AND NONPARAMETRIC MODELS

Mathematical models obtained by the process described above have parameters that depend on physical materials used in the structure of the system. Hence, they are called *parametric models* or *explicit models*.

It often happens that a model based on physical behavior contains a number of unknown parameters even if the structure is derived from physical laws. Experimentation, which can help by inferring the values of these parameters from measurements, is an essential part of the model validation process.

The two most common parametric models refer to either the time domain description of the system, normally in the form of systems of linear difference equations, or the frequency domain description as rational transfer functions.

Other types of models make a minimum number of assumptions appropriate for the purpose of the model. These models are generally referred to as *nonparametric models* or *implicit models*. As with parametric models, nonparametric models may characterize the behavior of the system in either the time domain or the frequency domain.

The impulse response of a linear physical system, which is a statement of the time domain output of the system when the input is an impulse, is of basic importance. If the impulse response is known, either explicitly, from a parametric model, or implicitly, from a nonparametric model, then the response of the system to any other type of input signal may be predicted. The frequency domain equivalent of the impulse response is called the *frequency response function* sometimes simply called the *frequency response*. This model is a representation of the way in which a system responds to sinusoidal signals of different frequencies. Consequently, when the system is linear, it can be used to predict the response of the system to any signal that can be modeled by a number of composite frequency components.

When the process of modeling has reached an appropriate stage, the next stage involves the problems of identification. Identification is concerned with the synthesis and analysis of signals used in inferential measuring instruments. Inferential measurement, also called *identification*, is defined as the second problem of measurement in **Article 5, Humans in the Real World, Volume 1** and also by McGhee *et al*. (2001). DSP is widely used in identification.

10 MATERIALS SCIENCE FOR SENSORS AND SENSING

From thermodynamic modeling, it has become even more apparent that Materials Science plays a major role in Sensor Science, quite apart from the pivotal role it plays in the general science of machines. This role, although predominantly concerned with those physical effects that may be used to realize the sensing or transduction function, is also important from other points of view. Examples are mechanical rigidity or flexibility, chemical inertness or sensitivity in adverse environments, relative ease of manufacture or processing and so on (Cahn, Haasen and Kramer, 1993; de Podesta, 1996; Gardner, 1994) – *see* **Article 117, Materials in Measuring Systems, Volume 2**.

Because of the vast amount of published information on materials, it is essential to discriminate between those general aspects of materials science and the more specific requirements of sensor science. This may be achieved using a scheme of classification based upon the COMETMAN classification of measurand energy form.

It has been noted above that Knowledge Engineering, or KE (Charniak and McDermott, 1985; Finkelstein, 1994) is an area of growing importance in Sensor Science. It is likely that this has much promise for the resolution of many cognitive issues in Sensor Science even though it is an area where there are a number of outstanding problems. Some proposals concerned with these issues have been made (Cropley, 1994, 1997; Cropley and Cropley, 1994; Ferris and Sydenham, 1994) – *see also* **Article 160, Smart Sensor System Features, Volume 3**.

RELATED ARTICLES

Article 2, Organization of Instrument Science, Volume 1; **Article 6, Substructure of Human–Machine Systems, Volume 1**; **Article 16, Reduction of Influence Factors, Volume 1**; **Article 108, Transducer Fundamentals, Volume 2**; **Article 111, Classification of Sensors, Volume 2**; **Article 112, Systematic Description of Sensors, Volume 2**; **Article 116, Selection of Sensors, Volume 2**.

REFERENCES

Abdullah, F., Finkelstein, L., Khan, S.H. and Hill, W.J. (1994) Modelling in Measurement and Instrumentation – An Overview. *Measurement*, **14**(1), 41–45.

Agousti, A.T. (1996) *Introduction to Fiber Optic Sensors*, Chapman & Hall.

Ballentyne, D.W.G. and Lovett, D.R. (1980) *A Dictionary of Named Effects and Laws in Chemistry, Physics and Mathematics*, Chapman & Hall, London.

Baltes, H. Göpel, W. and Hesse, J. (eds) (1996) *Sensors Update* (in 2 Volumes), VCH Publishers, Weinheim.

Barney, G.C. (1988) *Intelligent Instrumentation*, Prentice Hall, UK.

Bau, H.H., de Rooij, N.F. and Kloeck, B. (1994) *Mechanical Sensors*, Vol 7 of Göpel, W. *et al*. (1989).

Bellanger, R. (1989) *Digital Processing of Signals*, John Wiley & Sons, Chichester.

Bentley, J.P. (2004), *Principles of Measurement Systems*, Prentice-Hall.

Boll, R. and Overshot, K.J. (1990) *Magnetic Sensors*, Vol 5 of Göpel, W. *et al*. (1989).

Bradley, D.A., Dawson, D., Burd, N.C. and Loader, A.J. (1991) *Mechatronics*, Chapman & Hall, New York.

Cahn, R.W., Haasen, P. and Kramer, E.J. (eds) (1993) *Materials Science and Technology: A Comprehensive Treatment* (in 18 Volumes), VCH Publishers (UK) Ltd., Cambridge, MA.

Carr, J.J. (1993) *Sensors, Transducers and Measurement*, Prentice Hall, Englewood Cliffs, NJ.

Chapman, P. (1995) *Smart Sensors*, ISA, Research Triangle Park, NC.

Charniak, E. and McDermott, D. (1985) *Introduction to Artificial Intelligence*, Addison-Wesley, Reading, MD.

Ciureanu, P. and Middelhoek, S. (eds) (1992) *Thin Film Resistive Sensors*, A. Hilger, Bristol, CT.

Cluley, J.C. (1974) *Electronic Equipment Reliability*, MacMillan, London.

Cook, S.C. (1993a) A Methodology for Formulating Measuring Instrument Specifications. *Measurement*, **11**(1), 1.

Cook, S.C. (1993b) A Knowledge-Based System for Computer-Aided Production of Measuring Instrument Specifications. *Measurement*, **11**(3), 235.

Cosofret, V.V. and Buck, R.P. (1992) *Pharmaceutical Applications of Membrane Sensors*, CRC Press, Boca Raton, FL.

Crocker, M.J. (ed.) (1998) *Handbook of Acoustics*, John Wiley & Sons, Chichester.

Cropley, D.H. (1994) Qualitative Analysis and Optimisation of Measurement Systems, *ACTA IMEKO XIII*, **2**, 849.

Cropley, D.H. (1997) *Information and Semiotics in Measurement*, PhD thesis, University of South Australia.

Cropley, D.H. and Cropley, A.J. (1994) Cognition in Electronic Sensing Systems from the Viewpoint of Psychology, *ACTA IMEKO XIII*, **2**, 843.

Dadswell, N.G. and Dean, R. (1995) *Optical Sensors for Power Industry Applications*, ERA Technology, Leatherhead.

Daly, H.V. and Linsley, E.G. (1970) Taxonomy, in *Encyclopaedia of the Biological Sciences*, 2nd edn (ed. P. Gray), Van Nostrand Reinhold, New York, 920.

de Podesta, M. (1996) *Understanding the Properties of Matter*, UCL Press Limited, London.

De Sa, A. (1990) *Principles of Electronic Instrumentation*, Edward Arnold, London.

Everett, H.R. (1995) *Sensors for Mobile Robots*, A.K. Peters.

Ferris, T.L.J. and Sydenham, P.H. (1994) Epistimological Aspects of Measurement, *ACTA IMEKO XIII*, **2**, 865.

Finkelstein, L. (1994) Intelligent and Knowledge Based Instrumentation – An Examination of Basic Concepts. *Measurement*, **14**(1), 23.

Finkelstein, L. and Finkelstein, A.C.W. (1983) Review of Design Methodology. *Proceedings of IEE Pt A*, **130**(4), 213.

Finkelstein, L. and Grattan, K.T.V. (1994) *Concise Encyclopaedia of Measurement and Instrumentation*, Pergamon, Oxford.

Finkelstein, L. and Watts, R.D. (1983) Fundamentals of Transducers – Description by Mathematical Models, in *Handbook of Measurement Science*, Vol. 2 (ed. P.H. Sydenham), John Wiley & Sons, Chichester (pp. 747–793).

Finkelstein, L., Ginger, R., El-Hami, M. and Mirza, M.K. (1993) Design Concept Generation for Instrument Systems – A Knowledge Based Approach. *Measurement*, **11**(1), 45.

Fraden, J. (1996) *Handbook of Modern Sensors: Physics, Designs, and Applications*, American Institute of Physics, Woodburg, New York.

Frank, R. (1996) *Understanding Smart Sensors*, Artech House, Norwood, MA.

Gardner, J.W. (1994) *Microsensors: Principles and Applications*, John Wiley & Sons, Chichester.

Godfrey, K.R. (ed.) (1993) *Perturbation Signals for System Identification*, Prentice Hall, Hemel Hempstead.

Göpel, W., Hesse, J. and Zemel, J.N. (1989) *Sensors: A Comprehensive Survey* (in 8 Volumes), VCH Publishers, Cambridge, MA.

Göpel, W., Jones, T.A., Kleitz, M., Lundström, I. and Seiyama, T. (eds) (1989) *Chemical and Biochemical Sensors*, Vols 2 and 3 of Göpel, W. Hesse, J. and Zemel, J.N.

Hauptmann, P. (1993) *Sensors: Principles and Applications*, translated by T. Pownall, Prentice Hall, Hemel Hempstead.

Henderson, I.A., McGhee, J. and El Fandi, M. (1997) *Data Measurement*, ISBN 0 9531409 0 3, Industrial Control Centre, University of Strathclyde, Glasgow.

Henderson, I.A., McGhee, J., Szaniawski, W. and Domaradzki, P. (1991) Incorporating High Reliability into the Design of Microprocessor Based Instrumentation. *IEE Proceedings A*, **138**(2), 105.

Henry, M.P. and Clarke, D.W. (1993) The Self-Validating Sensor: Rationale, Definitions and Examples. *Control Engineering Practice*, **1**(4), 585–610.

Hoyland, A. (1994) *Reliability Theory: Models and Statistical Methods*, John Wiley & Sons, New York.

Janata, J. (1989) *Principles of Chemical Sensors*, Plenum Press, New York.

Karube, I. (1994) *On-line Sensors for Food Processing*, Gordon and Breach, Switzerland.

Kerlin, T.W., Hashemian, H.M. and Petersen, K.M. (1982) Response Characteristics of Temperature Sensors Installed in Processes, *ACTA IMEKO*.

Kress-Rogers, E. (ed.) (1993) *Instrumentation and Sensors for the Food Industry*, Butterworth-Heinemann, Oxford.

Krohn, D.A. (1992) *Fiber Optic Sensors: Fundamentals and Applications*, 2nd edn, ISA, Research Triangle Park, NC.

Kwaaital, Th. (1993) The Fundamentals of Sensors. *Sensrs and Actrs*, **39**, 103–110.

Lerch, R. (1994) *Sensors for Measuring Sound*, Chapter 16 of Bau *et al*. (1994).

Lion, K.S. (1969) Transducers: Problems and Prospects. *IEEE Transactions*, **IECI-16**, 2–5.

Ludeman, L.C. (1987) *Fundamentals of Digital Signal Processing*, John Wiley & Sons.

Lyn, M.R. (ed.) (1996) *Handbook of Reliability Engineering*, McGraw-Hill, New York.

Lynnworth, L.C. (1994) *Ultrasonic Nonresonant Sensors*, Chapter 8 of Bau *et al*. (1994).

Maenaka, K. (1994) *Hall Sensors*, Chapter 6 of Bau *et al*. (1994).

Maskell, W.C. (1996) *Gas Sensors*, Chapman & Hall.

McGee, T.D. (1988) *Principles and Methods of Temperature Measurement*, John Wiley & Sons, New York.

McGhee, J., Henderson, I.A. and Baird, A. (1997) Neural Networks Applied for the Identification and Fault Diagnosis of Process Valves and Actuators. *Measurement*, **20**(4), 267–275.

McGhee, J., Henderson, I.A. and Sydenham, P.H. (1999) Sensor Science – Essentials for Instrumentation and Measurement Technology. *Measurement*, **25**, 89–113.

McGhee, J., Henderson, I.A., Kulesza, W. and Korczynski, M.J. (2001) *Measurement Data Handling: Theoretical Technique*, ISBN 83-7283-007-X, The Technical University of Lodz, Lodz.

McGlone, P., McGhee, J. and Henderson, I.A. (1998) *Performance Evaluation of a Differential Pressure Cell*, *IEEE Transactions on Instrumentation and Measurement*, **47**(5) 1271–1276.

Michalski, L., Eckersdorf, K., Kucharski, J. and McGhee, J. (2001) *Temperature Measurement*, 2nd edn, John Wiley & Son, Chichester.

Michalski, L., Eckersdorf, K., Kucharski, J., Sankowski, D. and Urbanek, P. (1993) Experimental Computerised In-situ Identification of Thermocouple Sensors, in *Proceedings of TEMPMEKO 93, 5th International Symposium on Temperature and Thermal Measurement in Industry and Science*, Prague (pp. 107–111).

Middelhoek, S. and Audet, S. (1989) *Silicon Sensors*, Academic Press, London.

Moseley, P.T. and Tofield, B.C. (eds) (1987) *Solid-State Gas Sensors*, Adam Hilger, Bristol, CT.

Neubert, H.K.P. (1973) *Instrument Transducers*, 2nd edn, Clarendon Press, Oxford.

Ngo, T.T. (ed.) (1987) *Electrochemical Sensors in Immunological Analysis*, Plenum Press, New York.

Nicholas, J.V. and White, D.R. (2001) *Traceable Temperatures*, John Wiley & Sons, Chichester.

O'Connor, P.D.T. (2002) *Practical Reliability Engineering*, 4th edn, John Wiley & Sons, Chichester.

Pallas-Areny, R. and Webster, J.G. (1991) *Sensors and Signal Conditioning*, John Wiley & Sons, New York.

Patton, R.J., Frank, R.N. and Clark, R.N. (1989) *Fault Diagnosis in Dynamic Systems: Theory and Applications*, Prentice Hall, Englewood Cliffs, NJ.

Ramakumar, R. (1993) *Engineering Reliability: Fundamentals and Applications*, Prentice Hall, Englewood Cliffs, NJ.

Regtien, P.P.L. (1992) *Instrumentation Electronics*, Prentice Hall, Hemel Hemstead.

Ruocco, S.R. (1987) *Robot Sensors and Transducers*, Open University Press, Milton Keynes, MA.

Sandquist, G.M. (1985) *Introduction to Systems Science*, Prentice Hall, Englewood Cliffs, NJ.

Scholz, J. and Ricolfi, T. (1990) *Thermal Sensors*, Vol 4 of Göpel, W. *et al*. (1989).

Schubert, J. (1984) *Physikalische effekte*, Physika-Verlag, Weinheim, BDR.

Sinclair, I.R. (2000) *Sensors and Transducers*, 3rd edn, Butterworth-Heinemann, Oxford.

Soloman, S. (1994) *Sensors and Control Systems in Manufacturing*, McGraw-Hill, New York.

Stein, P.K. (1969) The Engineering of Measurement Systems. *Journal Metals*, **231**, 40.

Sydenham, P.H. (ed.) (1982) *Handbook of Measurement Science*, *Vol 1 Theoretical Fundamentals*, John Wiley & Sons, Chichester.

Sydenham, P.H. (ed.) (1983) *Handbook of Measurement Science*, *Vol 2 Practice Fundamentals*, John Wiley & Sons, Chichester.

Sydenham, P.H. (1994a) Disciplined Design of Sensor Systems – Part 1. Types of Design. *Measurement*, **14**(1), 73–80.

Sydenham, P.H. (1994b) Disciplined Design of Sensor Systems – Part 2. Knowledge Based Systems in Sensor and Sensor Systems Design. *Measurement*, **14**(1), 81–88.

Sydenham, P.H. (2003) *Systems Approach to Engineering Design*, Artech House.

Sydenham, P.H. and Thorn, R. (eds) (1992) *Handbook of Measurement Science*, *Vol 3 Elements of Change*, John Wiley & Sons, Chichester.

Sydenham, P.H., Hancock, N.H. and Thorn, R. (1992) *Introduction to Measurement Science and Engineering*, Wiley.

Sze, S.M. (ed.) (1994) *Semiconductor Sensors*, John Wiley & Sons, New York.

Taylor, R.F. (1996) *Handbook of Chemical and Biological Sensors*, Adam Hilger, Bristol, CT.

Usher, M.J. and Keating, D. (1996) *Sensors and Transducers: Characteristics, Applications, Instrumentation*, Macmillan, Basingstoke.

Van Duyn, D.C. (1993) *Multi-Signal Domain Modelling of Solid-State Transducers*, Delft University Press, Delft.

Van Duyn, D.C. and Middelhoek, S. (1990) Information Transduction in Solid-State Transducers: A General Thermo-Dynamic Systems Approach. *Sensrs and Actrs*, **A-21-A23**, 25.

Van Putten, A.F.P (1996) *Electronic Measurement Systems: Theory and Practice*, Institute of Physics, London.

Von Engelhardt, W. and Zimmermann, J. (1988) *Theory of Earth Science*, Cambridge University Press, Cambridge, MA.

Wagner, E., Dendliker, R. and Spenner, K. (eds) (1994) *Optical Sensors*, Vol 6 of Göpel, W. *et al*. (1989).

Webster, J.G. (ed.) (1988) *Tactile Sensors for Robotics and Medicine*, John Wiley & Sons, New York.

Weinberg, G.M. (1975) *An Introduction to General Systems Thinking*, John Wiley & Sons, New York.

Wolfbeis, O.S. (ed.) (1991) *Fiber Optic Chemical Sensors and Biosensors*, CRC Press, Boca Raton, FL.

Wolffenbuttel, R.F. (ed.) (1996) *Silicon Sensors and Circuits: On-chip Compatibility*, Chapman & Hall, London.

Ylilammi, M. (1989) Thermodynamics of Sensors. *Sensrs and Actrs*, **18**, 167–178.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

108: Transducer Fundamentals

Paul P.L. Regtien
University of Twente, Enschede, The Netherlands

1 Definition of a Transducer 717
2 Categorizing Sensors 718
3 Terminology of Sensors 720
4 Sensor Performance 721
5 Role and Use of Actuators 723
References 724

1 DEFINITION OF A TRANSDUCER

A transducer is an essential part of any information processing system that operates in more than one physical domain. These domains are characterized by the type of quantity that provides the carrier of the relevant information. Examples are the optical, electrical, mechanical, and magnetic domains. A transducer is the part of the instrument that converts information about a measurand from one domain into another, ideally without information loss. So, a transducer can formally be defined as *a device that converts one form of energy into another*, with the intention of preserving information.

A transducer has at least one input and one output. In measuring instruments, where information processing is performed by electrical signals, either the output or the input is of electrical nature (voltage, current, resistance, capacitance, etc.), whereas the other is a nonelectrical signal (displacement, temperature, elasticity, etc.). A transducer with a *nonelectrical input* is an *input transducer*, intended to convert a nonelectrical quantity into an electrical signal so as to measure that quantity. A transducer with a *nonelectrical output* is called an *output transducer*, intended to convert an electrical signal into a nonelectrical quantity so as to control that quantity. So, a more explicit definition of a transducer is *an electrical device that converts one form of energy into another*, with the intention of preserving information.

According to common terminology, these transducers are also called *sensor* and *actuator* respectively (see Figure 1).

So, a sensor is an input transducer and an actuator is an output transducer. It should be noted, however, that this terminology is not standardized. In literature, other definitions are found. In particular, some authors make a clear differentiation between a *sensor* and a (input) *transducer*, stressing a distinction between the element that performs the physical conversion and the complete device, for instance, a strain gauge (transducer) and a load cell with one or more strain gauges (sensor), or even vice versa. Modern sensors not only contain the converting element but also contain part of the signal processing (analog processing such as amplification and filtering, Analogue-to-Digital-conversion and even some digital electronics). Many such sensors have the electronics integrated with the transducing part onto a single chip. Present day sensors may have a bus compatible output, implying full signal conditioning on board. There is a trend to also include transmission electronics within the device, for instance, in biomedical applications.

Signal conditioning may be included

- to protect the sensor from being loaded or to reduce loading errors;
- to fit the sensor output to the input range of the Analogue-to-Digital Converter (ADC);
- to enhance the S/N (signal-to-noise ratio) prior to further signal processing;
- to generate a digital, bus compatible electrical output;
- to transmit measurement data for wireless applications.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

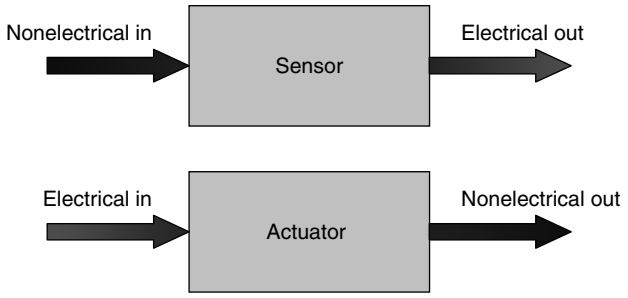


Figure 1. Input transducer (sensor) and output transducer (actuator).

In conclusion, the boundaries between sensor and transducer, as proclaimed in many sensor textbooks, are disappearing or losing their usefulness: the user buys and applies the sensor system as a single device, with a nonelectrical input and an electrical (analog, digital, bus compatible) output.

2 CATEGORIZING SENSORS

A sensor (or input transducer) performs the conversion of information from the physical domain of the measurand to the electrical domain. Many authors have tried, more or less successfully, to build up a consistent categorization of sensors (*see also* **Article 112, Systematic Description of Sensors, Volume 2**). It is not easy to create a consistent systematic description encompassing all sensor principles. There is at least consensus on a division into two groups of sensors: direct and modulating sensor types (see Figure 2). The distinguishing property is the need for auxiliary energy.

Direct sensors do not require additional energy for conversion. As information transport cannot exist without energy transport, a direct sensor withdraws the output energy directly from the measurement object. As a consequence, loss of information about the original state of the object may occur. There might also be energy loss, for instance, heat. An important advantage of a direct sensor is its freedom from offset: at zero input, the output is essentially zero. Examples of direct sensors are the thermocouple and the piezoelectric force and acceleration sensor.

Indirect sensors or *modulating sensors* use an additional energy source that is modulated by the measurand; the sensor output energy mainly comes from this auxiliary source, and just a fraction of energy is withdrawn from the measurement object. In this respect, modulating sensors do not significantly load the measurement object and hence are more accurate than direct sensors. Most sensors are of the modulating type, for instance, all resistive and capacitive sensors and many inductive sensors.

There are many sensors on the market, over a million types worldwide. A categorization of sensors would help make the proper choices and also make it easier to understand, but a useful basis for categorization is difficult to define. There are various possibilities such as

- according to the measurand,
- according to the conversion principle,
- according to the domain of the measurand,
- according to application fields.

All of them have their limitations. For instance, the number of measurands is rather large, making the first option not very practical. Figure 3 gives an overview of the most common physical quantities for which sensors are available, after Middelhoek and Noorlag (1981). However, the list is not complete. Moreover, many quantities can be measured by a variety of sensor types. For example, position can be measured using resistive, capacitive, inductive, acoustic, and optical methods.

The second option, according to the conversion principle, is often used for the reason that the sensor performance is mainly determined by the physics of the underlying principle of operation. On the other hand, a particular type of sensor might be suitable for a variety of physical quantities and in many different applications. For instance, a magnetic sensor of a particular type could be applied as a displacement sensor, a velocity sensor, a tactile sensor, and so on. For all these applications, the performance is limited by the physics of this magnetic sensor, but the limitations manifest in completely different ways.

A closer look at the various conversion effects may lead to the observation that the electrical output of a sensor depends either on a material property, or the geometry, or a movement. Figure 4 tabulates these three phenomena, for various types of sensors. The figure gives the material parameter, the geometric parameter, and the velocity induced parameter, together with associated sensors.

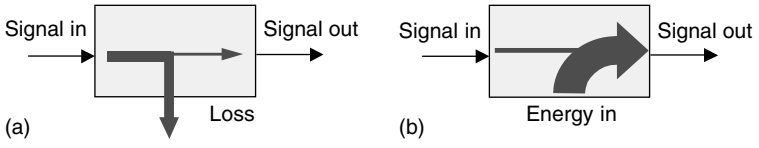


Figure 2. (a) Direct sensor and (b) modulating sensor.

1. Mechanical, solids
acceleration
angle
angular velocity
area
diameter
distance
elasticity
expansion
filling level
force
gradient
hardness
height
length
mass
moment
movement
orientation
pitch
position
pressure
proximity
rotation
roughness
shape
tension
torque
torsion
velocity
vibration
weight

2. Mechanical, fluids
density
flow direction
flow velocity
level
pressure
rate of flow
viscosity
volume

3. Thermal
enthalpy
entropy
temperature
thermal capacity
thermal conduction
thermal expansion
thermal radiation

4. Optical
color
light polariaztion
light wavelength
luminance
luminous intensity
reflection
refractive index

5. Acoustic
sound frequency
sound intensity
sound polarization
sound pressure
sound velocity
time of flight

6. Nuclear radiation
ionization degree
mass absorption
radiation dose
radiation energy
radiation flux
radiation type

7. Chemical
cloudiness
composition
concentration
electrical conductivity
humidity
impurity
ionization degree
moisture
molar weight
particle form
particle size
pH
polymerization degree
reaction rate
redox potential
thermal conductivity
water content

8. Magnetic and electrical
capacity
charge
current
dielectric constant
electric field strength
electric power
electric resistance
frequency
inductivity
magnetic field strength
phase
pulse duration
signal distortion

9. Time
time

Figure 3. Parameters for which sensors are on the market (after Middelhoek and Noorlag (1981)).

Type	Material property	Geometry	Relative movement
Resistive	Resistivity (piezoresistor,LDR)	Relative length (potentiometer; metal strain gauge)	
Capacitive	Permittivity (fluid level sensor)	Relative electrode distance capacitive displacement (LVDC)	
Magnetic	Permeability (magnetoresistor)	Distance source-detector (magnetic displacement sensor)	Induction (magnetic velocity sensors)
Inductive		Inductance, self-inductance, mutual inductance reluctance (inductive displacement sensors, LVDT, resolver)	Induction (inductive velocity sensors)
Optical	Index of refraction; absorptivity (fibre optic sensors)	Distance transmitter–receiver (intensity modulation sensors; interferometer; TOF sensor) transmissivity, reflectivity (optical encoder, tachometer)	Doppler frequency (Doppler velocimeter)
Acoustic	Acoustic impedance	Distance transmitter–receiver (TOF displacement sensors)	Doppler frequency (Doppler velocimeter)
Piezoelectric	Polarization (piezoelectric sensors)	Deformation (piezoelectric sensors)	

Figure 4. Three groups of parameters, with examples of sensors.

A categorization based on the domain of the measurand is too coarse, and domain definitions are not unambiguous.

Finally, an application field provides no restricted set of sensors, since in each field (biomedical, automotive, agriculture), almost all types of sensors could be applied.

3 TERMINOLOGY OF SENSORS

We have defined a sensor as a device that performs the conversion of information from the physical domain of the measurand to the electrical domain. For instance, a position sensor converts position information into an electrical signal (a voltage, current, etc.). Obviously, a position sensor measures position. However, there are different names for different applications and situations, for instance,

Distance sensor: measures the length of the straight line between two defined points;

Position sensor: measures the coordinates of a specified point of an object in a specified reference system;

Displacement sensor: measures the change of position relative to a reference point;

Proximity sensor: (i) determines the sign (positive or negative) of the linear distance between an object point and a fixed reference point;
(ii) a contact free displacement or distance sensor for short distances (down to zero);

Level sensor: measures the distance of the top level of a liquid or a granular substance in a container with respect to a specified horizontal reference plane;

Angular sensor: measures the angle of rotation relative to a reference position;

Encoder: displacement sensor (linear or angular) with digital output:

Tilt sensor: measures the angle relative to the earth's normal;

Tachometer: measures the rotational speed;

Vibration sensor: measures the motion of a vibrating object in terms of displacement, velocity or acceleration;

Accelerometer: measures acceleration.

Sensors for the measurement of force and related quantities are as follows:

Pressure sensor: measures pressure difference, relative to either vacuum (absolute pressure), a reference pressure or ambient pressure;

Force sensor: measures the (normal and/or shear) force exerted on the active point of the transducer;

Torque sensor: measures torque (moment);

Force–torque sensor: measures both forces and torques (up to six components);

Load cell: force or pressure sensor, for measuring weight;

Strain gauge: measures linear relative elongation (positive or negative) of an object, caused by compressive or tensile stress.

Many sensors have been given names according to their operating principle or construction. Examples are as follows:

LVDT: or linear variable displacement transformer, a device that is basically a voltage transformer, with a linearly movable core;

Gyroscope: a device for measuring the angular velocity, based on the gyroscopic effect occurring in rotating or vibrating structures;

Hall sensor: measures magnetic field on the basis of the Hall effect, after the American physicist Edwin Hall (1855–1938).

Some sensors use a *concatenation* of conversion steps. A displacement sensor combined with a spring can act as a force sensor. The measurand (force) is converted to a displacement, which in turn is converted into an electrical signal. In combination with a calibrated mass, a displacement sensor can serve as an accelerometer: the measurand (acceleration) is converted into a force (by the inertial effect), the force into a displacement (by the spring), and the displacement into an electrical signal. The performance of such transducers not only depends on the original sensor but also on the added components: in the case of the accelerometer, it depends on the spring compliance and the seismic mass respectively.

Information about a particular quantity can also be obtained by *calculation* or by additional electronic *signal processing* using relations between quantities. The accuracy

of the result not only depends on the errors in the quantities that are measured directly but also on the accuracy of the parameters in the model that describes the relation between the quantities involved. For instance, in an acoustic distance measurement, the distance is calculated from the measured time-of-flight (with associated errors) and the sound velocity. An accurate measurement result requires knowledge of the acoustic velocity of the medium at the prevailing temperature.

Speed and acceleration can be measured using a displacement sensor, by differentiating its output signal once or twice respectively, and vice versa: by integrating the output signal of an accelerometer, a velocity signal is obtained, and by a second integration, a position signal is obtained. Obviously, the performance of the final result depends on the quality of the signal processing. The main problem with differentiation is the increased noise level (in particular in the higher frequency range), and integration may result in large drift due to the integration of offset.

4 SENSOR PERFORMANCE

Imperfections of a sensor are usually listed in the data sheets provided by the manufacturer. These sensor specifications inform the user about deviations from the ideal behavior. The user has to accept technical imperfections, as long as they do not exceed the specified values.

Any measuring instrument, and hence any sensor, has to be fully specified with respect to its performance. Unfortunately, many data sheets show lack of clarity and completeness. Gradually, international agreements about formal error descriptions are being established. An extensive description of measurement errors and error terminology can be found in IOS (1995). Further, there is an international standard on transducer nomenclature and terminology ISA (1975). Finally, various international committees are working toward a uniform framework for specifying sensors.

The characteristics that describe sensor performance can be classified into four groups:

- static characteristics, describing the performance with respect to very slow changes;
- dynamic characteristics, specifying the sensor response to variations in time and in the measurand (the quantity that has to be measured);
- environmental characteristics, relating the sensor performance after or during exposure to specified external conditions (pressure, temperature, vibration, radiation);
- reliability characteristics, describing the sensor life expectancy.

First, we define some common specifications that apply to sensors:

- sensitivity;
- nonlinearity and hysteresis;
- resolution;
- accuracy;
- offset and zero drift;
- noise;
- response time;
- frequency response.

4.1 Sensitivity

The sensitivity of a sensor is defined as the ratio between a change in the output value and the change in the input value that causes that output change. Mathematically, the sensitivity is expressed as $S = \partial y/\partial x$, where x is the input signal (measurand) and y is the output (an electrical signal). Usually, a sensor is also sensitive to changes in quantities other than the intended input quantity, such as the ambient temperature or the supply voltage. These unwelcome sensitivities should be specified as well, for a proper interpretation of the measurement result. To have a better insight in the effect of such unwanted sensitivities, it is often related to the sensitivity of the measurement quantity itself.

Example 1 The sensitivity of a particular displacement sensor with voltage output is specified as $10\,\text{mV mm}^{-1}$. Its specified temperature sensitivity is $0.1\,\text{mV K}^{-1}$. Since 0.1 mV corresponds to a displacement of 10 mm, the temperature sensitivity can also be expressed as $10\,\text{mm K}^{-1}$. A temperature rise of 5 °C results in an apparent displacement of 50 mm.

Example 2 The sensitivity of a particular type of temperature sensor is $100\,\text{mV K}^{-1}$, including the signal conditioning unit. The signal conditioning part itself is also sensitive to (ambient) temperature and appears to create an extra output voltage of 0.5 mV for each °C rise in ambient temperature (not necessarily the sensor temperature). So, the unwanted temperature sensitivity is $0.5\,\text{mV K}^{-1}$ or 0.5/100 = $5\,\text{mK K}^{-1}$. A change in ambient temperature of ±10 °C gives an apparent change in sensor temperature equal to ±50 mK.

4.2 Linearity and hysteresis

If the output y is a linear function of the input x, the sensitivity S does not depend on x. In the case of a nonlinear transfer function $y = f(x)$, S does depend on

the input or output value. Often, a linear response is preferred to reduce computational burden in, for instance, multisensor control systems. In that case, the sensitivity can be expressed with a single parameter. The transfer of a sensor with a slight nonlinearity may be approximated by a straight line, to specify its sensitivity by just one number. The user should be informed about the deviation from the actual transfer; this is specified by the nonlinearity.

The linearity error of a system is the maximum deviation of the actual transfer characteristic from a prescribed straight line. Manufacturers specify linearity in various ways, for instance, as the deviation in input or output units: $\Delta x_{\max}$ or $\Delta y_{\max}$, or as a fraction of FS (full scale): $\Delta x_{\max}/x_{\max}$. Nonlinearity should always be given together with a specification of the straight line. The following definitions are in use:

- *Terminal nonlinearity*: based on the terminal line–a straight line between 0 and 100% theoretical full-scale points.
- *End-point nonlinearity*: based on the end-point line–the straight line between the calibrated end points of the range; coincides with the terminal (theoretical) line after calibration of zero and scale.
- *Independent nonlinearity*: referring to the best-fit straight line, according to a specified error criterion, for instance, the line midway between two parallel lines enclosing all calibration points; if the least-square error criterion for the best-fit straight line is used, this linearity error is as follows.
- *Least-square nonlinearity*: based on the least square line, the line for which the summed squares of the residuals is minimized.

Hysteresis is the maximum difference in output signal when the measurand first increases over a specified range, and next returns to the starting value. The traveled range should be specified because hysteresis strongly depends on it.

4.3 Resolution

The resolution indicates the smallest detectable increment of the input quantity. When the measurand varies continuously, the sensor output might show discontinuous steps. The value of the corresponding smallest detectable change in the input variable is the resolution: $\Delta x_{\min}$. Sometimes this parameter is related to the maximum value $x_{\max}$ that can be processed (full-scale value), resulting in the resolution expressed as $\Delta x_{\min}/x_{\max}$.

Example 3 The resolution of a particular type of wire-wound linear potentiometer with range 10 cm is specified as 10^{-4}; assuming this is relative to the full-scale value, it means that the output changes discontinuously in steps equivalent to input displacements of 10 mm.

Example 4 A particular type of optical encoder that has a resolution of 14 bit. The smallest change in angle that can be detected by this encoder is $2\pi/2^{14} \approx 1.9 \times 10^{-4}$ rad or 0.022°.

4.4 Accuracy

Formally, the accuracy reflects the closeness of the agreement between the actual measurement result and a true value of the measurand. The accuracy specification should include relevant conditions and other quantities. Many sensor manufacturers specify the sensor performance in terms of accuracy. This specification should be viewed with suspicion, because it may or may not include particular imperfections of the sensor (nonlinearity, hysteresis, drift), and may be only valid under strict conditions.

4.5 Offset and zero drift

Most sensors are designed such that the output is zero at zero input. If the transfer characteristic does not intersect the origin (x,y = 0,0) the system is said to have offset. The offset is expressed in terms of the input or the output quantity. Specifying the input offset is preferred to facilitate a comparison with the measurand.

Example 5 The sensitivity of a particular type of force sensing system is $0.1\,\mathrm{V\,N^{-1}}$. At zero force, the output appears to be 3 mV. The (input) offset of this system is the output offset divided by the sensitivity, so 0.03 N.

A nonzero offset arises mainly from component tolerances. Offset compensation can be performed in the interface electronics or the signal processing unit. Once adjusted to zero, the offset may nevertheless change, owing to temperature variations, changes in the supply voltage or aging effects. This relatively slow change in the offset is called *zero drift*. In particular, the temperature induced drift (the *temperature coefficient* or t.c. of the offset) is an important item in the specification list.

Sometimes a system is deliberately designed with offset. Many industrial transducers have a current output ranging from 4 to 20 mA. This facilitates the detection of cable fractures or a short circuit, producing a zero output clearly distinguishable from a zero input.

4.6 Noise

Electrical noise is also specified in terms of the input quantity, to show its effect relative to that of the measurand. *White noise* (noise with constant power over a wide frequency range) is usually expressed in terms of *spectral noise power* (W/Hz), *spectral noise voltage* (V/$\sqrt{}$Hz) or *spectral noise current* (A/$\sqrt{}$Hz). Thermal noise is an example of 'white noise'.

Another important type of noise is *1/f noise* (one-over-f noise), a collection of noise phenomena with a spectral noise power that is proportional to f^{-n}, with $n = 1$ to 2.

Quantization noise is the result of quantizing an analog signal. The rounding off results in a (continuous) deviation from the original signal. This error can be considered as a 'signal' with zero mean, and a standard deviation determined by the resolution of the AD converter.

4.7 Response time

The response time is associated with the speed of change in the output upon a stepwise change of the measurand. The specification of the response time always needs to be accompanied by an indication of the input step (for instance, FS – full scale) and the output range for which the response time is defined, for instance, 10 to 90%. Creep and oscillations may make the specification of the response time meaningless or at least misleading.

4.8 Frequency response

The sensitivity of a system depends on the frequency or rate of change of the measurand. A measure for the useful frequency range is the *frequency band*. The upper and lower limit of the frequency band are defined as those frequencies for which the output signal has dropped to half the nominal value, at constant input power. For voltage or current quantities, the criterion is $1/2\sqrt{2}$ of the nominal value. The lower limit of the frequency band may be zero; the upper limit has always a finite value. The extent of the frequency band is called the *bandwidth* of the system, expressed in Hz.

All specification items only apply within the operating range of the system, which should also be specified correctly. It is given by the measurement range, the required supply voltage, the environmental conditions and possibly other parameters.

Despite the specified limitations of sensors, a sensing system can be configured in such a way that the effect of some of these limitations are eliminated or at least reduced. Various possibilities of error reduced designs are described in **Article 16, Reduction of Influence Factors, Volume 1**.

5 ROLE AND USE OF ACTUATORS

Similar to the definition of a sensor (or input transducer) an actuator (or output transducer) can be defined as *a device that converts information from the electrical domain to another physical domain*, with preservation of information. Actuators and sensors can be considered as each other's opposites. Actually, many transduction principles are reversible (notably the direct conversion types), which means that the same device can act as both – a sensor as well as an actuator (for instance, a piezoelectric sensor and actuator). However, an actuator should, in general, be able to deliver output energy, whereas a sensor operates best when the amount of energy taken from the measurement object is close to zero. Reversibility does not imply that both options can be combined in a single device.

Since an actuator has to produce output energy, its construction and dimensions differ largely from that of sensors. There are far less physical conversion principles suitable for actuation than for sensing. However, the variability in types is larger, according to the application and the required energy.

Some actuator types, grouped according to the physical domain of the output are as follows:

- *Mechanical output*: induction motor; permanent magnet electromotor; stepper motor, piezoelectric actuators (linear and rotating movements), relay;
- *Optical output*: light emitting diode (LED), solid state laser diode (SSLD); incandescent lamp;
- *Thermal output*: electric heater; (thermo)electric cooler;
- *Acoustic output*: piezoelectric ultrasound transmitter; electrostatic ultrasound transmitter.

As with sensors, actuators too can operate either in a direct way or in a modulating mode. Most actuators are of the direct type: no auxiliary energy is required. Examples of modulating actuators are an oscilloscope (the electrical signal modulates the deflection of an electron beam and hence the position of the light spot on the screen) and an LCD (the electrical signal modulates the transmission of an auxiliary light source by influencing the transmittance of the crystal material).

The performance of a sensor system can sometimes be improved by including the sensor in a feedback loop. The electric output of the sensor is amplified and supplied to an actuator, having the inverse conversion characteristic. So its output is of the same nature as that of the measurand, or at least is closely related.

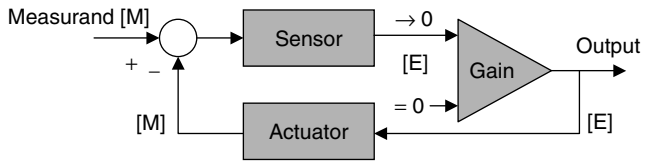


Figure 5. Feedback sensor configuration; [M] denotes the domain of the measurand, [E] the electrical domain.

In the steady state, and assuming stability and a high loop gain, the input of the sensor is compensated (made zero) by the feedback action, see Figure 5. The compensating signal is delivered by the actuator, and hence its input equals the output of the measurement system. Since the input signal of the sensor is zero, its static characteristics (sensitivity, nonlinearity, limited range) are irrelevant to the measurement result. The only requirement is zero offset. The performance of the measurement completely relies on the characteristics of the actuator. Further details of feedback sensing systems are discussed in more detail in the article **Article 113, Force-feedback Sensors, Volume 2**.

Like with sensors, also some actuators operate on the basis of a concatenation of conversions. Examples are as follows:

- *Thermal microactuator*: a particular element of the microstructure that is heated by an electric current upon which it expands, producing a displacement of the tip.
- *Reed switch*: a coil around the switch that is activated by a current, producing a magnetic field, which closes the contacts of the switch.
- *Incandescent lamp*: electric current heats up a filament that emits visible light above a particular temperature.

In the past decades, sensors tended to get smaller and actuators larger. Development in microtechnology, however, allows the creation of microactuators as well see, for instance, Tabib-Azar (1997). Consequently, such microactuators can produce little energy, and are therefore only useful as part of a microsystem, including (micro)sensors, signal processing, and actuation.

The role of an actuator in a measurement system can be

- indication of the measurement result (optical display),
- registration of the measurement result (magnetic or optical head; driving a plotter pen),
- control of the measurement system (scanning in 1, 2, or 3 dimensions),
- control of a process to perform measurements (for instance, in wear detection, fatigue tests)
- control of the sensing part of the measurement system (feedback sensors).

REFERENCES

Middelhoek, S. and Noorlag, D.J. (1981) Three-Dimensional Representation of Input and Output Transducers. *Sensors and Actuators*, **2**, 29–41.

Tabib-Azar, M. (1997) *Microactuators – Electrical, Magnetic, Thermal, Optical, Mechanical, Chemical and Smart Structures*, Kluwer Academic Publishers, Boston, MA.

IOS (1995) *Guide to the Expression of Uncertainty in Measurement*, International Organization for Standardization, Geneva.

ISA (1975) S37.1: Electrical Transducer Nomenclature and Terminology.

109: Structure and Energy in Sensor Systems

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Boundary Perspective of Sensor Systems 725
2 Sensors, Transducers, and Modifiers – Terminology and Definitions 725
3 Representing an Instrument Stage by Signal/Energy Notation 728
4 Classification of Transducers by Information and Energy 728
5 The Classification of Energy 729
6 Spatial Modeling of Transducers and Sensors 731
7 The Transducer Space 732
8 The Sensor Effect Tetrahedron for Modulating Sensors 733
9 The Sensor Effect Tetrahedron for Self-supporting Cross-converting Sensors 734
Related Articles 735
References 735

1 BOUNDARY PERSPECTIVE OF SENSOR SYSTEMS

When analyzing sensor systems, the Systems Engineering approach described in **Article 5, Humans in the Real World, Volume 1**; and **Article 14, Hierarchical Aspects of Measurement Systems, Volume 1**, can fruitfully be applied by combining the ideas of the boundary view of a measurement system and the information and energy tetrahedron as described in **Article 6, Substructure of Human–Machine Systems, Volume 1** to obtain Figure 1. This approach is consistent with that used in the application of the Systems Engineering boundary view to the measurement system described in **Article 68, Setting the System Boundaries, Volume 2**. The existence of intrinsic and extrinsic boundaries allows the discrimination between internally and externally arisen effects respectively. In this way, internally caused errors, which give rise to 'eigen' errors, or characteristic errors (Solopchenko, 1994), are noted to be qualitatively different from externally arising effects. From this boundary view, it can also be seen that inputs and outputs, which can be subdivided by energy form, can also be classified as within the resource, information, or contamination groups of boundary impacts.

Interfering influences – which arise from inside the sensor, are due to either parameter sensitivity or internal fundamental fluctuations. Parameter sensitivity is caused by the inexorable aging of all of the components that make up the sensor structure. Fundamental fluctuations occur as a result of equipartition of energy between the thermodynamic state and the states of the other energy forms. External interference is caused by disturbance power flows or by other forms of direct contamination. Highlighting the various sources of contaminating inputs provides a clear view of the limits on the ultimate resolution of a sensor in executing its primary function. In the case of sensor systems, the resource output is also the information output. It is the influence effects that make good design difficult to achieve.

2 SENSORS, TRANSDUCERS, AND MODIFIERS – TERMINOLOGY AND DEFINITIONS

Through the study of the various blocks used in routine practice, it has become clear that two prime kinds of energy

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

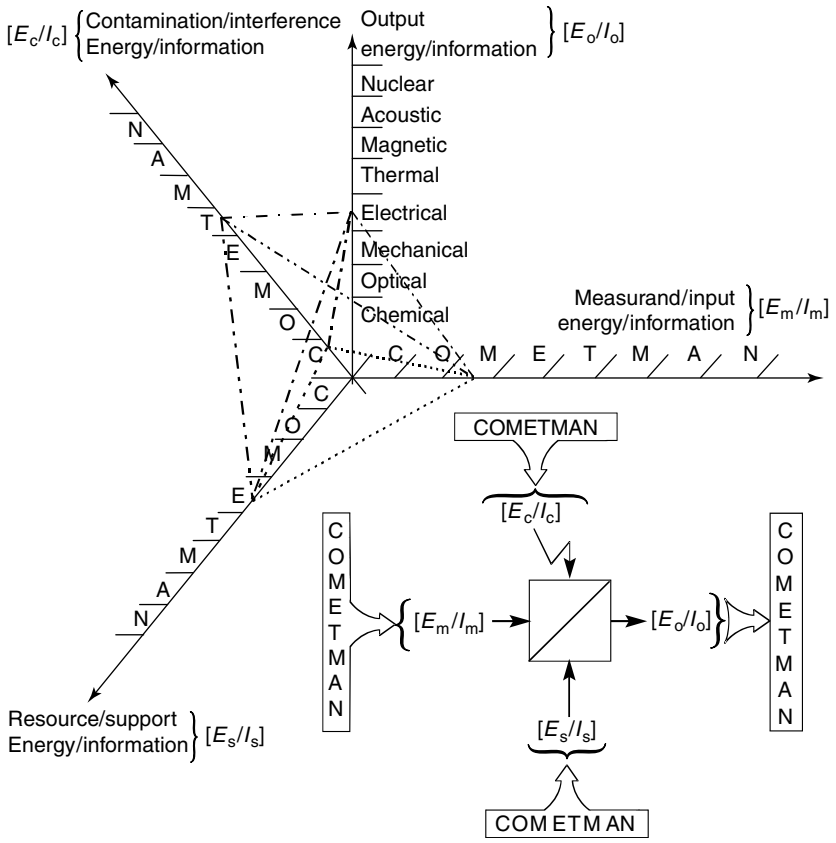


Figure 1. Boundary perspective of a sensor system and its influencing effects.

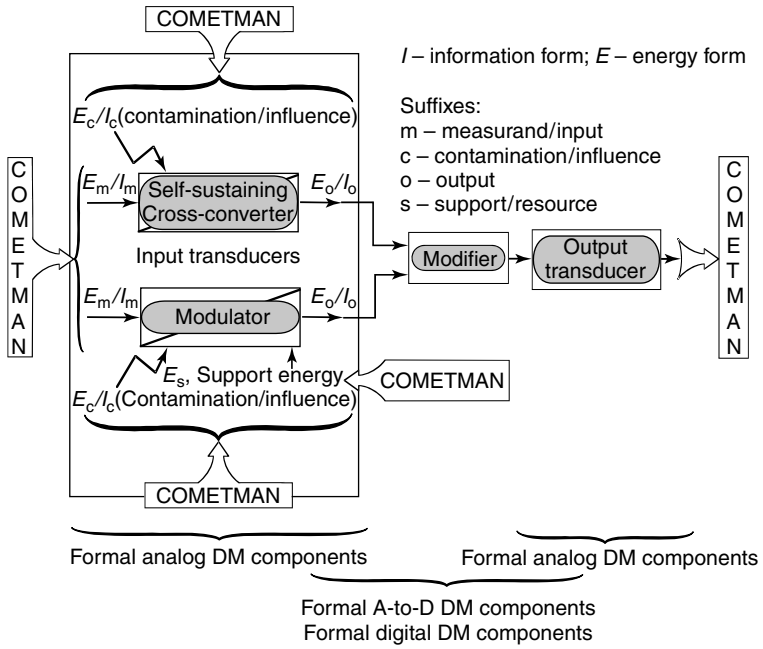


Figure 2. Functional block diagram of a basic instrumentation system illustrating modifier elements and input and output transducers.

handling operational structures are evident in instrumentation systems. Figure 2 illustrates the most basic structure of an instrumentation system made up of input and output transducers and modifiers (Middelhoek and Noorlag, 1981; Middelhoek and Hoogerwerf, 1986). The relationship between this structure and Formal Data Measurement Components, or FDMCs, outlined in **Article 14, Hierarchical Aspects of Measurement Systems, Volume 1**, is also given in this figure.

A transducer is a device that transforms the input signal of one energy form into an output of another energy form, doing so in such a way that a prescribed relationship is maintained between the input and the output signal.

A sensor is an input transducer involving the process of measuring those variables that characterize the flow of information and energy.

An example is the mercury-in-glass thermometer in which heat energy is converted into mechanical energy, forcing the mercury to move along the glass as it expands or contracts.

Another example is the bellows pressure sensor, where an increase in pressure causes the bellows to change its length.

The most commonly used sensors and transducers, which are especially suitable for automation and control purposes, use information carrying signals in the electrical energy form.

It is also important to distinguish between those sensors that do not require the supply of energy from any source other than the system under measurement, and those that do. The former are now commonly referred to as *self-generators* while the latter are referred to as *modulators*. They are also misleadingly referred to as *passive* and *active* respectively.

Self-supporting cross-converting transducers are energy converters, which do not require the supply of energy for support or excitation purposes.

Referring to this species of sensor as a self-generator does not really express the completeness of the functional energy transformation, which they perform. Hence, in some respects, calling this group of sensors self-generators could be viewed as incorrect. Strictly speaking, they should be regarded as converters of an input energy form to a different output energy form. For this reason, it would be of benefit if they were referred to by a name with a more incisive description than self-generators. Such a new name, which should preferably be short, should also clearly and unambiguously convey the idea that no auxiliary energy form is required to support the flow of information in the energy transduction process involved. McGhee *et al.* (1999) have proposed such a renaming, which attempts to express a complex phenomenon by a concise linguistic synonym. In making this renaming proposal, previous contributions to sensor modeling have been taken into account. Stein (1969), Middelhoek and Noorlag (1981) and Middelhoek and Hoogerwerf (1986) use the term *self-generators*. In other places, this group is also referred to as *passive* (Fraden, 1997) or *self-sustaining* (Areny and Webster, 1991).

All of these titles share common shortcomings associated with ambiguity and imprecision. To propose a replacement title, which conveys the important aspects previously stated, requires a clear view of the principles of sensing and the effects used in the sensing process.

Operating principles, including sensor effects, which have been described by Ylilammi (1989), Van Duyn and Middelhoek (1990) and Kwaaital (1993), represent a useful way of classifying sensors. From these principles, it is clear that self-supporting cross-converting sensors convert energy from one form at the input of the sensor to another different form at its output without the use of a supporting energy source. For this reason, they may be called *cross-converters*. However, since some types of modulating sensors also act as cross-converters, the use of the term cross-converter to describe a self-generator will lead to ambiguity. For this reason, McGhee *et al.* (1999) proposed that self-generators should be renamed as self-sustaining cross-converting sensors, or as self-sustaining cross-converters for short. It is likely that this description is probably the most concise but still an adequately complete description is possible.

Examples of self-sustaining cross-converters, misleadingly referred to as passive in various places, are

- the thermocouple, which generates a voltage proportional to input temperature;
- the photovoltaic cell, which gives an electrical output when exposed to light;
- the mercury-in-glass thermometer, whose mercury column expands when exposed to increasing temperature.

The other kind of sensor structure, which needs an excitation or auxiliary supply of energy for its operation, is called a *modulating type*. During their operation, the energy flow is modulated by the input measurand. They are also misleadingly classified as active. Examples of the modulating group of sensors are

- all resistance-based sensors, where electrical energy must be applied to allow the resistance to be measured;
- the photodiode, where a bias electrical current is provided by an electrical support energy source;
- the ultrasonic Doppler flow sensor where an acoustical interrogating energy form is modulated by the flow velocity.

It is recommended that the misleading reference to modulating sensors as active and the naming of self-supporting cross-converting sensors as passive should be avoided in sensor nomenclature. This position is adopted to remove any ambiguity that the terms introduce.

In contrast to other kinds of instrument subsystems, modifiers alter, or modify, the input energy form to give an output in the same energy form. They do not convert the energy to another form. An electronic filter, for instance, processes an electrical, or more commonly electronic, energy signal by modifying the distribution of the frequency content of the signal. It still provides an electrical energy output. Other examples of modifiers are electrical transformers, mechanical levers, and gearboxes, hydraulic rams, and heat exchangers.

Since this article is concerned with sensors, the reader is referred elsewhere for descriptions of modifiers and output transducers (Middelhoek and Noorlag, 1981; Middelhoek and Hoogerwerf, 1986).

3 REPRESENTING AN INSTRUMENT STAGE BY SIGNAL/ENERGY NOTATION

Figure 3(a, b) illustrates representative block diagrams for self-supporting cross-converting and modulating types of sensors with their input and output notation.

It is possible to write down the functional relationship between the signal/energy forms of the output, $[S_o/e_o]$, in Figure 3 in terms of the respective forms for the measurand, $[S_m/e_m]$, the support/resource, $[S_s/e_s]$, and the contamination/interference, $[S_c/e_c]$, inputs. Hence, in the case of a modulating sensor, the functional relationship between the output Signal/Energy form and the three inputs as shown in Figure 3 can be written as

$$\left[\frac{S_o}{e_o}\right] = f\left(\left[\frac{S_m}{e_m}\right], \left[\frac{S_s}{e_s}\right], \left[\frac{S_c}{e_c}\right]\right) \qquad (1)$$

In the case of a self-sustaining cross-converter, there is no support energy. Under these conditions, (1) can be rewritten in the form

$$\left[\frac{S_o}{e_o}\right] = f\left(\left[\frac{S_m}{e_m}\right], \left[\frac{O_{S_s}}{O_{e_s}}\right], \left[\frac{S_c}{e_c}\right]\right) \qquad (2)$$

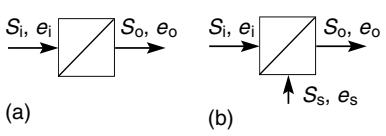


Figure 3. Block diagram and notation for the two types of sensors: (a) self-supporting cross-converting sensors and (b) modulating sensors.

where the term $[O_{S_s}/O_{e_s}]$ indicates that there is no support signal/energy form for self-sustaining cross-converter sensors.

4 CLASSIFICATION OF TRANSDUCERS BY INFORMATION AND ENERGY

Classification in sensing and transduction follows the same general route outlined in **Article 2, Organization of Instrument Science, Volume 1**. This scientific approach to grouping uses the four main orders or problems of classification to assist in the ordering.

In the First Order, known as *Generality or Resemblance*, the question of those relations describing likeness between sensors is considered.

Sensors and transducers must also be grouped on the basis of the Second Problem or Order of taxonomy, which is concerned with their Composition or Collectivity.

While this order is concerned with the relationship of the parts or elements of a sensor system to the whole system and vice versa, the Third Order of classification, called *Hierarchy*, when applied to sensors and transducers, takes account of the relation of rank between the heads or central members of the constituent groups of the elements comprising sensor systems.

These should then be related in the order of composition. In this process, the place occupied by each concrete sensor is assessed in every order relative to other sensors making up both the same order and different orders. Thus, the uniqueness of each sensor in the ordering is distinguished by its similarity with and difference from every other sensor.

Perhaps the most important Fourth Order in the classification of sensors hinges upon the notions of kinship through the relations describing ascent, descent, and collaterality.

As with the orders of Composition and Hierarchy, the last problem of classification just described, which is also known as *Genealogy and Evolution* in the biosciences, is also concerned with the actual sensors to be classified.

These latter orders, when applied for sensor systems, emphasize the importance of a functional and structural hierarchy.

Taxonomy in Sensor Science may thus be regarded as possessing the six main features of its objectives and functions, its materials and activities, together with the methods of discrimination used and hierarchical ordering (**Article 2, Organization of Instrument Science, Volume 1**).

Taxonomy is useful for proving the newness claimed for new sensors. It places a sensing stage into a set position on the basis of its properties. If sensors were so classified, it would show up those that are actually of the same type –

yet appear as a new kind. Too many sensors are constantly being reinvented for lack of knowledge that one already has been reported of the same kind.

5 THE CLASSIFICATION OF ENERGY

Consistent with the previously described orders of classification, it can be seen that the appropriate grouping of sensors must be on the basis of function, structure, information/energy form, and signal output. A good way of grouping transducers uses the transduction principle involved, and the form of the measurand, as shown in Table 1.

Kurt Lion (1969) produced such a classification similar to that in Table 1. In this classification, it is possible to relate the physical effect in the energy domain of the measurand, which allows a transduction into another energy form. Thus, a resistance transducer for measuring displacement is classified as a resistance displacement transducer. Other classification examples are a pressure bellows or a force diaphragm. Notice that all of the diagonal entries in this table correspond to modifiers as defined above.

Questions about energy and information, which are highly relevant in measurement systems, are of fundamental importance and of immediate interest in this article. Energy may be considered on either a macroscopic or microscopic level. There are a number of different, equally valid ways of classifying energy.

To understand the overall classification of energy, it is important to realize that there is a general tendency in physics to represent the largest number of phenomena by the smallest number of models (McGhee *et al*., 1999, 2001).

Physical considerations lead to the classification of gravitation, atomic, magnetic, and electric as possibly being the most representative, irreducible groups.

The energy acronym **GAME**, shown in Table 2, gives a convenient shorthand for these four basic energy forms. Because they lie at the top of the hierarchical grouping of energy domains, they are not very informative from the point of view of either physics or measurement systems.

The GAME energy domains can also be written as the expanded subdomains also given in Table 2. Each of these subdomains is associated with at least one characterizing physical phenomenon. The subdomains of the gravitation and atomic groups of energy forms can be seen to be similar.

Differentiation between both groups is made on the basis of distinguishing potential energy characteristics and kinetic energy manifestations.

Table 1. Examples of transducers and their logical grouping for different input and output signals.

OUT \ IN	Radiation	Mechanical	Thermal	Electrical	Magnetic	Chemical
Radiation	Geometric-optical system filter	Interferometer; Photoelastic transducer	Thermal radiation transducer	Kerr cell; Electro-luminescent transducer, LED	Faraday cell	Transducer based on emission & absorption spectroscopy
Mechanical	Crooke's radiometer; Radiation pressure system	Gravity balance; Ballistic pendulum; Bellows; Gear box	Thermometers; Bimetallic strip; Dilatation thermometer	Electrometer; Electrostrictive transducer; Piezoelectric transducer	Magnetometer; Magnetostriction transducer	Hydrometer; Electrodeposition cell
Thermal	Bolometer	Adiabatic system; Friction calorimeter	Mixing calorimeter; Heat exchanger	Thermal converter	Eddy-current transducer	Combustion calorimeter
Electrical	Photoelectric, Photoresistive & Photogalvanic cells; Photodiode	Resistive; Inductive; Capacitive; Piezo resistor	Thermistor; Thermoelement; Seebeck effect	Transformers; Langmuir probe; Charge collectors; Transistor	Magnetoresistance; Hall-effect transducer	Transducers based on potentiometry, conductimetry, polarography; pH-meter
Magnetic	Curie-effect radiation meter	Magnetoelastic transducer	Paramagnetic transducer device	Electromagnetic transducer; Coil	Magnetic recorder	Magnetic resonance system
Chemical	Photographic emulsion	Ultrasonic chemical transducer	Thermal dye indicators	Electrolytic integrator; Electrical plating	Paramagnetic oxygen analyzer	Chemical system modifiers

Table 2. A conventional classification scheme for energy forms.

Energy forms			
Physics (GAME acronym)	Expanded physical forms	Sensor measurement forms	Characteristic physical phenomenon
Gravitation	Gravitational		Attraction between masses
	Mechanical	Mechanical	Mass in motion has equivalent energy forces, displacements
Atomic	Nuclear	Chemical	Binding energy within nuclei
	Atomic		Binding energy of nuclei and electrons
	Molecular		Bond energy of atoms in molecules
	Thermal	Thermal	Kinetic energy of molecules
Magnetic	Magnetic	Magnetic	Magnetic fields
Electric	Electric	Electrical	Electric fields, potentials and currents
	Electromagnetic radiation	Radiation	Radio waves, through UV to X-, γ-and cosmic rays

The gravitational subgroup is connected with natural forces of attraction between masses.

In a similar way, the bonding energy within nuclei, within atoms and within molecules is also connected with natural forces of attraction but on a microscopic scale. Atomic forces can also be repulsive. This contrasts with the macroscopic scale of gravitational attraction.

Mechanical forces, which are associated with physical motion, cause kinetic energy effects.

The classification of energy using the GAME acronym and its expanded groups may be regarded as the most precise, since the grouping results from the application of the formal principles of taxonomy. It provides an entirely consistent physical view.

Van Dijck (1964) describes another way of grouping energy forms by extracting six from the extended group of nine of the GAME acronym. This abbreviated group of six, which can be identified as being of particular relevance to measurement systems (McGhee *et al*., 1996, 1999, 2001), are called *sensor and measurement energy forms*, in Table 2. These six have been used previously in classifying instrument systems.

This group of six energy forms, used by Kurt Lion (1969) and Middelhoek and coworkers (1981, 1986), are

1. radiant energy,
2. mechanical energy,
3. thermal energy,
4. electrical energy,
5. magnetic energy, and
6. chemical energy.

Gravitational energy, which appears to be unimportant here, is grouped within the mechanical energy form since mechanical variables are more important in measurement.

All of the microscopic potential energy forms are grouped together within the domain of chemical energy. The separation of thermal from the other atomic forms of energy is due to its kinetic energy aspects.

The magnetic form is self explanatory due to its particular manifestation.

Electric circuit variables are distinguished from the entire radiation group.

Although the grouping of energy by the forms in Table 2 is perfectly acceptable, the individual groups are not easily remembered. For this reason, the method used in this article recalls the organization of energy classification given by Peter Stein (1969) in his presentation of the transducer space.

His scheme of energy classification grew from his wide experience as a teacher of measurement and instrumentation. He perceived that the sensor measurement grouping of energy forms given in Table 2, although very close to the central heads of the GAME grouping are not easily remembered. Thus, arose his scheme for classifying energy forms using the memory-aiding (energy acronym COMETMAN, which is summarized in Table 3 (McGhee *et al*., 1999, 2001).

The second column of Table 3 provides the cross-link between it and Table 2. Whereas the previous methods of classification are based upon rigidly linking the energy by similarity, COMETMAN is specifically formulated as a memory aid.

The acronym arises from classifying by Chemical (i.e. molecular), Optical (i.e. radiation), Mechanical (i.e. including gravitation and mass), Electrical (i.e. both electric circuit forms and electromagnetic radiation), Thermal, Magnetic (i.e. including electrical and atomic forms), Acoustic (i.e. correctly speaking within the mechanical group) and Nuclear (i.e. strictly speaking within the atomic group) energy forms.

Once again, the important point to emphasize here is not whether one form of classification is more correct than

Table 3. A classification scheme for energy forms based upon the COMETMAN acronym.

Energy forms			
Modified acronym COMETMAN	Expanded physical forms (GAME acronym)	Physical variables	Human senses
Chemical	Molecular (**A**tomic)	Concentration, reaction rate, redox potential, biological properties, clinical analysis	Smell, Taste
Optical	Electromagnetic radiation (**E**lectric)	Intensity, phase, wavelength, amplitude, transmittance, polarization	Sight
Mechanical	Mechanical (**G**ravitation)	Position, velocity, acceleration, mass, density, force, stress, pressure, moment, torque, shape, roughness, orientation, viscosity	Touch
Electrical	Electric (**E**lectric)	Current, voltage, resistance, inductance, capacitance, permittivity	Pain sensors or nociceptors
Thermal	Thermal (**A**tomic)	Heat flux, temperature, thermal conductivity, thermal capacitance, specific heat, expansion	Touch and sight
Magnetic	Magnetic (**M**agnetic)	Field strength, magnetic moment, permeability, flux density	None
Acoustic	Mechanical (**G**ravitation)	Sound pressure	Hearing
Nuclear	Nuclear, atomic (**A**tomic)		Nociceptors

another. It is the convenience of the most easily remembered grouping that tends to be used, for convenience is an important ingredient in any scheme of classification (Thomson, 1911).

In this classification of energy form, the optical form is extracted from the radiation group, the acoustic from the mechanical group, and the nuclear from the atomic/chemical group, to allow the formation of the acronym.

The inputs of a sensor system then become the energy and information form of the input, the resource energy form, supporting the flow of information, and the energy and information form of the contamination. Thus, the resource and contamination classes have a possible total of eight energy form inputs each.

6 SPATIAL MODELING OF TRANSDUCERS AND SENSORS

Sensor systems have been clearly placed within human–machine supersystems (**Article 5, Humans in the Real World, Volume 1**). This allows sensor systems to be introduced as information machines for measurement (**Article 107, Principles of Sensor Science, Volume 2**) where the importance of sensors in the overall view of measurement was also emphasized. The importance of the boundary view of systems, and consequently of sensor systems, including the concepts of extrinsic and intrinsic sources of interference is highlighted in **Article 68, Setting the System Boundaries, Volume 2**. These sources of interference and their discrimination are essential in fulfilling the primary aim of measurement as stated by Solopchenko (1994). This primary aim is '... taking account of errors, their sources and their quantification ...'. This article describes a spatial tetrahedron method for modeling and for predicting the possible energy sources of interfering effects. Errors are covered in depth in **Article 48, Common Sources of Errors in Measurement Systems, Volume 1**; **Article 49, General Characterization of Systematic and Stochastic Errors, Volume 1**; **Article 50, Errors in Signal Systems, Volume 1**; **Article 51, Errors in Digital Signal Systems, Volume 1**; **Article 52, Error Models, Error Budgets and their Calculation, Volume 1**; **Article 53, Calculation and Treatment of Errors, Volume 1**; **Article 54, Explanation of Key Error and Uncertainty Concepts and Terms, Volume 1**; and **Article 55, Uncertainty Determination, Volume 1**.

A complete spatial representation for the boundary inputs and outputs of a generalized sensor system requires a multidimensional space. As the boundary view of human–machine supersystems reveals (**Article 6, Substructure of Human–Machine Systems, Volume 1**; McGhee *et al.*, 1996; Sandquist, 1985), there is a total of three groups of inputs and three groups of outputs. Hence, the representation becomes too complex for graphical illustration on a two-dimensional paper surface. Previously, Kurt Lion (1969), Peter Stein (1969) and Middelhoek and coworkers 1986, 1981) and McGhee *et al*. (1999, 2001) realized this requirement.

Lion formulated the most basic spatial model by concentrating on the input/output relations already outlined

in Table 1. A highly complex space is immediately reduced to a two-dimensional problem.

Although this model does not include either the support energy inputs or the interference inputs, it is still a convenient starting point for focusing upon transduction as an energy transforming operation. Each one of the input and output groups may be specified as information-bearing signals or energy flowing forms.

Stein (1969) extended the two-dimensional representation of Lion to three dimensions using what he called the *transducer space*. This uses a three-space to illustrate sensing and transduction. Adding a third axis to represent the range of possible support energy forms resulted in a three-dimensional Cartesian space. Each of the axes in this space uses the *COMETMAN* energy form classification. Some time later, Middelhoek and coworkers used the sensor measurement forms of energy classification of Table 2 in the three-space sensor effect cube. A detailed description of spatial transducer modeling is given later in this article.

The 3-D transducer space has one significant disadvantage. In this diagram, no account is taken of contaminating inputs. Consequently, the influence of interference from whatever source is not transparent. A spatial representation for the boundary inputs and outputs of a sensor system, in its most general form, requires a multidimensional space. Any spatial representation, which does not take all inputs and outputs into consideration, has serious shortcomings.

As there is a total of three groups of inputs and three groups of outputs, the representation becomes very complex for graphical illustration on paper surfaces. In spite of this, McGhee *et al*. (1999) extended the transducer space using the concept of the sensor effect tetrahedron, which is also described in some detail later in this article. This model is an adaptation of the information and energy tetrahedron used in system modeling (**Article 6, Substructure of Human–Machine Systems, Volume 1**; McGhee *et al*., 1996, 2001).

Other contributors in the field of measurement have also realized the importance of spatial modeling for simplifying the complex problems of classification in sensors such as in the case of optical sensors. Initial attempts which give good basic insight into optical sensor classification, utilized the three-dimensional structure of the transducer space and the sensor effect cube but named the axes of the representation differently. Subsequently, a more rigorous analysis gave an expanded classification called the *Tri-Angular Pyramid Sensor Representation*, or TAPSER, scheme (El-Hami *et al*., 1993).

The sensor effect tetrahedron, which is described below, is an appropriate combination of the transducer space/sensor effect cube, but using the COMETMAN energy classification, and the multifaced objects from the TAPSER space.

7 THE TRANSDUCER SPACE

It is possible to use the COMETMAN classification of energy to visualize the wide range of possible input and output transducers. To develop this approach, it is necessary to realize that any modulating sensor can be represented by the energy/signal notation in the [*x, y, z*] space, where the measurand energy form, x, modulates the support energy form, z, to produce a variation in the output energy form, y, bearing some relationship to the measurand energy form, x. This model is represented mathematically by

$$y = f(x, y) \tag{3}$$

The representation described by (3) can be drawn in a pseudoisometric view on a two-dimensional page as shown in Figure 4.

Each of the axes in this space is labeled along its length with the COMETMAN acronym to give a three-dimensional, or three-space, energy, and/or signal form of the inputs and outputs. In this way, the space can now provide an aid in understanding the interacting relationship between the support energy form and the input and output information carrying signals in their range of possible energy forms. This is now the three-dimensional transducer space, as it was called by Peter Stein, representing the input signal in its energy form along the conventional x-axis, the support energy form along the conventional z-axis and the output energy along the conventional y-axis as shown in Figure 3. Two examples are used to illustrate the use of the transducer space for modeling energy/information flow in sensors and transducers.

Two examples illustrate the application of Figure 4. The first example uses self-supporting cross-converters, called *self-generating transducers* by many other sensor scientists, and modifiers, which operate without an auxiliary energy source. They are represented in the vertical x-y plane.

Consider, for example, the thermocouple. Thermal information flows from the thermal energy input on the x-axis to the information output on the y-axis transduced to the electrical energy form. The appropriate block diagram also appears in Figure 4.

Modulating types of sensors, as the second example, can be shown in the $[x, y, z]$ Cartesian transducer space in Figure 4 depending upon which support energy source is used.

For example, a photoconductance sensor would have its input information at $x =$ Optical, its output at $y =$ Electrical and its support energy at $z =$ Electrical. The radiating optical input energy incident upon the sensor modulates the flow of electrical energy from the support energy port to the output energy port. For this reason,

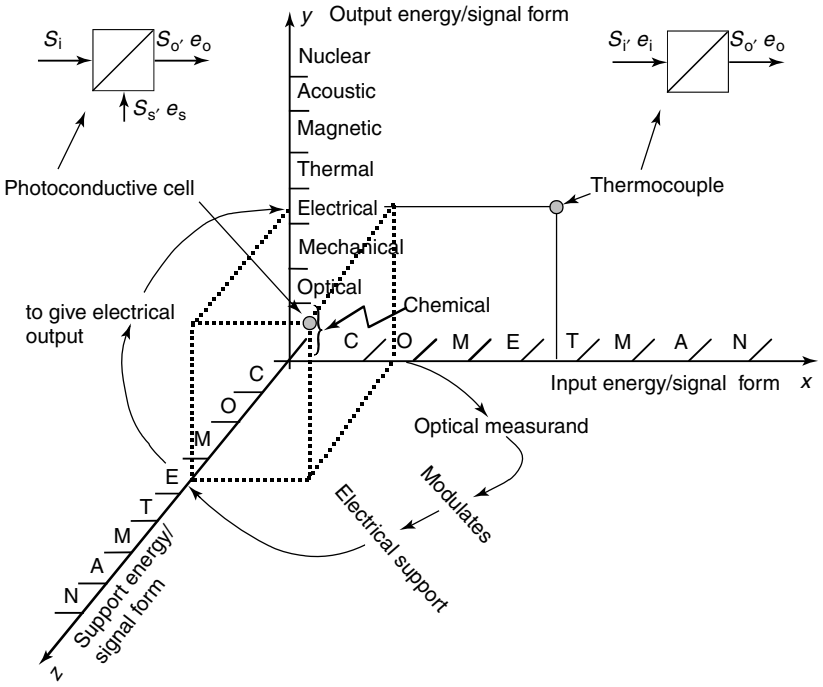


Figure 4. The transducer space – a three-dimensional (three)-space for transducers by energy/information.

modulating sensors of this type bring the advantage of much higher power gain to measurement systems than do self-supporting cross-converters.

The energy/information space diagram, which may also be interpreted as a signal/energy space, allows the prediction or visualization of transducer possibilities. It is clear that there are numerous different possibilities as feasible candidates, when proposing a transduction principle to meet a sensing need. While all known possible combinations of input, output, and support energy forms can be located in this 3-D representation, there are still many positions where practical devices have not been proposed so far.

In accordance with Figure 3, it can be seen from the block diagrams at the top of Figure 4 that any instrument system building block, has three possibilities on inputs and outputs. The self-supporting cross-converting sensor has an input signal, S_m, with its associated input self-supporting or excitation energy, e_m. At the same time its output signal is S_o, and output energy e_o. The difference between the self-supporting and modulating types of cross-converters is a matter of how e_m is supplied to the system as shown in Figure 4.

In self-supporting cross-converting sensors, the system under measurement supports the information flow by supplying energy to the sensor. Modulating sensors, which extract negligible energy from the system under measurement, modulate the support energy input. The energy symbol, e_m, is removed to indicate that negligible energy is drawn from the system under measurement. Inserting the energy symbol, e_S, in the diagram models the effect of modulation of e_S by S_m.

8 THE SENSOR EFFECT TETRAHEDRON FOR MODULATING SENSORS

Starting from the boundary limit viewpoint of a sensor system given in Figures 1, 2 and 3, it is possible to represent a sensor and a transducer effect space, which is more complex than that proposed by either Lion (1969), Stein (1969) or Middelhoek and his coworkers (1981, 1986), but is still useful. In a sensor system, the main output is an information output, which also corresponds to the resource output.

Contamination is also produced at the output. This contamination is the information output when there is cross-coupling into some other information-bearing channel.

It is clear that the output of a sensor system may be regarded as either the resource output, the information output, or even a contaminating source.

Although it is essential to design any system so that the contamination it produces is negligible, this question will not be considered further for the present purposes. Hence, the output energy/information form for a sensor system is given more emphasis in this figure than the contamination produced by the sensor system.

The proposed representation consists of one axis normal to a base plane as shown in Figure 5. Along the base plane of this spatial representation introduce a topological

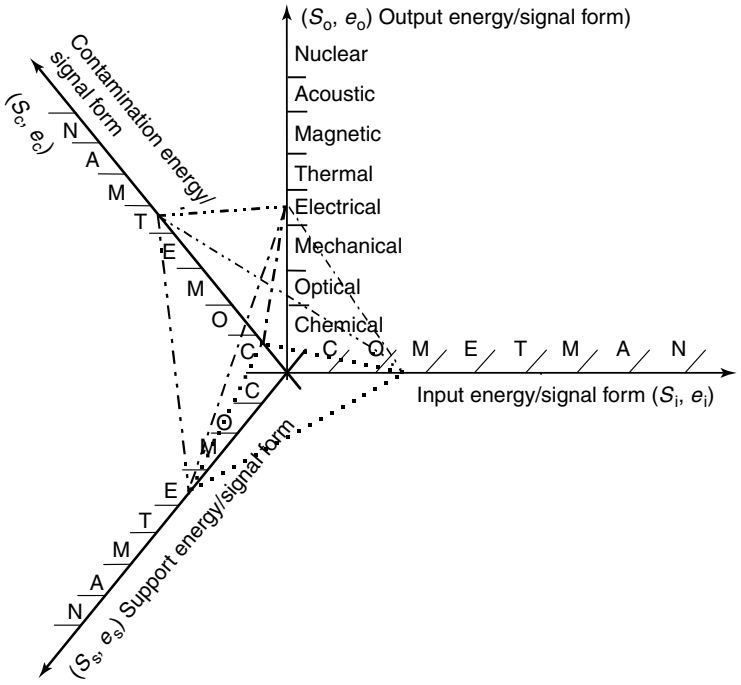


Figure 5. The information and energy tetrahedron for a modulating sensor.

grouping consisting of three axes set mutually at 120° spacing. Each of these axes, taken as representing one of the three input groups with their COMETMAN information or energy classification, clarifies, and justifies the way in which the acronym aids memory. The organization of the three axes on this base plane is a collapsed, two-dimensional representation of the input three-space consisting of the information, resource, and contamination inputs.

This organization of inputs may be regarded further as the input triangle. Such a name is appropriate as illustrated by the input triangle given by the dotted lines in Figure 5. Here, the support energy form is electrical, the contamination energy/information form is chemical, and the measurand energy form is mechanical using a sensor like a resistance strain gauge.

A strain gauge, which measures a mechanical energy domain variable, has been chosen as a typical, but not exclusive, example. It is straightforward to represent resistance transducers for variables in other measurand energy domains. The chemical contamination could be responsible for parasitic electrochemical potentials at the contacts in the electrical circuit.

Each point on the input triangle axes may be joined to an energy/information form on the output axis. When the output is in the form of an information-bearing electrical energy form, the dash-dot line combinations shown in Figure 5 allow the link between all three input forms and the output form to be clearly seen. The geometrical shape, which is outlined by the dotted and the dash-dot lines, has one base plane surface and three inclined surfaces. This represents a tetrahedron or four-faced object. This way of representing the boundary impacts on sensor systems may be called the *information/energy tetrahedron* or the *sensor/transducer effect tetrahedron*.

Illustration of the manner in which the contamination inputs influence the output may be extended to cover contamination from other energy forms. The dot-dot-dash lines in Figure 5 show the case when the contaminating energy form is due to the influence of temperature on the wire of the strain gauge. The deep perception allowed by the sensor/transducer effect tetrahedron given in Figure 5, which is fairly straightforward, is still realistic and comprehensive.

It is quite important to note that the sensor effect tetrahedron extends the transducer space by including the possibility for predicting the energy forms of the various contaminating sources that may be present.

McGhee *et al*. (1999) illustrate the application of the sensor effect tetrahedron to a modulating photodiode, a modulating ultrasonic Doppler flow sensor and self-supporting cross-converting thermocouples.

9 THE SENSOR EFFECT TETRAHEDRON FOR SELF-SUPPORTING CROSS-CONVERTING SENSORS

So far the conditions in Figure 5 correspond to the case of modulating types of sensors and transducers.

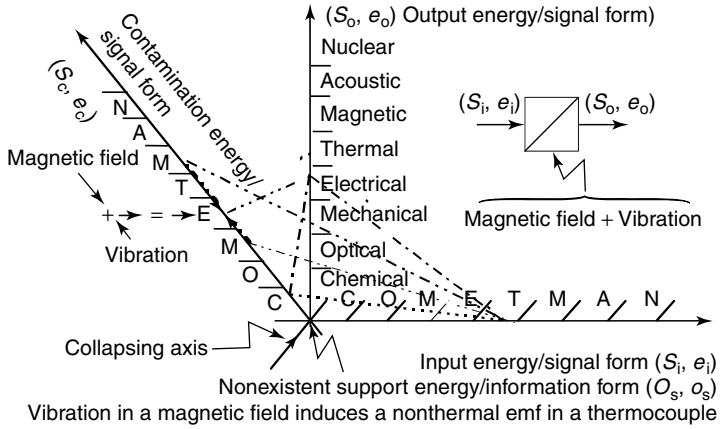


Figure 6. The sensor effect tetrahedron for self-supporting cross-converting sensors using a thermocouple as an example.

Self-supporting cross-converting sensor possibilities may also be represented by the sensor/transducer effect tetrahedron as shown in Figure 6 using a thermocouple for illustrative purposes. In self-supporting cross-converting sensors, there is no support energy form. Consequently, the support energy axis of the input triangle collapses to the origin of the system of axes in the tetrahedron. This origin, usually representing the number *zero* in a numerical space, may be regarded as the point defining the nonexistence of a support information/energy form. Under these conditions, (2) is the mathematical representation for the sensor.

The case for the sensor effect tetrahedron representation of a thermocouple is illustrated in Figure 6 for two types of interference/contamination sources. The chemical interference shown may give rise to a series mode parasitic electrochemical emf in the electrical circuit of the thermocouple as in the previous example for the modulating sensor. A situation illustrating the interaction of different energy forms to produce a propagated contaminating signal is also shown in Figure 6.

If the thermocouple sensor assembly, including its lead wires, are mounted on a structure that is vibrating in an electromagnetic field, the sensor effect tetrahedron allows the possibility of an induced series mode interfering electrical signal to be predicted and visualized. The powerful predictive property of the sensor effect tetrahedron is well illustrated by this example.

The representation of self-supporting cross-converting sensors and transducers shown in Figure 6 may be called the *self-supporting sensor tetrahedron*.

This model of a collapsing four-space may also be viewed as forming a three-space model, when the input and contaminating axes are drawn mutually perpendicular. This is analogous to the effect, which occurs in Stein's transducer space for self-supporting cross-converting sensors given in Figure 4. Whereas this transducer space collapses to a plane, or two-space, from a three-space or 3-D space, in the case of self-supporting cross-converting transducers, the sensor effect tetrahedron collapses to a three-space for the same kind of sensor. The different results of the collapsing are due to the fact that the sensor effect tetrahedron represents the input information/energy groupings using the idea of the input triangle illustrated in Figures 5 and 6 with the added transparency resulting from the inclusion of the interfering energy source forms.

RELATED ARTICLES

Article 41, Terminology and Classification of Measurement Systems, Volume 1; **Article 108, Transducer Fundamentals, Volume 2**; **Article 110, Signal/Energy Matrix Modeling, Volume 2**; **Article 111, Classification of Sensors, Volume 2**; **Article 112, Systematic Description of Sensors, Volume 2**.

REFERENCES

Areny, R.P. and Webster, J.G. (1991) *Sensors and signal conditioning*, John Wiley & Sons, New York.

El-Hami, M., Finkelstein, L., Grattan, K.T.V. and Palmer, A.W. (1993) Expert System Application Using Optical Fibre Sensor Classification Data. *Sensors and Actuators*, **A-39**, 181.

Fraden, J. (1997) *Handbook of Modern Sensors*, American Institute of Physics, New York.

Kwaaital, T. (1993) The Fundamentals of Sensors. *Sensors and Actuators*, **39**, 103–110.

Lion, K.S. (1969) Transducers: Problems and Prospects. *IEEE Transactions*, **IECI-16**, 2–5.

McGhee, J., Henderson, I.A., Korczynski, M.J. and Kulesza, W. (1999) The Sensor Effect Tetrahedron: An Extended Transducer Space. *Measurement*, **24**, 217–236.

McGhee, J., Henderson, I.A., Kulesza, W. and Korczynski, M.J. (2001) *Measurement Data Handling: Theoretical Technique*, ISBN 83-7283-007-X, The Technical University of Lodz, Lodz.

McGhee, J., Korczynski, M.J., Kulesza, W. and Henderson, I.A. (1996) *Scientific Metrology*, ISBN 83 904299 9 3, A.C.G.M. Lodart, Lodz.

Middelhoek, S. and Hoogerwerf, A.C. (1986) Classifying Solid-State Sensors: The 'Sensor Effect Cube'. *Sensors and Actuators*, **10**, 1.

Middelhoek, S. and Noorlag, D.J.W. (1981) Three Dimensional Representation of Input and Output Transducers. *Sensors and Actuators*, **2**, 29.

Sandquist, G.M. (1985) *Introduction to Systems Science*, Prentice Hall, Englewood Cliffs, NJ.

Solopchenko, G.N. (1994) Formal Metrological Components of Measuring Systems. *Measurement*, **13**, 1–12.

Stein, P.K. (1969) The Engineering of Measurement Systems. *Journal of Metals*, **21**, 40.

Thomson, A.J. (1911) *Introduction to Science*, Williams & Norgate Ltd, London.

Van Dijck, J.G.R. (1964) *The Physical Basis of Electronics*, Centrex, Eindhoven/Macmillan, London.

Van Duyn, D.C. and Middelhoek, S. (1990) Information Transduction in Solid-State Transducers: A General Thermodynamic Systems Approach. *Sensors and Actuators*, **A-21–A23**, 25.

Ylilammi, M. (1989) Thermodynamics of Sensors. *Sensors and Actuators*, **18**, 167–178.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

110: Signal/Energy Matrix Modeling

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Principles of Signal/Energy Matrix Modeling 737
2 Case Studies in Signal/Matrix Modeling of Sensors – Shaft Torque Measurement 738
3 Case Studies in Signal/Energy Modeling – Level Measurement 739
Related Articles 740
References 740
Further Reading 740

1 PRINCIPLES OF SIGNAL/ENERGY MATRIX MODELING

The basic ideas of classification in sensors and transducers allow a visualization of the diverse possibilities for realizing measuring functions.

Another very important aid in visualizing how the energy and signal are transduced uses the concept of a signal/energy flow matrix (Middelhoek and Noorlag, 1981).

Figure 1 illustrates this for two different kinds of temperature sensors. One of these is a mercury-in-glass thermometer and another is a modern digital thermometer.

The fundamental block diagram representation of input and output sensing/transducing elements and modifying elements, shown in Figure 2 of **Article 109, Structure and Energy in Sensor Systems, Volume 2**, is used to construct the matrix shown in Figure 1. Above this block diagram in Figure 1, is a matrix of rows and columns with the left-hand column indicating the COMETMAN classification scheme for input energy/signal form. The energy domain of the temperature input signal is, of course, the thermal domain.

Consider first when the sensor is a mercury-in-glass thermometer. The input transduction process transforms the thermal energy into mechanical energy, due to the cubic expansion of the mercury in the bulb of the thermometer. This expansion is propagated through the thermometric fluid, the mercury, by modifying the original mechanical cubic expansion into a change in the length of the mercury column. This column length belongs to the group of mechanical variables.

Radiation, which is incident upon the column, is modulated by the reflective, refractive, and absorptive properties of the mercury column so that the human eye perceives the change in length. This perception is by a noncontacting form of sensing through the 108 videceptors in the human sense of seeing, which are mentioned with the other human senses in Figure 2 of **Article 5, Humans in the Real World, Volume 1**.

If the thermometer is a digital electronic unit, the thermal energy is transduced to the electrical domain using a suitable electrical temperature sensor, which could be any one within the contacting group in the classification of temperature sensors. Modification of the electrical energy at the electrical output port of resistive sensors could use a Wheatstone bridge circuit.

Further modification could be performed with an electronic instrumentation amplifier. Subsequently, the analog signal at the output of the instrumentation amplifier is converted to a digital form using an analog-to-digital converter, or ADC. The digital output of the ADC is then encoded and applied to a light emitting diode, or LED. The LED is a photon generator since it transduces its digital electrical input signal to visible radiation at its output.

It is also worthwhile illustrating the application of the signal/energy matrix model of sensors using two other

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

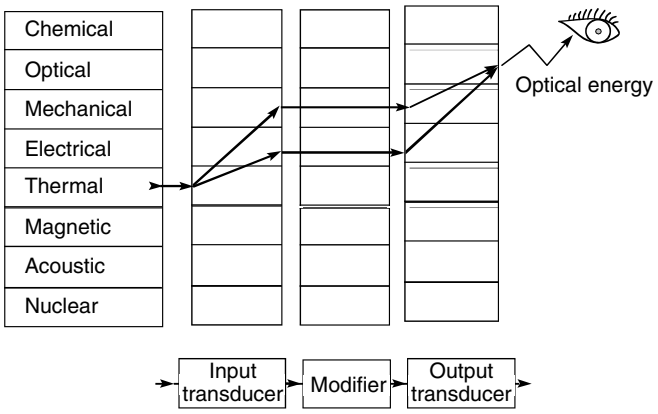


Figure 1. Generalized but simplified energy regime for a mercury-in-glass thermometer and a modern digital thermometer.

examples, which will be presented as case studies. The first example illustrates two methods of measuring mechanical torque, while the second demonstrates two different transduction principles for level measurement.

2 CASE STUDIES IN SIGNAL/MATRIX MODELING OF SENSORS – SHAFT TORQUE MEASUREMENT

The instantaneous value of axial torque on a shaft is to be measured and displayed using an appropriate method. Simultaneously, storage of the measured torque signal is required on a computer diskette or magnetic tape for further analysis and processing. Only two methods of measuring torque will be considered in the proposed design. In the first method, a bonded strain gauge is to be used as the sensor, while in the second an optical method is recommended.

The block diagrams of two torque-measuring instruments are given in Figures 2 and 3. In Figure 2, strain gauges are used to sense changes in applied mechanical force by transducing mechanical force to electrical resistance.

To measure axial torque, a rosette of strain gauges (McGhee *et al*., 1996) should be bonded to the shaft under measurement. Mechanical torque is converted into strain on the shaft surface. The strain gauges bonded to the shaft at angles of 45° to the perpendicular, are under strain. Two of them are under tension and two are under compression. The strain gauges both change their dimension depending upon changes in the forces on the shaft surface.

Dimensional changes are transduced to changes in the resistance of the strain gauges. These resistance variations are then measured, after suitable conditioning, by applying a Wheatstone bridge. The output of the bridge is amplified, converted into a digital signal in an ADC, whose output is fed to a driver for an LED or an LCD display element. The LED operates on the principle of optical photon generation.

In the case of the LCD (liquid crystal display), the electrical input modulates the structure of the liquid crystal whose orientation acts as a modulator of ambient photons. Recording on computer disk or magnetic tape is a transduction from the electrical energy form to the magnetic energy form. The block diagram of the whole instrument is also shown in Figure 2. As an exercise, the reader should write a short description of the energy transductions and signal flows that are shown in Figure 2.

In the optical method of torque measurement represented in Figure 3, the principle of optical modulation is proposed

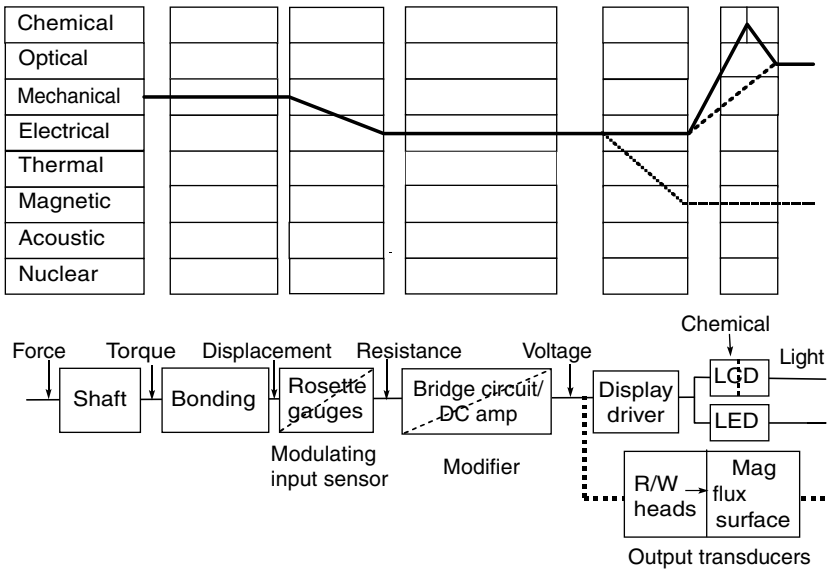


Figure 2. The signal/energy domain matrix and block diagram of a proposed method for axial torque measurement using strain gauges.

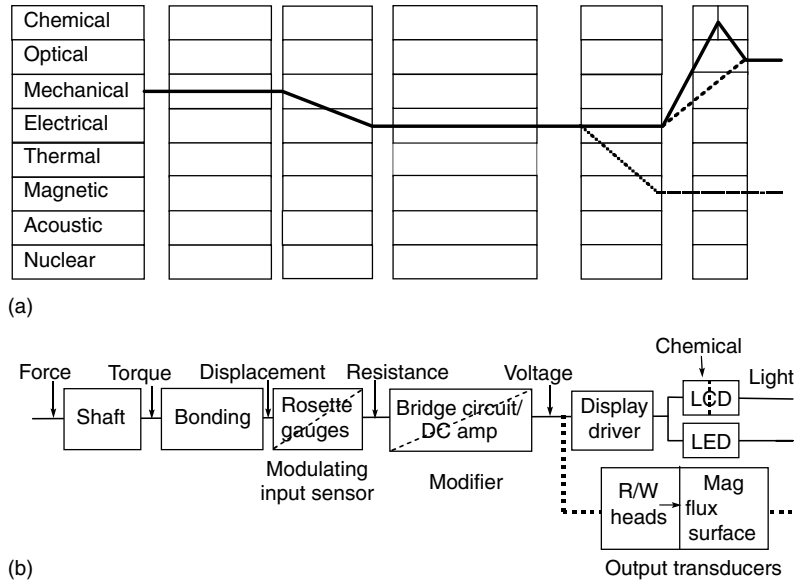


Figure 3. (a) The signal/energy domain matrix and (b) block diagram of a proposed method for axial torque measurement using torque to modulate a light beam.

as a suitable solution. Two rotating circular vanes, each having a regular array of small holes at a fixed radius from their centers, are attached to a shaft with a separation of l mm between them. Light sources are attached adjacent to the outside face of one vane. An array of stationary photodetectors is mounted alongside the outside of the other vane.

The light from the sources passes through the holes in the first vane. If the second vane is aligned properly, there will be a maximum amount of light incident on the photodetectors. When a torque is applied to the shaft, misalignment of the arrays of holes in the vanes will occur. This misalignment causes a reduction in the light that is incident on the photodetectors. The applied shaft torque, a rotary mechanical energy form, modulates the amount of light that is incident upon the photodetectors.

3 CASE STUDIES IN SIGNAL/ENERGY MODELING – LEVEL MEASUREMENT

It is required to measure the level of fluid in a container, to display the measured value, and to record it upon either

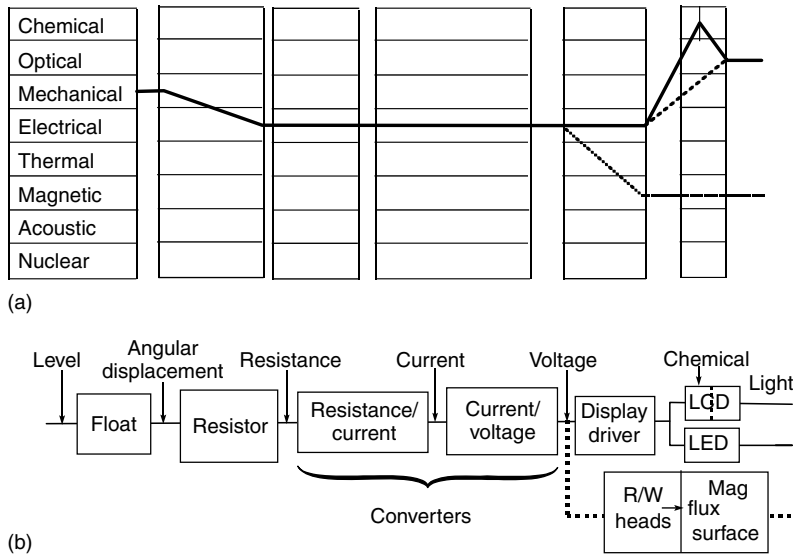


Figure 4. (a) The signal/energy domain matrix and (b) block diagram of a proposed method for level measurement using a float and a rotary potentiometer.

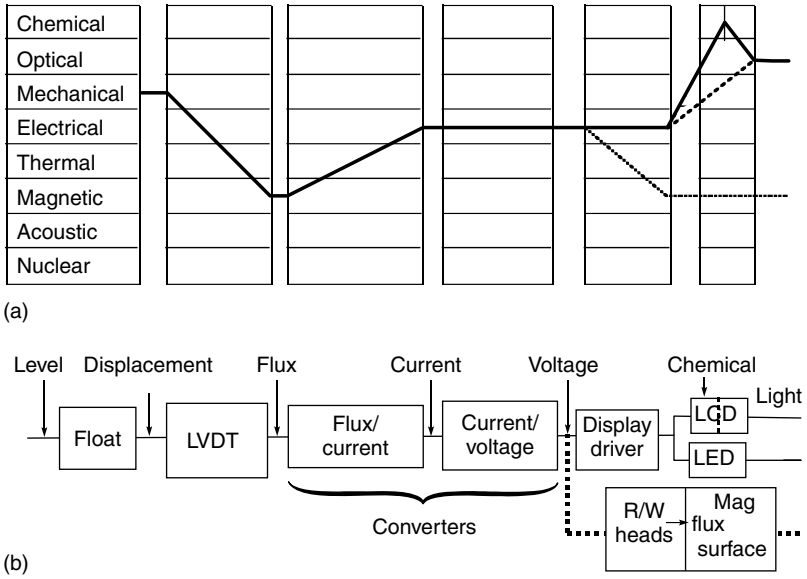


Figure 5. (a) The signal/energy domain matrix and (b) block diagram of a proposed method for level measurement using a float and a linear variable differential transformer, or LVDT.

a magnetic diskette or a magnetic tape. The signal/energy domain flow diagrams and block diagrams are shown for each respective system in Figures 4 and 5. Since it is essential to have some primary sensing element to transduce the level, both the proposed methods of level measurement use a float as a primary sensing element. The first also uses a rotary potentiometer and the second uses a linear variable differential transformer (LVDT). As the schematics are quite explanatory, it is left to the reader to complete a written description of both – the matrix and the block diagram.

Further explanations of the method are given in Sydenham, Hancock and Thorn (1992).

RELATED ARTICLES

Article 109, Structure and Energy in Sensor Systems, Volume 2; **Article 111, Classification of Sensors, Volume 2**; **Article 112, Systematic Description of Sensors, Volume 2**.

REFERENCES

McGhee, J., Henderson, I.A., Kulesza, W. and Korczynski, M.J. (1996) *Scientific Metrology*, ISBN 83-904299-9-3, printed by A.C.G.M. LODART, Lodz.

Middelhoek, S. and Noorlag, D.J.W. (1981) Three Dimensional Representation of Input and Output Transducers. *Sensors and Actuators*, **2**, 29.

Sydenham, P.H., Hancock, N. and Thorn, R. (1992) *Introduction to Measurement Science and Engineering*, Wiley, Chichester.

FURTHER READING

Middelhoek, S. and Hoogerwerf, A.C. (1986) Classifying Solid-State Sensors: The 'Sensor Effect Cube'. *Sensors and Actuators*, **10**, 1.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

111: Classification of Sensors

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Need for Classification of Sensors 741
2 Classification Rules in General 741
3 Classification by Energy 742
4 Applying the Energy Classification 743
5 Three-space Modeling of Transducers and Sensors 743
6 The Sensor Effect Tetrahedron 745
7 Signal/Energy Matrix Modeling of Sensors and Transducers 747
8 General Principles of Temperature Sensor Classification 748
9 Temperature Measuring Ranges of Temperature Sensors 749
Related Articles 750
Reference 750

1 NEED FOR CLASSIFICATION OF SENSORS

New ideas for sensors are constantly being invented – and reinvented. This is not efficient for in many cases the knowledge on a particular sensor form is already available and is redeveloped instead of being advanced with each new application and experience.

With the existence of hundreds of potential transduction principles, and numerous ways in which they can be implemented, it is possible to form thousands of sensor forms that are often little different to others yet to the new inventor, seem novel and a contribution.

For example, a simple electrical strain gauge mounted on a thin circular diaphragm can be used to measure many variables. Table 1 lists many of the methods used.

What is needed is a universal way to classify sensors as this assists retrieval of information regarding their existence, reveals new combinations, and links potential users to deeper knowledge about them.

In the life sciences, classification and taxonomy has been a key branch of scholarship for centuries. It was needed to assist understanding of the myriad of living objects. With respect to sensors, however, there has not been much attention paid to setting up a serious universal classification schema. This article outlines the main approaches that can be taken – the energy classification that has a universal application but does not cover the form of sensors, and how temperature sensors have been classified as an example of a species taxonomy.

2 CLASSIFICATION RULES IN GENERAL

Classification in sensing and transduction follows the scientific approach for grouping objects using the four main orders of classification to assist in the ordering.

In the First Order, known as Generality or Resemblance, the question of those relations describing likeness between sensors is considered.

Sensors and transducers must also be grouped on the basis of the Second Problem or Order of taxonomy, which is concerned with their Composition or Collectivity.

While this order is concerned with the relationship of the parts or elements of a sensor system to the whole system and vice versa, the Third Order of classification, called *hierarchy*, when applied to sensors and transducers,

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

Table 1. Use of simple instrumented diaphragm to measure different measurands.

Sensor	Method of using the diaphragm
Pressure gauge	Strain gauge on round diaphragm in sealed chamber under pressure
Small deflection meter	Diaphragm used with pin resting on it to deflect diaphragm
Sound pressure meter	Sensitive form of pressure gauge above with large cone connected to the diaphragm
Volume sensor	Chamber with pressure diaphragm at bottom measuring force exerted by height of fluid or solid material
Accelerometer	Flexibility of diaphragm with added mass forms seismic system
Level sensor	Chamber with pressure diaphragm at bottom measuring force exerted by height of fluid or solid material

takes account of the relation of rank between the heads or central members of the constituent groups of the elements comprising sensor systems. These should then be related in the order of composition.

In this process, the place occupied by each concrete sensor is assessed in every order relative to other sensors making up both the same order and different orders. Thus, the uniqueness of each sensor in the ordering is distinguished by its similarity with, and difference from, every other sensor.

Perhaps the most important Fourth Order in the classification of sensors hinges upon the notions of kinship through the relations describing ascent, descent, and collaterality.

As with the orders of Composition and Hierarchy, the last problem of classification just described, which is also known as Genealogy and Evolution, is also concerned with the actual sensors to be classified. These latter orders, when applied for sensor systems, emphasize the importance of a functional and structural hierarchy. Taxonomy in Sensor Science may thus be regarded as possessing the six main features of its objectives and functions, its materials and activities, together with the methods of discrimination used and hierarchical ordering.

3 CLASSIFICATION BY ENERGY

Consistent with these orders of classification, it can be seen that the appropriate universal grouping of sensors must be on the basis of:

- function
- structure
- energy form
- signal output.

Table 2. A scheme of classification for energy forms.

Energy form	Characteristic physical phenomenon
Electromagnetic radiation	Radio waves, through to cosmic rays
Gravitational	Attraction between masses
Mechanical	Lifting, stretching, and movement
Thermal	Kinetic energy of molecules
Electrical	Electric and magnetic fields, currents
Molecular	Bond energy of atoms in molecules
Atomic	Binding energy between nuclei and electrons
Nuclear	Binding energy within nuclei
Mass	Relativistic energy

Energy and signal are of immediate interest. Energy may be considered on either a macroscopic or microscopic level.

Serious consideration of the fundamentals of physics shows that the most basic starting point might be at its four prime forces – electric, magnetic, gravitational, and atomic. These, however, are too much removed from practice to be useful to designers of sensing systems.

There are a number of different, equally valid ways of classifying the energy found in the physical world. For example, a comprehensive but not exclusive grouping of nine various types of existing energy forms is given in Table 2; it stems from experience.

From this group of nine, an abbreviated group of six can be identified as being of general relevance to instrument systems. These six, used by Kurt S. Lion and later by Simon Middelhoek and coworkers, are

1. radiant energy
2. mechanical energy
3. thermal energy
4. electrical energy
5. magnetic energy
6. chemical energy.

Transducers can also be classified according to the transduction principle involved and the form of the measurand. Thus, a resistance transducer for measuring displacement is classified as a resistance displacement transducer. Other classification examples are pressure bellows or a force diaphragm. In the classification, grouping on the basis of transduction principle has been developed by a number of people. Kurt S. Lion once produced such a classification similar to that shown in Table 2. In this classification, it is possible to relate the physical effect in the energy domain of the measurand that allows a transduction into another energy form. The topic is also discussed under the modeling of sensors in **Article 112, Systematic Description of Sensors, Volume 2**.

4 APPLYING THE ENERGY CLASSIFICATION

Although these methods of grouping by energy form are perfectly acceptable, they are not easily remembered in a consistent manner. For this reason, another possibility is organization by energy classification given by Peter K. Stein in his presentation of the transducer space.

In this scheme, the energy forms are classified with the memory-aiding acronym COMETMAN. Whereas the previous methods of classification are based upon a more rigid linking of the energy by similarity, COMETMAN is specifically designed as a memory aid. The acronym comes from classifying by

- **C**hemical (i.e. molecular)
- **O**ptical (i.e. radiation)
- **M**echanical (i.e. including gravitation and mass)
- **E**lectrical
- **T**hermal
- **M**agnetic (i.e. including electrical and atomic forms)
- **A**coustic (i.e. correctly speaking within the mechanical group)
- **N**uclear energy forms.

Once again, the important point to emphasize here is not whether one form of classification is more correct than another. Rather, it is from the convenience of that which is most easily remembered.

5 THREE-SPACE MODELING OF TRANSDUCERS AND SENSORS

It is possible to use this COMETMAN classification of energy to visualize the wide range of possible input and output transducers from two approaches.

The first approach uses a three-dimensional or three-space energy and/or signal form of the inputs and outputs to provide an aid to understanding the interacting relationship between the form of support energy used and the input and output information carrying signals in their range of possible energy forms.

This three-dimensional transducer space represents the input signal in its energy form along the conventional x-axis, the support energy form along the conventional z-axis, and the output energy along the conventional y-axis as shown in Figure 1.

Self-generating transducers and modifiers, which operate without an auxiliary energy source, lie in the vertical x–y plane. Consider, for example, the thermocouple in which thermal information flows from the thermal energy input on the x-axis to information output on the y-axis transduced to the electrical energy form.

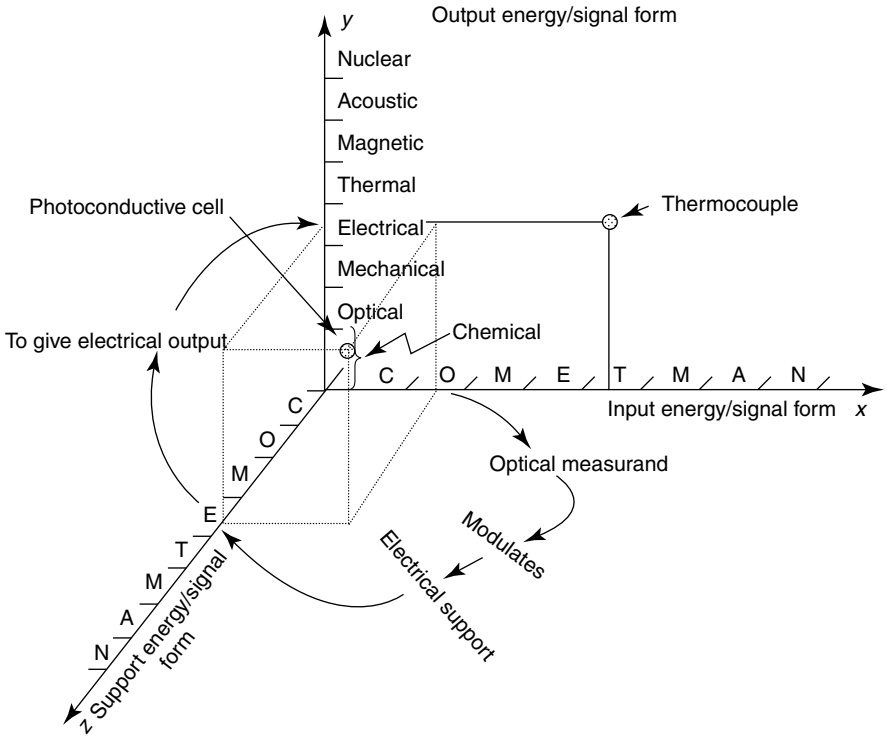


Figure 1. The three-dimensional (three) space for transducers by energy/information.

Table 3. Examples of transducers and their logical grouping for different input and output signals.

OUT \ IN	Radiation	Mechanical	Thermal	Electrical	Magnetic	Chemical
Radiation	Geometric-optical system filter	Interferometer; Photoelastic transducer	Thermal radiation transducer	Kerr cell; Electro-luminescent transducer, LED	Faraday cell	Transducer based on emission & absorption spectroscopy
Mechanical	Crooke's radiometer; Radiation pressure system	Gravity balance; Ballistic pendulum; Bellows; Gear box	Thermometers; Bimetal strip	Electrometer; Electrostrictive transducer; Piezoelectric transducer	Magnetometer; Magnetostriction transducer	Hydrometer; Electrodeposition cell
Thermal	Bolometer	Adiabatic system; Friction calorimeter	Mixing calorimeter; Heat exchanger	Thermal converter	Eddy-current transducer	Combustion calorimeter
Electrical	Photoelectric, Photoresistive & Photogalvanic cells; Photodiode	Resistive; Inductive; Capacitive; Piezo resistor	Thermistor; Thermoelement; Seebeck effect	Transformers; Langmuir probe; Charge collectors; Transistor	Magnetoresistance; Hall-effect transducer	Transducers based on potentiometry, conductimetry, polarography; pH-meter
Magnetic	Curie-effect radiation meter	Magnetoelastic transducer	Paramagnetic transducer device	Electromagnetic transducer; Coil	Magnetic recorder	Magnetic resonance system
Chemical	Photographic emulsion	Ultrasonic chemical transducer	Thermal dye indicators	Electrolytic integrator; Electrical plating	Paramagnetic oxygen analyzer	Chemical system modifiers

Modulating types of sensors can be shown in the [x, y, z] 'transducer space' in Figure 1 according to which support energy source is used. A photo-conductance sensor would have its input information at $x =$ optical, its output at $y =$ electrical, and its support energy at $z =$ electrical. The radiating optical input energy incident upon the sensor modulates the flow of electrical energy from the support energy port to the output energy port. For this reason, modulating sensors of this type bring the advantage of much higher power gain to measurement systems than do self-generators.

The energy/information space diagram allows the prediction or visualization of transducer possibilities. It is clear that there are numerous different possibilities as feasible candidates when proposing a transduction principle to meet a sensing need. While all known possible combinations of input, output, and support energy forms can be located in this three-dimensional representation, there are still many positions where practical devices have not been proposed. Table 3 lists many forms in terms of the energy classification.

An instrument system building block, therefore, has three possibilities for inputs and outputs. The self-generating sensor has an input signal S_i, with its associated input self-generating or excitation energy e_i. At the same time, its output signal is S_o and output energy e_o. The difference between the self-generating and modulating types is a matter of how e_i is supplied to the system as shown in Figure 2.

In self-generating sensors, the system under measurement supports the information flow by supplying energy to the sensor.

Modulating sensors, which extract negligible energy from the system under measurement, modulate the support energy input. The energy symbol e_i is removed to indicate that negligible energy is drawn from the system under measurement. The energy symbol e_s is inserted in the diagram to model the effect of modulation of e_s by S_i.

It is clear that any instrument stage can be represented by the energy/signal notation in the [x, y, z] transducer

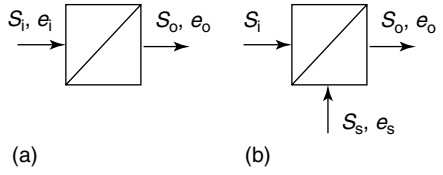


Figure 2. Block diagram and notation for (a) self-generating and (b) modulating sensors.

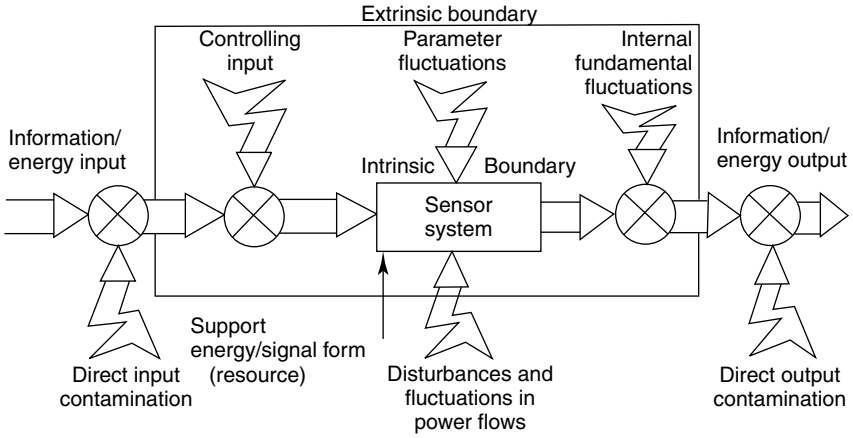


Figure 3. Boundary perspective of a sensor system and its influencing effects.

space, in which the measurand energy form, x, modulates the support energy form, z, to produce a variation in the output energy form, y, bearing some relationship to the measurand energy form, x. These possibilities are all included in Figure 2. The diagonals in the blocks represent the general transduction from one energy form to another. The relationships on Figure 1 and Figure 2 may be written as a mathematical relation with the form

$$y = f(x, y) \tag{1}$$

6 THE SENSOR EFFECT TETRAHEDRON

The three-dimensional transducer space has one significant disadvantage. In this diagram, no account is taken of contaminating inputs. Consequently, the influence of interference from whatever source is influencing the system is not transparent. A spatial representation for the boundary inputs and outputs of a sensor system, in its most general form, requires a multidimensional space. As there is a total of three groups of inputs and three groups of outputs, the representation becomes too complex for graphical illustration.

This situation can be understood by starting from the various perspectives at the boundaries of a sensor system, as are given in Figure 2, Figure 3, and Figure 4. It is possible to represent a sensor and transducer effect space that is more complex than the schemes proposed previously.

In spite of the increased complexity, this sensor effect space is still useful, since it also includes all of the possibilities for contamination inputs.

In a sensor system, the main output is an information output, which also corresponds to the resource output. Although contamination is also produced at the output, it is not necessary to consider it for the present purposes.

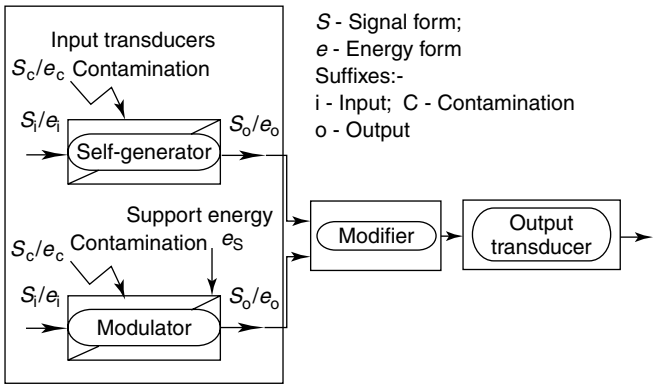


Figure 4. Functional block diagram of a basic instrumentation system illustrating modifier elements and input and output transducers.

Hence, the output energy/signal form for a sensor system is given more emphasis in this figure than the contamination produced by the sensor system at its output.

The proposed representation consists of one axis normal to a base plane as shown in Figure 5. Along the base plane of this spatial representation, introduce a topological grouping consisting of three axes set mutually at 60° spacing. Each of these axes, taken as representing one of the three input groups with their COMETMAN information or energy classification, clarifies and justifies the way in which the acronym aids memory. The organization of the three axes on this base plane is a collapsed, two-dimensional representation of the input three-space consisting of the information, resource, and contamination inputs. This organization of the inputs may be regarded further as the input triangle. Such a name is appropriate as illustrated by the input triangle given by the dotted lines in Figure 5. Here, the support energy form is electrical, the contamination energy/information form is chemical, and the measurand

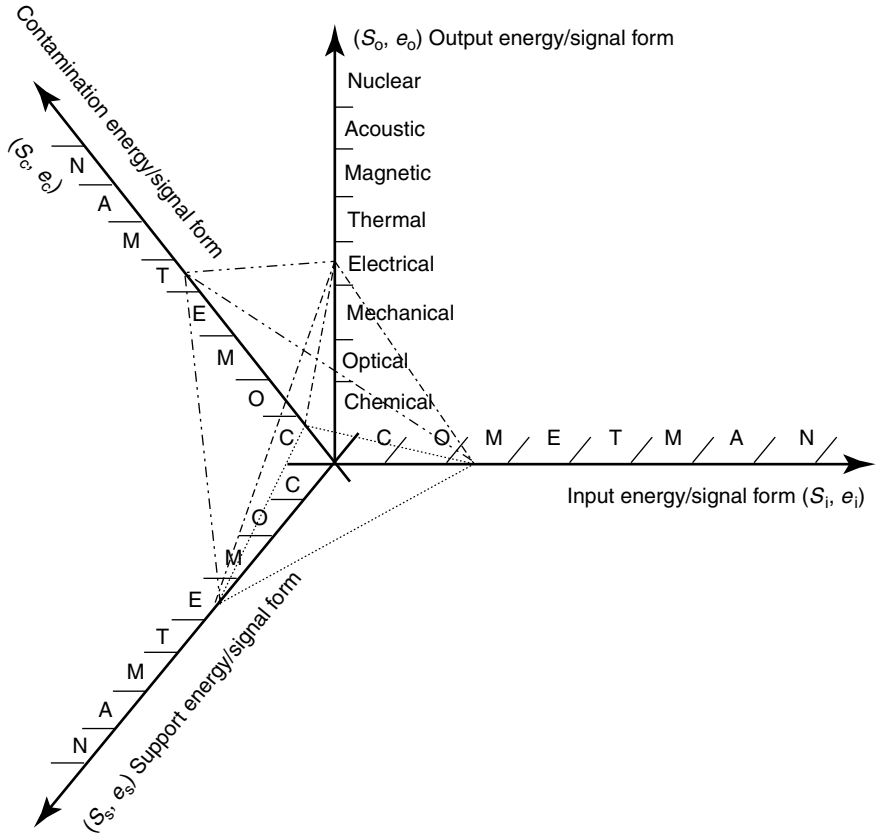


Figure 5. The information and energy tetrahedron for a modulating sensor.

energy form is mechanical, using a sensor such as a resistance strain gauge.

The chemical contamination could be responsible for parasitic electrochemical potentials at the contacts in the electrical circuit. Each point on the input triangle axes may be joined to an energy/information form on the output axis. When the output is in the form of an information bearing electrical signal, the dash-dot line combinations shown in Figure 5 allow the link between all three input forms and the output form to be clearly seen. The geometrical shape, which is outlined by the dotted and the dash-dot lines, has one base plane surface and three inclined surfaces. This represents a tetrahedron or four-faced object. This way of representing the boundary impacts on systems may be called the *information/energy tetrahedron* or the *sensor/transducer effect tetrahedron*.

Illustration of the manner in which the contamination inputs influence the output may be extended to cover contamination from other energy forms. The dot-dot-dash lines in Figure 5 show the case when the contaminating energy form is due to the influence of temperature on the wire of the strain gauge. The deep perception allowed by the information/energy tetrahedron given in Figure 5, which is fairly straightforward, is still realistic and comprehensive.

Extending (1) to include the effects of contamination is easily understood. Hence, the functional relationship between the output energy/signal forms can be written as

$$\left[\frac{S_o}{e_o}\right] = f\left(\left[\frac{S_i}{e_i}\right], \left[\frac{S_s}{e_s}\right]\right), \left[\frac{S_c}{e_c}\right] \tag{2}$$

So far, the conditions in Figure 5 correspond to the case of modulating types of sensors and transducers, which are represented in Figure 2 and Figure 4.

Self-generating sensor possibilities may also be represented by the sensor/transducer effect tetrahedron as shown in Figure 6 using a thermocouple for illustrative purposes.

In self-generating sensors, there is no support energy form. Consequently, the support energy axis of the input triangle collapses to the origin of the system of axes in the tetrahedron. This origin, usually representing the number zero in a numerical space, may be regarded as the point defining the nonexistence of a support information/energy form. Under these conditions, (2) can be rewritten in the form

$$\left[\frac{S_o}{e_o}\right] = f\left(\left[\frac{S_i}{e_i}\right], \left[\frac{O_s}{o_s}\right]\right), \left[\frac{S_c}{e_c}\right] \tag{3}$$

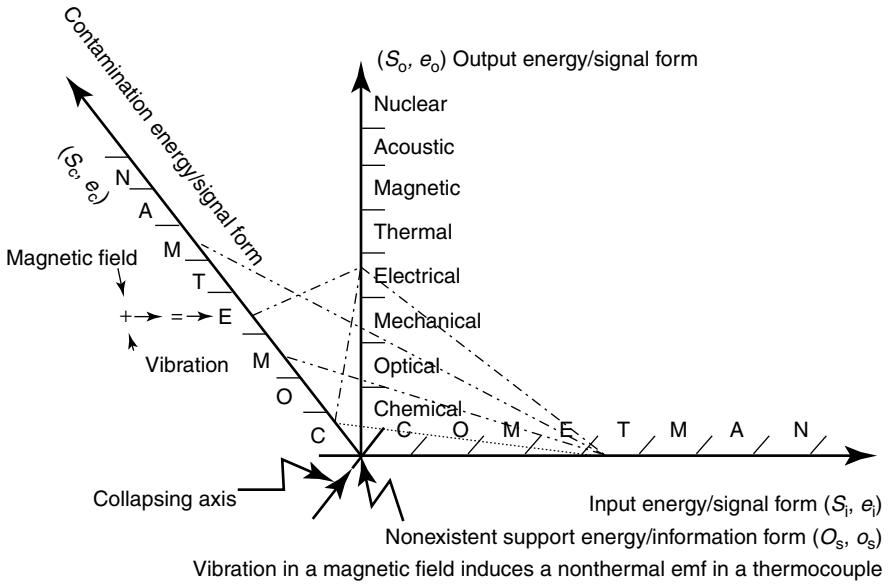


Figure 6. The sensor effect tetrahedron illustrating self-generating sensors and transducers using a thermocouple as an example.

where the term $[O_s/o_s]$ indicates that there is no support signal/energy form for self-generating sensors.

In the case of the thermocouple, illustrated in Figure 4, the measurand energy form is thermal. The chemical interference shown once again may give rise to a series mode parasitic electrochemical emf in the electrical circuit of the thermocouple. A situation, illustrating the interaction of different energy forms to produce a propagated contaminating signal, is also shown in Figure 6. If the thermocouple sensor assemblies, including its lead wires, are mounted on a structure, which is vibrating in an electromagnetic field, the sensor effect tetrahedron allows the possibility of an induced series mode interfering electrical signal to be predicted and visualized. The powerful predictive property of the sensor effect tetrahedron is well illustrated by this example.

The representation of self-generating sensors and transducers shown in Figure 6 may be called the *self-generator tetrahedron*. This model of a collapsing four-space may also be viewed as forming a three-space model, when the input and contaminating axes are drawn mutually perpendicular. This is analogous to the effect that occurs in Stein's transducer space for self-generating sensors given in Figure 1. Whereas this transducer space collapses to a plane, or two-space, from a three-space or three-dimensional space, in the case of self-generating transducers, the sensor effect tetrahedron collapses to a three-space for the same kind of sensor. The different results of the collapsing is due to the fact that the sensor effect tetrahedron represents the input information/energy groupings using the idea of the input triangle illustrated in Figure 5 and Figure 6.

7 SIGNAL/ENERGY MATRIX MODELING OF SENSORS AND TRANSDUCERS

The basic ideas of classification in sensors and transducers allow a visualization of the diverse possibilities for realizing measuring functions. Another very important aid in visualizing how the energy and signal are transduced uses the concept of a signal/energy flow matrix. Figure 7 illustrates this for two different kinds of temperature sensor. One of these is a mercury-in-glass thermometer and another is a modern digital thermometer. The fundamental block diagram representation of input and output sensing/transducing

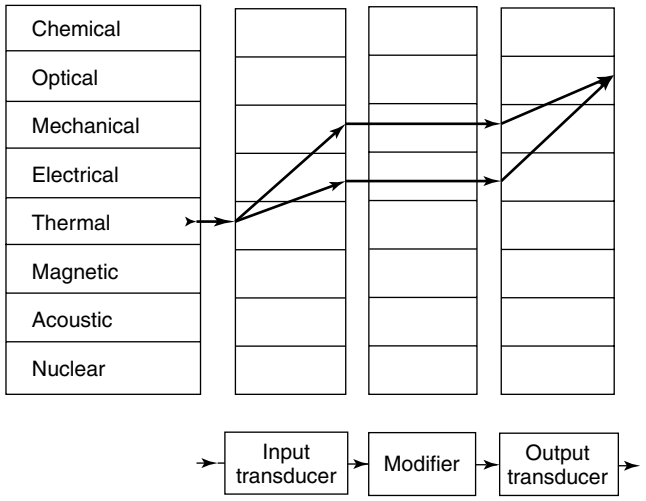


Figure 7. Generalized but simplified energy regime for a mercury-in-glass thermometer and a modern digital thermometer.

elements and modifying elements, shown in Figure 4, is used to construct the matrix shown in Figure 7. Above this block diagram in Figure 7 is a matrix of rows and columns, with the left-hand column indicating the COMETMAN classification scheme for input energy/signal forms. The energy domain of the temperature input signal is of course the thermal domain.

Consider first when the sensor is a mercury-in-glass thermometer. The input transduction process transforms the thermal energy into mechanical energy, because of the cubic expansion of the mercury in the bulb of the thermometer. This expansion is propagated through the thermometric fluid, the mercury, by modifying the original mechanical cubic expansion into a change in the length of the mercury column. This column length belongs to the group of mechanical variables. Radiation, which is incident upon the column, is modulated by the reflective, refractive, and absorptive properties of the mercury column so that the human eye perceives the change in length. This perception is by a noncontacting form of sensing through the 10^8 videceptors in the human sense of seeing.

If the thermometer is a digital electronic unit, the thermal energy is transduced to the electrical domain using a suitable electrical temperature sensor, which could be any one of the contacting group in the classification of temperature sensors given in Figure 8, explained later in Section 8. Modification of the electrical energy at the electrical output port of resistive sensors could use a Wheatstone bridge. Further modification could be performed with an electronic instrumentation amplifier. Subsequently, the analog signal at the output of the instrumentation amplifier is converted to a digital form using an analog-to-digital converter (ADC). The digital output of the ADC is then encoded and applied to an LED (light emitting diode). The LED, which is a photon generator, transduces its digital electrical input signal to visible radiation at its output.

8 GENERAL PRINCIPLES OF TEMPERATURE SENSOR CLASSIFICATION

While little systemic classification work has been contributed for the sensor family as a whole, a significant foundation showing the general rules of classification has been developed for temperature. It provides an example that illustrates the characteristics and scope for other categories.

Elements for measuring temperature extend the human faculties to sense hotness relations between bodies or entities in the real world. This functional extension of the human faculties is held in common with other instruments for measurement, calculation, communication, and control. From the classification perspective, the tree classification diagram of Figure 8 illustrates benefits that the science of taxonomy can bring to the study of temperature sensors. It is obvious from Figure 8 that there are diverse methods for temperature measurement. Only contacting sensors with heat transfer by convection or conduction will be considered in detail. None of the wide variety of noncontacting sensors are considered here.

To provide breadth of perspective, it is essential to introduce some kind of ordering so that similarities between each kind may be identified without in any way diminishing their important differences.

The main function of taxonomy, or classification, is to discover the relationship between different entities that unites them. In the first instance, at the metaphysical, theoretical, or fictitious level, it is necessary to deal with the

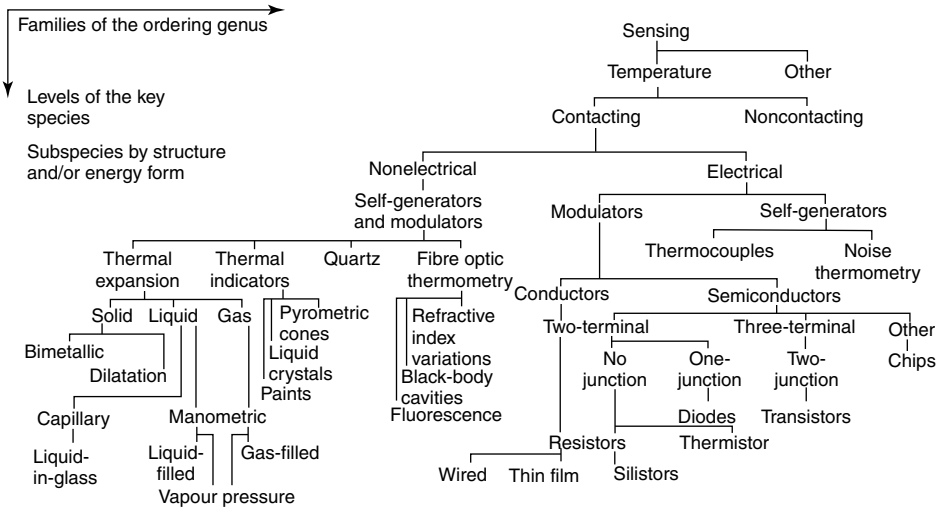


Figure 8. Classification of temperature measuring instruments/sensors by structure and/or energy form.

generality or resemblance of specific methods of temperature measurement. This is concerned with establishing the relationship of likeness.

The second important problem in the classification of temperature sensors should analyze two relationships. The first is that between those similar subspecies of sensing methods. The second is the relationships between one part of a temperature sensor with other parts, which together collectively contribute to the composition of the sensor. It is also necessary to construct a hierarchical ordering based upon the relationship between the heads or central members of groups of sensors.

Composition and collectivity play a role in this ordering, as they relate the places occupied in each order, relative to other sensors of the same order. Another last, but not least, ingredient in ordering temperature sensors is concerned with the *kinship* of one sensor type with another type. The relationships of ascent, descent, and collaterality are the essential factors in this aspect of temperature sensor ordering. Ascending and descending relationships place the sensor by structure and energy form relative to the head or central members of principal groups of sensors. Collaterality is that relationship which one type of sensor has with another at the same hierarchical level in the classification.

Applying the above principles, it can be seen that temperature sensors are a specific genus of sensor within the much wider family of sensors. This is emphasized by the branch that shows that other genus of sensor exist. Also, in accordance with the general classification of sensors by their energy relationships given earlier, temperature sensors may belong to the groups of sensor known as either *self-generators* or *modulators*.

The generality and resemblance level in temperature sensor ordering is best approached by comparison with the functions of the human senses. Humans can detect hotness relations by looking at an object, by approaching it, or by touching it.

Neither looking at nor approaching an object requires physical contact to sense its hotness. Touching an object to sense its hotness requires physical contact. Thus, the senses of touch, by contact, and sight or proximity, with no contact, are those used by humans to sense hotness relations.

It is apparent that temperature sensors may be based upon heat transfer by contacting or noncontacting methods. This grouping at the species level of the hierarchy of temperature sensors, classifies them on the basis of these heat transfer mechanisms. This approach to temperature measurement is recognized as the direct form of measurement in the thermal field.

Other possibilities for the noncontacting species use inferential methods. The principle of this technique applies an external energy as an interrogating medium in the measuring scheme for the purposes of information capture about the abilities of the body under measurement to store, dissipate, transmit, or transform thermal energy.

9 TEMPERATURE MEASURING RANGES OF TEMPERATURE SENSORS

It is also possible to classify temperature sensors on the basis of the temperature range of application. Such a

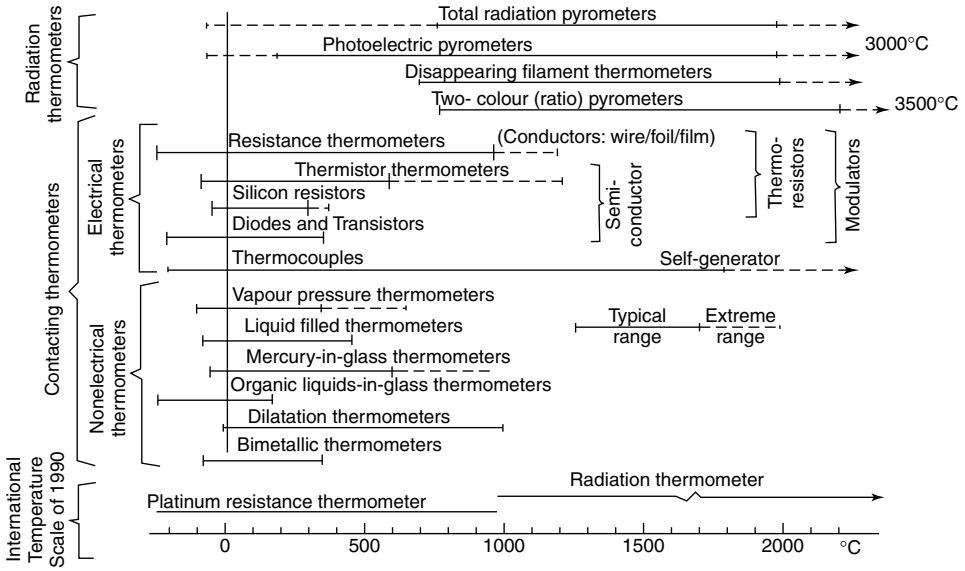


Figure 9. Classification of temperature measuring instruments/sensors by measuring range.

classification is given in Figure 9. An abbreviated form for the temperature ranges of the standard thermometers of ITS-90 is included in this diagram for the purposes of comparison.

A simple classification tree on the various types of elastic flexure used in instruments has been published, Sydenham (1984).

RELATED ARTICLES

Article 108, Transducer Fundamentals, Volume 2; **Article 109, Structure and Energy in Sensor Systems, Volume 2**; **Article 110, Signal/Energy Matrix Modeling, Volume 2**; **Article 112, Systematic Description of Sensors, Volume 2**; **Article 114, Models of the Sensor Interface, Volume 2**; **Article 115, Designing the Sensor Interface, Volume 2**.

REFERENCE

Sydenham, P.H. (1984) Elastic Design of Fine Mechanism in Instruments. *Journal of Physics E: Scientific Instruments*, **17**, 922–30.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

This article is based in part on a chapter by J. McGhee *et al* originally published in Measurement Data Handling: Hardware Techniques Vol. 2, J. McGhee *et al*, Technical University of Lodz, 2001.

112: Systematic Description of Sensors

Paul P.L. Regtien
University of Twente, Enschede, The Netherlands

1 Principles of Systematic Description of Sensors 751
2 Systematic Description Based on Energy Domains 752
3 Systematic Description of Physical Quantities 753
4 Systematic Description Based on Thermodynamic Laws 754
5 A Unified Approach to Transducer Modeling 756
References 757

1 PRINCIPLES OF SYSTEMATIC DESCRIPTION OF SENSORS

A sensor performs the exchange of information (thus energy) from one domain to another and therefore it operates on the interface between different physical domains. Several frameworks have been developed for a systematic description of sensors. Basically, they are based on either of two different approaches. The first method follows a categorization on the basis of the various energy domains. However, there is no clear definition of an energy domain, leaving room for different opinions on this description. The other approach is based on a categorization of physical quantities. Although more fundamental, this method also leaves scope for discussions. Obviously, the two approaches are connected to each other, since quantities describe particular phenomena in which energy or conversion of energy is involved.

A common categorization according to physical effects is based on thermodynamic principles. It is the most fundamental, and covers most of the conversion effects that occur in physics. The description is, in particular, useful in the field of material research and optimization of sensor materials. Derived from the thermodynamic approach is the Bond graph notation (Blundell, 1982) with a division of variables into two groups called *effort and flow variables*. This method is not only useful for the description of sensors but has great significance in the design of all kinds of technical systems, irrespective of the domain type.

A further distinguishing property for a subdivision of sensors into just two categories is the need for auxiliary energy (*see* **Article 108, Transducer Fundamentals, Volume 2**). Sensors that need no auxiliary energy for their operation are called *direct sensors* or *self-generating sensors*. Sensors that use an additional energy source for their operation are called *indirect sensors*, *modulating sensors*, or *interrogating sensors*. The names direct and indirect are rather obscure; the name self-generating is misleading: neither the energy nor the information is generated by the sensor for it comes from the measurand.

The terms modulating and interrogating refer to the fact that the measurand affects a specific material property, which in turn is interrogated by an auxiliary quantity. Most sensors belong to this group: all resistive, capacitive, and inductive sensors are based on a parameter change (resistance, capacitance, inductance) caused by the measurand. Likewise, most displacement sensors are of the modulating type: displacement of an object modulates optical or acoustic properties (transmission, reflection, interference), where light or sound is the interrogating quantity. The distinction between these two groups can be useful when considering offset and load errors of sensors **(Article 108, Transducer Fundamentals, Volume 2**). In the last section of this article, a universal model of a sensor is presented that avoids any misinterpretation due to badly chosen names.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

The operation of sensors is governed by

- physical effects,
- material properties, and
- system layout.

Physical effects may either act within one-energy domain or cross domain boundaries. The latter describe the fundamental operation of sensors. Material properties determine the quantitative characteristics of the conversion process, and hence form a guide to the proper material choice for a particular sensor. Finally, with an appropriate sensor layout, the performance of the sensor can further be optimized.

2 SYSTEMATIC DESCRIPTION BASED ON ENERGY DOMAINS

A systematic representation of sensor effects based on energy domains involves a number of aspects. First, the energy domains have to be defined. Next, the energy domains should be allocated to both the sensor input and output. Finally, since many sensors are of the modulating type (**Article 108, Transducer Fundamentals, Volume 2**) also, the domain of the auxiliary quantity should be considered.

From a physical point of view, the nine energy forms that can be distinguished are (Middelhoek and Audet, 1989)

- electromagnetic radiant energy
- gravitational energy
- mechanical energy
- thermal energy
- electrostatic and electromagnetic energy
- molecular energy
- atomic energy
- nuclear energy
- mass energy.

This classification is rather unpractical for the description of sensors. Stein has proposed a system of eight distinct energy forms:

- Chemical
- Optical
- Mechanical
- Electrical
- Thermal
- Magnetic
- Acoustic/fluid
- Nuclear

forming the acronym COMETMAN (Stein, 1963).

Lion reduced this to only six domains, and adopted the term *signal domain* (Lion, 1969). These six domains are

- radiant
- mechanical
- thermal
- magnetic
- chemical and
- electrical.

The number of domains is actually a rather arbitrary choice, so for practical reasons we continue with the system of six domains and remain calling them *energy domains*.

Information contained in each of the six domains can be converted to any other domain. These conversions can be represented in a 6×6 matrix. Figure 1 shows that matrix, including some of the conversion effects. An input transducer or sensor performs the conversion from a non-electrical to the electrical domain (the shaded column), and an output transducer or actuator performs the conversion from the electrical to another domain (the shaded row in Figure 1).

In↓ Out→	**Radiant**	**Thermal**	**Electrical**	**Magnetic**	**Mechanical**	**Chemical**
Radiant	Luminescense	Radiation heating	Photoconductivity	Photomagnetism	Radiation pressure	Photochemical process
Thermal	Incandescense	Thermal conductivity	Seebeck effect; Pyroelectricity	Curie–Weiss law	Thermal expansion	Endothermal reaction
Electrical	Injection luminescense	Peltier effect	PN junction effect; Ohm's law	Ampere's law	Converse piezoelectricity	Electrolysis
Magnetic	Faraday effect	Ettinghausen effect	Hall effect	Magnetic induction	Converse magnetostriction	
Mechanical	Photoelastic effect	Friction heat	Piezoelectricity	Magnetostriction	Gear	Pressure induced reaction
Chemical	Chemoluminescense	Exothermal reaction	Volta effect		Explosive reaction	Chemical reaction

Figure 1. Conversion effects. (Reproduced from 'Physics of Silicon Sensors', Middelhoek *et al*, Copyrigtht (1989), with permission from Elsevier.).

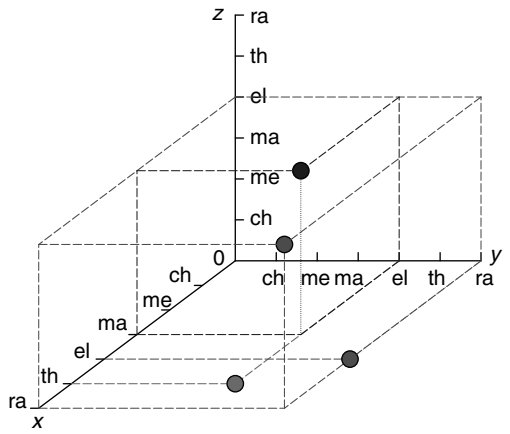


Figure 2. Sensor cube with four transducer types; thermocouple: [th, el, 0], Hall sensor: [ma, el, el], LED: [el, ra. 0] and LCD: [ra, ra, el].

This two-dimensional representation can be extended to three dimensions, when the interrogating energy domain is included. This gives 216 energy-triplets. To get a clear overview of all these possible combinations, they can be represented in a 3D Cartesian space, the sensor cube shown in Figure 2. The three axes refer to the input energy domain, the output energy domain, and the domain of the interrogating input.

Each of the 216 elements of the $6 \times 6 \times 6$ matrix represents one or more conversion effects. When restricting to electrical transducers, we have 5 direct input transducers, 5 direct output transducers, 25 modulating input transducers, and 25 modulating output transducers. To facilitate notation, the transducers can be indicated by indices, like in crystallography, the so-called *Miller indices*: $[x, y, z]$. The x-index is the input domain, the y-index, the output domain and the z-index, the domain of the interrogating quantity. With these three indices, a transducer can be typified according to the energy domains involved. Some examples are

- direct input transducer: thermocouple: [th, el, 0]
- modulating input transducer: Hall sensor; [magn, el, el]
- direct output transducer: LED: [el, rad, 0]
- modulating output transducer: LCD: [rad, rad, el].

These transducers are also visualized in Figure 1. The value of such a representation is rather limited. It may serve as the basis of a categorization for overviews, or as a guide in the process of sensor selection. This also shows why there are so many alternatives for sensing a given variable, and why the uncontrolled generation of so-called new sensors is hard to keep in clear perspective.

3 SYSTEMATIC DESCRIPTION OF PHYSICAL QUANTITIES

Several attempts have been made to set up a consistent framework for quantities and material properties. Most of these descriptions are based on energy considerations, or more precisely, on the first and second law of thermodynamics.

For each energy domain, two variables can be defined in such a way that their product equals the energy (or a derived energy quantity) of the domain type, the conjugate quantities. For instance:

Mechanical:	Tension $\boldsymbol{T}$ (N m^{-2}) and deformation $\boldsymbol{S}$ [−]
	Shear tension τ (N m^{-2}) and shear angle χ [−]
Electrical:	Field strength $\boldsymbol{E}$ (V m^{-1}) and dielectric displacement $\boldsymbol{D}$ (C m^{-2})
	Field strength $\boldsymbol{E}$ (V m^{-1}) and electric polarization $\boldsymbol{P}$ (C m^{-2})
Magnetic:	Magnetic induction $\boldsymbol{B}$ (Wb m^{-2}) and magnetic field $\boldsymbol{H}$ (A m^{-1})
Thermal:	Temperature Θ (K) and entropy σ (J K^{-1} or J Km^{-3}).

The variables ***E, D, B, H, T***, and ***S*** are vector variables, whereas σ and Θ are scalars (so often denoted as $\Delta\sigma$ and $\Delta\Theta$). Note that the dimension of the product of each domain pair is J m^{-3} (energy per unit volume) in all aforementioned cases (note a possible confusion with respect to the thermal domain).

The variables in this list show a remarkable analogy: they belong to either of two classes: through-variables and across-variables. To explain this classification, we first introduce the term *lumped element*. A lumped element symbolizes a particular property of a physical component; that property is thought to be concentrated in that element between its two end points or nodes. Exchange of energy or information only occurs through these terminals.

A *through-variable* is a physical quantity that is the same for both terminals of the lumped element.

An *across-variable* describes the difference with respect to a physical quantity between the terminals. They are also called *intensive variables* and *extensive variables* respectively.

In an electronic system, current is a through-variable, voltage (or potential difference) is an across-variable. Therefore, through-variables are called *generalized I-variables* and across-variables are called *generalized V-variables*. However, this is just a matter of viewpoint. It is perfectly justified to call them *generalized forces* and *displacements*.

In the above groups of variables, ***T, E***, and Θ are across-variables. On the other hand, ***S, D***, and σ are through-variables. They are related to each other through physical material properties or system layout.

For each domain, we can define a pair of *conjugate variables*. In the preceding examples, the product was energy density (J m^{-3}). It is also possible to find other pairs of variables whose product is, for instance, energy or power. A few examples are

mechanical: pressure P (N m^{-2}) and volume V (m^3);
mechanical: force F (N) and distance x (m);
electrical: voltage V (V) and charge Q (C);

all with product energy (J); these pairs are energy conjugate variables.

Other examples are

mechanical: force F (N) and velocity v (m s^{-1});
electrical: voltage V (V) and current I (A = C s^{-1});

both with product power (W); these pairs are power conjugate.

For other domains, such pairs can be defined as well, but with sometimes rather unusual quantities.

From the last examples, we observe regularity in the various domains. Within one domain, we can distinguish state variables and rate variables. They are related as

$$X_{\text{rate}} = \frac{\mathrm{d}}{\mathrm{d}t} X_{\text{state}} = \dot{X}_{\text{state}} \text{ or } X_{\text{state}} = \int X_{\text{rate}}\,\mathrm{d}t \qquad (1)$$

The most obvious cases are

$$I = \frac{\mathrm{d}}{\mathrm{d}t} Q \text{ or } Q = \int I\,\mathrm{d}t \qquad (2)$$

$$v = \frac{\mathrm{d}}{\mathrm{d}t} x \text{ or } x = \int v\,\mathrm{d}t \qquad (3)$$

4 SYSTEMATIC DESCRIPTION BASED ON THERMODYNAMIC LAWS

On the basis of laws on energy conservation, several frameworks for a systematic description of material properties and physical effects and hence for the description of sensors can be built up.

According to the first law of thermodynamics, the energy content of an infinitely small volume of an elastic dielectric material changes by adding or extracting heat $\mathrm{d}Q$ (J) and by work $\mathrm{d}W$ (J) exerted upon it is given by

$$\mathrm{d}U = \mathrm{d}Q + \mathrm{d}W\,(\mathrm{J}) \qquad (4)$$

The work $\mathrm{d}W$ is the sum of the different energy forms involved, for instance, electrical and mechanical energy: $V\,\mathrm{d}Q$ and $F\,\mathrm{d}x$. Since $\mathrm{d}Q = \Theta\,\mathrm{d}\sigma$ (J) (temperature and entropy), we can generalize the energy change as

$$\mathrm{d}U = \sum_i A_i\,\mathrm{d}B_i \text{ (J)} \qquad (5)$$

where A_i is an extensive variable and B_i the conjugated intensive variable. So if we consider only the thermal, electrical, and mechanical domains, the energy change is governed by the equation

$$\mathrm{d}U = \Theta\,\mathrm{d}\sigma + V\,\mathrm{d}Q + F\,\mathrm{d}x \text{ (J)} \qquad (6)$$

When energy per unit of volume is considered, this equation reads as

$$\mathrm{d}U = T\,\mathrm{d}S + E\,\mathrm{d}D + \Theta\,\mathrm{d}\sigma \text{ (J m}^{-3}\text{)} \qquad (7)$$

(Note that for entropy we have used the same symbol but the dimension is now J Km^{-3}.) Apparently, in this equation only through-variables affect the system. If, on the other hand, only across-variables affect the energy state of the system, the equation for the energy change per unit volume is

$$\mathrm{d}G = -S\,\mathrm{d}T - D\,\mathrm{d}E - \sigma\,\mathrm{d}\Theta \text{ (J m}^{-3}\text{)} \qquad (8)$$

where G is the Gibbs potential that can be found from the free energy U by a Legendre transformation. We continue the discussion with the last expression because the resulting parameters are more in agreement with experimental conditions. After all, it is easier to have the across-variables as inputs or independent quantities (temperature; electric field, or voltage; force) and to measure the resulting through-variables or dependent quantities (strain; dielectric displacement or current).

The energetic state of a system can be described by pairs of conjugate variables, the pairs being determined by the energy forms that have to be regarded for a specific situation. However, the system configuration or the material couples the conjugate variables of each pair. Examples within one domain are

$T = c \times S$ (Hooke's law, with c, the mechanical elasticity)

$P = \varepsilon_0 \chi_e E$ (with χ_e, the electrical susceptibility)

$J = \mu_0 \chi_m H$ (with χ_m, the magnetic susceptibility)

In general, the extensive variable, A_i and the intensive variable, B_i within one domain are connected according to

$$A_i = \alpha_i B_i \text{ or } B_i = \beta_i A_i \tag{9}$$

where α_i is a (generalized) modulus and β_i is a (generalized) compliancy. These are the state equations of the system.

As an example, in the mechanical domain, Hooke's law can also be written as

$$T = c \times S \text{ or } S = s \times T \tag{10}$$

where c is the modulus of elasticity (or Young's modulus) and s, the compliance.

Likewise, variables in one domain are linked to variables in another domain, expressing the cross-effects. So

$$A_i = \gamma_{ij} B_j \text{ or } B_j = \gamma_{ji} A_i \tag{11}$$

where γ_{ij} and γ_{ji} represent material properties linking domains i and j.

The aforementioned equations form the basis for a formal description of sensors, since it is now possible to describe physical effects crossing domain boundaries. We illustrate the concept with a system in which only the mechanical, the electrical, and the thermal domains are involved. So the starting point is the equation:

$$\mathrm{d}G = -S\,\mathrm{d}T - D\,\mathrm{d}E - \sigma\,\mathrm{d}\Theta (\mathrm{J\ m^{-3}}) \tag{12}$$

The through-variables S, D and σ in this equation can be written as partial derivatives of the Gibbs potential:

$$\begin{aligned} S(T, E, \Theta) &= \left(\frac{\partial G}{\partial T}\right)_{\Theta,E} \\ D(T, E, \Theta) &= \left(\frac{\partial G}{\partial E}\right)_{T,\Theta} \\ \sigma(T, E, \Theta) &= \left(\frac{\partial G}{\partial \Theta}\right)_{T,E} \end{aligned} \tag{13}$$

From these equations, we can derive the various material and sensor parameters. To that end, the variables S, D, and $\Delta\sigma$ are approximated by linear functions, that is, we take only the first term of the Taylor series expansion in the points $T = 0$, $E = 0$ and $\Theta = \Theta_0$ of the functions $S(T, E, \Theta)$, $D(T, E, \Theta)$, and $\sigma(T, E, \Theta)$:

$$\begin{aligned} \mathrm{d}S &= \left.\frac{\partial S}{\partial T}\right|_{E,\Theta} \mathrm{d}T + \left.\frac{\partial S}{\partial E}\right|_{\Theta} \mathrm{d}E + \left.\frac{\partial S}{\partial \Theta}\right|_{E} \Delta\Theta \\ \mathrm{d}D &= \left.\frac{\partial D}{\partial T}\right|_{E,\Theta} \mathrm{d}T + \left.\frac{\partial D}{\partial E}\right|_{\Theta} \mathrm{d}E + \left.\frac{\partial D}{\partial \Theta}\right|_{E} \Delta\Theta \end{aligned} \tag{14}$$

$$\mathrm{d}\sigma = \left.\frac{\partial \sigma}{\partial T}\right|_{E,\Theta} \mathrm{d}T + \left.\frac{\partial \sigma}{\partial E}\right|_{\Theta} \mathrm{d}E + \left.\frac{\partial \sigma}{\partial \Theta}\right|_{E} \Delta\Theta$$

Combination of (13) and (14) results in

$$\begin{aligned} S &= \left(\frac{\partial^2 G}{\partial T^2}\right)_{\Theta,E} T + \left(\frac{\partial^2 G}{\partial T \partial E}\right)_{\Theta} E + \left(\frac{\partial^2 G}{\partial T \partial \Theta}\right)_{E} \Delta\Theta \\ D &= \left(\frac{\partial^2 G}{\partial E \partial T}\right)_{\Theta} T + \left(\frac{\partial^2 G}{\partial E^2}\right)_{\Theta,T} E + \left(\frac{\partial^2 G}{\partial E \partial \Theta}\right)_{T} \Delta\Theta \\ \Delta\sigma &= \left(\frac{\partial^2 G}{\partial \Theta \partial T}\right)_{E} T + \left(\frac{\partial^2 G}{\partial \Theta \partial E}\right)_{T} E + \left(\frac{\partial^2 G}{\partial \Theta^2}\right)_{E,T} \Delta\Theta \end{aligned} \tag{15}$$

Now we have a set of equations connecting the (dependent) through-variables S, D, and σ with the (independent) across-variables T, E, and Θ. The second-order derivatives represent material properties; they have been given special symbols.

The constant variables are put as superscripts, to make place for the subscripts denoting orientation (applied in nonisotropic materials).

$$\begin{aligned} S &= s^{E,\Theta} T + \mathrm{d}^{\Theta} E + \alpha^{E} \Delta\Theta \\ D &= \mathrm{d}^{\Theta} T + \varepsilon^{\Theta,T} E + p^{T} \Delta\Theta \\ \Delta\sigma &= \alpha^{E} T + p^{T} E + \frac{\rho}{T} c^{E,T} \Delta\Theta \end{aligned} \tag{16}$$

These superscripts denote constancy with respect to the indicated parameters, for instance, $s^{E,\Theta}$ is the compliance at zero electric field E and constant temperature Θ.

The nine associated effects are displayed in Table 1. Table 2 shows the corresponding properties. The parameters for just a single domain are ε, c_p, and s. The other parameters (p, α, and d) denote cross-effects.

Note that direct piezoelectricity and converse piezoelectricity have the same symbol (d) because the dimensions are equal (m $\mathrm{V^{-1}}$ and C $\mathrm{N^{-1}}$). The same holds for the pair pyroelectricity and converse pyroelectricity, as well as for thermal expansion and piezocaloric effect. Further, the second order derivatives in (15) correspond with the parameters γ_{ij} of (11). Since the order of differentiation can be reversed, the parameters γ_{ij} and γ_{ji} are equal.

Table 1. Nine physical effects corresponding to the parameters in (16).

Elasticity	Converse piezoelectricity	Thermal expansion
Direct piezoelectricity	Permittivity	Pyroelectricity
Piezocaloric effect	Electrocaloric effect	Heat capacity

Table 2. Symbols, property names, and units of the effects in Table 1.

Symbol	Property	Unit
s	Compliance, elasticity	m^2N^{-1}
d	Piezoelectric constant	$mV^{-1} = CN^{-1}$
α	Thermal expansion coefficient	K^{-1}
p	Pyroelectric constant	$Cm^{-2}K^{-1}$
ε	Permittivity, dielectric constant	Fm^{-1}
$c^{E,T}$	(Specific) heat capacity	$Jkg^{-1}K^{-1}$

Equations (7) and (8) can be extended just by adding other couples of conjugate quantities, for instance, from the chemical or the magnetic domain. Obviously, this introduces many other material parameters. With three couples we have nine parameters, as listed in Table 1. With 4 couples of intensive and extensive quantities we get 16 parameters, so 7 more (for instance, the magnetocaloric effect, expressed as the partial derivative of entropy to magnetic field strength).

5 A UNIFIED APPROACH TO TRANSDUCER MODELING

As a consequence of the energetic basis for the sensor description in the preceding sections, the information exchange between a sensor (or measuring system) and its environment (input devices, output devices, auxiliary sources, interfering sources) always implies two quantities: an across-variable and an associated through-variable. Energy (and thus information) enters or leaves the system through a pair of terminals making up a port. We distinguish input ports and output ports. A direct sensor can be described by a two-port model or a four-terminal model (Figure 3a). The input port is connected to the measurand and the output port corresponds with the electrical connections of the sensor. Likewise, a modulating sensor can be conceived as a system with three ports: an input port, an output port, and a port through which the auxiliary energy is supplied (Figure 3b).

In these models, the across-variables are denoted by E (effort) and the through-variables by F (flow) respectively. For each domain, the two variables E and F are defined in such a way that their product equals the energy of the domain type or a time or space derivative of energy. The subscripts x, y, and z are chosen in accordance with the sensor cube from Figure 1.

Direct sensors provide the information about the measurand as an output signal, an energetic quantity. Modulating sensors contain the information as the value of material property, or a geometric quantity, not an energetic signal. The information enters the system through the input port,

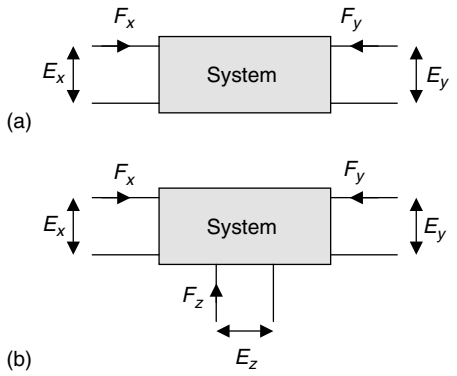


Figure 3. Port models of a sensor: (a) two-port (direct sensor) and (b) three-port (modulating sensor).

where the measurand affects specific material or geometric parameters. To extract the information from such a sensor, it has to be interrogated using an auxiliary signal. The information stored in the sensor is available latently, in the latent information parameters or LIP (Stein, 1992). These parameters are modulated by the input signal, and interrogated by the auxiliary or interrogating input.

At zero input, the LIPs of a modulating sensor have initial values, set by the material and the construction. Generally, the input has only a small effect on these parameters, resulting in relatively small deviations from the initial values.

Note that direct sensors also have LIPs set by materials and construction. They determine the sensitivity and other transfer properties of the sensor. So, the input port of all sensors can be denoted as LIP-input port. As a consequence, any sensor can be described with the three-port model of Figure 3(b). Only the functions of the ports may differ, notably the LIP-input port and the interrogating-input port.

According to the unified transducer model as introduced by Stein, an input port can be controlled either by design (it has a fixed value) or by the environment (the measurand or some unwanted input variable). So we have four different cases (Figure 4).

LIP input \ INT. input	Design controlled	Environment controlled
Design controlled	**Source**	**Direct sensor**
Environment controlled	**Modulating sensor**	**Multiplying devices**

Figure 4. Unified transducers.

The characteristics of these four cases are briefly reviewed.

1. *Design-controlled LIP-input and design-controlled interrogating input*. All inputs are fixed. This type represents a signal or information source, for instance, a standard, or a signal source with a constant or predetermined output. The output is totally determined by the construction and the materials that have been chosen. Any environmental effect on the output is (ideally) excluded.
2. *Design-controlled LIP-input and environment-controlled interrogating input*. Since the latent information parameters are fixed by design, the output only depends on what is connected to the interrogating input. When this is the measurand, the transducer behaves as a direct sensor.
   Examples are as follows:
   - *Thermocouple sensor*: the Seebeck coefficient is fixed by the choice of the materials.
   - *Piezoelectric accelerometer*: the sensitivity is fixed by the seismic mass and the piezoelectric properties of the crystal.
3. *Environment-controlled LIP-input and design-controlled interrogating input.* The measurand affects particular material properties or geometric parameters. These changes are interrogated by a fixed or well defined signal at the interrogating input. The transducer behaves as a modulated sensor.
   Examples are as follows:
   - *Strain gauge bridge*: strain alters the resistance of the strain gauge(s); a bridge voltage converts this resistance change into an output voltage.
   - *LVDT* (*Linear Variable Differential Transformer)*: a displacement of an object connected to the moving core of the LVDT will change the transfer ratio of the differential transformer. An AC signal on the primary coil acts as an interrogating quantity.
   - *Hall sensor*: the measurand is a magnetic induction field that acts on moving charges imposed by a fixed (or known) current applied to the interrogating input.
4. *Environment-controlled LIP-input and environment-controlled interrogating input:* These are multiplying transducers: the output depends on the quantities at both inputs, often in a multiplicative relation. For instance, a Hall sensor could act as such when the interrogating input is not a fixed current (by design) but a current that is related to just another measurand.

It is important to note that any practical transducer shows all four types of responses to varying extents. A strain gauge (a modulating transducer) produces, when interrogated, an output voltage related to the strain-induced change in resistance. But the circuit can also generate spurious voltages caused by capacitively or magnetically induced signals. A thermocouple (a direct transducer) produces an output voltage proportional to the measurand at the interrogating input. If, however, the material parameters change due to (for instance) strain or nuclear radiation (inputs at the LIP port) the measurement is corrupted.

Since just one response is desired, other responses should be minimized by an appropriate design. This universal approach helps identify such interfering sensitivities and their relative importance.

REFERENCES

Blundell, A.J. (1982) *Bond Graphs for Modelling Engineering Systems*, Ellis Horwood Publishers, Chichester.

Lion, K. (1969) Transducers: Problems and Prospects. *IEEE Transaction on Industrial Electronics and Control Instruments*, **16**, 2–5.

Middelhoek, S. and Audet, S. (1989) *Silicon Sensors*, Academic Press, London.

Stein, P.K. (1963) Classification System for Transducers and Measuring Systems, in *Symposium on Environmental Measurements: Valid Data and Logical Interpretation*, 4–6 September; US Department of Health Education and Welfare, 1964 (pp. 65–84).

Stein, P.K. (1992) *The Unified Approach to the Engineering of Measurement Systems – Basic Concepts*, Stein Engineering Services, Phoenix, AZ.

113: Force-feedback Sensors

Barry E. Jones
Brunel University, Uxbridge, UK

1 Feedback-measuring System 758
2 General Properties of Feedback Systems 758
3 Stability Considerations 759
4 Examples of Functional Architecture 760
5 Force and Torque Balance 761
6 Block Diagram and Damping 761
7 Examples of Force-feedback Sensors 762
8 Other Feedback-measuring Systems 763
Related Articles 763
References 763

1 FEEDBACK-MEASURING SYSTEM

There are only a few structural schemes employed in the construction of instruments and instrument systems, and one such scheme is the use of a feedback. A feedback system can be thought of as a system that tends to automatically maintain a prescribed relationship of one system variable to another by comparing functions of these variables and using the difference as a means of control. The main characteristic of a feedback system is its closed-loop structure (*see also* **Article 18, Nature and Scope of Closed-loop Systems, Volume 1**). A measuring system in which feedback is the basic structural arrangement is called a *feedback-measuring system*, and a systematic analysis, description, and review of such systems has been undertaken by Jones (1979).

A general block diagram of such a system is shown in Figure 1. Here, the output signal (usually electrical) is converted to a form (usually nonelectrical, e.g. force) suitable for comparison with the quantity to be measured (e.g. force). The resultant error is usually transduced into electrical form and amplified to give the output indication. Normally, a sensor and associated circuit has a nonelectrical input and an electrical output, for example, a thermistor, strain gauge, and photodiode, whereas a so-called inverse transducer or precision actuator has an electrical input and a low-power nonelectrical output: for example, a piezoelectric crystal, translational, and angular moving-coil elements can be used as inverse transducers. The sensor, inverse transducer, and usually the amplifier must be close to the point of measurement, whereas the indicator may be some distance away. The feedback loop must have sufficient negative gain and the system must be stable. The system is driving fairly low-power devices at its output, and the inverse transducer essentially determines the characteristics of the system, although noise connected with the sensor and amplifier input stage may well be important. In practice, the measurand may not be directly connected to the null or balance point, and there may well be one or more primary sensing elements inserted between points X and Y as in Figure 1.

2 GENERAL PROPERTIES OF FEEDBACK SYSTEMS

A measuring system may have many components or elements and it is convenient to consider each such element as a block with its own input–output relationship that is its own transfer function. The blocks representing the various elements of a system are connected to use their functional relationship within the system, thus producing a block diagram for the system.

A generalized block diagram of a simple feedback system is shown in Figure 2. The diagram illustrates negative

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

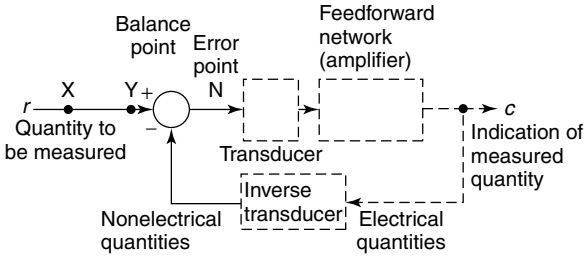


Figure 1. Diagram of a feedback-measuring system: - - - -, electrical units or signals; ——, mechanical units or signals. (Reproduced from *Feedback in Instruments and its Applications*, Jones, 1979, with permission of Institute of Physics Publishing Ltd.)

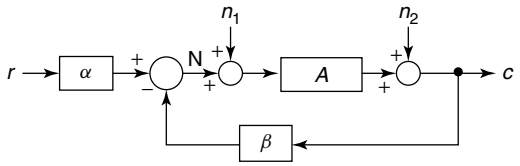


Figure 2. Generalized block diagram of a simple feedback system. (Reproduced from *Feedback in Instruments and its Applications*, Jones, 1979, with permission of Institute of Physics Publishing Ltd.)

feedback, in that the gain around the loop has negative polarity. In this case, the measurand r and the output indication c are in phase, but if in the practical system they are in antiphase, then the amplification A will have a negative sign associated with it and both inputs to the comparator will have positive signs.

In Figure 2 a fraction α of the quantity to be measured (r) is fed in and a fraction β of the output is fed back; α, β and A are transfer functions, while n_1 and n_2 are noise sources at the input and output of the amplifier respectively. Assuming $n_1 = n_2 = 0$, the following equation expresses the situation of Figure 2:

$$\alpha r - \beta c = \frac{c}{A}$$

or

$$\frac{c}{r} = \frac{\alpha}{\beta}\frac{A\beta}{1 + A\beta} \tag{1}$$

The ratio c/r is called the *closed-loop gain*, while $A\beta$ is the *open-loop gain* and both are transfer functions. Provided that $A\beta \gg 1$, $c/r \approx \alpha/\beta$, and c/r is largely independent of A and is simply determined by transfer functions α and β. Thus, the sensitivity of such a system can be changed by altering α or β. In some practical feedback-measuring systems, the feedback also makes α and β well defined. Actually, α and β may alter, but so long as their ratio remains fixed, c/r is well defined.

The main properties of this general system can be summarized as well-defined gain (giving accuracy), wide bandwidth (giving fast response), low or high output impedance (allowing remote indication), reduction in noise effects and extraneous disturbances, and the property of inversion. Negative feedback can be employed to produce instrument high input impedance, and unity-gain feedback often allows noncontact measurement. Measurement devices providing a frequency output usually employ feedback, and feedback can be used to enhance the characteristics of frequency-selective networks.

3 STABILITY CONSIDERATIONS

It is important to investigate the output response of a system, particularly a feedback-measuring system, when the input changes rapidly. This transient response characterizes the stability of the system. A system is defined as stable if its impulse response approaches zero as time approaches infinity. A system is defined as being unstable, if, with zero input, the output increases indefinitely. If the output of a system has continuous oscillation of constant peak-to-peak amplitude, the system is considered to be neutrally stable. Consideration of the degree of stability of a system often provides valuable information about its behavior. Clearly, it is desirable that feedback-measuring systems are stable and respond satisfactorily to rapid changes in the quantity to be measured.

It should be noted that in general for a feedback system, designs for accuracy and stability tend to be incompatible. From (1) accuracy required $|A\beta|$ to be high, while for second- and third-order systems, high $|A\beta|$ produces less damped and unstable systems respectively. In practice, when it is necessary, the open-loop frequency characteristics of a feedback system are tailored such that at quasi-static frequencies, the gain is very high to give the required accuracy at these frequencies, and at frequencies near to unity gain, the loop phase shift is less than -180° to give the required damping characteristics. Feedback-measuring systems often use a combination of mechanical and electrical means of damping. There are a number of useful texts available on the design of systems with control loops, for example, by Jones (1979), and by DiStefano, Stubberud and Williams (1990); *see also* **Article 19, Dynamic Behavior of Closed-loop Systems, Volume 1**.

Systems incorporating feedback involve comparison of two physically similar variables and production of a minimum or null (at point N in Figure 1). In feedback-measuring systems, pervariable (or 'through' variable) balance occurs with force, torque, current, and heat flow, while in the case of transvariables (or 'across' variables), voltage, temperature, and displacement balances are common.

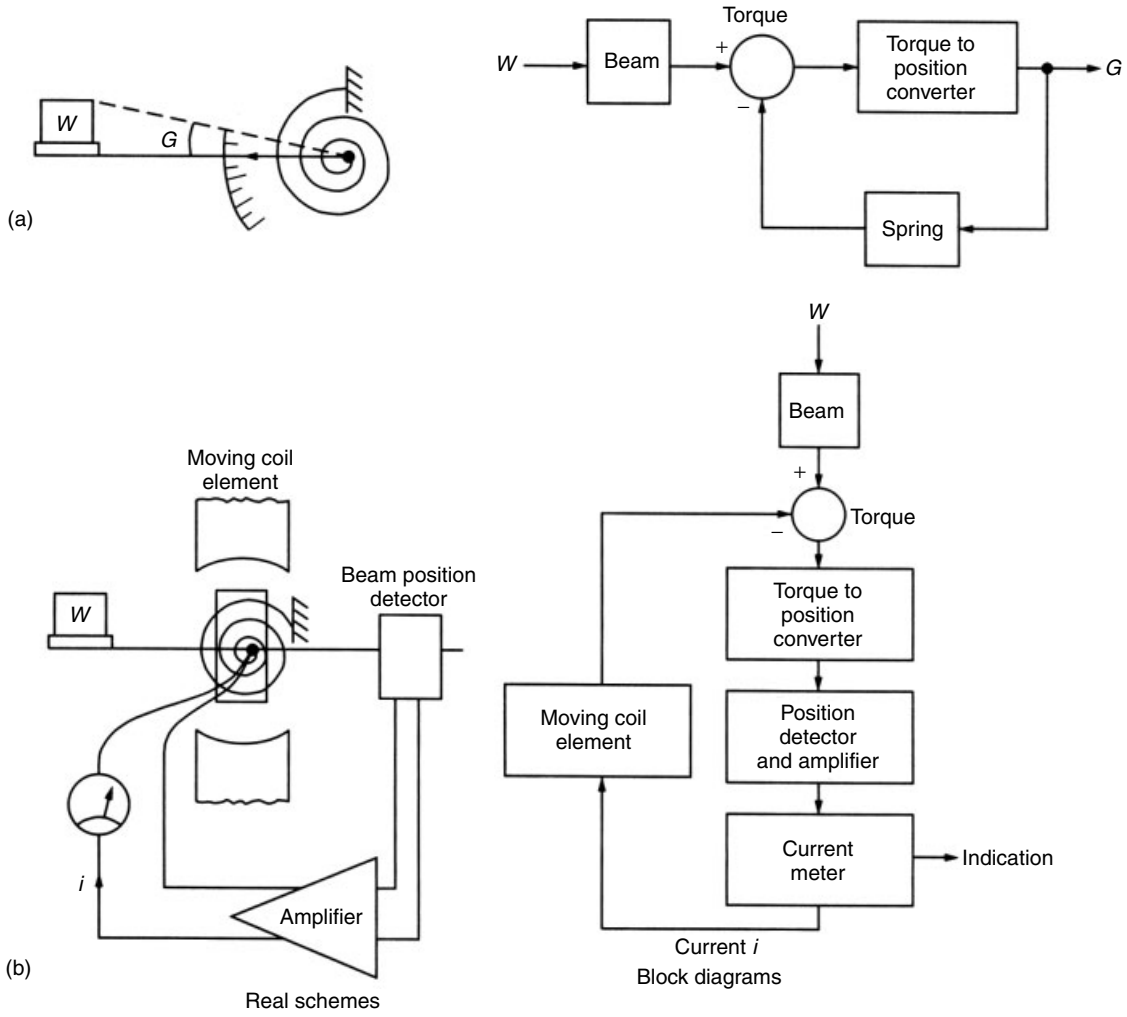


Figure 3. Measurement of small weights: (a) simple system and (b) more complex system. (Reproduced from *Instrument Systems: Functional Architectures from Concise Encyclopedia of Measurement and Instrumentation*, Finkelstein and Grattan, 1994, with permission of Elsevier.)

4 EXAMPLES OF FUNCTIONAL ARCHITECTURE

A simple system for measuring small weights is shown in Figure 3(a). The weights W are placed on a plate fixed to the end of a pivoted beam such as to produce a torque in an anticlockwise direction. A spiral spring fixed about the pivot produces a torque in a clockwise direction proportional to the beam angular deflection G. At some point the torques will balance, and G is an indication of W. The block diagram for this system shows that there is inherent feedback, in that the mechanical indication G produces a torque via the spring constant (feedback path), which subtracts from the torque produced by the weights on the beam (feedin path) to give an error torque. Linearity between G and W is very dependent on the linearity between G and the spring torque.

A torque balance that results in very little beam movement and, therefore, greater accuracy is shown in Figure 3(b). Here, the clockwise torque is produced by the current i in a moving-coil element fixed about the pivot point. Without weights the beam remains stationary and i is zero. When weights are applied, the beam tends to move in an anticlockwise direction, and this movement is detected by sensors, for example, photodiodes, which produce an electrical signal; this is amplified to give current i, which, in turn, produces the counter torque to almost maintain beam balance. As there is little angular movement of the moving-coil element, the current-torque relationship is linear and the electrical output i is proportional to W. The output meter can be some distance from the point of measurement.

Because there is little mechanical movement using this method of measurement, it can be used to measure more rapid changes in W than the fully mechanical scheme of Figure 3(a).

5 FORCE AND TORQUE BALANCE

Force and torque-balance systems are very common and are employed, for example, in precision weighing, for the measurement of acceleration, pressure, flow, level, electrical power, and high voltage. These quantities can be converted to a force or torque; for example, acceleration of a fixed seismic mass produces a force on the mass proportional to acceleration, and a dynamometer movement produces torque on a shaft proportional to electrical power in the movement. The relationship between the quantity to be measured and force may be nonlinear.

The main methods of force and torque balance are shown in Figure 4, where F is the force to be measured and F_f is the balancing or feedback force generated in the particular instrument. For the beam and shaft, F and F_f are converted to torques T and T_f respectively. In each case, the element being used (seismic mass, diaphragm, beam or shaft) is displaced by a small amount in the direction in which F acts before the equilibrium balance $F = F_f$ is achieved. This displacement is the information required by the instrument to generate F_f.

6 BLOCK DIAGRAM AND DAMPING

The block diagram of a basic force (or torque) balance instrument is given in Figure 5(a). The input creates a force (or torque) via, say, a constant K_1, while the output creates the balancing force (or torque) via a precision actuator; if this is of a moving-coil type, it will have the transfer function of a single-order system as shown, where T_1 is the time constant (s is the Laplace complex variable). The mechanical moving element has a second-order transfer function, where m is the mass (or moment of inertia), k is the spring stiffness, and d is the viscous damping (inherent or artificial). K_d is the sensitivity of the displacement sensor and K_a is the gain of an amplifier, while the term $1 + T_a s$ is the transfer function of a (ideal) phase-advance network, providing rate compensation and feedback damping.

In fact this network is likely to be essential to stabilize the system, and this is illustrated by the typical Bode magnitude plots in Figure 5(b). Curve A is the system open-loop gain measured in the region of unity open-loop gain, with $T_a = 0$ *and* $K_a = 1$. The curve has a slope of slightly more than -12 dB/octave at the unity-gain frequency of 10 Hz, and the system is unstable in the closed loop. A suitable phase-advance characteristic is shown as curve P (in practice it is limited as shown by the dotted line), and with this phase-advance in the loop, the system open-loop gain is given by curve B. It is clear that the slope is now less than -12 dB/octave at unity gain and the closed-loop system is stable. A resultant closed-loop step response is shown in Figure 5(c), indicating that the damping is still fairly small. The steady state gain of the instrument or static characteristic $c/r = K_1/K_f$, provided that the open-loop static gain $K_d K_a K_f/k \gg 1$; the accuracy of such instruments is usually better than 1% of full scale. This static gain is chosen to give just sufficient accuracy, because as its value is increased stabilizing the instrument becomes more difficult. Most feedback-measuring instruments incorporate a stability network of some form. Having demonstrated the need and use of such networks, we will not show them in subsequent diagrams. The detailed stability analysis of automatic weight balances has been given by Smirnova

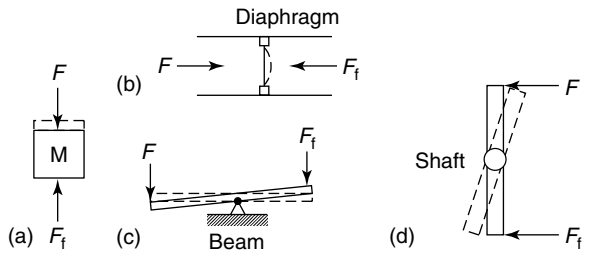


Figure 4. Force balance with: (a) a seismic mass M, (b) a diaphragm; torque balance with (c) a beam, and (d) a shaft. Broken line, position before application of F, when $F = F_f = 0$. (Reproduced from *Feedback in Instruments and its Applications*, Jones, 1979, with permission of Institute of Physics Publishing Ltd.)

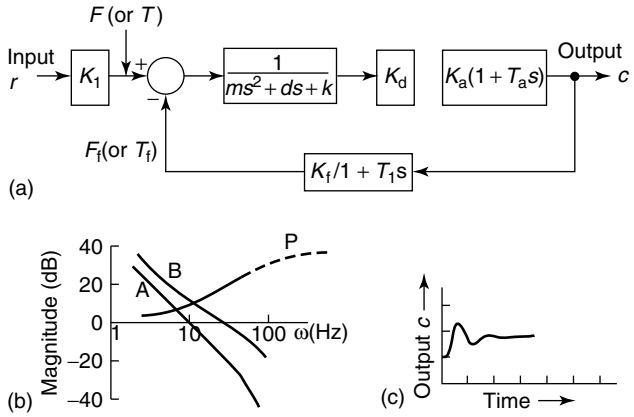


Figure 5. Basic force (or torque) balance instrument: (a) block diagram, $K_f/1 + T_1 s$ (b) open-loop Bode magnitude plots, and (c) step response (time scale 20 ms per division). (Reproduced from *Feedback in Instruments and its Applications*, Jones, 1979, with permission of Institute of Physics Publishing Ltd.)

(1960), and Gal'perin and Kolesov (1972). An example of an automatic chemical balance has been given by Smith and Stevens (1959).

7 EXAMPLES OF FORCE-FEEDBACK SENSORS

Various force- and torque-balance instruments are shown schematically in Figure 6, and illustrate the methods of balance, some typical devices used, and the wide application of this form of balance. The arrangement of Figure 6(a) can be used to transmit force information over a long distance. An input movement on a force spring, due to a measurand, creates a torque on a beam with a flexure pivot. Any movement of the beam is detected by the position detector and a current is created in the moving-coil actuator (or force motor) to produce the balancing torque. The output DC is linearly related to the input movement. This two-wire transmitter has a 'live zero' current output, and its sensitivity can be altered by changing the spring stiffness.

A pneumatic pressure transmitter is shown in Figure 6(b). The pressure unit is completely sealed and P_2 could be the pressure at the bottom of a petrol tank, enabling the level of petrol in the tank to be monitored. The output of

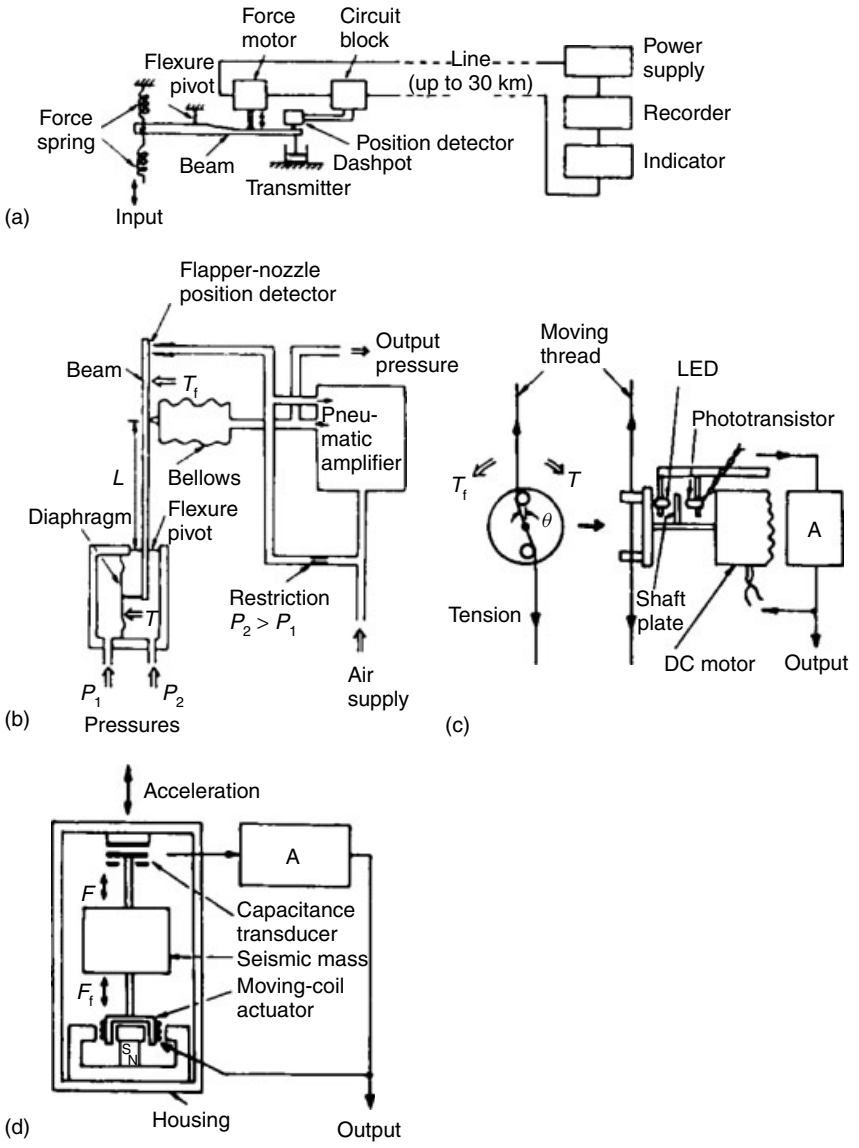


Figure 6. Various force- and torque-balance instruments: (a) two-wire transmitter, (b) pneumatic pressure transmitter, Flexure pivot, $T_1 \Rightarrow T_f$ (c) thread-tension monitor, and (d) seismic-mass accelerometer. (Reproduced from *Feedback in Instruments and its Applications*, Jones, 1979, with permission of Institute of Physics Publishing Ltd.)

this transmitter is air pressure, and the sensitivity of the transmitter is inversely proportional to length L. Various forms of feedback transmitter are used in industry.

The monitor of Figure 6(c) employed in the textile industry uses the torque of a DC motor to balance torque created by tension in the thread; an optical transducer is used to determine the angular position of the motor shaft and the monitor is described by Auckland and Hawke (1978).

In the accelerometer of Figure 6(d), the inertial force derived from the acceleration is balanced by the feedback force created by current in the moving-coil actuator; this has been described by Macdonald (1961).

It should be noted that the moving-coil actuators (both translational and angular) make use of the electromagnetic force created when a current-carrying conductor is at right angles to a magnetic field. The force is at right angles to both the conductor and the field, and is proportional to both the current i_f and field flux density; as a consequence, force-balance instruments employing such actuators usually have linear characteristics with a current as the output. The ratio T_f/i_f may vary from about 0.1 to 50 mN mA^{-1}, while the ratio F_f/i_f may vary from about 0.1 to 20 N A^{-1}.

8 OTHER FEEDBACK-MEASURING SYSTEMS

Feedback is widely used in instruments, and may well be the main structural element of a measuring system, linking the output indication with the measurand. Using feedback, it is possible to improve accuracy and speed of measurement, reduce the effect of interfering and modifying inputs, and allow remote indication and noncontact measurement. The property of inversion assists instrument design, and can assist in provision of digital indication. In general, the main disadvantages are increased complexity, size and cost, but the advent of small cheap integrated circuits has reduced these. The development of transducers and accurate precision actuators seems certain to extend the development and use of feedback-measuring systems.

Besides force and torque balance, other physically similar variables can be balanced in sensory systems: electrical current (and magnetic flux), heat flow, voltage, pressure, temperature, displacement, phase, and radiation, are described by Jones (1979).

Measuring devices where the output is in frequency form usually employ feedback to maintain oscillation. Feedback is used in instruments for control purposes, and the operation of the feedback is similar in both the control and measuring applications; the design for accuracy and stability is basically the same.

RELATED ARTICLES

Article 18, Nature and Scope of Closed-loop Systems, Volume 1; **Article 19, Dynamic Behavior of Closed-loop Systems, Volume 1**; **Article 109, Structure and Energy in Sensor Systems, Volume 2**.

REFERENCES

Auckland, D.W. and Hawke, J.R. (1978) A Thread-Tension Transducer Using Torque-Balance About the Axis of a Motor, *Proceedings of Transducer '78 Conference*, Trident International Exhibitions, Tavistock (pp. 1–4).

DiStefano, J.J., Stubberud, A.R. and Williams, I.J. (1990) *Feedback and Control Systems*, *Schaum's Outline Series*, 2nd edn, McGraw-Hill, New York.

Gal'perin, L.N. and Kolesov, Yu.R. (1972) Theory of Automatic Balances with an Electrical Compensator. *Measurement Techniques*, **15**, 546–549.

Jones, B.E. (1979) Feedback in Instruments and its Applications. *Journal of Physics E: Scientific Instruments*, **12**, 145–158.

Macdonald, W.R. (1961) in *Flight Test Instrumentation* (ed. M.A. Perry), Pergamon Press, Oxford (pp. 15–23).

Smirnova, N.A. (1960) The Theory of Automatic Beam Scales with Electrical Balancing. *Measurement Techniques*, **10**, 844–850.

Smith, B.O. and Stevens, J.W. (1959) Quick-Acting Chemical Balance *Journal of Scientific Instruments*, **36**, 206–209.

114: Models of the Sensor Interface

Qing Ping Yang
Brunel University, Uxbridge, UK

1 Types of Models and their Characteristics 767
2 Information and Knowledge Flow 769
3 Energy and Mass Flow Interaction 769
4 Semiotic Issues 770
5 Set Theoretical Expression of Interface Characteristics 770
6 Allowance for Influence Effects 771
7 Practical Sensor Interface Modeling 771
References 772

1 TYPES OF MODELS AND THEIR CHARACTERISTICS

1.1 Types of models

There are several different types of models for a measurement system interface. The simplest model (Figure 1) is based upon the traditional concept of measurement, which is defined as the comparison of the measurand with some fixed standard unit.

According to the representation scheme, there are physical, linguistic, pictorial, and mathematical models. Mathematical models are by far the most important because of the nature of measurement. In general, a measuring instrument may be described as

$$Z = f(X, Y, W) \tag{1}$$

where Z – output; X – input measurand(s); Y – interference or influence quantities; and W – latent parameters.

According to the level of abstraction, models may be used to represent power flow, signal flow, or information flow. Models may be functional or object oriented, for example, computer-based models implemented in functional or object-oriented programming languages. According to whether the model is based upon the physical attributes and structures or the observed data, functional models may be further classified as data models or physical models.

This section will mainly compare functional and object-oriented models. Other types of models are discussed elsewhere – **Article 13, Signals, Information and Knowledge, and Meaning, Volume 1**; **Article 71, Modeling Methodology, Volume 2**; **Article 104, Models of the Measurement Process, Volume 2**; and **Article 105, Modeling with LabVIEW™, Volume 2**.

1.2 Model characteristics

1.2.1 Functional models

Traditionally, measurement systems are usually modeled with a functional representation. This kind of model typically consists of three basic stages, that is, input stage including a sensor or transducer, signal processing stage and output stage, each of them represented as an input–output functional block, as shown in Figure 2.

Functional representations are well established in measurement and other branches of science. This kind of model is often used as the basis for classification of measuring instruments. They are very useful for determining system static and dynamic characteristics by means of the transfer characteristic and transfer function of each element – *see* **Article 57, Static Considerations of General Instrumentation, Volume 1**.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

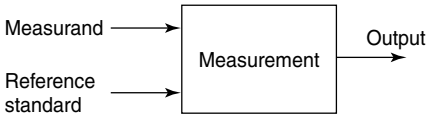


Figure 1. The simplest comparison model.

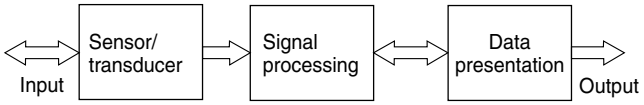


Figure 2. Functional model.

Common to the use of modeling in various problem domains, the intrinsic limitations of a functional model arise from its low level of abstraction. Its reusability is very poor as it is intended for specific systems use. Despite wide use of structural or modular design concepts, it is rather inflexible and often difficult to modify. The size of this kind of model, especially a computer-based one, often increases exponentially with complexity.

As has originated mainly from control engineering, functional models of measurement systems have often failed to address some issues that are of theoretical and practical importance to measurement, for example, operating space of measurement discussed in Stein (1992, 1996). As a result, incomplete descriptions of measuring systems are widely used, and the distinctiveness of measurement science has been seriously obscured. This is probably one of the important reasons measurement science has attracted little theoretical interest when compared with control engineering.

1.2.2 Object-oriented models

The object-oriented approach views a system as a collection of discrete objects that contain both data structure and behavior. The data structures and behavior or operations of an object are defined by a class. An object is an instance of a class. Objects can interact with other objects.

The development of object-oriented methods has been closely associated with computer programming. Use of object orientation first appeared in the development of a discrete event simulation language (Simula) in Norway, Bjørn (1989). As they began to mature in the late 1980s, object-oriented methods have spread to other areas, for example, system analysis and design. Indeed, they have become a general modeling approach in recent years, Rumbaugh (1991).

The power of the object-oriented approach lies largely in its closeness to the natural view of the real world. By raising the level of abstraction from the function to the object level, it focuses on the real-world aspects of a system, providing a better model of the problem space. Its consistence with the real world also provides a unifying platform for all the related activities, for example, analysis, design, and implementation.

Essential characteristics of the object-oriented approach include information hiding or encapsulation, inheritance and piecewise refinement. Objects encapsulate both their data structures and operations and they thus work at a higher level of abstraction. This makes object-oriented models more stable than conventional functional models because changes in operations or functions are decoupled and are localized within objects.

Inheritance comes from the class hierarchy structure, in which a class can have its own data structure and operations, and can also inherit them from its parent class(es). Although the concept originated in artificial intelligence, its inherent existence in class hierarchy structure is unique to object-oriented approach. The reusability of similar objects is one of the most important benefits of object-oriented methods.

The feature of piecewise refinement results from object-orientation and, in particular, its separated external and internal views of an object. Together with encapsulation, it provides a powerful mechanism to achieve high productivity, good maintainability, system integrity, and reliability.

Yang and Butler (1998) have demonstrated the application of object-oriented modeling for a generic measurement system, based upon a generalized measurement concept. A new framework was proposed in Yang and Butler (1997). Depicted in Figure 3 is the architecture of the object-oriented model for measurement systems. It captures important relations in a measurement system and embraces five objects or subsystems, that is, measured object, measuring instrument, human observer, referencing standard, and operating environment. The model represents a measurement system at three system levels: internal (measuring instrument), operational, and environmental.

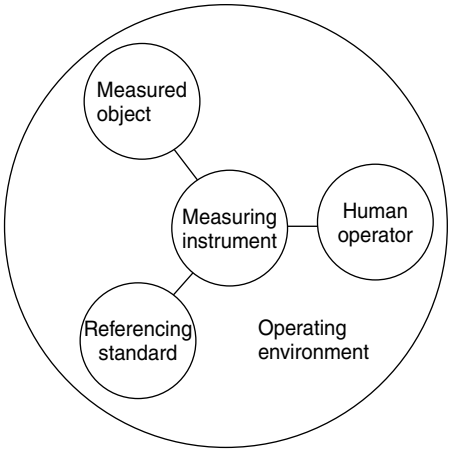


Figure 3. The architecture of an object-oriented model.

2 INFORMATION AND KNOWLEDGE FLOW

A measuring instrument or system may usually be conceptually viewed as a measuring chain, which is defined as a series of elements constituting the path of measurement signal from the input to the output. Sensor interface is the key element of a measuring chain.

The measuring chain can be modeled at different levels of abstraction. At the top level is information and knowledge flow, as the purpose of any measurement is to acquire information and knowledge about the measured system or process. The measuring chain can be viewed as the information path, handling sensing, processing, transmission, storage, and display of information. Although this chain is represented by different terms and blocks, they are invariably of this form.

The information-flow model of a sensor interface represents only the relevant information operation performed by the sensor interface. It typically takes the form of flow or block diagram.

As an information carrier, signal flow always accompanies information flow in a measuring chain. Signal flow is at the intermediate level of abstraction and is commonly used in the analysis, design, and evaluation of measuring instruments. Signal-flow models represent a measuring chain or process as transformed signals in various signal domains, for example, time domain, frequency domain, and time-frequency domain.

Since a signal results from the change or modulation of the energy state, useful signals and unwanted ones (noises) always exist together because of energy interactions. It is thus necessary to apply some form of signal conditioning and signal processing techniques to reduce the effect of noise and to increase the signal-to-noise ratio. Further processing including compensation and correction is also often required. The information is finally decoded or determined with reference to calibration and measurement scale.

Computer tools, both hardware and software, for modeling and processing signals are highly developed. There are many packages commercially available for handling the entire process of signal flows, for example, MATLAB® (http://www.mathworks.com/products/matlab/), and LabVIEW® (http://amp.ni.com/niwc/labview/lv.jsp) – *see* **Article 105, Modeling with LabVIEW™, Volume 2**.

3 ENERGY AND MASS FLOW INTERACTION

At the lowest level of abstraction of the measuring chain is energy or power flow. The system behavior of a measuring instrument and, indeed, a physical system can be modeled on the basis of the energy and power interactions.

One methodology first proposed by Paynter (1961) has been further developed by many researchers. The model is typically presented in a graphic form called a *bond graph* that represents a physical system in terms of power bonds connecting the system elements to so-called *junction structures* that model different types of energy constraints.

The system elements are generally classified into several standard idealized types, namely, storage elements (C-element and I-element), resistor, sources, transformer, and gyrator. There are two types of junction structures, namely, 0- and 1-junctions corresponding to the two generalized Kirchhoff's laws. They are generalized elements and domain-independent, and can describe different systems (e.g., mechanical, electrical, hydraulic, acoustical, thermodynamic, material) in the same way, using analogies in both, their governing mathematical equations and physical concepts.

The bonds, normally labeled and directed, are the energy connections between the two power ports of two interacting elements. Each port represents the physical interaction with other elements or the system environment. The energy flow along each bond or port has a physical dimension of power that can be described with a pair of terminals or variables, flowing in opposite directions.

Such a combination can be found in various physical domains, for example, voltage and current in electrical domain, force and velocity in mechanical domain. One of them may be classified as *effort* variable (e.g., voltage, force), the other as *flow* variable (e.g., current, velocity).

A bond graph has many properties of object-oriented modeling. It can also be easily converted to signal-flow graph and block diagram. An example bond graph for an RLC circuit is shown in Figure 4.

A power-flow model can be used to completely explain the behaviors and characteristics of a physical system such as transducers because it encompasses all the significant energy flows related to the system. System equations governing variables at various ports or elements can be systematically derived from a bond graph. The procedure

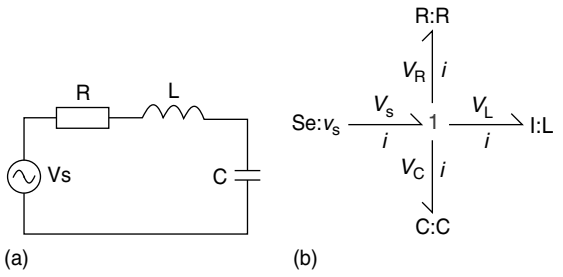


Figure 4. (a) RLC circuit and (b) corresponding bond graph.

may be algorithmized and performed by computer software, for example, ENPORT®, ARCHER®, Camp-G®, and so on.

The modeling of instruments based upon this methodology is given in Finkelstein and Watts (1982). A similar approach was presented in Stein (1992, 1996) as a three-port six-terminal transducer model.

4 SEMIOTIC ISSUES

Semiotics is the study of signs and symbols in various fields, especially language. Its theory explains how the meaning arises during the perception and interpretation of signs or sensory data. Since measurement can be regarded as a symbolic measurement process, there are similar semiotic issues concerned with measurement.

Different models can be used to explain the relationship between signs and their meaning. The most popular one is the triadic model or semiotic triangle developed by Charles Sanders Peirce (1839–1914), as shown in Figure 5.

The semiotic triangle based upon the triadic model consists of three elements:

1. the sign or symbol presented to an observer;
2. the meaning or the concept formed by the observer as the result of the perception of the sign; and
3. the object or referent the sign is referring to.

The meaning of a sign is not only dependent upon the sign itself but also influenced by the observer's experience with the referent. The semiotic model is obviously applicable to an empirical measurement process in which the measurand is the 'object' or 'referent', the measured signal is the 'sign', and the information about the measurand or the measured object can be regarded as the 'meaning' of the signal.

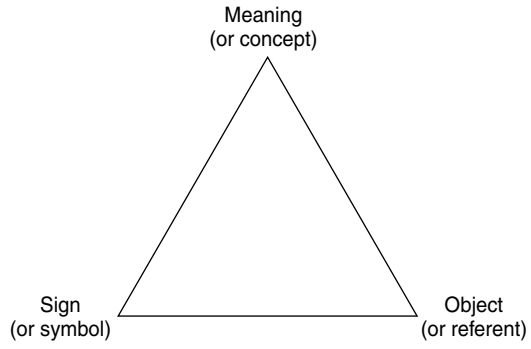


Figure 5. Semiotic triangle.

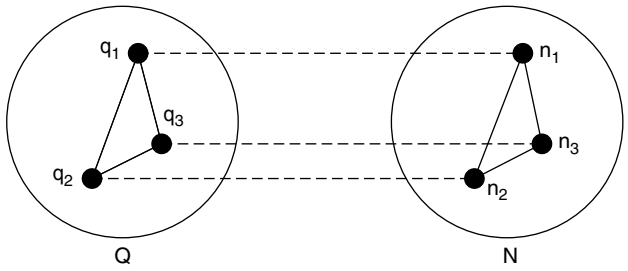


Figure 6. Mapping from observed space to number space.

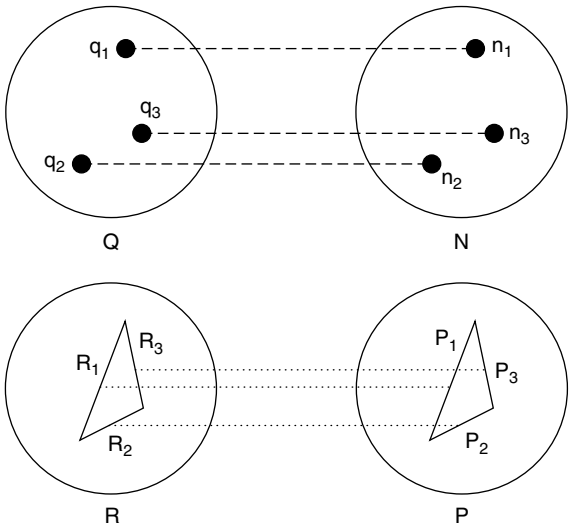


Figure 7. Mapping of individual observation value and the set of empirical relations.

Indeed, the relationship between the human perception and machine perception as shown in the previous block (Figures 6 and 7) is a triadic model, and the 'measurement triangle' in Figure 10 (*see* **Article 13, Signals, Information and Knowledge, and Meaning, Volume 1**) is also a semiotic triangle.

Measurement is primarily concerned with how measurement symbols (signals or data) are related to the observed object(s), whereas semiotics centers on how the symbols are interpreted. They are represented as two sides of the triangle, but they share the common third side, that is, the direct experience of the human observer.

5 SET THEORETICAL EXPRESSION OF INTERFACE CHARACTERISTICS

The nature of a transducer and indeed a measurement process may also be described using the representational

theory of measurement based upon set theory. Measurement may be viewed as a mapping process from an observed space (Q) to a number space (N), as shown in Figure 6.

The possible observation values of the measurand can be represented as a set, for which a set of relations exist. The measurand can thus be regarded as an empirical relational system, including the set of observation values and the set of relations.

According to the representation theory of measurement, as discussed in Pfanzagl (1968), Krantz *et al*. (1971), Narens (1985), and Finkelstein (1982), the nature of measurement is to map the above empirical relational system to a numerical relational system in which there are a set of numbers and a set of relations defined on the set of numbers. There are two mappings involved, one from the set of individual observational values (Q) onto the set of numbers (N), and the other from the set of empirical relations (R) onto the set of numerical relations (P). Together, these sets and mapping define a scale of measurement, Finkelstein (1982). The two mapping processes are also shown in Figure 7.

Given a scale of measurement in addition to the measurement result as numbers, the information or knowledge about the measurands and their relations can be determined.

The above discussions are based upon numerical representation, which can be easily extended to a symbolic representation.

6 ALLOWANCE FOR INFLUENCE EFFECTS

As the sensor interface interacts with other objects including a measured object, reference object, environment, and even possibly a human observer, the energy connections prevailing will undoubtedly influence the behavior and characteristics of the sensor interface. Most models, based upon traditional functional representation, represent well the internal functional aspects of a sensor interface, but are weak in their modeling of the operational and, to some extent, environmental interactions.

In the modeling of measurement systems, functional models tend to be oversimplified, neglecting some important aspects of a measurement system, for example, the coupling interactions between the instrument and the measured object, although some improvement has been made by Finkelstein (1994) and Sydenham *et al*. (1994).

A better approach is object-oriented modeling as it can easily accommodate these external influences. In addition to the conceptual framework, Yang and Butler (1998) have developed an object-oriented model for a generic measurement system. For example, the interactions and influence effects in a measurement system may be modeled as shown in Figure 8.

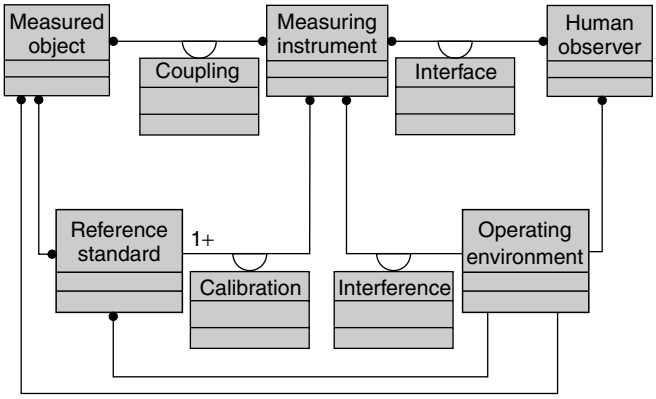


Figure 8. Object-oriented model of measurement systems.

It should be mentioned that the bond-graph model is essentially object oriented, and as such it can also conveniently model the various influence effects.

7 PRACTICAL SENSOR INTERFACE MODELING

A number of practical techniques are available for modeling a sensor interface. They broadly fall into two categories.

The first one is to model the system functions based upon observed or experimental data. It represents the functional relationship between the inputs and outputs of the sensor interface without explicit correspondence between the model parameters and physical attributes.

Traditional techniques of doing the above include regression models, time series models, and system identification. More recently, artificial neural networks are also used for modeling sensor interfaces. It has been proved that a feed-forward artificial neural network with one hidden layer can approximate arbitrary continuous mapping, Cybenko (1989) and Funahashi (1989). An example of three-layered feed-forward artificial neural network is shown in Figure 9.

These traditional techniques and artificial neural networks are normally applied using computer software packages, for example, MATLAB® and its relevant toolboxes.

The second category of modeling techniques is based upon physical modeling, which relates model parameters to physical attributes and structures.

Simple models may be developed manually, but computers have to be employed for more complicated modeling. The models can be developed using general programming languages or special software packages. LabVIEW

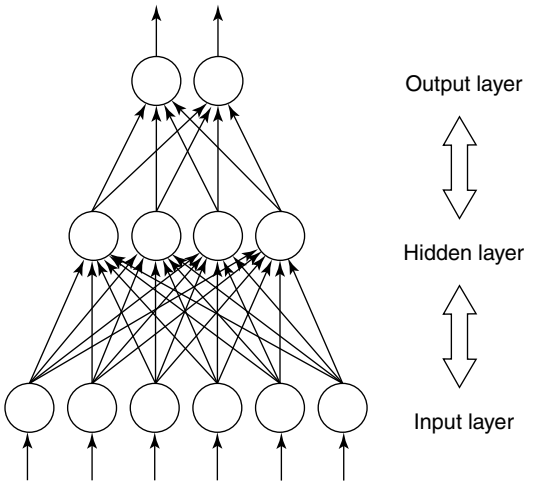


Figure 9. A three-layered feed-forward neural network.

from National Instruments and HP-VEE® (Hewlett-Packard Visual Engineering Environment) are good examples for instrument modeling – *see* **Article 105, Modeling with LabVIEW™, Volume 2**. They are essentially a graphical programming language specially developed for instrumentation and are intended for virtual instrumentation applications.

Another well-established technique is finite element method (FEM) as a mathematical simulation of the actual physical structure. The model is divided into a number of regions known as elements, each of them defined by a pattern of nodes. The method defines an approximation within each element, with appropriate continuity conditions imposed on the interelement boundaries. FEM can be applied to linear and nonlinear structural, thermal, dynamic, electromagnetic, and flow analysis. A number of software packages are commercially available, for example, ANSYS®, EMFlex®, and so on. These make the task relatively easy, for the mathematics is hidden, allowing practitioners to apply the practical inputs needed.

As described in Section 3 here, bond graphs have been successfully applied to build power-flow models. A bond-graph model can be both power and information oriented. There are several software packages for general bond-graph modeling, for example, ENPORT®, ARCHER®, Camp-G®, BondLab®, SYMBOLS®, Modelica®, and so on. One particular package, developed specifically at City University, London for transducer modeling, was MEDIEM (Multi-Energy Domain Interactive Element Modeling).

Further information on the difficulties and shortcomings of modeling of the measurement systems interface is found in the keynote and in other papers of the IMEKO TC-7 Conference held in Krakow, Poland, June, 2002. Key papers are 'in production' with the *Measurement Jnl* at the time of writing.

REFERENCES

Bjørn, K. (1989) *Object-Oriented Programming with Simula*, Addison-Wesley, Wokingham.

Cybenko, G. (1989) Approximation by Superposition of a Sigmoidal Function. *Mathematical Control Signal Systems*, **2**, 303–314.

Finkelstein, L. (1982) Theory and Philosophy of Measurement, in *Handbook of Measurement Science*, *Volume 1 Theoretical Fundamentals* (ed. P.H. Sydenham), John Wiley & Sons, Chichester (pp. 1–30).

Finkelstein, L. (1994) Measurement and Instrumentation Science – An Analytical Review. *Measurement*, **14**, 3–14.

Finkelstein, L. and Watts, R.D. (1982) Mathematical Models of Instruments – Fundamental Principles, in *Instrument Science and Technology*, Vol. 1 (ed. B.E. Jones), Adam Hilger Ltd, Bristol, CT (pp. 9–27).

Funahashi, K. (1989) On the Approximation Realization of Continuous Mappings by Neural Networks. *Neural Networks*, **2**(3), 183–192.

Krantz, D.R., Luce, R.D., Suppes, P. and Tversky, A. (1971) *Foundations of Measurement*, Academic Press, New York.

Narens, L. (1985) *Abstract Measurement Theory*, MIT Press, Cambridge, MA.

Paynter, H.M. (1961) *Analysis and Design of Engineering Systems*, MIT, Cambridge, MA.

Pfanzagl, J. (1968) *Theory of Measurement*, Physica-Verlag, Würzburg, Vienna.

Rumbaugh, J. (1991) *Object-Oriented Modelling and Design*, Prentice Hall, Englewood Cliffs, NJ.

Stein, P.K. (1992) *The Unified Approach to the Engineering of Measurement Systems-Basic Concepts*, Stein Engineering Services Inc., Phoenix, AZ.

Stein, P.K. (1996) The Unified Approach to the Engineering of Measurement Systems for Test & Evaluation – A Brief Survey, in *Proceedings IEEE Instrumentation and Measurement Technical Conference*, Brussels (pp. 1–28).

Sydenham, P.H. *et al.* (1994) Unsolved Problems of Measurement – An International Study, in *Proceedings XIII IMEKO World Congress*, Turino (pp. 827–831).

Yang, Q. and Butler, C. (1997) On the Framework of Measurement Science, in *Proceedings XIV IMEKO World Congress*, Vol. 5, Tampere (pp. 101–106).

Yang, Q. and Butler, C. (1998) Object-Oriented Model of Measurement Systems. *IEEE Transactions on Instrumentation and Measurement*, **47**(1), 104–107.

115: Designing the Sensor Interface

Qing Ping Yang
Brunel University, Uxbridge, UK

1 Deciding the Input and Output Forms of Energy 773
2 Transformation Representation Methodology 774
3 Specifying the System Parameters 775
4 Progressive Development Using the Transducer Conversion Logic 776
5 Use of Tools to Support Sensor Interface Development 777
References 777

1 DECIDING THE INPUT AND OUTPUT FORMS OF ENERGY

The information in a transducer is carried and conveyed by signals, which in turn reside in certain power or energy form. Signals can be regarded as the factors of energy or power.

Signals may be classified into various domains according to their energy forms, that is, electrical, mechanical, thermal, radiant, magnetic, and chemical. The above six signal domains, as proposed by Middlehoek and Noorlag (1981), are most commonly used, although it is possible to add additional domains. Some example signals in these six domains are given below.

Electrical: resistance, capacitance, inductance, current, voltage, and so on;
Mechanical: motion and dimensional: displacement, position, strain; length, height, width, thickness, area, volume; velocity; acceleration, vibration, force, load, torque, and so on;
pressure, sound, flow, and so on;
Thermal: temperature, heat, and so on;
Radiant: intensity, phase, frequency (wavelength), polarization, power, and so on;
Magnetic: field intensity, and so on;
Chemical: concentration, crystal structure, aggregation condition, and so on.

A transducer has energy connections with some other elements and the environment. Each connection interface is usually called *power port*. Intrinsically (without considering environmental interactions), a transducer has two or three power ports, that is, input, output, and modulation (or sometimes called *interrogation*), although modulation is not always necessary. Two-port and three-port transducers are shown in Figures 1 and 2.

A two-port transducer is also called *self-generating* or a passive transducer, and a three-port transducer is called *modulating* or active transducer (Sydenham, Hancock and Thorn, 1989; Middlehoek and Noorlag, 1983).

For example, a piezoelectric transducer is a two-port transducer and a strain gauge is a three-port transducer.

Since each power port of a transducer may be in one of the six energy domains, a two-port transducer has $6^2 = 36$ combinations, each as a transducing or conversion mechanism, for example, a piezoelectric transducer has direct conversion from mechanical to electrical.

Similarly, a three-port transducer has $6^3 = 216$ possible conversion mechanisms. For example, a strain gauge connected to a bridge circuit has a conversion from mechanical to electrical, which is modulated by electrical power. With both two-port and three-port transducers combined, there are 252 possible conversion mechanisms.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

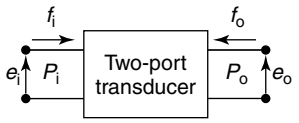


Figure 1. Two-port (self-generating or passive) transducer.

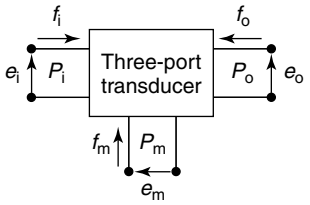


Figure 2. Three-port (modulating or active) transducer.

2 TRANSFORMATION REPRESENTATION METHODOLOGY

The input, output energy, and, if necessary, the modulation energy, together define a conversion mechanism or transduction principle of a transducer.

2.1 3D coordinates and matrix representation

A useful way of classifying transducers and physical effects is to represent the conversion mechanism in 3D space, with the energy form of input, output, and modulation as the three axes (e.g. x, y, and z axes, respectively), and each conversion principle or physical effect can thus be shown as a 'sensor effect cube' (i.e. a cube region in 3D coordinate system), as presented in Middlehoek and Noorlag (1981).

Consequently, each element of a measuring chain may be expressed as 3D coordinates, for example, $[x, y, z]$, where x, y, and z are one of the six energy forms, (Middlehoek and Noorlag, 1981).

On the basis of the 3D classification, Middlehoek and Noorlag (1981) have further proposed a methodology for describing a measuring instrument by joining the 3D coordinates of each stage, with additional lines drawn to show the energy source/flow as well as the signal flow.

Although it is a novel and interesting idea, one has to constantly check to see which element and which coordinate the lines are from, and to which element and which coordinate they go to. With all the coordinates quite similar or the same because only the energy form is given, it is quite difficult to interpret a conversion process.

A simpler approach shows each transducer as a transformation column (Figure 3), with the left and right vertical edges corresponding to the input and output energy form respectively, as presented in Middlehoek and Noorlag (1981) and Sydenham, Hancock and Thorn (1989). A line may be drawn to connect the input and output domains, representing the transformation from one energy to another.

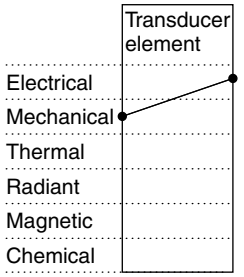


Figure 3. Matrix representation of a transducer. (Reprinted from Sensors and Actuators, Middlehoek, S. and Noorlag, D.J.W., Vol 2, Three-Dimensional Representation of Input and Putput Transducers, page 27, Copyright (1981), with permission from Elsevier.)

All the elements of a measuring chain can be shown as such columns and connected together as a matrix. The conversion lines may be joined together in the transformation matrix of a measuring chain, (Middlehoek and Noorlag, 1981).

For example, assume that a pressure transducer has a mechanical diaphragm to convert the measured pressure to displacement, which is measured by an intensity-modulated transducer. The light intensity is first converted to electrical current by a photodiode and then to a voltage with an amplifier. The corresponding matrix representation is shown in Figure 4.

It is quite straightforward to use this kind of matrix representation to describe conversion processes and energy flows. However, it does not include the modulation input directly, and the representations of energy flow and signal flow are not integrated.

Both the coordinates and matrix approaches proposed are mainly graphic presentations, and, as such, it is neither easy nor efficient to handle these representations using computers. *See also* **Article 109, Structure and Energy in Sensor Systems, Volume 2**.

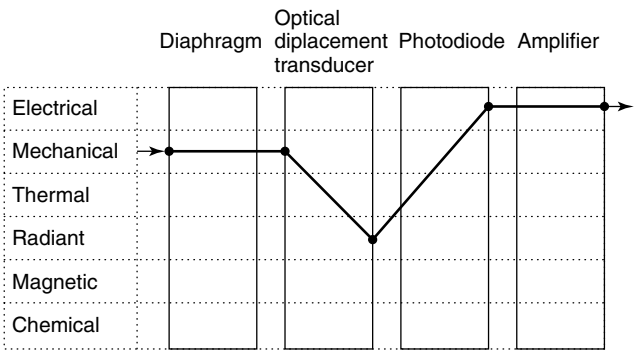


Figure 4. A pressure transducer using matrix representation.

2.2 Transducer conversion logic representation

A better representation is presented here, which combines energy and signal flows. It is compact, highly efficient, and easy to use. It also has a natural link to signal flow diagram and block diagram. This approach may be generally referred to as transducer conversion logic or language.

According to this approach, a two-port (or self-generating) and three-port (or modulating) transducer or element can be represented as

$$x \rightarrow y \tag{1}$$

and

$$x|z \rightarrow y \tag{2}$$

respectively, where x, y, and z are input, output, and modulation of a transducer respectively. They take a form Ds, where s is the signal concerned, D is one of six energy forms, which may be identified with the initials, that is,

E = Electrical

Me = Mechanical

T = Thermal

R = Radiant

Ma = Magnetic

C = Chemical

For example, a piezoelectric transducer has a direct conversion process

$$\mathrm{Me}.F \rightarrow \mathrm{E}.q \tag{3}$$

which converts mechanical force to electrical charge. A strain gauge connected to a bridge circuit can be represented as

$$\mathrm{Me}.\varepsilon \rightarrow \mathrm{E}.R|\mathrm{E}.S_v \rightarrow \mathrm{E}.v \tag{4}$$

where the strain is converted to electrical resistance, which is then converted to a voltage output with voltage modulation.

Of course, longer names can also be used for signals, if necessary. But as the signal domain is specified, it is quite easy to understand standard symbols for variables or signals that are commonly used. In fact, the same symbol, but in different domains, can be easily resolved. This makes the representation more powerful.

The above conversion expression for a three-port or modulating transducer can be expanded to model a more complicated process in which the modulation itself is converted from another source, that is,

$$x|z \rightarrow y \tag{5}$$

where $z = x_1 \rightarrow z_1$ or $z = x_1|y_1 \rightarrow z_1$.

The combined conversion can be expressed as

$$x|(z \leftarrow x_1) \rightarrow y \tag{6}$$

or

$$x|(z \leftarrow x_1|z_1) \rightarrow y \tag{7}$$

Take the above strain gauge connected to a bridge circuit for example; assume the voltage supply is converted from an optical light source, which in turn is controlled by an electric current source, then the process can be expanded as

$$\mathrm{Me}.\varepsilon \rightarrow \mathrm{E}.R|(\mathrm{E}.S_v \leftarrow \mathrm{R}.I|\mathrm{E}.S_i) \rightarrow \mathrm{E}.v \tag{8}$$

where R.I is the light intensity in radiant (R) domain, E.S_v is the voltage source, and E.S_i is the current source.

3 SPECIFYING THE SYSTEM PARAMETERS

The best way to specify the system parameters for a measurement system is to follow a systematic approach, given by Bosman (1978) and the framework proposed by Yang and Butler (1997, 1998). The latter is based upon systems approach and object-oriented modeling.

Figure 5 shows the general model architecture. It captures several important aspects of a measurement system. It encompasses five subsystems: measuring instrument,

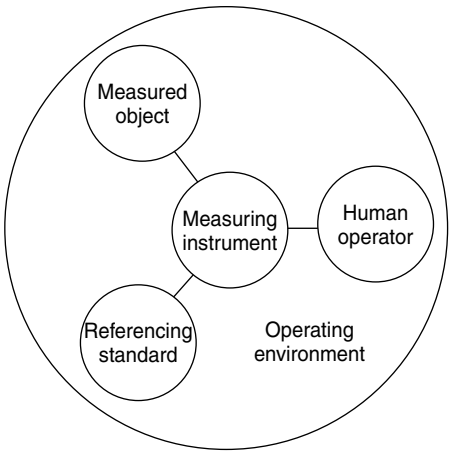


Figure 5. Object-oriented model for system specification.

Table 1. Quality matrix for system specification.

	Internal	Operational	Environmental
Features	Size Weight Principle Material	Data format Coupling method Calibration method Human interface	Power supply Operating environment
Capability	Measurand(s) No. of channels Intelligence	Networking interface	Temperature compensation
Performance	Resolution Measuring range Dynamic range Repeatability Uncertainty Frequency response Drift	Flexibility (object dependence) User friendliness Calibration frequency	Cross sensitivity EMC

measured object (measurand), referencing standard, human operator, and operating environment. The measurement results can be affected by all the five subsystems, with the instrument as the most important one. The system may be considered at three different system levels, that is, internal, operational, and environmental. At each level, important attributes, capability, and performance can be specified. The specification may be presented as a quality matrix as shown in Table 1, which was first used by Yang (1995).

4 PROGRESSIVE DEVELOPMENT USING THE TRANSDUCER CONVERSION LOGIC

The transducer conversion logic or language presented in Section 2 has the key features of object-orientation, such as hierarchy and inheritance, piecewise refinement, and information hiding. Thus, it is essentially an object-oriented approach to description and modeling of a measuring chain and conversion processes.

Since the transducer conversion logic has a very simple syntax and consists of only texts, it can be easily handled and processed by computers. For many transducers, typical conversion mechanisms can be used as default or options, which makes it much easier to automate analysis and design of transducers or instruments.

In addition, the representation can be easily converted to a signal flow or block diagram. The use of the transducer conversion logic can be further demonstrated, taking the above pressure transducer, for example.

The measured pressure is applied to a mechanical diaphragm, whose deformation is sensed by an intensity-modulated optical fiber sensor. The light intensity is then converted to electrical current by a photodiode and finally converted to a voltage with a transimpedance amplifier. The measuring chain may be first represented as

$$\mathrm{Me}.p \rightarrow \mathrm{Me}.\varepsilon \rightarrow \mathrm{R}.I \rightarrow \mathrm{E}.i \rightarrow \mathrm{E}.v \tag{9}$$

where p is the measured pressure; ε is the diaphragm deformation; I is the output light intensity from the optical fiber sensor; i is the current output from the photodiode; and v is the amplifier voltage output.

More details about the light intensity modulation can be added as follows:

$$\mathrm{Me}.\varepsilon|\mathrm{R}.S_I \rightarrow \mathrm{R}.I \tag{10}$$

where S_I is the light source, and

$$\mathrm{Me}.\varepsilon|(\mathrm{R}.S_I \leftarrow \mathrm{E}.S_i) \rightarrow \mathrm{R}.I \tag{11}$$

where S_i is the current source.

The latter expression shows that the mechanical deformation is converted to light intensity, with the light source powered by an electric current source. Note that more details can be simply added without affecting the structure.

Assume the photo detector is voltage biased, then we have

$$\mathrm{R}.I|(\mathrm{E}.v = \mathrm{const}) \rightarrow \mathrm{E}.i \tag{12}$$

The expanded chain can thus be expressed as

$$\mathrm{Me}.p \rightarrow \mathrm{Me}.\varepsilon|(\mathrm{R}.S_I \leftarrow \mathrm{E}.S_i) \rightarrow \mathrm{R}.I|\mathrm{E}.S_v \rightarrow \mathrm{E}.i \rightarrow \mathrm{E}.v \tag{13}$$

In fact, this expression may be easily converted to a block diagram as shown in Figure 6.

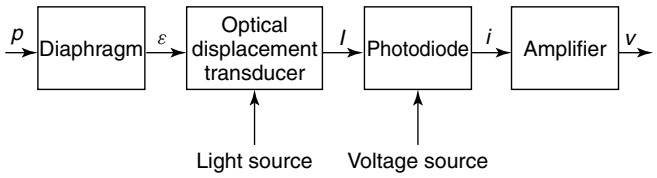


Figure 6. Block diagram of the pressure transducer.

5 USE OF TOOLS TO SUPPORT SENSOR INTERFACE DEVELOPMENT

The complete process of sensor interface development starting from requirement and specification encompasses several important stages, namely, design (including modeling, simulation, optimization), construction, test, and evaluation. Various tools and methods, especially computer-aided tools, are needed during the development process.

Computer-aided design (CAD) is widely used in engineering design, and there are many commercial packages, for example, AutoCAD®, Pro/ENGINEER®, SolidWorks®. There are also software packages for both design and manufacturing, that is, CAD/CAM (computer aided manufacturing) tools. In addition, CAD packages are also available for electronic or optical design.

Modeling and simulation have become increasingly important in product development. As discussed in the previous article, mathematical modeling may be carried out using some software packages, for example, MATLAB®, LabView®, and so on. These models may be implemented as traditional procedural algorithms or as the more recent object-oriented programs – *see also* **Article 105, Modeling with LabVIEW™, Volume 2** and **Article 71, Modeling Methodology, Volume 2**.

The finite element method (FEM) is a very powerful tool and can be applied to linear and nonlinear structural, thermal, dynamic, electromagnetic, and flow analysis. A number of software packages are commercially available, for example, ANSYS®, EMFlex®, and so on.

The modeling can also be based upon the use of bond graphs. A number of software packages have been developed for this kind of modeling, namely, ENPORT®, ARCHER®, Camp-G®, BondLab®, SYMBOLS®, Modelica®, and so on. One special package developed specifically for transducer modeling is the MEDIEM (multi-energy domain interactive element modeling, Sydenham, Hancock and Thorn, 1989).

Once the transducer models are created, simulation and optimization can be further performed. The criteria for optimization need to be clearly identified and defined, and optimization generally involves analysis of a number of different design configurations and/or parameters. The process can be quite computing intensive and time consuming.

Alternatively, it may be easier to follow the design of experiment methodology (DOE) to experimentally model and optimize the design. One popular methodology is the robust design developed by Taguchi (1986), consisting of system design, parameter design, and tolerance design. There are also software tools available for the use of design of experiment. The design of an experiment may be used together with the above theoretical modeling.

REFERENCES

Bosman, D. (1978) Systematic Design of Instrumentation Systems. *Journal of Physics E: Scientific Instruments*, **11**, 97–105.

Middlehoek, S. and Noorlag, D.J.W. (1981) Three-Dimensional Representation of Input and Output Transducers. *Sensors and Actuators*, **2**, 29–41.

Middlehoek, S. and Noorlag, D.J.W. (1983) Silicon Micro-Transducers, in *Instrument Science and Technology*, Vol. 2 (ed. B.E. Jones), Adam Hilger Ltd, Bristol, CT (pp. 33–44).

Sydenham, P.H., Hancock, N.H. and Thorn, R. (1989) *Introduction to Measurement Science and Engineering*, John Wiley & Sons, Chichester.

Taguchi, G. (1986) *Introduction to Quality Engineering – Designing Quality into Products & Processes*, Asian Productivity Organization, Tokyo.

Yang, Q. (1995) Basics and Theory of Measurement, in *Manufacturing Measurement*, Part 1 (ed C. Butler and Q. Yang), Chapter 2, Brunel University, Uxbridge (pp. 1–16).

Yang, Q. and Butler, C. (1997) On the Framework of Measurement Science, in *Proceedings XIV IMEKO World Congress*, Vol. 5, Tampere (pp. 101–106).

Yang, Q. and Butler, C. (1998) Object-Oriented Model of Measurement Systems. *IEEE Transactions on Instrumentation and Measurement*, **47**(1), 104–107.

116: Selection of Sensors

Paul P.L. Regtien
University of Twente, Enschede, The Netherlands

1 The Design Process 778
2 The Requirements 778
3 Selecting the Measurement Principle 779
4 Selecting the Sensing Method 779
5 Sensor Selection 779
References 780

1 THE DESIGN PROCESS

The design of a technical system involves making choices on the basis of criteria (from a list of requirements), availability of parts and materials, financial resources, and time. These aspects play a significant role when designing a measurement system.

Blanchard and Fabrycky (1998) distinguish six major phases of the design process: (a) conceptual design; (b) preliminary design; (c) detail design and development; (d) production/construction; (e) operational use/maintenance; (f) retirement – see also (Sydenham, 2004). Thus, sensor selection is a crucial activity in the systems design process, as it will make a great impact on the production of the measurement instrument and the performance during its entire lifetime and may even have consequences related to disposal.

Design methods have evolved over time, from purely intuitive (as in art) to formal (managerial). The process of sensor selection is somewhere in between: it is an act of engineering, in which the design is supported by advanced tools for simulating system behavior based on scientific knowledge. The basic attitude is (still) the use of know-how contained in the minds of people and acquired through experience.

2 THE REQUIREMENTS

Sensor selection means meeting requirements. Unfortunately, these requirements are often not known precisely or in detail, in particular when the designer and the user are different persons. The first task of the designer is to get as much information as possible about the future applications of the measurement instrument, all possible conditions of operation, the environmental factors, and the specifications with respect to quality, physical dimensions, and costs.

The list of demands should be exhaustive. Even when not all items are relevant, they must be indicated as such. This will leave more room to the designer, and will minimize the risk of being forced to start all over again at a later date. Rework is an expensive process and should be avoided where possible by reducing errors as early as possible in the systems engineering life cycle process. The requirements list should be made in such a way as to enable unambiguous comparison with the final specifications of the designed instrument. Once the designer has a complete idea about the future use of the instrument, the phase of the conceptual design can start.

Before thinking about sensors, the measurement principle has to be considered first. For the instrumentation of each measurement principle, the designer has a multitude of sensing methods at his or her disposal. For realization of a particular sensor method, the designer has to choose the optimal sensor type out of a vast collection of sensors offered by numerous sensor manufacturers.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

3 SELECTING THE MEASUREMENT PRINCIPLE

The design process is illustrated by using an example of a measurement for a single, static quantity: the amount of fluid in a container. The first question to be answered in this case is in what units the amount should be expressed – volume or mass. It is important to have a sound understanding of the physics involved and the circumstances of the situation. These may influence the final selection of the sensor.

Figure 1 shows the various measurement principles that could be used in this case:

A: the tank placed on a balance, to measure its total weight;
B: a pressure gauge on the bottom of the tank;
C: a gauging-rule from top to bottom with electronic readout;
D: level detector on the bottom, measuring the column height;
E: level detector from the top of the tank, measuring the height of the empty part.

Obviously, many more principles can be used to measure a quantity that is related to the amount of fluid in the tank.

In the conceptive phase of the design, as many principles as possible should be considered, even unusual or unorthodox ones. On the basis of the list of demands and not as a 'hunch', it should be possible to find a suitable candidate principle from this list, or at least delete many of the principles, on the basis of arguments.

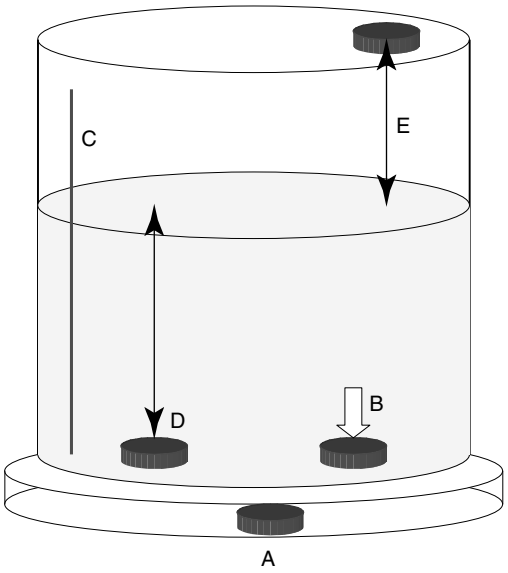


Figure 1. Measurement principles for the amount of fluid in a reservoir.

For instance, where the fluid must remain in the tank during measurement, principles based on volume or mass flow are excluded. If the tank contains a chemically aggressive fluid, a noncontact measurement principle is preferred, placing principles B, C, and D lower on the list, and so on.

Also, method A can possibly be eliminated because of its high costs for large containers.

In this way, the conceptual design ends up with a set of principles having related pros and cons, ranked according to the prospects of success.

4 SELECTING THE SENSING METHOD

After having specified a list of candidate principles, the next step is to find a suitable sensing method for each of them. In the example in Figure 1, we will further investigate principle E, a level detector placed at the top of the tank. Again, a list of the various possible sensor methods is made. This may be

1. a float, connected to an electronic readout system;
2. an optical time-of-flight measurement;
3. an optical range measurement;
4. an electromagnetic distance measurement (radar);
5. an acoustic time-of-flight (ToF) measurement; and so on.

As in the conceptual phase, these methods are evaluated using the list of demands, so not only the characteristics of the sensing method but also the properties of the measurement object (liquid level) and the environment should be taken into account. For the tank system, the acoustic ToF method could have an excellent chance because of being contact-free; or just the contrary, for instance, because of possible high temperatures. In this phase, it is also important to consider methods to reduce such environmental factors (*see* **Article 16, Reduction of Influence Factors, Volume 1**). Anyhow, this phase ends up with a list of candidate sensing methods with merits and demerits with respect to the requirements.

5 SENSOR SELECTION

The final step is the selection of the components that make up the sensing system. Here, a decision has to be made between a commercially available system and the development of a dedicated system. The major criteria are costs and time: both are often underestimated when a new development is considered.

In this phase of the selection process, sensor specifications become important. Sensor providers publish specifications in data sheets or on the Internet. However, the accessibility of such data is still poor, making this phase of the selection process critical and time consuming, in particular for nonspecialists in the sensor field.

Computer-aided sensor selection programs are under development and are partly realized, but up to now their use has been limited. A prerequisite for a general, successful tool facilitating sensor selection is the continuous availability of sensor data on the Internet, in a more standard format, and regularly updated. A good start is www.globalspec.com.

Obviously, the example of the level sensor is greatly simplified here. Usually, the selection process is not that straightforward. Since the sensor is often just one element in the design of a complex technical system, close and frequent interaction with other design disciplines is necessary.

More detail is given on the systems engineering process in **Article 63, Outline of Systems Thinking, Volume 2**; **Article 64, Executing A Measuring System Design, Volume 2**; **Article 66, Phases of System Life Cycle, Volume 2**; **Article 67, Principle of Concept of Operations (ConOps), Volume 2**; **Article 68, Setting the System Boundaries, Volume 2**; **Article 69, Requirements Allocation, Volume 2**; **Article 65, Life Cycle Concept, Volume 2**; and **Article 70, Measuring System Design Methodologies, Volume 2**.

REFERENCES

Blanchard, B.S. and Fabrycky, W.J. (1998) *Systems Engineering and Analysis*, Prentice Hall, Upper Saddle River, NJ.

Sydenham, P.H. (2004) *Systems Approach to Engineering Design*, Artech House.

117: Materials in Measuring Systems

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Role of Materials in Sensor Construction 781
2 Metals 782
3 Ceramics 782
4 Plastics 784
5 Influence Effect on Materials 785
6 Selecting a Material 786
7 A Rationale for Selecting Passive Materials 787
8 Altering the Active and Passive Behavior 787
9 An Example in Material Selection 788
10 Sources of Knowledge on Materials 790
References 790
Further Reading 790

1 ROLE OF MATERIALS IN SENSOR CONSTRUCTION

A report of the Fellowship of Engineering, RFE (1983), classified materials as having two main roles.

In their active role, they are 'hosts of potentially useful scientific phenomena, and always bring about a conversion of energy (or information) from one form to another'. They can provide this conversion using a gas (such as in a filled thermometer), fluid (liquid crystals in a display), or a solid (piezo-electric material in an accelerometer).

In the sensors of measuring systems, energy conversions are concerned with the provision of information about the physically existing world; modulation of an energy link carries the required information from the system being studied to the observer.

In their passive role, materials provide the capability to support, connect, control, contain, or display the sensing phenomenon provided by the active components of a constructed object.

All passive materials are active to some degree in that they respond to one or more forms of external energy. They are not, however, regarded as transducer materials unless the active transduction action is deliberately enhanced and controlled. For example, normal passive structural steel has mechanical hysteresis that varies far more than it does in the active instrumentation alloys processed to make precision springs and load cells.

Another characteristic often exhibited in a sensor material is reversibility of the transducer action. For example, piezoelectric material will produce electric charge when its shape is deformed; conversely, when an electric charge is supplied to it, the shape changes. Not all materials exhibit this.

Sensor materials ideally provide a linear, or at least reproducible, energy transformation. Efficiency of the energy transformation is usually less important than its all-important fidelity.

The concern of technology for producing sensor materials is to enhance the efficiency of the information transformation action. A requirement is also that the transduction action remains constant regardless of time and the in-service level of influencing variables and operational signals.

Some sensors rely on loss of some of their mass to produce signals. Examples include many chemical sensors where intimately related mass/energy transformations take place; and special, electrically resistive, tape used to monitor the abrasive level of magnetic recording tapes, whose resistance rises as the tape is worn away. In these cases, the design aim is to prolong the useful sensing life, usually

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

by increasing the sensing sensitivity, thereby requiring less mass consumption per unit time.

By the traditional design methodology, a sensor is created from a selected mixture of different active and passive materials, their integration chosen to enhance the sensing effect while attenuating undesirable internal and external influence effects.

In contrast, the modern integrated microelectronic form of sensor makes use of a common base material that is modified to enhance the sensing parameters while adequately controlling the nonsensing ones. In this microelectronic sensor form, highly pure silicon is doped, with a range of materials, at minute levels of impurity to make it sensitive to one of a wide range of measurands in certain positions in its bulk. By use of masks and other processing techniques – *see* **Article 162, Principles of MEMS, Volume 3**; **Article 163, Uses and Benefits of MEMS, Volume 3**; **Article 164, Principles of MEMS Actuators, Volume 3** – the doping and etching can be made to produce both active and passive areas in the two-dimensional and three-dimensional geometry of a single chip of the host material. Here, the technology is required to also control the passive properties.

Another class of integrated sensor is the fiberoptic sensor, wherein a basic element – the optical fiber – is modified by a range of methods to form many different sensors.

Many transduction principles in use today were discovered as physical and chemical effects in the nineteenth century (Sydenham, 1979). The twentieth century has been a period of enhancement of the passive and active uses of materials; their use in sensing systems requires extensive interdisciplinary understanding.

Three main groupings of material types that find application in measuring systems are

- metals
- ceramics, including semiconductors here for convenience
- plastics.

Depending on composition, processing, or application arrangement, these variously form either active or passive components.

Optical materials are covered in **Article 210, Optical Transducers for Chemical Measurements, Volume 3**; **Article 89, Optical Materials, Volume 2**; and **Article 91, Light Sources and Detectors, Volume 2**.

Brief mention is also needed of the many natural materials that were once commonly used, such as wood, leather, bone, ivory, mother of pearl, tortoiseshell, and the like. In general, these are rarely used now because of their relatively high cost, problems of supply, legal restrictions on species protection, and lack of uniformity of properties of the materials.

2 METALS

This group comprises the common, relatively inexpensive, base metals and alloys such as steel, brass, aluminum, and copper. Also used are special metal alloys such as Invar™, Ni-Span C™, Nichrome™, and others sold under proprietary names.

Literature on the common metals is widely available, especially on those used for structural purposes. Information on properties of the specialist metals, however, must usually be obtained directly from the alloy makers.

The common metals have found time-honored acceptance as passive engineering constructional materials due to their ease of working, wide temperature operating range, general availability, well-known history of performance, and range of useful structural properties.

The engineering use of metals is covered extensively in terms of analysis of their upper working load limits. Their use in measuring systems, however, is often more concerned with allowable deflection and stiffness under a working load.

Table 1 is a representative list of the metals used in instrument construction. Typical applications are given to indicate, not prescribe, which to use.

3 CERAMICS

The ceramic group of materials (glasses and semiconducting materials are included here but are not strictly defined as such) provide for such functions as thermal insulation or conduction, high-temperature mountings, and can provide very useful transduction processes for use as both sensors and actuators. While the use of ceramics is less common in instruments, they can be vital because they augment metals and plastics by providing properties the latter two cannot (Ichinose, 1987). Table 2 lists representative ceramics used in instrument construction along with some typical applications.

Silicon semiconductor material is the key to the advance of many miniature sensors and actuators, *see* **Article 162, Principles of MEMS, Volume 3**; **Article 163, Uses and Benefits of MEMS, Volume 3**; and **Article 164, Principles of MEMS Actuators, Volume 3**.

More information is provided on the ultrasonic use of ceramics in **Article 118, Ultrasonic Sensors, Volume 2**; **Article 119, Ultrasonic Instrumentation Principles, Volume 2**; and **Article 120, Ultrasonic Instrumentation Design, Volume 2**.

Table 1. Metals commonly used in instrument construction.

Metal	Application
Commonly available passive metals (often alloyed)	
Aluminum	Panels, mountings, leads, fittings
Brasses (copper–zinc alloys)	Diaphragms, mounts, contacts, springs
Bright steel (often hardened)	Frame, panels, shafts, gears, machined parts, pins, clips, springs, cams, knife-edges, pulleys, chains, flexure strips
Bronzes	Cast frames, cast mechanisms, bearings, gears, cams
Copper	Leads, printed circuit connections, connectors contacts, heat sinks, heat conductors, shielding
Iron	Castings, temperature compensation
Nickel	Corrosion resistant needs, temperature match with glasses
Platinum, Silver, and Gold	Contacts, electrodes, corrosion resistant layers
Solders (silver, tin–lead)	Connections, contact materials
Stainless steel	Springs, support frames, panels, shafts, flexure strips
Special passive alloy groups	
Beryllium–copper and phosphor–bronze (copper–tin)	Springs, sealing diaphragms
Copper–nickel, such as Monel	Corrosion resistance
Copper–nickel–chromium, such as Mu-Metal	Magnetic shielding
Nickel–iron, such as invar, and Ni Span-C	Temperature stable parts Minimal mechanical hysteresis
Nickel–chromium–iron, such as Hastellow C	Diaphragms, springs
Nickel–chromium, such as Nichrome	Electrical heaters
Silicon–iron	Transformer laminations
Active metals for sensing applications	
Bimetal strips	Temperature actuators and detectors
Electrical resistance alloys	Stable resistors, strain gauge foils
Load cell materials	Diaphragms, load cell structures
Magnetic alloys	Thermomagnetic effects, magnetic detectors
Nickel alloys	Magnetostriction
Thermally sensitive electrical resistance alloys and thermocouple alloys	Temperature sensors

Table 2. Some ceramic materials used in instrument construction (From Ichinose, 1987).

Ceramic type	Application
Passive constructional roles	
Alumina	• Heat resistance support
Beryllia	• Heat conducting support
Fused silicas	• Low TEMPCO parts
Graphite	• Heat resistant support, colloidal lubricant
Hydroxyl apatite	• Artificial bones and tooth root replacement
Manganese oxide	• Insulators, IC substrates
Silicon carbide	• Cutting tools, low-wear surfaces
Zirconia	• High-temperature mechanisms.
Active sensing roles	
Alumina	• Sodium batteries
Barium titanate (a PZT)	• Piezoelectric sensing and actuation
Ferrites	• Recording heads, RF transformer cores
Silicon dioxide	• Optical fibers
Stable zirconia	• Oxygen detection
Tin oxide	• Gas sensor
Zinc oxide–beryllium oxide	• Voltage-dependent resistor.

Source: Reproduced from *Introduction to Fine Ceramics: Applications in Engineering*, Ichinoise (1987) © John Wiley & Sons Limted.

A useful introductory text on fine ceramic materials is Ichinose (1987) from which much of this explanation is extracted. More detailed data must be sought from the makers because ceramics, like plastics, are materials with very complex formulae; numerous controlled production factors decide the final performance.

Some instrument ceramics are made with reasonably straightforward, specialized processes that convert natural materials such as sand and clays into the final form. Examples here include the manufacture of fused silica (needing electric arc refining of special sands) and insulating porcelains (liquid pug is cast in molds, followed by drying, and then high-temperature firing).

The advanced materials, so-called hi-tech forms, are much more complex to manufacture. These fine ceramics require to be manufactured using highly refined raw materials, rigorously controlled composition, and strictly regulated forming and sintering.

Production of ceramic powder, the starting material of many fine ceramics, is made with several processes, each being complex and needing expensive plant investment. The powder is formed into the required shape by one of five main sintering processes such as

- die casting
- rubber mold pressing
- extrusion molding
- slip casting
- injection molding

The item is then sintered to bring about the minimum system energy through densification of the powder. These processes require high temperatures (1400–1800 °C) to be present while the pressing action is in place. Final mechanical properties are decided by some twenty major parameters, each needing strict control of several parameters.

To illustrate the complex productions involved, consider making silicon 'chip' sensors. This is a long, many-stage process that requires a plant that is so expensive that many countries cannot afford to have such facilities. The process starts by pulling a pure silicon crystal, followed by zone refining it by sweeping it with heat. This is vital to obtain the highest purity material.

The quite large crystal is then sliced and polished to form thin wafers of pure silicon. These are subjected to several similar cycles of masking, doping, and developing to form the miniature electronic and mechanical components within the silicon. Vacuum deposition is then used to add the metal connections.

The silicon die is then tested and mounted on a support package, electrical leads being bonded from the die to the connection pins.

Finally the product is tested, packaged, and labeled.

Most silicon microelectronic manufacturing plants are built to make electronic circuits and, as such, are not always capable of providing the processes needed for certain sensor manufacturing needs – micromachining needed in MEMs forms, for example, requires special additional processes.

Once a ceramic or semiconductor item is made, it cannot be easily remanufactured, so these components must be ready formed and processed for specific applications. The high costs of manufacturing plant must be supported by large-volume production; the designer of small-runs of sensors has to make use of stock proprietary forms.

4 PLASTICS

Plastics have steadily displaced metals in many instrument application areas, such as in frames, in cases, and in many precision mechanisms. Caren (1988) summarizes their place well.

Plastics permit a greater amount of structural design freedom than any other material. Plastic parts can be large or small, simple or complex, rigid or flexible, solid or hollow, tough or brittle, transparent, opaque or virtually any color, chemical resistant or biodegradable, and materials can be blended to achieve any desired property, or combination of properties. The maximum temperature of operation of selected plastics is constantly being raised, being now well over the boiling point of water.

However, success in their use is very much affected by the part's design and processing. The designer's knowledge of all of these variables can profoundly affect the ultimate success or failure of a consumer or industrial product.

Plastics fall into two major groupings

- thermoplastics, which can be melted and reused;
- thermosets, which are chemically changed by the manufacturing process and cannot be remelted

Typically, thermoplastics are shaped by melting, forcing the (often highly viscous) liquid into the desired shape, and allowing it to freeze. Injection molding (for bulk material) and thermoforming (sheet material) are processes of this type.

Thermosets are compressed into a mold and heated, the heating process changing the chemical structure and forming the molding shape. However, these simplistic divisions are blurred, and injection-moldable thermosets are available. The heating required to set off the thermoset is also often aided by catalysts (as in dough molding) and the chemical change can be brought about by catalysis alone with no external heat as in composite lay-up (for example, fiberglass) construction.

Crystalline thermoplastics are generally stronger, more chemically resistant, and often self-lubricating. Examples are polyethylene, nylon, polypropylene, and acetal. Amorphous thermoplastics such as polystyrene vinyl (PVC), acrylic, and polycarbonate are more easily formed, decorated, and fastened and can be transparent.

Many plastic materials can be formed as foams with closed or open cell structure. Polystyrene foam is familiar as a rigid, yet light and strong, packaging material. Polyurethane foams can be rigid though soft for applications like cushioning material in instrument carry cases. Materials can also be made to foam during molding, forming a solid external skin with a foam core. This provides structural benefits (high strength to weight) and material savings as well as molding benefits from a reduced cycle time.

Individual plastic materials are supplied with a wide range of properties affected by the molecular weight of the bulk material, additives for modification of production performance and final product properties, and coloring agents. Suppliers have ranges of product to suit different processes and provide specialty blends for specific requirements.

An important subset of thermoplastics is the elastomer group. These are plastics with rubber-like elastic properties. SBR (styrene butadiene rubber) is used for protective boots,

Table 3. Plastics used in instrument construction. (Compiled from industry catalogs – a guide only – manufacturers must be consulted in making a detailed choice).

Common plastic name	Application
Passive constructional roles	
ABS	• Telephone handsets, high-quality cabinets, PCB plugs, transparent covers – is platable
Acetals	• Gears, pawls, links, cams, cranks – is platable
Acrylics	• Optical lenses and covers
Cellulosics	• Blister packaging, poor durability
Epoxys	• Circuit boards, hard surfaces
Fluoroplastics (includes PTFE)	• Electrical insulation, bearings, valves, linings, implants
Nylons	• Gears, snap-on parts, hinges, guide rolls, cams, plug connectors, brush holders, print wheels, hot-water valve bodies, switches, coil formers press buttons, bearings
Polycarbonates	• Tough, transparent structures, lenses, safety enclosures, cases
Phonelics	• Low-cost molded parts
Polyimids	• High temperature, low-creep, instrument parts
Polyesters	• Resin for glass, carbon and aramid fiber-based composites, containers, films
Polyethylenes	• Nozzles, containers, cable clamps, cases, cable insulation
Polystyrenes	• Lamp shades, diffusers, instrument frames, panels, video cassette parts, foam thermal insulation and packaging
Polypropylenes	• Appliance housings, cable sleeves, sterilizable parts
Polyurethane	• Cushion soft to glass hard needs, fascias, tubing, cords
Poly (vinyl chloride)	• Cable ducts, pipes insulation on cables switch covers, medical parts
Silicones	• Flexible seals, protective covers, molded seals *Active sensing roles*
Polyvinylidene fluoride (PVDF)	• Piezo electric
Polyacetylene (and others)	• Electrical conduction

shock-absorbing mounts, and grommets. Polyurethane (a thermoset) can be made in hardness ranging from glass hard to soft and rubbery. It is prepared by mixing two liquids and finds applications in cast-in-place elastic elements. Silicone rubbers provide rubbery consistency with excellent chemical resistance, although if used in contact with sensitive items (such as foil strain gauges), a grade should be chosen which does not emit acetic acid on setting.

Some combinations of polymers can also be mixed together to form copolymers (where the chains of the different materials cross-link) or alloys in which the ingredients remain discrete. ABS is a typical terpolymer (a copolymer of three ingredients) while PC/PVC is an alloy.

While the range of plastic materials available presents a large spectrum of opportunity for the designer, the possibilities added by additives, fillers, and reinforcers increase this enormously. These can modify finished part strength, impact resistance, density, and color. The performance of the material in production can also be affected. Materials can be added to plastics in bulk, during molding or by hand lay-up.

In general, the measuring instrument and sensor designer need only be aware of the possibilities arising from the use of additives, leaving detailed recipes to the plastics supplier and manufacturer. Such additives as biocides, fungicides, heat and light stabilizers, and antistatic agents can be quite vital to the success of a product, especially those that relate to dimensional stability and to life.

Fillers are mainly inorganic materials such as talc, kaolinite, feldspar, and glass microspheres. They can improve processability, reduce shrinkage, increase stiffness, and reduce cost.

Reinforcers are fibrous materials including various glass fibers, aramid, carbon, and more exotic materials. Their major use is increasing strength and stiffness, often anisotropically, providing finished products, which, on a strength-to-weight basis, are often comparable with high-strength steels.

All additives can affect surface finish and subsequent decorative processes, those added for physical property improvement frequently adversely affecting the finish.

The selection of plastics is a highly skilled task. Plastics manufacturers, however, provide extensive data and design services including electronic forms of adviser, Endura (2004) and Rapra (2004) being examples of guides. Kroschwitz (1990), Rubin (1988) provide greater and more generic detail.

Table 3 is provided as a first guide to the type of plastics commonly used for the construction of various instrument parts.

5 INFLUENCE EFFECT ON MATERIALS

As explained elsewhere – *see* **Article 16, Reduction of Influence Factors, Volume 1** – all materials respond, to some degree, to a range of external influence effects.

Table 4. Physical effects give rise to unexpected activity in materials.

Term describing effect of	Input variable on (most are reversible)	Output variable
Gas law effect (at constant volume)	Pressure	Temperature
Electrochemical effect	Electric current	Material decomposition
Electromagnetic effect	Electric current	Force
Electrostatic effect	Electric voltage	Force
Faraday effect	Magnetic field	Polarization angle
Hall effect	Magnetic field	Current/voltage
Magnetoresistive effect	Magnetic field	Electrical resistance
Magnetostrictive effect	Magnetic field	Length
Photoelastic effect	Strain	Optical polarization
Piezoelectric effect	Force	Charge
Piezoresistive effect	Force	Electrical resistance
Poisson effect	Longitudinal load	Transverse length
Pyroelectric effect	Heat	Charge
Rectifying effect	Electric current	Flows only one way
Resistance–effect	Strain	Electrical resistance
Self-heating effect	Electrical power	Temperature
Shear modulus effect	Shear load	Length
Temperature velocity effect	Temperature	Velocity of sound
Thermoelastic effect	Temperature	Elasticity
Thermoelectric effect	Temperature	Voltage
Thermoresistive effect	Temperature	Electrical resistance
Young's modulus effect	Tensile load	Length

These can change shapes of passive material support structures and sensitivities of sensing active materials and as such can considerably impact sensor and instrument performance.

Table 4 is provided to assist in identifying likely effects. It lists often-met transduction effects that may give unwanted activity in a material normally regarded as passive or that may alter the active performance. Due attention to such, often unexpected, factors is needed to reach a satisfactory design outcome. Many of the effects are reversible. Generally, only one or two would be expected to be significant in a chosen material but as the system sensitivity is increased more effects become significant.

6 SELECTING A MATERIAL

Metal and ceramic materials tend to be used in instruments to provide design solutions where plastics are unsuited, for instance, where very high temperature and high mechanical stability are needed.

Another factor that dominates the choice of material is the volume of the production run. For small volume runs, tooling costs associated with the often preferred use of plastics are usually very high compared with the cost of computer-controlled machining of metal parts. Ceramics, usually requiring very elaborate manufacturing processes to obtain the forms needed, are often used in stock shapes with minimal machining. Their use is somewhat inhibited by their manufacturing difficulties existing for small volume production.

It might be thought that as materials have been developing since the time of early man, the designer always has at hand materials with the right combination of properties to suit any task. Unfortunately, this is far from the case because the number of well-defined and controlled material properties required is great and each material application will usually involve the need for several particularly stringent properties at the same time. This will be seen in the illustrative case study given in Section 9.

Measuring instruments are made in production volumes ranging from one-off to many thousands but rarely in the numbers seen in consumer electronic goods. Plastic materials, and the production processes required to convert them into products, are generally more suited to high-volume production. For example, injection molding is one of the most widely used processes but the simplest molding dies can cost around $10 000. It is an elegant and versatile process but production runs of less than several thousand inevitably carry a high tool amortization cost per part.

Die-cast metals, once very popular, also can be used but these are falling from favor except where plastics cannot cope because of lack of a key property such as the need to have metal shielding. Even there, however, special electrically resistive, sprayed on and plated coatings can be used to shield the contents of plastic cases.

7 A RATIONALE FOR SELECTING PASSIVE MATERIALS

The first step toward material selection for the passive part of a sensor is to set up the system architecture that will satisfy the operational requirement for the measurement system. This means deciding the functions needed and how they are to be provided as the various needed organic modules.

The module connecting the observed process to the subsequent signal processing and the output actuating device both need an active sensing material, whereas the mountings and following signal processing assembly are made from sufficiently passive materials.

These building blocks, in turn, are made from a combination of parts. This reticulation process eventually identifies the basic components needed, calling for readymade or new parts to be made from appropriate materials.

Crucial in this design process is the simultaneous identification of which salient physical effects, the influence parameters, need to be allowed for.

The passive role of materials in instruments can be found by considering the generalized mechanical restraining purpose of the passive support providing components.

Reuleaux, in 1875, defined a machine as a combination of resistant forces so arranged that by their means the mechanical forces of nature can be compelled to do work accompanied by certain determined motions. As instrument and sensor design is highly interdisciplinary, there are usually many such resistant forces and determined motions to be considered.

In mechanical design, the main system requirement aims to allow movements, in a three-dimensional space, of as many of the six possible degrees of freedom as are needed. These freedoms, for an extended object in space, can occur as three translations and three rotations. A support framework has to hold all six stable so that parts mounted on it are constrained within the required dimensional limits. A slideway allows one degree of translation; a pivot one degree of rotation and so forth.

Two approaches can be used to allow the necessary freedoms in a mechanical system. The first is to make use of materials, selected in shape and type, to form a structure that is effectively nonelastic. This method is known as kinematic mechanical design, *see* **Article 80, Principles of Fine Mechanics – Kinematic and Elastic Designs, Volume 2**; **Article 81, Principles of Fine Mechanics – Systems Considerations, Volume 2**; **Article 82, Kinematical Regime – Members and Linkages, Volume 2**; **Article 83, Kinematical Regime – Fasteners, Bearings, Volume 2**; and **Article 84, Kinematical Regime – Rotary Motion, Volume 2**. These are formed to allow the needed freedoms and constrain the others. An example is a ball bearing where the rollers and tracks are made to exhibit as little deformation as possible while allowing one-dimensional slide action.

The alternative is to purposefully provide certain parts with elasticity so that movements can occur at intended places. This is known as *elastic mechanical design*, *see* **Article 85, Elastic Regime of Design – Design Principles, Volume 2**; **Article 86, Elastic Regime of Design – Spring Systems, Volume 2**; and **Article 87, Elastic Regime of Design – Plates and Bimorphs, Volume 2**. An example is a flexure strip hinge used to allow limited rotation.

In general, kinematic design has the widest applicability but when relatively small movements are needed, its options often cannot compete with the cost-effectiveness and small excursion fidelity of the elastic method.

Once the overall constraint of freedoms strategy is decided, a suitable physical framework is evolved to support the various parts needed to provide the functions. A collection of basic parts set up to provide a particular function is termed a *mechanism*. This, in turn, is formed of members, linkages, joiners, bearings, and parts that provide transfer motion. The classic overview of the detailed design of such parts for fine instruments is given in Trylinski (1971), with a shorter account given in Trylinski (1983).

Parts and roles identified by this top-down design process then allow the choice of material. As can be seen from the abridged list of material properties given in Table 4, the choice of a suitable material can be complex and often impossible to totally satisfy.

8 ALTERING THE ACTIVE AND PASSIVE BEHAVIOR

An often-met difficulty is that the material that must be used is not passive enough to some external influences. Conversely, for sensing use, it may be desirable to increase the activity level. Several methods can be used.

First, unwanted material effects can often be reduced by placing the active part, or the whole unit, in a suitably controlled environment.

Second, some form of compensation can be incorporated. For example, to render a mechanical length formed from a metal bar less variant with temperature, the bar can be mounted end to end with another bar of material that expands in the opposite direction. By use of materials with different TEMPCOs and different lengths, it is possible to set up a system that has closely stable length. In practice, this concept usually needs to apply an opposing parallel action because few materials have negative TEMPCOs.

Another compensation method is to measure the influencing variable and from knowledge of the systematic nature of the activity, change a value in the system at some appropriate place.

The principle of these two compensation methods can be applied to systems in any energy regime. Often, transducers are used to convert energy forms in order to make the comparison and compensatory actuation. An example of this type is the use of a second resistance strain gauge placed in one arm of a Wheatstone bridge interrogation circuit such that its temperature-induced changes cancel out those of the sensing gauge. This second gauge is mounted such that it is not subjected to any strain but is subjected to the same temperature as the sensing gauge. Feedback in instruments is discussed in **Article 113, Force-feedback Sensors, Volume 2**.

There is sometimes a need to increase the effective activity of a material. A commonly used method is to place more than one element in an arrangement that adds or multiplies the effect. An example is the use of a set of series-connected single element thermocouples to form a thermopile. Another example is where the cold junction of a Peltier cooling cell is mounted on a second unit, the temperature difference then being more than one unit alone can provide.

These methods add the effects. A mechanical, hydraulic, pneumatic, or optical lever system can be used to increase a small deflection by simple multiplication.

9 AN EXAMPLE IN MATERIAL SELECTION

As an example of the necessary interplay between the active and passive roles of materials forming a component, consider the materials needed to form a spring and its support framework for a precise and stable weighing balance that is little influenced by its ambient temperature.

The measurement purpose of the spring element (the clearly defined active component) is to produce a reproducible and repeatable elongation proportional to the mass hung on its end.

If the material from which the spring is made changes length with temperature, it will exhibit a displacement error for a fixed load as its ambient temperature changes. This suggests, on first considerations (but see later), that the spring should be made from a material that has a zero thermal coefficient of expansion. Figure 1 illustrates how this coefficient varies in typical instrument materials,

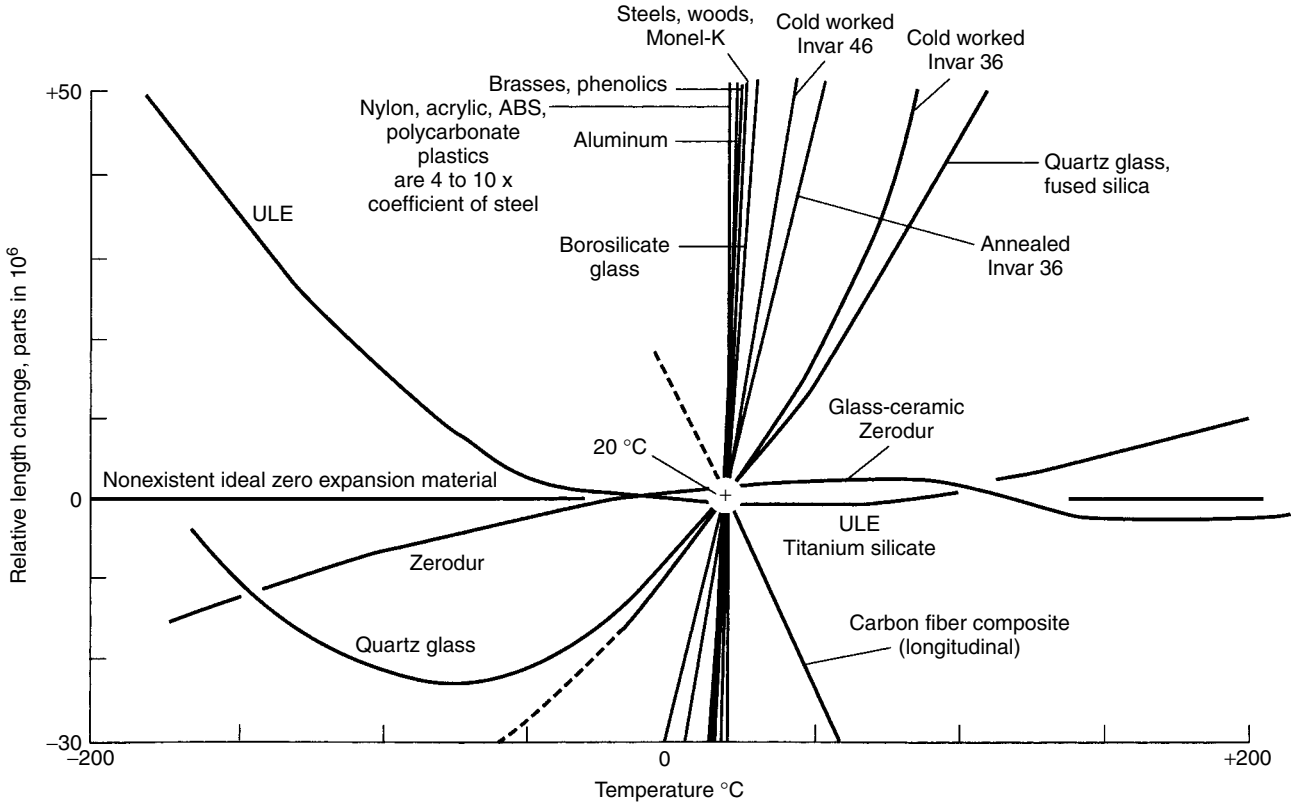


Figure 1. Thermal expansion of selected metals, plastics, and ceramic materials.

showing that some virtually zero coefficient materials are, indeed, available.

However, measurement of spring displacement is made by comparing the length of the spring with a reference support framework that should at first sight be made of a passive material – but again see later.

What is actually required, however, is that the framework and the spring have the same thermal expansion coefficient so that their changes in length track each other as temperature changes. As it is expedient to make a frame from steel, it does appear feasible to use a steel spring, despite its being quite active to temperature effects, because the framework will change length to the same degree as the spring. However, there is another aspect still to be covered.

A vital physical property needed of the measuring spring material is that its modulus of elasticity remains adequately constant with variation of operating temperature – this is expressed as its thermoelastic coefficient or (TEC).

Study of the properties of the steel family shows that its TEC is relatively large. Thus, a steel spring is prone to temperature error in that its weighing sensitivity increases as the temperature rises. For this reason, precision weighing springs made of steel are not suitable unless a temperature correction is applied.

The material needed ideally has a thermal coefficient of expansion similar to structural materials, yet also has a low TEC. No entirely suitable material has yet been developed – a common situation to be faced in instrument design. The nickel–iron–chromium alloys, however, provide some scope. Figure 2 shows how the TEC varies with alloy composition for this family. Note how the alloying ratios are so critical.

Even if the above needs could be met, there is still another important parameter to be considered – the spring should return to the same displacement position when a mass being weighed is removed. This is its mechanical hysteresis. Steels are quite poor in this regard as they obtain some offset with load cycling. Special alloys, such as Ni-Span C™, have been developed that exhibit insignificant mechanical hysteresis, but, unfortunately their TEMPCO and TEC are not at the preferred values.

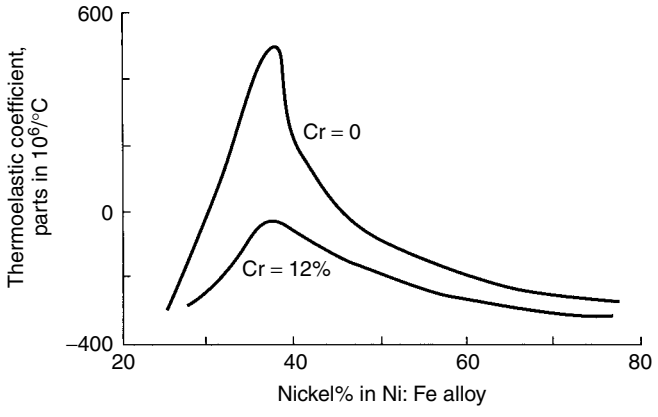


Figure 2. Thermoelastic coefficient variation versus alloying composition for the Ni-Fe-Cr alloys. (Copyright Inco Europe Ltd.)

To make material selection still more complex, the properties mentioned above all may change with history of use, with temperature of operation, and sometimes with level of the present and past history of stress. Additionally, they are usually time-variant, often in nonsystematic ways.

The same considerations apply for the support frame design as that provides the positional reference framework. They are, however, usually of less significance.

In practice, precision springs can be made to only certain performance limits. Use of some form of active or passive compensation can then be used to squeeze a little more sensing performance from the spring. These material limits were gradually met in the evolving design of spring-mass clock movements, finally being replaced by better defined elastic properties of the piezoelectric quartz crystal now commonly used in electronic timekeepers.

The design and operation of highly precise and stable springs is a matter of setting up a system that recognizes the sources of likely error arising from material activity and somehow operates with all of the material restraints in a manner that optimizes the operation. In the case of the spring, this may mean tightly controlling the temperature to keep the spring at its optimum operational condition.

An example of the lengths that are needed is seen in the practical design of a top-performance spring-mass geophysical gravity meter, which is, in effect, an ultrasensitive spring balance that can be carried in the field over rough terrain! To retain the best operational performance, it is necessary to precisely temperature control the spring at all times. The temperature is selected as that where the complex combination of material properties is optimized.

Thus, it is seen that the design of a sensor is not just the procedure of selecting an active material and setting it up within a system of passive supporting and containing materials. To get the best performance sensor design invariably needs complex interaction of material properties with both the active and the passive parts being carefully designed.

The problem of interaction between material properties will be found in any design regime. For example, in optical lens design, obtaining a certain magnification power for a given spectral bandwidth will be a matter of juggling between availability of a few materials having limited ranges of refractive index and varying wavelength dependent transmission losses, *see* **Article 89, Optical**

Materials, Volume 2; and **Article 91, Light Sources and Detectors, Volume 2**.

10 SOURCES OF KNOWLEDGE ON MATERIALS

10.1 Books

Budinski, K.G. and Budinski, M.K. (2001) *Engineering Materials: Properties and Selection*, Prentice Hall, Englewood Cliffs, NJ.

Crane, F.A.A. and Charles, J.A. (1996) *Selection and Use of Engineering Materials*, Butterworths, London.

Crawford, R.J. (1998) *Plastics Engineering*, Butterworth-Heinemann.

Farag, M. (1989) *Selection of Materials and Manufacturing Processes for Engineering Design*, Prentice Hall, Englewood Cliffs, NJ.

Flinn, R.A. and Trojan, P.K. (1986) *Engineering Materials and Their Applications*, Houghton Mifflin, Boston, MA, International Student Edition, (1992), Wiley, New York.

Harper, C.A. (2000) *Modern Plastics Handbook*, McGraw-Hill Professional, New York.

Harwood, F. (1987) *Fine Ceramics: Technology and Applications*, Norton Krieger Publishing Company.

Jones, D.R.H. and Ashby, M. (1996) *Engineering Materials*, Vol. 1, Butterworth-Heinemann, New York.

Kalpakjian, S. and Schmid, S. (2002) *Manufacturing Processes for Engineering Materials*, Prentice Hall, Englewood Cliffs, NJ.

Phule, P.P. and Askeland, D.R. (2002) *The Science and Engineering of Materials*(with CD-ROM), Brooks Cole.

10.2 Journals

Measurement Science and Technology (formerly *Journal of Physics E: Scientific Instruments*) Institute of Physics. UK.

Measurement IMEKO, Elsevier; Lausanne.

Sensors and Actuators, A – Physical, B–Chemical, Elsevier; Lausanne.

REFERENCES

Caren, S. (1988) Product design, basic parameters, basic requirements, in *Handbook of Plastic Materials and Technology*, (ed. I.I. Rubin), Wiley, New York.

Endura (2004) http://www.endura.com/material1.htm.

Ichinose, N. (1987) *Introduction to Fine Ceramics*, Wiley, Chichester.

Kroschwitz, J.I. (1990) *Concise Encyclopedia of Polymer Science and Engineering*, Wiley, Chichester.

Rapra (2004) http://www.rapra.net/technicalservices/plastics rubberselection.htm.

RFE (1983) *Modern Materials in Manufacturing Industry*, Fellowship of Engineering, UK.

Rubin, II. (1988) *Handbook of Plastic Materials and Technology*, Wiley, New York.

Sydenham, P.H. (1979) *Measuring Instruments: Tools of Knowledge and Control*, Peter Peregrinus, Stevenage.

Trylinski, W. (1971) *Fine Mechanism and Precision Instruments; Principles of Design*, Pergamon Press, Oxford.

Trylinski, W. (1983) *Mechanical Regime of Measuring Instruments*, in *Handbook of Measurement Science*, Vol. 2, (ed P.H., Sydenham), Wiley, Chichester.

FURTHER READING

Caren, S. (1990) Product design, basic parameters, basic requirements, in *Handbook of Plastic Materials and Technology*, (ed. I.I., Rubin), Wiley, New York.

Rubin, I.I. (1990) *Handbook of Plastic Materials and Technology*, Wiley, New York.

This article is based in part on a chapter by W.E. Duckworth, D.D. Harris and P.H. Sydenham originally published in Handbook of Measurement Science Vol 3, P.H. Sydenham and R. Thorn, Wiley, 1992.

118: Ultrasonic Sensors

Peter J. Lesniewski
University of South Australia, Mawson Lakes, South Australia, Australia

1 Overview of Ultrasonic Techniques in Sensing 791
2 Theory of Ultrasound 793
3 Key Ultrasound Principles Used in Sensing 793
4 Overview of Sensors Using Ultrasonic Technique 795
5 Use of Ultrasound Principles in Sensing in Solids 796
6 Use of Ultrasound Principles in Sensing in Liquids 797
7 Use of Ultrasound Principles in Sensing in Gases 798
8 Limitations of Ultrasonic Methods of Sensing 799
References 800

1 OVERVIEW OF ULTRASONIC TECHNIQUES IN SENSING

Although there are a few different sensing principles for ultrasonic waves, the focus here is on the leading, linear piezoelectric transducers. Individual measurands are represented by the following categories of the transmitted wave parameters:

- propagation or time delay (representing medium flow, wave velocity, range/distance);
- attenuation or complex frequency response (characterizing materials using spectrometry and transmission tomography);
- doppler shift (representing velocity of individual objects);
- resonant frequency (affected by load e.g. deposited mass or fluid level in single frequency and interdigital (IDT) surface acoustic wave sensors (SAW));
- statistical/averaging (e.g. cross-correlation flow measurement, steady state monitoring in alarm/proximity sensors, absorption estimated from reverberation time or scattering, and Doppler or spectral characterization of scatterers such as fish or blood particles).

Influences by other variables are minimized through compensation, numerical correction, noise/drift reduction, and evaluation of methodology. An overview is given by Hauptmann *et al*. (1998).

Piezoelectric transducers are typically highly resonant. Increasing their bandwidth (BW) reduces sensitivity (increasing losses) and is difficult. The upper frequency range is limited by the molecular granularity of physical medium (low-loss ZnO films use GHz range), but common applications use frequencies from 40 kHz to a few MHz. At high frequency f, the wavelength $\lambda = c/f$ is short and the particle velocity high, allowing small transducer size, microelectronic implementation, and use of wave filters and lens for focusing. For typical indicative values of the wave velocity c for several materials, refer to Table 1 in Section 3.

The acousto-optic effect and laser/microwave excitation of piezoelectrics and the glow-discharge effect addressed further, extend measurement principles.

Conversion between the magnetic field and mechanical vibrations is available through the magnetostrictive effect. Unlike Cr-Ni-Co-Fe-based alloys and nickel, it is the new giant magnetostriction of Terfenol (Tb–Fe alloy), rare

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

Table 1. Comparison of piezoelectric data for various materials.

Piezoelectric material → parameter (SI units) ↓	ZnO p	CdS p	AlN p	SiO_2 p Quartz	SPN f	$LiNbO_3$ f	$BaTiO_3$	PVDF	PZT	PZT4	PZT5	$LiTaO_3$	$LiSO_4$ p	$KNbO_3$
Piezoel. strain coefficient [$\times 10^{-12}$]							−58–78	23	110	−123	−171			
d_{31}								3						
d_{32}/laterally clamped	10.6	10.3	5–8	2–3	127	40	149, 190	−33		289	374, 593		15, 16	
d_{33}								−22						
d_t														
Piezoel. stress constant [$\times 10^{-3}$]							−5.5–5	216	10					
g_{31}								19						
g_{32}/laterally clamped	135	122	66–105	57	29	116	14.1–13	−339		25	25, 20		156, 175	
g_{33}								−207						
g_t														
Coupling coefficient														
k_{31}							0.194–0.21	0.12/1 kHz	0.3					
k_{33}	0.28	0.15	0.2	0.1	0.46	0.49	0.48	0.19/1 kHz						
k_t							0.38	0.29		0.5	0.49, 0.5		0.35, 0.33	
Acoustic impedance ρc [$\times 10^6$]														
All directional							30	3.9	30					
1directional								2.7						
Wave velocity c														
Longitudinal (transverse)	6400, 6330	4500, 4460	10 400, 10 700	5700, 5740	6940	7400	5470	2200 (1500)		4600	4350, 4560		5460	
Density [10^3] (i.e. vs water)														
ρ	5.68	4.82	3.26	2.65	4.46	4, 6.64	5.55, 5.7	1.78	7.5	7.5	7.75, 7.5		2.06	
Relative permittivity														
ε_{33}^T	8.84	9.53	8.6	4.58		39	1700	12	1200	1300	1700, 3400		10.3	
ε_{33}^S							1260			635	830			
Curie temperature/range, °C				575			<0, >115	>145		328	365, 193	450		410
Breakdown voltage field [V/μm]	50	50	200	50		50		80, 100						
Max oper. voltage field [V/μm]								30						
Elasticity const [$\times 10^{-12}$]														
C_{33}^K							9.5			15.5	18.8			
C_{33}^D							7.1			7.9	9.46			
C_{11}^K							9.1			12.3	16.4			
Mechanical tanδ/kHz								0.1						
Q factors: Q_E/Q_M							100/300			250, 500	50/75			
Electrical tanδ [$\times 10^{-3}$]								15/10 Hz 20/10 kHz						

Note: These values should serve as orientation and are not intended for design purposes. Design parameters should be obtained from suppliers of transducers or samples.

earth metals, and some compound semiconductors that offer competitive parameters. Magnetostrictive transducers exhibit significantly larger conversion than traditional piezoceramics, are more robust, heat insensitive, and do not require high voltage, high impedance electric port. Like piezoelectric, the magnetostrictive transducers are reversible (can be used as sensors or actuators). Their ultrasonic applications remain currently in the power area, such as ultrasonic cleaning. Ultrasonic magnetostrictive sensors are likely to emerge soon (http://www.public.iastate.edu/~terfenol/). More detail on ultrasonic methods is given in **Article 119, Ultrasonic Instrumentation Principles, Volume 2** and **Article 120, Ultrasonic Instrumentation Design, Volume 2**.

2 THEORY OF ULTRASOUND

Ultrasonic wave is defined as the vibration of medium particles exceeding the audio frequency (>18 kHz). Field variables (common with the acoustic wave) are particle velocity **v** (vector) and acoustic pressure p in fluids (a scalar) or stress tensor $[T]$, including shear forces in solids or viscous fluids. The ratio $p/|\mathbf{v}|$ or T_{ij}/v_i represents the medium impedance for the relevant vibration mode and the products represent complex power density per unit area (defined for a plane wave).

Ultrasonics in liquids exhibit specific effects, for example, molecular absorption, manifested by a sharp increase of losses in wave propagation at a frequency corresponding to the molecular relaxation time of a fluid (ranging from 30 kHz for carbon dioxide to 10 MHz for hydrogen).

Nonlinear phenomena result from high particle velocities or at high ultrasonic signal frequency or from high pressure level in focused/high energy fields (where either the particle velocity $|\mathbf{v}|$ cannot be regarded as being much smaller than the wave velocity, and the Euler equation governing the wave equation becomes nonlinear, or the flux continuity equation is nonlinear).

They include

- streaming (medium flowing away from transducer's face);
- cavitation (breaking the continuity in a vigorously vibrating liquid and the development of bubbles);
- fog formation (breaking molecules of a liquid medium away from its surface);
- audibility of modulated waves – used in focused beam audio transmission.

3 KEY ULTRASOUND PRINCIPLES USED IN SENSING

The widely used piezoelectric effect offers linear conversion between the stress tensor representing ultrasonic vibration and electric field K received as the output voltage. The mathematical formulation, similar to the one used for electrostatic transducers, represents the class of reciprocal transducers (whereby electromagnetic transducers are antireciprocal). The equivalent circuit uses the dynamic (impedance) type port, (the antireciprocal class uses the mobility type) or else a cumbersome gyrator is required.

The transduction process, best described by single mode relations, can be extended to the tensor form. Detailed, valid-today description of both was given by Mason in the 1940s. For a review of equivalent circuits, refer to Kino (1987). Formulation for magnetostrictive transduction is the same, except for the electric field being replaced by the magnetic field strength H.

Figure 1 presents a single mode transduction (specific orientation of electric and mechanical field) in a thin piezoelectric slice governed by (1a, b).

The voltage E, related to the electric field strength ($\mathrm{K} = E/x$), produces strain ($S = \Delta x/x$) if no external force is applied:

$$\Delta x = Ed \text{ or } \frac{\Delta x}{x} = \mathrm{K}d \tag{1a}$$

The stress in the material ($\mathrm{T} = F/A$) corresponds to free charge Q appearing between the electrodes, manifested as the polarization ($D = Q/A$), if no additional external

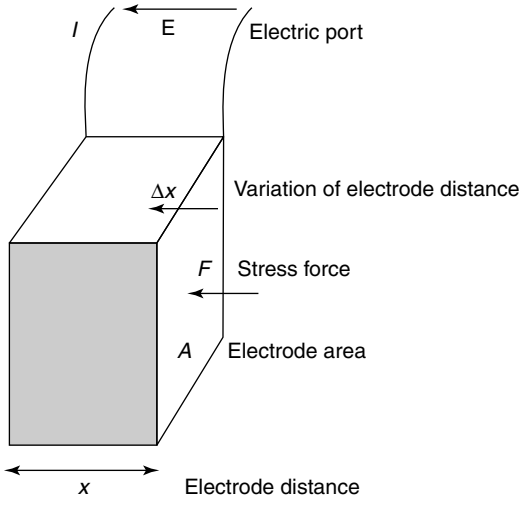


Figure 1. Operation of a piezoelectric transducer: stress F/A causes strain $\Delta x/A$, which produces voltage E (and vice versa).

polarization occurs:

$$Q = Fd \text{ or } \frac{Q}{A} = d\left(\frac{F}{A}\right) \tag{1b}$$

where d denotes the piezoelectric strain coefficient.

Using incremental notation, strain is expressed as the ratio of the material's infinitesimal displacement over the increment of the position along the chosen axis, that is, for a one-dimensional case: $S = \partial u_x/\partial x$. Because the displacement itself is a vector, its variation has to be related to a three-dimensional position – thus, strain is a size 3×3 tensor [S] relating two vectors:

$$\mathrm{d}\mathbf{u} = [S]\,\mathrm{d}\mathbf{w} \text{ where } \mathrm{d}\mathbf{u} = [\mathrm{d}u_x,\ \mathrm{d}u_y,\ \mathrm{d}u_z] \text{ and}$$
$$\mathrm{d}\mathbf{w} = [\mathrm{d}x,\ \mathrm{d}y,\ \mathrm{d}z] \tag{2}$$

Similarly, stress is the variation of the force vector over the infinitesimal spatial grid. For one-dimensional case: $T = \partial F_x/A\partial x$, but for three dimensions, including shear forces, a symmetric 3×3 stress tensor $[T]$ with six independent variables (linearity condition), is used:

$$\mathrm{d}\frac{\mathbf{F}}{A} = [T]\,\mathrm{d}\mathbf{w} \tag{3}$$

The relation between stress and the resulting strain is, in fluids, determined by the elasticity or compliance, but in solids, the elasticity constant $[c]$ is a 6×6 matrix defined by the general Hooke's law: $[c] = \partial[T]/\partial[S]$.

For isotropic solids, $[c]$ contains only three different components, all related to two Lamé constants (μ, λ).

These relations and the equations of motion lead to the wave equation for isotropic materials (4) that can be split into two simpler potential wave equations (5a, b), one for longitudinal (index l) and one for transversal (index t) wave, each propagating along z with a different velocity $c_l \geq c_t\sqrt{2}$.

$$\frac{\partial \mathbf{u}^2}{\partial t^2} = \frac{\lambda + \mu}{\rho}\nabla(\nabla \cdot \mathbf{u}) + \frac{\mu}{\rho}\nabla^2\mathbf{u} \tag{4}$$

$$\frac{1}{c_t^2}\frac{\partial^2 v_t}{\partial t^2} = \frac{\partial^2 v_t}{\partial z^2} \tag{5a}$$

$$\frac{1}{c_l^2}\frac{\partial^2 v_l}{\partial t^2} = \frac{\partial^2 v_l}{\partial z^2} \tag{5b}$$

These two waves, transversal and longitudinal, in piezoelectric materials are interrelated (disturbing one affects the other). To avoid dispersion and linear distortion, only one mode of propagation is facilitated in ultrasonic sensors.

The combined shear-longitudinal motion, in the presence of a boundary, forms waves propagating even slower than c_t – such as Lamb wave (in waveguides) and Rayleigh wave (on a surface) utilized in SAW transducers.

Properties of piezoelectric materials are described by a large number of parameters with significant redundancy (16 basic parameters in 4 pairs of constitutive equations). In the general tensor form, two essential pairs are

$$\left\{\begin{array}{l}[T] = [c^K][S] - [e]'[K] \\ [D] = [e][S] + [\varepsilon^S][K]\end{array}\right\} \tag{6}$$

$$\left\{\begin{array}{l}[S] = [s^K][T] + [d][K] \\ [D] = [d]'[T] + [\varepsilon^T][K]\end{array}\right\} \tag{7}$$

where,

$[c]$ matrix of elastic constants,
$[e]$ matrix of piezoelectric stress coefficients,
$[\varepsilon]$ matrix of dielectric constants,
$[d]$ matrix of piezoelectric strain coefficients,
$[s]$ matrix of compliance constants,

while symbol [..]' denotes the transpose of matrix [..].

Superscript within a matrix denotes that the parameter is measured with invariant voltage, that is, $K = 0$ meaning shorted electric terminals. Similarly, T denotes nulling the variation of internal stress, $T = 0$, whilst superscript S denotes zero displacement condition $S = 0$, that is, rigidly clamped transducer.

Other parameters in use include: piezoelectric stress constant $g = e^{-1}$, transmitting constant $h = e/\varepsilon^S$, transformation ratio $\alpha = d/C_m$ (where mechanical compliance $C_m = \mathrm{d}x/\mathrm{d}F = s.x/A$), and coupling coefficient $k = d(\varepsilon A)^{-1/2}$ that are important in assessing piezoelectric materials since it determines the fraction of the input transmitted between the mechanical and the electric port in an unloaded transducer. Values of k closer to unity are preferred except for such cases where the Q-factor is exceptionally high (Quartz). Note that $k_{33} \neq k_t$, since k_t is defined for transversely clamped operation.

Many transducers have one dominant mode of operation (either longitudinal/thickness '33' or shear/bender '31', '32' mode) but full optimization requires following strict relations in mathematical modeling for secondary effects. Table 1 lists the typical parameters of several piezoelectrics indicating discrepancies between independent measurements.

The piezoelectric stiffening phenomenon where piezoceramics become 'harder' after being poled results from (6, 7). To prove this, calculate $[c]$ or $[s]$ with and without piezoelectric effect, that is, nulling d or e and assuming $D = 0$ (open electric circuit). Consequently, the longitudinal wave velocity in poled materials is increased.

Ultrasonic fundamentals are covered well by Ristic (1983) and Kino (1987).

Improved modeling should avoid using inaccurate material constants – when conditions imposed on a sample's surface cannot be maintained throughout it. Piezoelectric transducer equivalent circuits based on Mason's description have been related to the general linear transducer theory for the general case of distributed parameter devices by Lesniewski and Thorn (1995) leading to a more universal transmission-line circuit.

SPICE simulations by Püttmer *et al.* (1997) using the Mason-based Redwood model, for impedance interferometer 1-MHz piezoceramic transducers showed that incorporating losses, helps reduce simulation error to 10% and in some cases below 0.1%.

The polar pattern of linear transducers in fluids (both sensors and actuators) is determined by the radiation theory of Kirchoff but in practical cases the working integral formula of Rayleigh (8) is used,

$$\Phi_{(x,y,z)} = -\frac{1}{2\prod}\int_{S} v_{n(x,y,0)}\frac{e^{-jkR}}{R}\,dS \qquad (8)$$

where for each frequency component ω (with $k = \omega/c$ being the wave number) of the velocity distribution $v_{n(x,y,0)}$ normal to the transducer's surface S, the resulting acoustic potential Φ is observed in the far field at a distance R from the source.

This formula is valid for flat baffled surfaces but it is often used for slightly curved ones in small enclosures. Resulting inaccuracies are of limited concern because nonuniformity of vibrations in transducers cannot be realistically determined. A better modeling, limited to analysis only, is offered by the finite element method.

4 OVERVIEW OF SENSORS USING ULTRASONIC TECHNIQUE

An ultrasonic wave can be generated in solid dielectrics containing electric dipoles through application of varying electric field (piezoelectricity in piezoceramics and piezocomposites). While crystals and ferroelectrics have natural permanent polarization created by their dipole or domain lattice, dipoles in electrostrictive materials are aligned permanently by applying high electric field (several kV), for example, by corona or contact poling, preferably cooling them down from the Curie point. Piezoceramics manufactured from precalcinated fine powders mixed with binders, dry pressed, baked, and fired, are finally poled and have electrodes deposited. Thin films (0.05–5 μm of $Ba_2Si_2TiO_8$, TeO_2, Si_3N_4, PZT (lead zirconate titanate), or ZnO on silicon) for hair or SAW microsensors are obtained by laser or microwave vacuum sputtering, but also heat only (200–500 °C) bonding of piezosensors on silicon is an inexpensive option since the 1980s.

The fundamental series resonant frequency (Figure 2) is well defined by the electromechanical parameters of the sensor and its load (housing and the medium). The parallel resonance, which involves electric capacitance of the material and connections, is less stable. Minute variations of series resonant frequency relate well to the load impedance (representing fluid density, phase transition, temperature, or deposits on the vibrating surface) and are measurable with reasonable accuracy. A self-resonating PZT sensor with frequency resolution of 0.1 Hz in the MHz range allows reliable measurement of 10–1000 Hz variation caused by mechanical loading such as deposition/growth of biochemical material (refer to Guibault and Luong, 1994).

Narrowband sensors are also suitable in single frequency systems where phase delay of a fixed frequency continuous (CW) or semi continuous signal represents the measurand (e.g. in noncontact surface profiling).

The Q-factor of a sensor (damping), which may be determined at its electric port, is useful in characterization of molecular absorption of solutions. In such measurements, many resonant frequencies (overtones) specific to the sensor's geometry are available, allowing frequency profiling

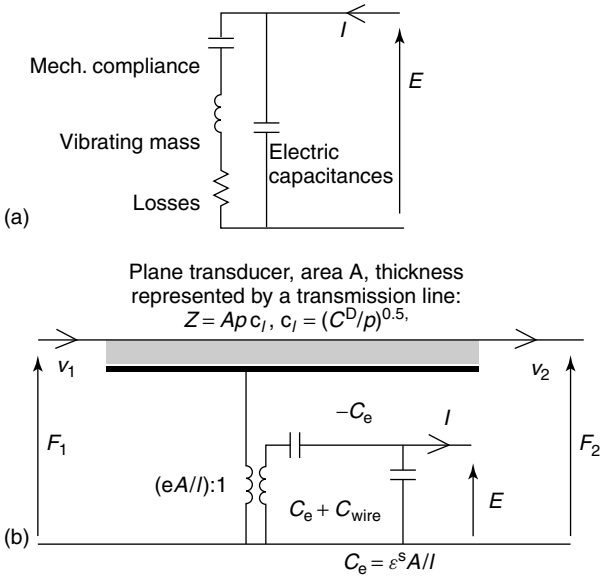


Figure 2. Equivalent circuits of single mode solid-state ultrasonic transducers, (a) simplified lumped network or electric impedance and (b) the exact transmission-line equivalent circuit developed by Mason in 1948 and verified by Redwood in 1956 and 1961.

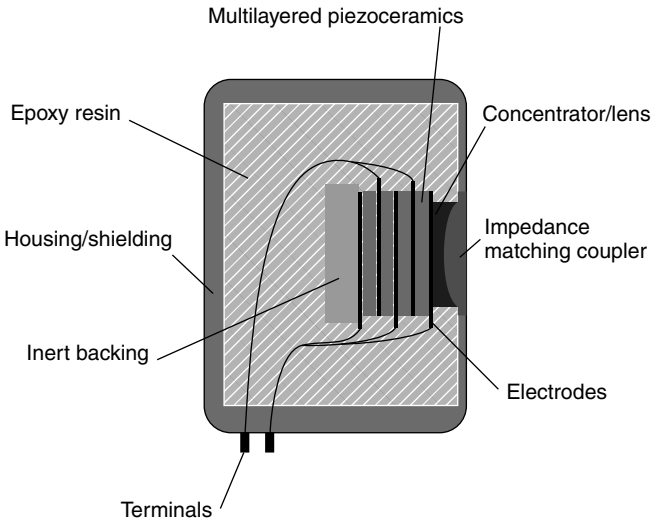


Figure 3. Cross section of a simple bidirectional transducer for NDT or medical imaging, utilizing a concentrator lens for beam focusing and offering an increased efficiency/sensitivity through using multilayer piezoelectric structure with damped housing.

of absorption (relaxation spectroscopy). For a detailed outline, refer to Edmonds (1981, Ch. 2, 3, 4, 8).

Wider BW (see Figure 3) sensors are needed in pulse-echo techniques, where a short transient burst represents either time delay (in ranging systems) and/or signal magnitude (in defectoscopy). The BW representing the difference between the maximum and minimum 3-dB angular cutoff frequency ($\Delta\omega = \omega_{max} - \omega_{min}$, typically ~50% of the center frequency) determines the duration ($t_{min} \cong 1/\Delta\omega$) of the sensor's electrical response to the shortest received transient indicating time delay error (if correlation techniques in signal detection are not used).

The group of broadband (BB) sensors use transient or wideband signals (chirps, modulated, or coded signals) in imaging systems, tomography, holography, and sonar; and requires flat frequency response with minimum distortion. It is possible to compensate/equalize for time-invariant linear distortion using optimized Wiener inverse filter if spectral components of interest have sufficient energy (above the noise level). For an inexpensive way of broadening BW of resonant transducers and references to the practical use of the Wiener filter, refer to Lesniewski and Thorn (2001).

Transducers for imaging applications (sonar, medical) appear as planar or circular arrays with electronic beamforming, but in single low power devices, for example, in nondestructive testing (NDT), lenses/concentrators (of aluminum, epoxy resin, magnesium) offer fixed focus. For details of a 4.5-MHz PZT array for medical imaging, refer to Turnbull and Foster (1992). Bronze, brass, or tungsten are still used for backing piezoceramics but reducing reverberation and cross talk in arrays is better served by lossy resin composites, possibly loaded with tungsten. Composite ceramics offer lower acoustic impedance and losses with good piezoelectric parameters. Refer to Patel and Nicholson (1991) and Richter, Reibold and Molkenstruck (1991) for performance comparison of conventional and composite transducers and to Silk (1984) for backing methods.

The most versatile surface acoustic wave, SAW sensors, developed in the 1970s utilize bulk acoustic wave (BAW) or thickness shear mode (TSM). They are usually microdevices with thin film piezoelectrics, deposited on silicon substrate, operating at frequencies from a few MHz to a few hundred MHz, either band pass or resonating at frequencies relating to IDT structure. Being extremely sensitive to any acoustic loading on the surface, responding to deposited mass, ionic and molar concentrations, temperature, pressure, and humidity, they are suitable as probes for conventional structural tests (http://www.sofratest.com/site/Sales/Catalog/Systems/SAW_Probes/kcu023a.pdf) as well as probes for microsensing (http://www.techtransfer.anl.gov/techtour/sensors.html).

Since they can be excited remotely by microwaves, mixed technology devices are possible. Because of their sensitivity, they require careful compensation against irrelevant variables. For a versatile overview, refer to Drafts (2001).

Piezoelectrics such as SiC and piezopolymers can also be excited by laser beams.

A quality ultrasonic sensor is also a small size capacitance microphone (such as the Bruel&Kjaer 4138 type with BW exceeding 100 kHz) or inexpensive 1/8 inch electret capsules usually offering BW of just over 50 kHz.

5 USE OF ULTRASOUND PRINCIPLES IN SENSING IN SOLIDS

Ultrasonic wall thickness/flaw sizing in metals uses the pulse-echo mode at a few MHz with single transmit–receive piezoceramic transducers (backed by dense lossy resin) and pads of oil, wax, or resin; or else fully immersed in liquid to facilitate coupling. For timber, concrete, and tires, the frequency is reduced by a decade due to higher losses. The actuator is separate from the sensor, although both may remain in one housing. Flaws are assessed according to the delay time of observed reflected pulse and its relative amplitude (refer to Krautkrämer and Krautkrämer, 1990) (http://www.krautkramer.com/).

For monitoring a machining tool (wear, cooling), Nayfeh, Eyada and Duke (1995) employed the pulse-echo methodology and precise measurements of a reflected 10-MHz burst using an ultrasonic transducer mounted within the tool holder. Similarly, Morton, Degertekin and Khuri-Yakub (1999) monitored baked photoresist by relating its temperature and thickness to the reflection coefficient in pulse reflection measurement using a GHz BW, ZnO-on-saphire transducer.

Single transducers for medical imaging tend to employ PZT in epoxy housing, but a copolymer PVDF-TrFE with lossy polymeric backing offers lower acoustic impedance, reducing the mismatch while maintaining a reasonably flat 120 MHz BW.

Research in constructing improved piezoelectric transducers for medical imaging has led to specific profiling of the surface, subdividing the front electrode, and/or graded poling (for state-of-the-art microfabricated transducers, refer to http://www.sensant.com/pro_advDiag.html). This has coincided with new solutions of the acoustic harmonic wave equation from 1987 (following a similar, earlier development in the electromagnetic wave theory), representing a new class of limited diffraction and localized waves (nondiffracting waves, X-waves, Bessel Beams, transient Bessel Bullets), which form very narrow beams, although they decay faster than conventional wave types.

An experimental pulse-echo 2-D imaging system with small limited diffraction beam transducer arrays (48 or 64 elements) has been developed by Lu (1997, 1998) (http://bioe.eng.utoledo.edu/Laboratories/UltrasoundLab/jian_yu_lu_lab.html).

It offered improved images of tissue phantoms up to a depth of 200 mm utilizing a new Fourier transform–based algorithm allowing high frame rate processing. Theoretical analysis of the new waves and an outline of other developments is offered by Salo, Fagerholm and Friberg (2000), a numerical analysis of Bessel Bullets by Stepanishen (1999), and an investigation of finite planar transducers by Chatzipetros *et al*. (1998).

6 USE OF ULTRASOUND PRINCIPLES IN SENSING IN LIQUIDS

Steady coupling and reduced acoustic impedance mismatch between the medium and the immersed transducer offers good interface; thus, solid-state ultrasonic transducers in liquids are readily used. Resonances are damped and the sensor's BW is larger than the BW in gases. The mismatch may be efficiently reduced by using novel composite materials (elastomer based piezorubbers or copolymers) suited for sonar and medical imaging (refer e.g. to Schlaberg and Duffy, 1994).

Sonar transducers are either small arrays of piezoceramic elements or single transducers with focusing concave face and/or lenses embedded in epoxy resin. They are used for bottom profiling or fish location ({echo-sounding}) using pulse-echo principle and narrowband signals (http://www.simrad.com/). A multibeam commercial unit Simrad EM3000 offers depth resolution of 1 cm within 200-m range at 300-kHz operating frequency (for minimum attenuation in seawater).

Other applications include velocimetry, where a measurand that uniquely influences ultrasonic wave velocity in a given liquid is found through time of flight (ToF) of an ultrasonic transient traveling over known distance. Measurement of fluid composition, its temperature, density, or concentration (water–air, water–alcohol, oils/fats, alcohols, molecules in a solution) using the above approaches has been reviewed by Povey (1997). The variation of wave velocity tends to be as high as 10% (e.g. for temperature variation of 30 °C), but effects on wave attenuation by a liquid medium, although under 1% and strongly influenced by frequency, are also measured (refer e.g. to examples by Harker and Temple, 1991). The use of ultrasonic sensors in velocity measurements is straightforward (using edge triggered timer with transient signal amplitude detection) as long as the sensor's low linear distortion allows it to maintain undistorted transient and good definition of the arrival time. If distortion is significant, then a reference transient propagating in a known reference liquid can be used and two distorted transients correlated to establish the time delay.

Imperfect sensors can still be used in velocimetry based on interferometry utilizing the superposition of continuous, sinusoidal signal with its delayed version (through a known length of transmission in an unknown fluid). This interference is manifested by resonances affecting the resulting frequency response of such systems where the wave velocity c can be found from the location of resonant frequencies: $f_n = n(c/l)$, or notch frequencies (antiresonances): $f_n = (2n + 1)(c/l)$, $n = 1, 2, 3, 4, 5$ and so on.

With a single transmit–receive transducer (where T and S1 would become the same transducer in Figure 4), the reflected wave travels twice the length of the fluid-filled tube terminated by a reflecting/rigid end. Examples of acoustic interferometers are given by Trusler (1991) and review of measurement approaches is given by McClements (1995). Errors in this method are caused by imprecise location of the resonance frequencies, which results from multipath transmission in the fluid tested. The transducer cannot radiate a perfect plane wave with

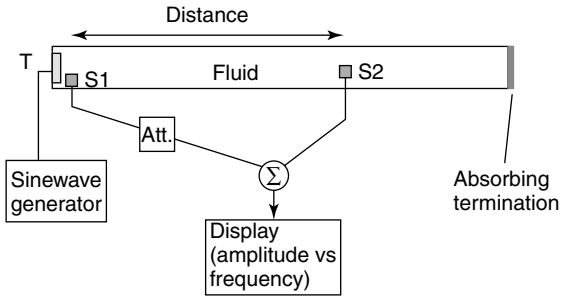


Figure 4. Principle of an ultrasonic interferometer with two independent sensing points (*Note*: (i) The sensor S1 may be omitted and the signal taken directly from the transmitter T. (ii) Only one transmit–receive transducer is used and the tube termination is fully reflecting in a single transducer system).

a constant phase over its body – therefore, the wave reaching the sensor may be interpreted as a sum of beams, some reflected at the tube walls and having different phases, and yet reaching the sensor with different phases.

Acousto-optic effects, such as diffraction of light, variation of refractive index or scattering, and frequency shifting by nonuniform medium density (and by surface waves in solids) created by a perpendicular ultrasonic standing wave are used for nonacoustic sensing of high-frequency ultrasonic waves. The deformed light beams are sensed by photodetectors or interferometers (the most popular commercial sensors being the Bragg Cells) (http://www.goochandhousego.com/page/home.html). Review of relevant sensing techniques is offered by Scruby and Drain (1990). Chiang, Chan and Gardner (1990) have outlined an improved sensor that can be used for detection of ultrasonic wave up to the MHz range (sufficient for medical applications) based on birefringence in optical fiber (ultrasonic field varying the refractive index and causing dispersion of differently polarized light beams).

7 USE OF ULTRASOUND PRINCIPLES IN SENSING IN GASES

Because of the low wave velocity in gases, ultrasonic sensors find application in flow velocity (v) measurement, by sensing the time delay, $t_d = \text{distance}/(c \pm v)$, between two transmission points along the gas flow. A two-path system with simultaneous transmission along and against the flow allows elimination of the wave velocity c by offering two such equations (Figure 5). A commercial five-path gas flowmeter has an error less than 0.5% within the range of 0.4 to 20 ms^{-1} – Fluenta, FMU700 (http://www.fmckongsbergsubsea.com/).

The delay time used for ranging/positioning (in robotics) and level control allows sensing distance to reflecting objects (locally flat, of size larger than signal wavelength with an abrupt change of acoustic impedance). Often 40-kHz center frequency, narrowband piezoceramic sensors with aluminum horn/scatterer are used to reduce impedance mismatch else polyvinylidene fluoride (PVDF) sensors are used – for an overview refer to Manthey, Kroemer and Magori (1992). The impulse response is of irregular ringing form having autocorrelation function with many lobes. Therefore, the limiting error reaches several microns and yet the wave velocity has to be determined. In simpler systems (Vargas *et al*., 1997), the sensed amplitude envelope of the signal leads to a pulse duration of some 100 μs with relative error reaching 3%. Another robotic system developed by Munro *et al*. (1990), with conventional 50-kHz transducers offered 10-m sensing range.

For the state-of-the-art selection, refer to: http://www.hexamite.com/.

Reduction of errors limiting ultrasonic ranging and/or imaging can be obtained through integrating results from a system producing uncorrelated errors.

In a robotic collision avoidance system, Chou and Wykes (1999) combined four-element 100 kHz cf^{-1} ultrasonic linear array (giving poor angular, but good range resolution, reaching 1.7 mm at a distance of some 140 mm) with a charge-coupled device (CCD) optical imager having different characteristics/limitations.

Abreu *et al*. (1999) combined electromagnetic an ultrasonic positioning 3-D system with one transmitter (pulse source: an electric spark) and three remote receivers (conventional resonant 40-kHz transducers) obtaining a resolution of 1 mm within a range of 5 m.

Combining a conventional 40-kHz ultrasonic CW Doppler system (utilizing 5-Hz frequency resolution) with a microwave motion detector unit has improved the elimination of false alarms in an intruder alarm system described by Ruser and Magori (1998).

An efficient transducer, developed for operation in air by Kielczynski, Pajewski and Szajewski (1990), comprised one or more axially vibrating PZT rings embedded in low impedance resin, transmitting vibrations to the air. The resulting device had a narrow beamwidth ($<4^\circ$) and bandwidth of 100–200 kHz (larger than the rings alone), although containing several resonances.

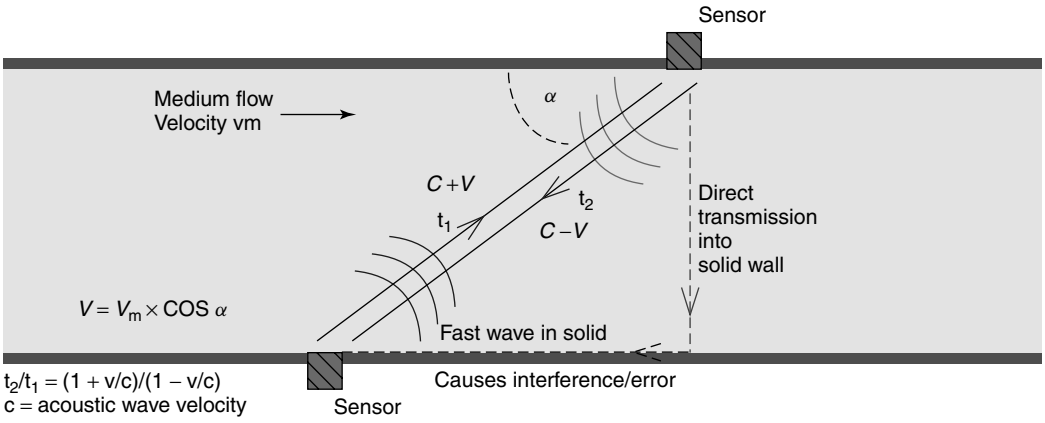


Figure 5. The principle of operation of a two-path ultrasonic flow velocity measurement (*Note*: the dotted path identifies the cross-talk transmission, interfering with the measured transient and causing errors).

An option that may prove successful is the use of sensors where glow-discharge is modulated by acoustic pressure – refer to Kocis and Figura (1996).

8 LIMITATIONS OF ULTRASONIC METHODS OF SENSING

General limitations are gradually lifted with the steady progress in technology (frequency range, automation) and material processing (material purity, profiling accuracy) but the following fundamental limitations remain.

8.1 Impedance mismatch

Affects any method involving transmission in gases, where the huge impedance ratio (7000:1 or worse) between a solid-state transducer and the medium introduces signal loss. In actuators, the peak signal level may be increased to some 100 dB at 1 m by reaching the maximum operating voltage of the piezoelectric material (electric breakdown or linearity limit), but in sensors impedance transformation is necessary. Wave filters offer any transformation but over a narrow frequency range, while horns increase the level/sensitivity by some 30 dB at the expense of reduced frequency range and a likely dispersive transmission (for horn theory refer to Ensminger, 1988 and for modeling, to Lesniewski, 1995). In the case of small size sensors with reduced source capacitance (as low as 10 pF), even a short connecting cable affects the response, thus, sensor-amplifier hybrids increase sensitivity by more than 10 dB (Zheng *et al.*, 1997). Contact between solid-state transducer and tissue in medical imaging is also a source of signal attenuation that cannot be avoided due to impedance discontinuity/instability.

8.2 Delay time

The usual assumption in signal transmission is that signal travels through a homogeneous and time-invariant medium. In fact, a flowing medium is unsteady and nonuniform, with bubbles or multiple phase, yet there are also reflected signals from instrument housing – leading to significant uncontrollable distortion of the received signal. This limits the accuracy of the delay measurement and therefore ultrasonic transducer needs only the BW sufficient to generate a test signal whose autocorrelation has a width comparable with the above error. Narrowing beamwidth may be justified in tomographic reconstruction where many sensors are used.

In signal reflection, narrow beam and wide, BW of transducers is most important, while limitations come from interfering reflections and imprecise determination of the wave velocity (a differential method with reference path is useful).

8.3 Polar and spatial response

A narrow response is useful in isolating interfering signals but it is costly and it negatively affects frequency response and sensitivity. If lenses or special techniques (digitally equalized/filtered arrays) are not used, then the main lobe is frequency dependent and sidelobes are created. Finally,

severely distorted interfering signals may be more difficult to handle (e.g. spikes turning into longer ringing).

REFERENCES

Abreu, M.J.M., Ceres, R., Calderon, L., Jimenez, M.A. and Gonzalez-de-Santos, P. (1999) Measuring 3D-Position of a Walking Vehicle Using Ultrasonic and Electromagnetic Waves. *Sensors and Actuators*, **75**, 131–138.

Chatzipetros, A.A., Shaarawi, A.M., Besieris, I.M. and Abdel-Rahman, M.A. (1998) Aperture Synthesis of Time-Limited X Waves and Analysis of their Propagation Characteristics. *Journal of the Acoustical Society of America*, **103**(5), 2287–2295.

Chiang, K.S., Chan, H.L. and Gardner, J.L. (1990) Detection of High-Frequency Ultrasound with a Polarization-Maintaining Fiber. *Journal of Lightwave Technology*, **8**(8), 1221–1227.

Chou, T.N. and Wykes, C. (1999) An Integrated Ultrasonic System for Detection, Recognition and Measurement. *Measurement*, **26**, 179–190.

Drafts, B. (2001) Acoustic Wave Technology Sensors. *IEEE Transactions on Microwave Theory and Techniques*, **49**(4), 795–802.

Edmonds, P.D. (ed.) (1981) *Ultrasonics*, Methods of Experimental Physics, Vol. 19, Academic Press, Orlando, FL.

Ensminger, D. (1988), *Ultrasonics*, Marcel Dekker, New York.

Guibault, G.G. and Luong, J.H.T. (1994) Piezoelectric Immunosensors and their Applications in Food Analysis, Chapter 7, in *Food Biosensor Analysis* (eds G. Wagner and G.G. Guibault), Marcel Dekker, New York (pp. 151–172).

Harker, A.H. and Temple J.A.G. (1991), Propagation and Attenuation of Ultrasound in Suspensions, in *Developments in Acoustics and Ultrasonics* (eds M.J.W. Povey and D.J. McClements), IOP Publishing, Bristol, UK; 1992 (pp. 19–31); Proceedings of the meeting organised by the IOP Physical Acoustics Group, Leeds, UK, 24–25 September 1991.

Hauptmann, P., Lucklum R., Püttmer, A. and Henning, B. (1998) Ultrasonic Sensors for Process Monitoring and Chemical Analysis: State-of-the Art and Trends. *Sensors and Actuators A*, **67**, 32–48.

Kino, G.S. (1987) *Acoustic Waves: Devices, Imaging and Analog Signal Processing*, Prentice Hall, Englewood Cliffs, NJ.

Kielczynski, P., Pajewski, W. and Szajewski, M. (1990) Ring Piezoelectric Transducers Radiating Ultrasonic Energy into the Air. *IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control*, **37**(1), 38–43.

Kocis, S. and Figura, Z. (1996) *Ultrasonic Measurements and Technologies*, Chapman & Hall, London.

Krautkrämer, J. and Krautkrämer, H. (1990), *Ultrasonic Testing of Materials*, Springer-Verlag, Berlin; New York.

Lesniewski, P. (1995) Discrete Component Equivalent Circuit for Webster's Horns. *Applied Acoustics*, **44**(2), 117–123.

Lesniewski, P. and Thorn, R. (1995) A Z-Matrix Model of the Thickness-Mode Piezoelectric Transducer. *Acoustics Letters*, **19**(3), 43–46.

Lesniewski, P.J. and Thorn, R. (2001) Generating Broadband Test Signals with Resonant Piezoelectric Transducers. *Measurement Science and Technology*, **12**, 723–735.

Lu, J. (1997) 2D and 3D High Frame Rate Imaging with Limited Diffraction Beams. *IEEE Trans. on Ultrason. Ferroel. and Freq. Control*, **44**(4), 839–856.

Lu, J., (1998) Experimental Study of High Frame Rate Imaging with Limited Diffraction Beams. *IEEE Trans. on Ultrason. Ferroel. and Freq. Control*, **45**(1), 84–97.

McClements, D.J. (1995) Chapter 4, in *Ultrasonic Characterization of Foods, in: Characterization of Food* (eds A.G. Gaonkar), Elsevier, Amsterdam; New York.

Manthey, W., Kroemer, N. and Magori, V. (1992) Ultrasonic Transducers and Transducer Arrays for Applications in Air. *Measurement Science and Technology*, **3**, 249–261.

Morton, S.L., Degertekin, F.L. and Khuri-Yakub, B.T. (1999) Ultrasonic for Photoresist Process Monitoring. *IEEE Transactions on Semiconductor Manufacturing*, **12**(3), 332–339.

Munro, W.S.H., Pomeroy, M., Rafiq, M., Williams, R., Wybrow, M. and Wykes, C. (1990) Ultrasonic Vehicle Guidance Transducer. *Ultrasonics*, **28**, 250–354.

Nayfeh, T.H., Eyada, O.K. and Duke, J.C. (1995) An Integrated Ultrasonic Sensor for Monitoring Gradual Wear On-Line During Turning Operations. *International Journal of Tools Manufacturing*, **35**(10), 1385–1395.

Patel, N.D. and Nicholson, P.S. (1991) Influence of Effective Diameter and Piezoelectric and Dielectric Properties on the Performance of Ultrasonic Transducers. *Ultrasonics*, **29**, 389–394.

Povey, M.J.W. (1997) *Ultrasonic Techniques for Fluids Characterization*, Academic Press.

Püttmer, A., Hauptmann, P., Krause, O. and Henning, B. (1997) SPICE Model for Lossy Piezoceramic Transducers. *IEEE Trans. on Ultrason. Ferroel. and Freq. Control*, **44**(1), 60–66.

Richter, K.P., Reibold, R. and Molkenstruck, W. (1991) Sound Field Characteristics of Ultrasonic Composite Pulse Transducers. *Ultrasonics*, **29**, 76–80.

Ristic, V.M. (1983) *Principles of Acoustic Devices*, John Wiley & Sons.

Ruser, H. and Magori, V. (1998) Highly Sensitive Detection with Combined Microwave-Ultrasonic Sensor. *Sensors and Actuators A*, **67**, 125–132.

Salo, J., Fagerholm, J. and Friberg, A.T. (2000) Unified Description of Nondiffracting X and Y Waves. *Physical Review E*, **62**(3), 4261–4275.

Schlaberg, H.I. and Duffy, J.S. (1994) Piezopolymer Composite Arrays for Ultrasonic Medical Applications. *Sensors and Actuators A*, **44**, 111–117.

Scruby, C.B. and Drain, L.E. (1990) *Laser Ultrasonics – Techniques and Applications*, Adam Hilger.

Silk, M.G. (1984) *Ultrasonic Transducers for Nondestructive Testing*, Adam Hilger.

Stepanishen, P.R. (1999) Acoustics Bessel Bullets. *Journal of Sound and Vibration*, **222**(1), 115–143.

Trusler, J.P.M. (1991) *Physical Acoustics and Metrology of Fluids*, Adam Hilger.

Turnbull, D.H. and Foster, F.S. (1992) Fabrication and Characterization of Transducer Elements in Two-Dimensional Arrays for Medical Ultrasound Imaging. *IEEE Trans. on Ultrason. Ferroel. and Freq. Control*, **39**(4), 464–474.

Vargas, E., Ceres, R., Martin, J.M. and Calderon L. (1997) Ultrasonic Sensor for Liquid-Level Inspection in Bottles. *Sensors and Actuators A*, **61**, 256–259.

Zheng, X.R., Lai, P.T., Liu, B.Y. and Cheng, Y.C. (1997) An integrated PVDF ultrasonic sensor with improved sensitivity using polyimide. *Sensors and Actuators A*, **63**, 147–152.

119: Ultrasonic Instrumentation Principles

Lawrence C. Lynnworth

Lynnworth Technical Services, Waltham, MA, USA; Formerly Chief Technologist at GE Panametrics

1 Active and Passive Transducers 802
2 Effects of Measurand on Ultrasonic Properties 802
3 Transducer Effects Used in Ultrasonic Instrumentation 803
4 Guiding Ultrasonic Waves to Sense a Measurand Remotely; Buffering an Ultrasonic Transducer 805
5 Container Walls as Sensors or Windows for Noninvasive Interrogation 806
6 Limitations on Precision and Accuracy 807
References 807
Further Reading 808

1 ACTIVE AND PASSIVE TRANSDUCERS

Transducers for launching or detecting ultrasonic waves utilize mechanical, electromagnetic, chemical, pneumatic, fluid dynamic, or thermal phenomena. Transducers for ultrasonic instruments are mainly in the electroacoustic category. Here, electrical energy is converted to acoustic energy, when transmitting; acoustic or mechanical wave energy is converted to an electrical signal, when receiving.

An important concern in the design of ultrasonic transducers and measuring instruments is the optimization of the system with respect to a reliable and accurate determination of the measurand – the parameter whose value is to be inferred or computed on the basis of 'ultrasonic' observations. In an ultrasonic experiment, one may observe transit time, amplitude of the received signal, oscillator frequency, oscillator strength, or other terms related to these. If more than one parameter is varying, say temperature and pressure, it may be necessary to observe more than one characteristic of the generated or received wave to resolve ambiguities. Generally speaking, one can say that each measurand (flow, temperature, pressure, etc.) influences one or more of the following: generation, propagation, and detection of ultrasound. For a fuller account of such influences and their measurement see Asher (1997), Auld (1990), Baker (2000), Brown and Lynnworth (2001), Hunt, Stubbs and Lee (2003), Kleppe (1989), Lynnworth (1989), Lynnworth and Mágori (1999), Youngdahl and Ellingson (1982), and Royer and Dieulesaint (1999).

2 EFFECTS OF MEASURAND ON ULTRASONIC PROPERTIES

Consider Figure 1. Without further study, it is not obvious as to which is the best way to proceed, where, say the object is to measure water temperature. Ultrasonic transducers and instruments have been available commercially for measuring signal characteristics related to temperature. Their operation is based on the temperature dependence of: (i) the transducer frequency; (ii) the speed of sound in water; (iii) the speed of sound in an intrusive sensor or probe; or (iv) the speed of sound in the beaker material.

In discussing the laboratory experiment depicted in Figure 1, the influence of pressure has been disregarded because this effect is relatively minute unless pressure is very high. Actually, the speed of sound depends on pressure in all the media involved. These include the solids (glass or metal beaker; quartz crystal, which we shall imagine to be embedded within the housed transducer assembly), the liquid (water), and the gas (air) that remain after all the water

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.

© 2005 John Wiley & Sons, Ltd.

ISBN: 0-470-02143-8.

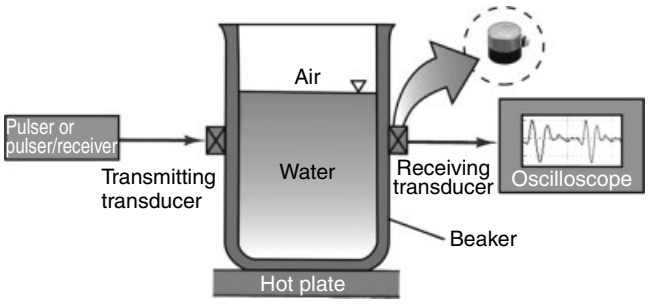


Figure 1. Ultrasonic effects of heating water in a beaker on which a diametrically-opposed pair of resonant piezoelectric (e.g. quartz) transducers are mounted. (Reproduced by permission of Panametrics, Inc.)

has boiled away. One might expect pressure effects to be negligibly small in an 'incompressible' liquid like water, and certainly negligible in a solid such as quartz. But this is not the case. For example, in the particular cut of quartz used by one manufacturer in their quartz pressure gauge, a change of pressure of 69 Pa (0.01 psi) was resolvable as a difference in its resonant frequency. In the case of air, pressure may be calculable if one can utilize the amplitude of the received signal, or the attenuation coefficient, as a measure of air pressure. Once pressure is approximately known, the temperature of dry air (or air of known relative humidity) can be calculated from the speed of sound.

The dependence of the speed of sound on temperature and density in a fluid such as oxygen is given in Figure 2 (Straty and Younglove, 1973). Their data suggests that, over the range studied, the *density* of fluid oxygen could be computed from measurements of speed of sound and temperature. The idea of determining the density of a *known*

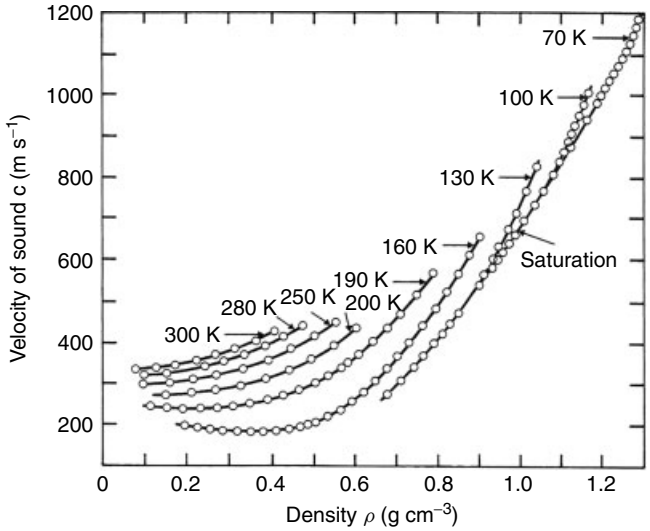


Figure 2. Velocity of sound versus density for saturated liquid and compressed fluid oxygen. (After Straty and Younglove, 1973.)

liquid at an approximately or accurately known pressure is illustrated for several liquids in Matson *et al*. (2002). Figures 3 and 4 suggest methods for determining air pressure or density.

3 TRANSDUCER EFFECTS USED IN ULTRASONIC INSTRUMENTATION

3.1 Piezoelectric, magnetostrictive, electrostatic, and electromagnetic mechanisms

The *piezoelectric* effect discovered by the Curie brothers over a century ago, involves a coupling between elastic

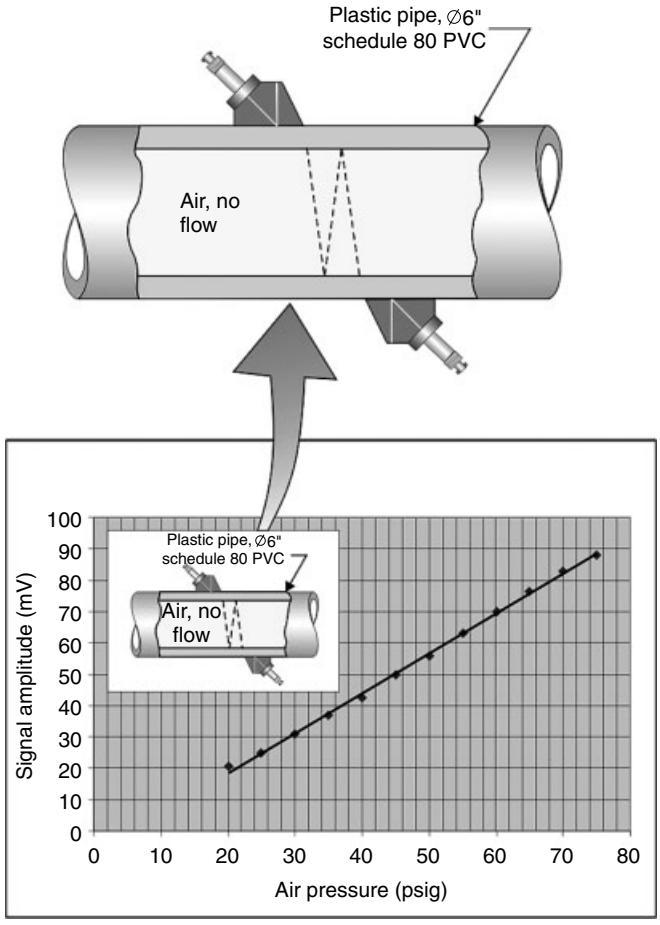


Figure 3. Example of air pressure measured at room temperature, and at no flow with clamp-on ultrasonic transducers. This measurement utilizes the principle that the amplitude of the received signal is proportional to the acoustic impedance of air and therefore, to a good approximation, proportional to the air pressure. Under turbulent flow, signals become jittery. At Mach numbers >0.1, beam drift further complicates the simple relationship depicted in the graph. Changes in temperature or gas composition also influence the pressure/amplitude relationship. (Reproduced by permission of Panametrics, Inc.)

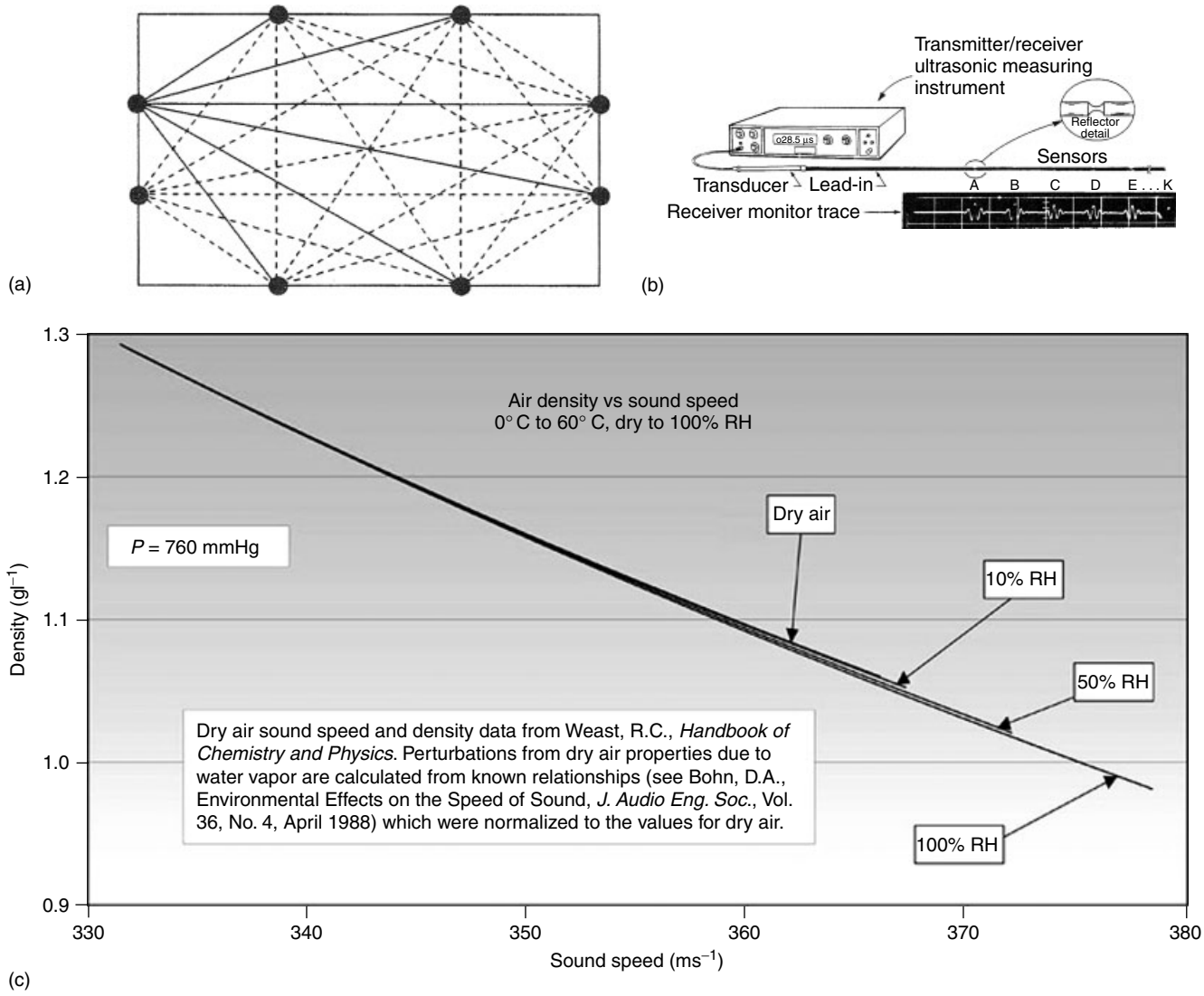


Figure 4. Ultrasonic thermometry using the speed of sound as a multipath indicator of temperature: (a) in the medium itself, (b) in an intrusive waveguide with one or more measuring zones, (c) extending concept (a) to density measurement based on the speed of sound in air. (Reproduced by permission of Panametrics, Inc.)

and dielectric phenomena. Piezoelectricity is found only in materials that have no center of symmetry. Quartz, lithium niobate, tourmaline, and other crystals exhibit the effect, but probably most industrial transducers, as used in testing and process measurements today, use poled ceramics as the electroacoustic element. PZT (lead zirconate titanate) and LM (lead metaniobate) are examples of such materials. As a transmitter, piezoelectric (ferroelectric) materials expand (or contract), or exhibit shear, in response to a driving voltage. As a receiver, an electrical signal is generated in response to an incoming pressure or shear wave. These effects can be observed over a wide temperature range, from nearly absolute zero up to the Curie point or the transformation temperature of the element, for example, 573 °C for quartz, and approximately 1200 °C for $LiNbO_3$. The usable frequency range extends down below the ultrasonic band into the audio and infrasonic range, and up to the gigahertz range.

Magnetostrictive devices are usually limited to frequencies below a few hundred kilohertz by the electrical skin effect, although operation into the megahertz range has been reported. Ni, Fe, Co, and their alloys are magnetostrictive, that is, they twist or change their dimensions in response to an applied magnetic field. Conversely, in response to a torsional, tensile, or compressional load, they generate a magnetic field that is typically sensed by a coil surrounding the magnetostrictor. As a compressional stress wave propagates along a magnetostrictive wire, one can imagine a

little bar magnet moving through the wire at the speed of sound. When that magnet passes through an encircling coil, a voltage is induced in that coil.

Electrostatic forces have been used in microphones for a long time. Micromachining has 'modernized' the options, leading to CMUTs (capacitive micromachined transducer). Many of their performance and manufacturability characteristics were summarized by Oralkan *et al.* (2002). Because of the low mass of the radiating membrane, CMUTs have been used in *air* transducers and *immersion* transducers.

Electromagnetic acoustic transducers (EMATs) utilize Lorentz forces in conductors. The Lorentz force is relatively small, but if the conductor is magnetostrictive, the combined effect is large enough to be utilized with essentially standard flaw detectors designed for ordinary piezoelectric transducers. Their chief advantage is the avoidance of any couplant. This can be important if the object under test has a rough surface, for example, a heavily oxidized iron pipe.

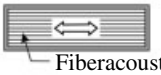
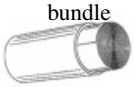
4 GUIDING ULTRASONIC WAVES TO SENSE A MEASURAND REMOTELY; BUFFERING AN ULTRASONIC TRANSDUCER

Ultrasonic waves can be guided along elastic structures such as rods, tubes, and plates; and through fluids in pipes or tubes. (In the ocean, acoustic waves can be guided in a sandwiched channel when the water above and below has a sound speed *higher* than that in the channel. But such waveguiding is beyond the scope of this section.) The sensing and buffering purposes of interest here include (i) using propagation in the waveguiding structure to sense one or more characteristics of the waveguide, such as its elastic moduli or its temperature; (ii) to sense a characteristic in the medium adjacent to the waveguide, such as the density or viscosity of the fluid in which the waveguide is immersed; or (iii) to buffer an ultrasonic transducer from a hot fluid into which the buffer radiates in order to measure some characteristic of that fluid, for example, speed of sound, attenuation coefficient, or flow velocity. Waveguides can convey ultrasound to and from a remote region where the transducer, if exposed to the harsh environment of that remote region, might fail because of thermal or chemical attack, or excess nuclear exposure. Another reason for using waveguides is because the specimen size or shape, for example, micrometer-diameter glass fiber, is such that it is awkward to introduce the interrogating energy directly from the transducer to that specimen. In other cases, the waveguide assists in mode-converting from extensional to torsional; from transverse shear to torsion; from extensional to flexural, and so on. Sometimes, multiple purposes are simultaneous goals.

Table 1 illustrates waveguides and buffers investigated by various authors between 1934 and 2003 in monostatic (pulse-echo) and bistatic (through-transmission, or pitch-catch) arrangements.

Figure 5, after Nguyen, Smart and Lynnworth (2004), shows pulse-echo and through-transmission torsional-wave interrogation of a noncircular portion of an otherwise-circular cross-section waveguide, in order to

Table 1. Buffers for cryogenic, high temperatures, or for other nasty fluids.

For *longitudinal* waves:	References	For *extensional* waves:	References
• Plain thick rod 	F (1948)	• Thin rod	B (1957)
• Threaded rod		• Marginally dispersive rod 	L (1992)
– Uniform pitch 	A & M (1977); Y & E (1978, 1982)	• Bundle of thin rods **or** concentric tubes  Fiberacoustic bundle	L, L & Z (1998); L & L (1999 & 2002)
– Non uniform pitch 	N & M (1993)	• Spiraled sheet 	H & K (2002)
– Knurl, undulating, other irregular surface			
• Clad rod			
– Cylindrical 	J & L (1998)	**For *torsional* waves**: see thin rod, above	B (1957); A & T (1957)
– Tapered 	J & L (1998)	**For *transverse shear* waves**: • Thin blade 	L, J & C (1996); L (2000)
• Hollow tube 	S & G (1934)	• Threaded rod - see diagram, above left, or **Article 120, Ultrasonic Instrumentation Design, Volume 1**, Figure 6, cell for 1977	A & M (1977)
		Clad rod-see diagrams, above left	J & L (1998)

Reproduced by permission of Panametrics, Inc.

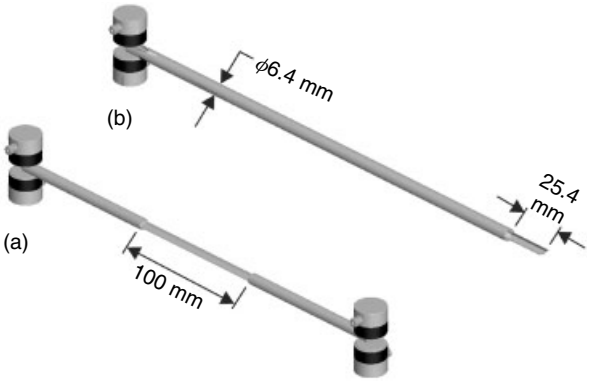


Figure 5. (a) Through-transmission configuration with two pairs of transverse shear transducers, one pair as transmitters and one pair as receivers, at each end of a waveguide. These generate and detect torsional waves by a mode conversion process (Nguyen, Smart and Lynnworth, 2004; Lynnworth, Cohen and Nguyen, 2004). The waveguide with the diamond cross section comprising its center section responds to liquid density (Kim and Bau, 1989; Shepard *et al*., 1999; Lower, Brow and Kurkjian, 2004) and (b) Pulse-echo configuration with one pair of transducers coupled to the circular cross-section end of the waveguide. (Reproduced by permission of Panametrics, Inc.)

sense the density of fluid surrounding the noncircular portion. This technique has a resolution on the order of $1\,\mathrm{mg\,cm^{-3}}$ and, therefore, while sensitive enough for many industrial liquid applications, it is not particularly useful for sensing the density of gases. Air, for example, has a density of $1.29\,\mathrm{mg\,cm^{-3}}$ at ordinary conditions. On the other hand, when the sensor portion is immersed in a low-density gas, it can sense the gas *temperature*, T, nearly independent of gas density or viscosity, especially if the T-sensor is a round rod or tube.

What happens if the sensor is immersed in a high-viscosity liquid? The torsional wave slows down in response to density *and* viscosity, besides being responsive to temperature. In this case, compensation is required to separate the variables of interest. Solutions may include using additional cross-sections and modes such that three independent ultrasonic measurements allow one to sort out the three parameters affecting propagation. If temperature is sensed by a thermocouple, only the two remaining variables need to be sorted out ultrasonically.

5 CONTAINER WALLS AS SENSORS OR WINDOWS FOR NONINVASIVE INTERROGATION

When categorizing ultrasonic instrumentation principles and techniques, it is convenient to organize them according to whether the sensor is the medium itself or an intrusive sensor. Here we use 'sensor' to mean the medium, material, or structure in which the generation, reception, or propagation of ultrasound is measured and in which the said generation, reception, or propagation is influenced by the state, characteristics, or value(s) of measurands in the path.

Referring again to Figure 1, if we think of water as its own sensor ('Propagation' column in Table 2), we must nevertheless acknowledge that the electroacoustic transducers might be legitimately called *sensors*. By designating water as the sensor, we can emphasize that the speed of sound in water is the basis of the way in which we obtain information on the water's temperature. The electroacoustic piezoelectric quartz transducers shown in this particular illustration could be replaced by an external laser generation and detection system, focused on opposite sides of a black beaker, yet the speed of sound in water could still serve as the basis for determining water temperature.

Finally, one must not overlook what might appear to be an intermediate category, that is, using the existing wall used as a sensor. As one example, the acoustic 'ringdown' or echo decay pattern in the wall of a vertical standpipe is often proposed to tell if liquid is present at that level. In the common example of a clamp-on flowmeter, the wall of the pipe provides an acoustically transparent window through which the interrogating beam gains access to the liquid in the pipe. The liquid is its own sensor. The pipe and suitable transducer assemblies make it possible to utilize that sensor

Table 2. Examples of heating effects on the generation, propagation, and detection of ultrasound in water, in the laboratory test of Figure 1.

	Heating effects on:		
Temperature (°C)	Generation	Propagation	Detection
Increase from 20 to 74	Frequency increases	Speed of sound increases; absorption decreases	Signal arrives progressively earlier
75 to 99	Frequency increases	Speed of sound decreases; absorption decreases	Signal arrives progressively later
100	Frequency increases	Propagation is largely blocked by boiling and then by low-impedance air after all the water boils away	Signal weakened by scattering due to boiling and may be undetectable

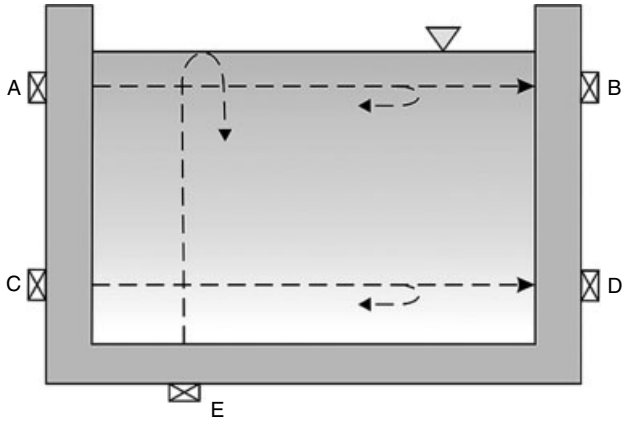


Figure 6. Liquid level obtained noninvasively from an upward-looking pulse-echo transducer, together with four sidewall-mounted through-transmission transducers to compensate for sound speed gradients as well as to have multiple discrete measuring heights (two in this example). Transducers B and D can be deleted if transducers A and C operate in pulse-echo mode, like transducer E. (Reproduced by permission of Panametrics, Inc.)

without invading or physically penetrating the boundaries of the fluid (see Figure 6).

6 LIMITATIONS ON PRECISION AND ACCURACY

6.1 Limitations on coupling, propagation, spatial resolution, and correlations among propagation and measurands

In the field of ultrasonic nondestructive testing (NDT), a variety of couplants – liquids, oils, pastes, gels, and elastomers – are available for transmitting the interrogating wave from the transducer assembly to the part being tested (see Figure 7). The couplant tends to match acoustic impedances as it replaces the low-impedance air gap that would otherwise block most of the transmission between a solid transducer and a solid test object. At temperature extremes, the choice of couplants is greatly diminished. Accordingly, in cryogenic or high-temperature situations, one may resort to pressure coupling (sometimes aided by a compliant film of metal) or to solid bonds of epoxy, solder, or ceramic cements. Coupling to liquid metals is sometimes facilitated by coating the probe so that it is wetted but not dissolved by the liquid metal.

Temperature extremes may be associated with severe sound speed gradients or other inhomogeneous conditions that block propagation. On the one hand, ultrasonic measurements have been made in the laboratory from nearly absolute zero to plasma temperatures of 20 000 K. On the other hand, practical measurements in cryofluids or merely

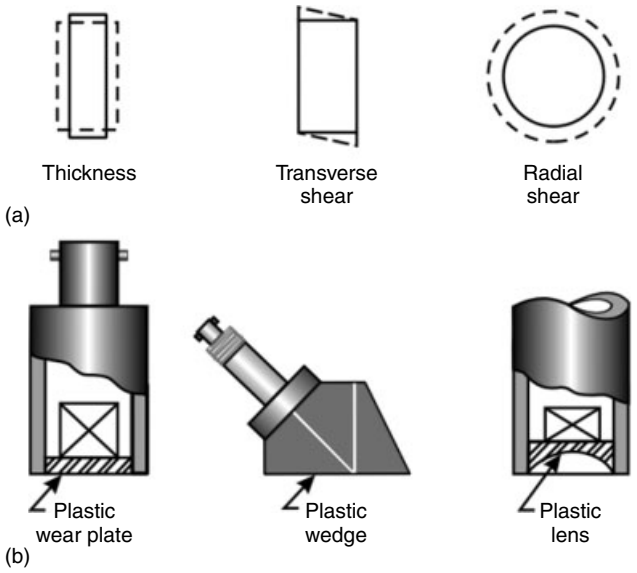


Figure 7. Examples of: (a) thickness mode, transverse shear (often called thickness shear), and radial mode crystals and (b) epoxy-sealed NDT transducers. (Reproduced by permission of Panametrics, Inc.)

in hot water near saturation conditions are often interrupted by 'flashing' or boiling because of the severe mismatch in acoustic impedance between liquid and gas phases.

Another limitation in ultrasonic measuring systems is spatial resolution. Apart from speckle and wavelength-dependent limits, the propagation characteristics of the medium between the transducer and the region under study influence resolution in that region. As in optical systems, diffraction limits the lateral resolution. Depth resolution is usually limited by the time duration of the interrogating pulse, or its bandwidth. Refraction (Snell's law) limits the angle that a beam can be aimed relative to an interface. This means noninvasive interrogations may be unable to examine all areas inside a bounding pipe or vessel.

Yet another limitation applies to the attempts to generalize the correlations between variables such as, say, strength and the speed of sound, molecular weight and the speed of sound, or Young's modulus and the speed of sound. Such correlations may be quite useful when judiciously applied to a limited class of materials, or to specimens of similar alloys of like cross-section. One must be alert, however, to avoid errors introduced by unwanted variables, such as moisture, temperature, pressure, boundary dimensions comparable to wavelength, and dispersion.

REFERENCES

Asher, R.C. (1997) *Ultrasonic Sensors for Chemical Process Plant*, Institute of Physics Publishing, London.

Auld, B.A. (1990) *Acoustic Fields and Waves in Solids*, 2nd edn, Krieger Publishing, Malabar Florida.

Baker, R.C. (2000) *Flow Measurement Handbook*, Cambridge University Press, Cambridge.

Brown, A. and Lynnworth, L.C. (2001) Ultrasonic Flowmeters, Chapter 20, in *Flow Measurement – Practical Guides for Measurement Control*, 2nd edn (ed D.W. Spitzer), ISA (pp. 515–573).

Hunt, W.D., Stubbs, D.D. and Lee, S.-H. (2003) Time-Dependent Signatures of Acoustic Wave Biosensors. *Proceedings of the IEEE*, **91**(6), 890–901.

Kim, J.O. and Bau, H.H. (1989) On Line, Real-time Densimeter-theory and Optimization. *Journal of the Acoustical Society of America*, **85**(1), 432–439.

Kleppe, J.A. (1989) *Engineering Applications of Acoustics*, Artech House, Boston, MA.

Lower, N.P., Brow, R.K. and Kurkjian, C.R. (2004) Inert Failure Strains of Sodium Aluminosilicate Glass Fibers. *Journal of Non-crystalline Solids*, **344**, 17–21.

Lynnworth, L.C. (1989) *Ultrasonic Measurements for Process Control, Theory, Techniques, Applications*, Academic Press.

Lynnworth, L.C., Cohen, R. and Nguyen, T.H. (2004) Clamp-On Shear Transducers Simplify Torsional and Extensional Investigations, *Proceedings Ultrasonics Symposium IEEE*, Montreal, August 2004.

Lynnworth, L.C. and Mágori, V. (1999) Industrial Process Control Sensors and Systems, Chapter 4, in *Ultrasonic Instruments and Devices: Reference for Modern Instrumentation, Techniques, and Technology* (E.P. Papadakis Guest Editor), 23 in the Series *Physical Acoustics*, Academic Press (pp. 275–470).

Matson, J., Mariano, C.F., Khrakovsky, O. and Lynnworth, L.C. (2002) Ultrasonic Mass Flowmeters Using Clamp-On or Wetted Transducers, in *Proceedings of the 5th International Symposium on Fluid Flow Measurement*.

Nguyen, T.H., Smart, C.D. and Lynnworth, L.C. (2004) Transverse Shear to Torsion Mode Converter and Applications. *Materials Evaluation*, **62**(6), 690–698.

Oralkan, Ö., Sanli Ergun, A., Johnson, J.A., Demirci, U., Karaman, M., Kaviani, K., Lee, T.H. and Khuri-Yakub, B.T. (2002) Capacitive Micromachined Ultrasonic Transducers: Next-Generation Arrays for Acoustic Imaging? *IEEE Transactions on UFFC*, **49**(11), 1596–1610.

Royer, D. and Dieulesaint, E. (1999) *Elastic Waves in Solids II*, Springer.

Shepard, C.L., Burghard, B.J., Friesel, M.A., Hildebrand, B.P., Moua, X., Diaz, A.A. and Enderlin, C.W. (1999) Measurements of Density and Viscosity of One- and Two-phase Fluids with Torsional Waveguides. *IEEE Trans UFFC*, **46**(3), 536–548.

Straty, G.C. and Younglove, B.A. (1973) Velocity of Sound in Saturated and Compressed Fluid Oxygen. *Journal of Chemical Thermodynamics*, **5**(5), 305–312.

Youngdahl, C.A. and Ellingson, W.A. (1982) Acoustic System for Monitoring Pressure Boundary Wear, Presented at *1982 Symposium on Instrumentation and Control for Fossil Energy Processes*, Houston, TX.

FURTHER READING

Andreatch, P. Jr. and Thurston, R.N. (1957) Disk-Loaded Torsional-Wave Delay Line. *Journal of the Acoustical Society of America*, **29**(1), 16–19.

Ao, X. (2003) *Clamp-On Steam/Gas Flow Meter*, U.S. Patent 6,626,049.

Apfel, J.H. (1962) Acoustic Thermometry. *Review of Scientific Instruments*, **33**(4), 428–430.

Araki, H. and Matsunaga, Y. (1977) *Ultrasonic Flow Meter*, U.S. Patent 4,014,21.

Bell, J.F.W. (1957) The Velocity of Sound in Metals at High Temperatures. *Philosophical Magazine*, **2**, 1113–1120.

Bohn, D.A. (1988) Environmental Effects on the Speed of Sound. *Journal of the Audio Engineering Society*, **36**(4), 1–9.

Frederick, J.R. (1948) Ultrasonic Measurement of the Elastic Properties of Polycrystalline Materials at High and Low Temperatures (Abstr.). *Journal of the Acoustical Society of America*, **20**, 586.

Heijnsdijk, A.M. and Klooster, J.M.van P. (2002) *Ultrasonic Waveguide*, U.S. Patent 6,400 (p. 648).

Hill, J. and Weber, A. (2002) Qualification of Ultrasonic Flowmeters of Natural Gas Using Atmospheric Air Calibration Facilities, in *Proceedings of the 20th North Sea Flow Measurement Workshop*, St. Andrews.

Hurd, J.C., Nguyen, T.H. and Lynnworth, L.C. (2004) Shear Wave Polarization Follows Twist of Rectangular Steel Bar. *Materials Evaluation*, **62**(1), 37–42.

Jen, C.K. and Legoux, J.G. (1998) *Clad Ultrasonic Waveguides with Reduced Trailing Echoes*, U.S. Patent 5,828,274.

Khuri-Yakub, B.T., Kim, J.H., Chou, C.-H., Parent, P. and Kino, G.S. (1988) A New Design for Air Transducers, in *1988 Ultrasonics Symposium Proceedings, IEEE* (pp. 503–506).

Lemmon, E.W., McLinden, M.O. and Friend, D.G. (2000) Thermophysical Properties of Fluid Systems, in *NIST Chemistry WebBook*, NIST Standard Reference Database Number 69 (eds W.G. Mallard and P.J. Linstrom), National Institute of Standards and Technology, Gaithersburg, MD 20899. URL: http://webbook.nist.gov.

Liu, Y., Lynnworth, L.C. and Zimmerman, M.A. (1998) Buffer Waveguides for Flow Measurement in Hot Fluids. *Ultrasonics*, **36**(1–5), 305–315.

Lynnworth, L.C. (1979) Ultrasonic Flowmeters, Chapter 5, in *Physical Acoustics – Principles and Methods*, Vol. 14 (eds W.P. Mason and R.N. Thurston), Academic Press (pp. 407–525).

Lynnworth, L.C. (1992) *Marginally Dispersive Ultrasonic Waveguides*, U.S. Patent 5,159,838.

Lynnworth, L.C. (2000) *Ultrasonic Buffer/Waveguide*, U.S. Patent 6,047,602.

Lynnworth, L.C. and Liu, Y. (1999) *Ultrasonic Path Bundle and Systems*, U.S. Patent 5,962,790; (2002) 6,343,511.

Lynnworth, L.C., Jossinet, G. and Chérifi, E. (1996) 300 °C Clamp-On Ultrasonic Transducers for Measuring Water Flow

and Level, in *1996 Ultrasonics Symposium Proceedings, IEEE* (pp. 407–412).

Nygaard, O.G.H. and Mylvaganam, K.S. (1993) Ultrasonic Time – Domain Reflectometry for Level Measurement in Molten Metals. *Technisches Messen* **60**(1), 4–14.

Sherratt, G.G. and Griffiths, E. (1934) The Determination of the Specific Heat of Gases at High Temperatures by the Sound Velocity Method, I-Carbon Monoxide. *Proceedings of the Royal Society (London), Series A*, **147**, 292–308.

Szabo, T.L. (2004) *Diagnostic Ultrasound Imaging – Inside Out*, Academic Press.

Thurston, R.N. (1978) Elastic Waves in Rods and Clad Rods. *Journal of the Acoustical Society of America*, **64**(1), 1–37.

Ting, V.C. and Ao, X. (2002) Evaluation of Clamp-On Ultrasonic Gas Flowmeters for Natural Gas Applications, in *Proceedings of the 20th North Sea Flow Measurement Workshop*, St. Andrews.

Weast, R.C. (1989) *Handbook of Chemistry and Physics*, CRC Press (pp. E-46, F-8).

Youngdahl, C.A. and Ellingson, W.A. (1978) Development of Ultrasonic Techniques for Remote Monitoring of Erosive Wear in Coal-Conversion Systems, in *1978 Ultrasonics Symposium Proceedings, IEEE* (pp. 305–310).

120: Ultrasonic Instrumentation Design

Lawrence C. Lynnworth
Lynnworth Technical Services, Waltham, MA, USA; Formerly Chief Technologist at GE Panametrics

1 Selection and Influence of New Electroacoustic Materials, Components, and Techniques 810
2 Special Considerations for Temperature and Pressure Extremes 812
3 Ultrasonic Transducers for Gases 812
4 Ultrasonic Transducers for Liquids – Wetted, Clamp-on, Hybrid 812
5 Ultrasonic Transducers for Nondestructive Testing and Evaluation of Solid Materials 816
References 819
Further Reading 820

1 SELECTION AND INFLUENCE OF NEW ELECTROACOUSTIC MATERIALS, COMPONENTS, AND TECHNIQUES

The pie chart of Figure 1 categorizes ultrasonic measurement or instrumentation applications according to the four principal process control measurands (flow, temperature, pressure, and liquid level). Other slices express explicitly or implicitly, combinations, NDT(nondestructive testing), and analyzer measurands (e.g. gas concentration, relative humidity, etc.). To keep this section of manageable size, let us focus on a specific task.

Consider the task of designing an ultrasonic system to measure without contact, that is, using air-coupled transducers, several characteristics of a fast-moving solid. If the measurement is important, there are likely to be competitors, so the task requires a solution that is better, faster, and cheaper than the competition's. The best solution is likely to be a compromise that nevertheless benefits from the designer's judicious choice of transducer material and mode, methods of data acquisition, signal processing, display features, output functions, and so on. The total ultrasound design task includes many factors (Figure 2). Design principles underlying a diagram of this type are discussed in books such as those by Asher (1997), Auld (1990), Baker (2000), Lynnworth (1989), Royer and Dieulesaint (2000), Rogers (1984), or others. In this section, we deal only with 'air' transducers, admittedly in an overview fashion.

It will be understood that the relatively *low-acoustic impedance of air* (compared to water or common engineering metals) has been a major barrier to ultrasonically testing and measuring materials in air, or even testing air itself, that is, measuring air flow or air temperature. Low-mass membranes and low acoustic–impedance electroacoustic materials are appropriate for ultrasonic measurements of air, gases, or solids tested by air-coupled transducers. NDT 'immersion transducers', developed many years ago for testing solids immersed in water, are sometimes used for biomedical applications in body tissue, technical applications in liquids, and in some gases other than air, as well as air. In other words, sometimes one transducer can be used to test solids, liquids, or gases. Optimizing for a specific application in one state of matter or another, however, may involve selecting or improving the electroacoustic material.

As far as impedance mismatch is concerned, bear in mind that the characteristic acoustic impedance of a gas, Z_{GAS}, equals the product of gas density times sound speed. This means Z_{GAS} increases roughly in proportion to pressure times the square root of gas MW (molecular weight). As a numerical example, the Z for argon at 10 bar is approximately thirty times higher than Z for helium at atmospheric pressure.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

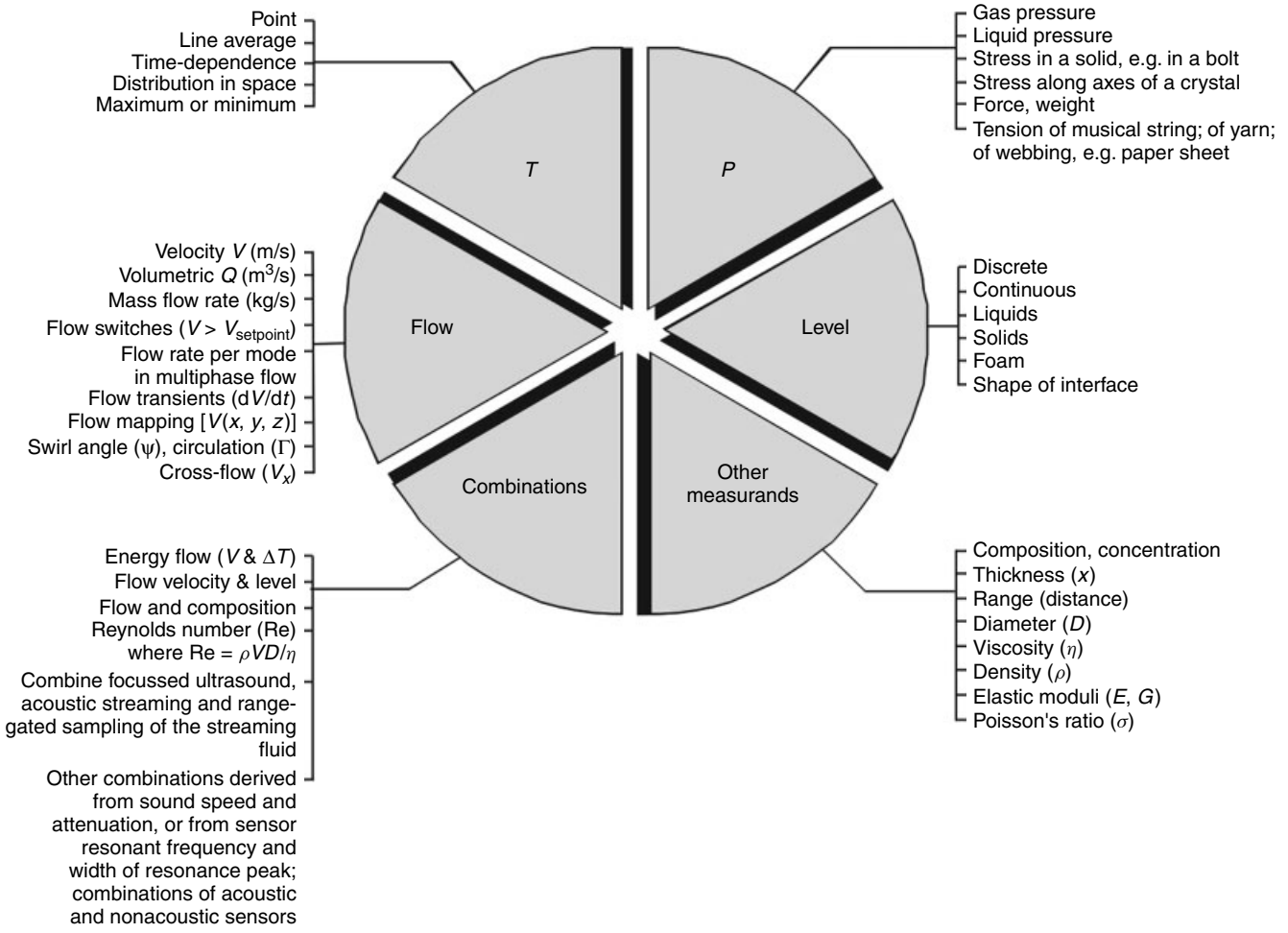


Figure 1. Pie chart categorizes ultrasonic low-intensity sensing, measurement and control applications. (Reproduced by permission of Panametrics, Inc.)

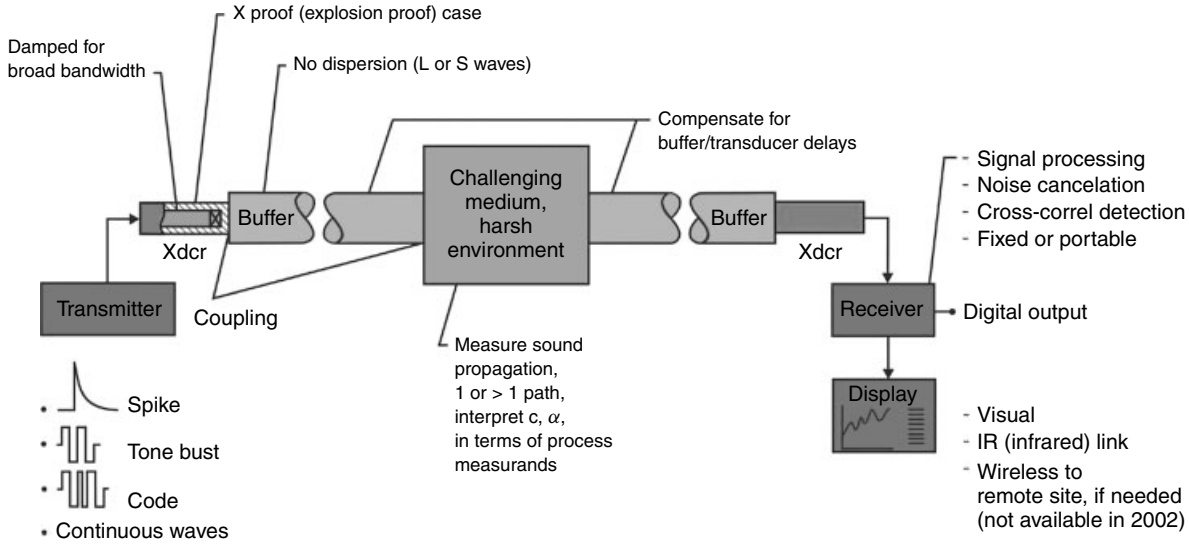


Figure 2. Factors and choices to be considered when designing an ultrasonic measuring system. (Reproduced by permission of Panametrics, Inc.)

New electroacoustic materials, new manufacturing techniques [e.g. layered manufacturing, Safari *et al.* (2002)], piezo polymer composites, CMUTs (capacitive micromachined ultrasonic transducers), polyvinylidene fluoride and copolymer piezoelectric sheet materials, and electrets comprise areas that continue to be dynamic.

2 SPECIAL CONSIDERATIONS FOR TEMPERATURE AND PRESSURE EXTREMES

Temperature extremes can cause failure of transducers in several ways: depolarization when the Curie point is exceeded; chemical reactions when hot; cracking or disbonding due to differential thermal expansion; couplant boils away or becomes embrittled at hot or cold extremes; transducer material becomes electrically conductive when hot; attenuation in the transducer backing or speed of sound in an adjacent wedge changes too much; soldered connections melt. High-purity gas applications may require that during vacuum bakeout the transducer, if wetted, neither outgases nor leaks. For use at high pressure, the transducer must neither deform nor leak. In a downhole tool for exploring geothermal reserves, transducers may be subjected simultaneously to high temperature and high pressure (e.g. $T = 100$ to 350 °C; $P = 20$ to 70 MPa). In testing red-hot steel, buffer rods or momentary contact may be able to isolate the transducer element from the high temperature of the product and the high pressure required for dry coupling. Lasers and emats (electromagnetic acoustic transducers) offer noncontact approaches that avoid couplants. Eventually, such methods may become routine in industry.

3 ULTRASONIC TRANSDUCERS FOR GASES

In contrast to liquids or solids, gases are of low density ρ, low speed of sound c, and consequently of low acoustic–impedance Z – see comparisons in Table 1.

This means that if a *high* acoustic–impedance transducer is used, it is much more difficult to transmit ultrasound into or out of a gas compared to liquids or solids. One generally seeks low-impedance transducers for gases.

Apart from low-mass membrane, ultrasonic transducers and low acoustic–impedance transducers, which are appropriate for air, there are several housed piezoelectric disk designs to be considered. These may be categorized as unmatched and matched. Unmatched designs may encapsulate a resonator (thickness, radial, or flexural mode) in an impervious housing. Matched designs, for which more details have been published, include a single $\lambda/4$ matcher having an impedance as close as practical to the geometric mean of the impedances of the piezoelement and the gas. Two-layer designs also exist (Figure 3). The simpler one-layer design has proved adequate in over a thousand flare gas flowmeters installed in pipelines, refineries, and chemical plants between 1983 and 2003. This design has also been used in high-accuracy flowmeters for natural gas, that is, Hill and Weber (2002).

Acoustic isolation of transmitter from receiver, to prevent cross talk, is often one of the most difficult design aspects of gas transducers used in pairs. Some approaches are illustrated in Lynnworth (1989), or in Lynnworth and Mágori (1999) utilizing transducers mounted on membranes; intentionally alternated impedance mismatches; small contact areas; time delays or damping. If the same transducer is used as transmitter and receiver, ringdown often becomes the key problem as far as coherent noise is concerned. Signals indicative of the measurand cannot be detected until coherent noise generated on transmission has subsided sufficiently or unless it can be subtracted. Some examples of isolating or damping transducers, and damping cross talk or ringing, are shown in Figure 4. In manufacturing broadband NDT transducers, the backing impedance matches that of the piezoelement; it is intentionally attenuative; and may have one or more antireflection surfaces (Papadakis and Fowler, 1972).

4 ULTRASONIC TRANSDUCERS FOR LIQUIDS – WETTED, CLAMP-ON, HYBRID

In this section, we introduce three categories of transducer designs for liquids, Figure 5. Principal characteristics are compared in Table 2.

Table 1. Comparison of density and two ultrasonic propagation characteristics of ultrasound for air, water, and stainless steel (longitudinal waves).

Medium (20°C, 10^5 Pa)	Density (ρ) (kg/m^3)	Speed of sound (c) (m/s)	Acoustic impedance (Z) (kg/m^{2}s)	(Mrayls)
Air	1.29	343	450	0.00045
Water	998	1482	1.48×10^6	1.48
Stainless steel 316	7833	5760	45.1×10^6	45.1

Table 2. Categories of transducer designs for liquids.

Type	Advantages	Disadvantages
Wetted	Beam path is controllable; used in the most accurate ultrasonic methods; usable over wide frequency range e.g. $0.1 < f < 10$ MHz. May comprise part of replaceable module, e.g. avionic flowmeter.	Requires penetration of pressure boundary; requires isolation valve or empty pipe for installing and removing; cavities distort flow.
Nonwetted clamp-on, strap-on, snap-on, epoxy-on, braze-on, weld-on	Noninvasive, removable unless bonded; easy to retrofit, no cavities. In many cases, it would appear to be the safest way to measure flow in the sense that the pipe integrity is not jeopardized.	Oblique beam path depends on liquid; tendency for acoustic short-circuit noise when conduit wall is thick; usually $f < 2$ MHz. Limited usually to liquids; sometimes usable with gases at high pressure, or gases at atmospheric pressure in special circumstances only.
Hybrid	Combines advantages of wetted and nonwetted approaches. In some cases, the 'window' can be welded, molded, or cast into a valve or pipe fitting to accommodate the occasional or permanent attachment of the transducer.	Requires initial penetration of pressure boundary if window is installed as a retrofit; introduces cavities.

Wetted designs, Figure 5(a), typically employ piezoelements resonant near 0.5, 1, or 2 MHz and may be narrowband or broadband depending on the backing impedance and on the window thickness. Low-impedance backings and thin windows lead to narrowband waveforms. Backings that match the element's impedance and thick metal windows lead to broadband waveforms. But ringing in a thick window can compromise the broadband response. Tapered pipe–threaded housings can be rated up to 20 Mpa (3000 psi). *Flanged* designs, however, are usually preferred for high pressure, depending in part on the fluid and the industry.

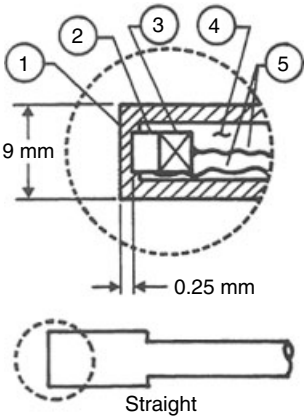
Nonwetted external or clamp-on types, Figure 5(b), are designed to fit in standard yokes or tracks. Sometimes they are permanently attached or integral with the spool. Tracks are available with scales so that the axial separation between transducers, as required in contrapropagation or tag flow measurements, can be set easily. For large pipes, magnets or wrap-around straps are commonly used to hold the yokes, tracks, or transducers. For small pipes, below 50-mm inside diameter, one or more transducers may be fixed temporarily or permanently within a given clamp-on or snap-on assembly. Such dual-transducer flowmeter assemblies, for example, became available in the 1980s from Transonic Systems for soft biomedical tubing having internal diameters down to 1 mm, and from Panametrics in 1990 for metal tubing from 10- to 50-mm inside diameter.

The two principal liquid process control uses of external nonwetted transducers have been to measure (i) liquid level, and (ii) flow in pipes. The same or similar transducers can often be used to measure thickness and integrity of pipes and pressure vessels.

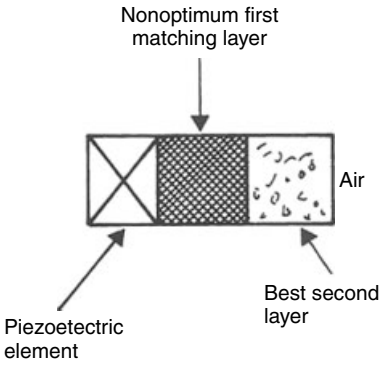
Hybrid flowmeter transducers, Figure 5(d, e) evolved first as a convenience for removability and second, and independently, as a remedy to the occasional nuisance and sometimes intolerable errors caused by uncertainty or drifting of the obliquely refracted beam launched in the liquid by a clamp-on angle beam transducer. A hybrid transducer consists of a flat-faced plug that can permanently maintain the pressure boundary while defining one terminus of a known fixed measuring path normal to the wetted face of the plug, and a removable nonwetted transducer that repeatedly mates with the said plug. A hybrid *flowcell* may be taken to mean a spool with transducers that do not comprise part of the pressure boundary. These worked with liquids first and later with gases. A few designs are collected in Figure 6.

In a round conduit of area A, if the flow profile were known precisely, then measurements at a point or along a chord could be readily converted to the area-averaged flow V_{AVG}, from which the volumetric flow rate Q could be calculated as AV_{AVG}. For example, if the flow is steady and *laminar,* the profile is parabolic and a velocity measurement on the axis, V_{AXIS}, would be exactly twice V_{AVG}. Along the midradius chord the chordal average equals V_{AVG}. A measurement along the tilted diameter, V_{DIAM}, is higher than V_{AVG} by 33.3% but is corrected by a meter factor K of value 0.750 in this instance. In other words, $Q = KAV_{\mathrm{DIAM}} = 0.750AV_{\mathrm{DIAM}}$.

For *turbulent* flow, if pipe relative roughness $\in_R$ is known, models exist for estimating the flow profile as a function of the Reynolds number Re. At $\mathrm{Re} = 10^6$, if the pipe is smooth, the diameter path yields V_{DIAM} about 5% higher than V_{AVG}, which is routinely corrected using $K = 0.95$, in this instance. The correction generally is slightly incorrect because of (i) uncertainty in $\in_R$ and its influence, and (ii) uncertainty in the effective beam diameter. *Multipath* flowmeters typically include two or more


1
2
3
4
5
9 mm
0.25 mm
Straight
90°
45°

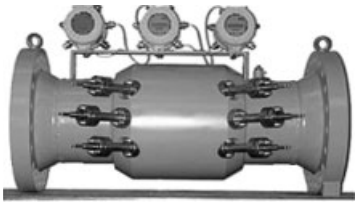
(a)


Nonoptimum first matching layer
Air
Best second layer
Piezoetectric element

(b)



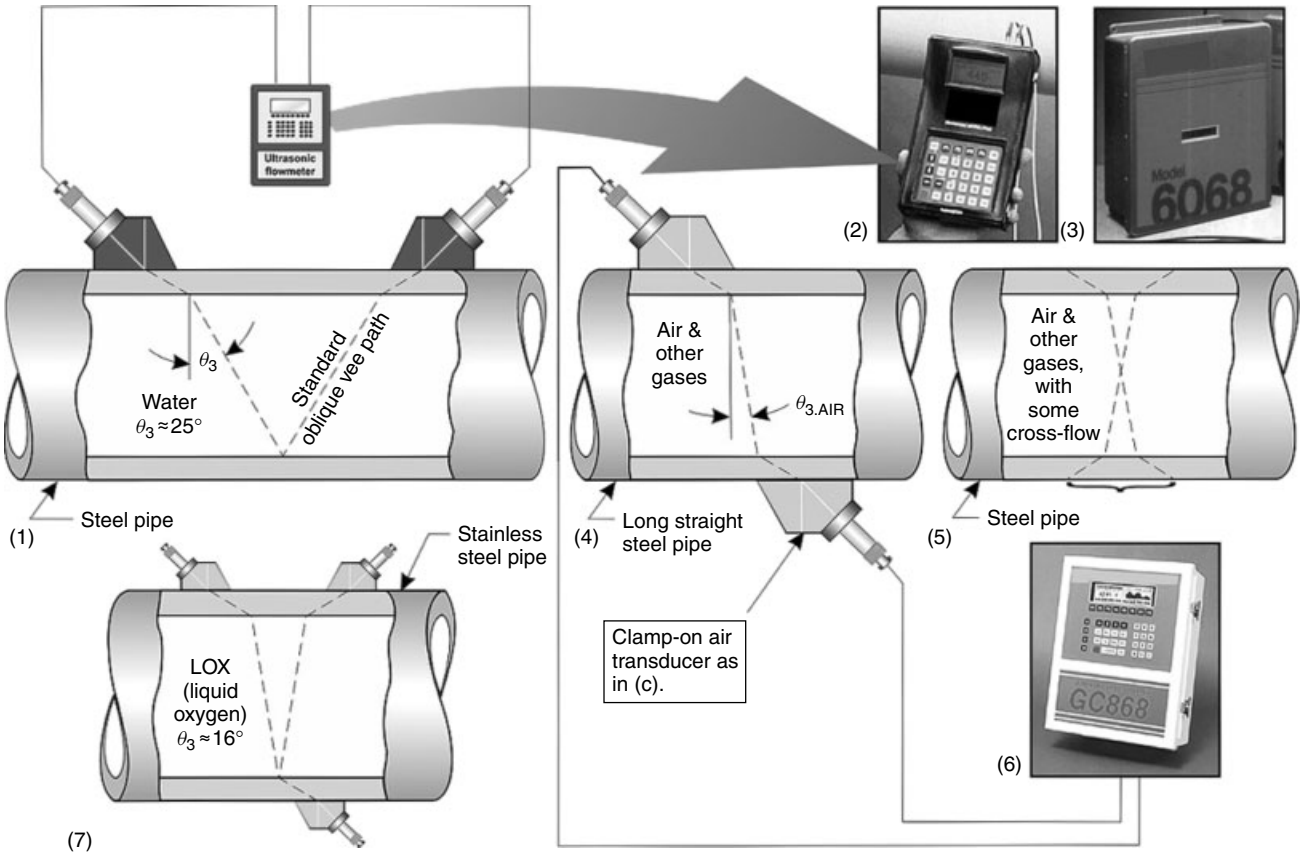
(c)



(d)



(e)


Ultrasonic flowmeter
(2)
(3)
Model 6068
θ_3
Water
$\theta_3 \approx 25°$
Standard oblique vee path
Steel pipe
(1)
Air & other gases
$\theta_{3.AIR}$
(4)
Long straight steel pipe
Air & other gases, with some cross-flow
(5)
Steel pipe
Stainless steel pipe
LOX (liquid oxygen)
$\theta_3 \approx 16°$
(7)
Clamp-on air transducer as in (c).
GC868
(6)

(f)

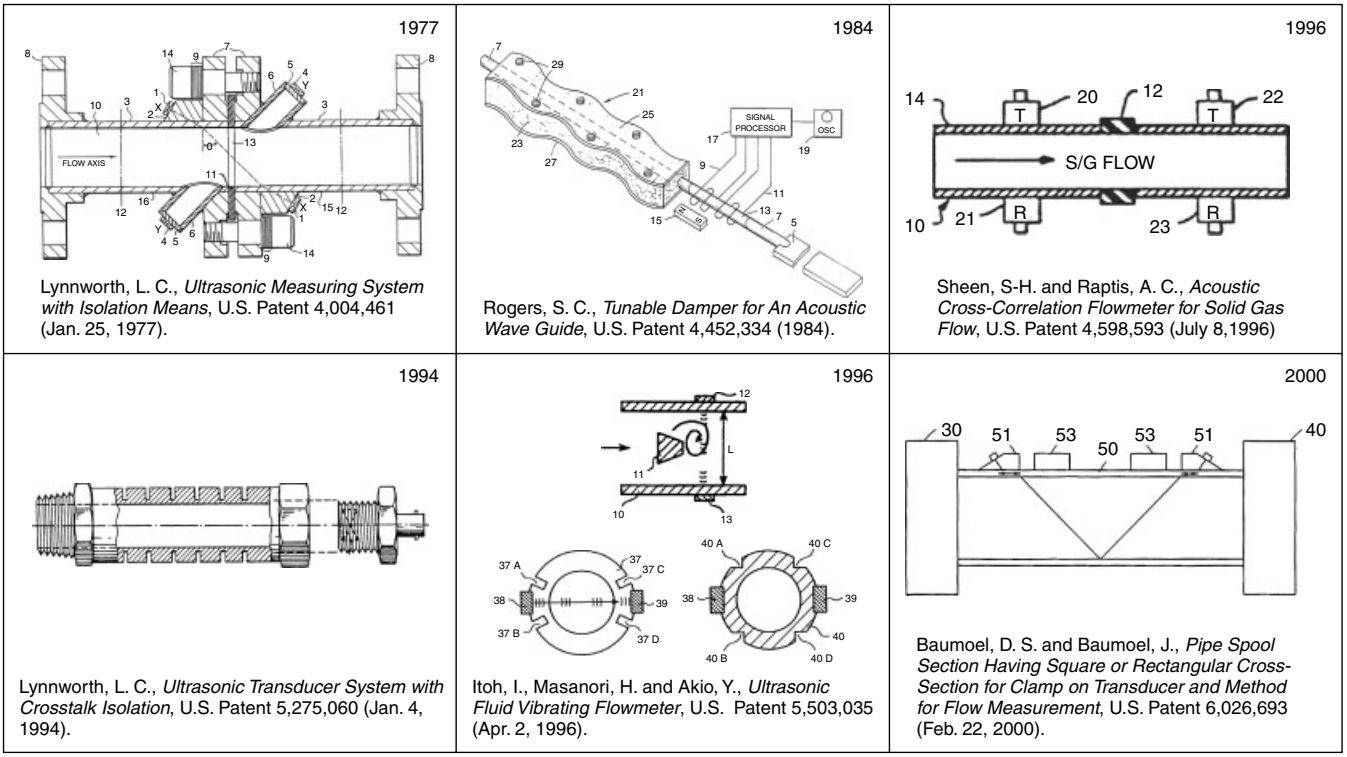


Figure 4. Isolation and damping methods that have been used when the object is to measure a gas, liquid, or solid medium. (Reproduced by permission of Panametrics, Inc.)

paths off the diameter so that a quadrature integration or other weighting procedure can obtain V_{AVG} substantially independent of the profile. For small pipes, the entire duct area can be insonified, as practiced in methods termed 100% area-averaging. In some Doppler flowmeters, range gating is employed, whereby measurements are obtained as a function of distance from the wall. Then the corresponding annular areas are multiplied by their average velocities and the products are integrated to yield $\mathrm{Q} = AV_{\mathrm{AVG}}$ (Takeda *et al.*, 2000). To enhance the interaction between the interrogating wave and the fluid, and thereby improve performance of Doppler or other scatter-based

Figure 3. Transducers for gases: (a) metallurgically sealed single-layer impedance-matched designs for flare gases according to a patent by Lynnworth, Fowler and Patch (1981); legend: 1 thin window, 2 quarter wave impedance matcher, 3 piezo element, 4 potting/backing material, 5 leadwires, (b) two-layer unencapsulated design for air, after Khuri-Yakub *et al*. (1988), (c) clamp-on air transducer, after Ao (2003); see Ting and Ao (2002), (d) gas flowmeter using quadrature integration of flow data obtained in three parallel planes, courtesy of RMG; see Hill and Weber (2002), (e) gas transducer similar to straight one shown in (a) and usable in air at atmospheric pressure and methane up to 100 bar, and (f) comparison of clamp-on paths. Path angles are calculated assuming a refracted angle θ_2 of 60° for a shear wave in the steel pipe. Refracted angle depends on sound speed c_3 in the fluid. Referring to items (1)–(3) commercial contrapropagation clamp-on flowmeters available since the early 1990s include Panametrics' PT868 and 6068 for measuring the flow of liquids. In *water*, the vee path usually works, and θ_3 is about 25° at room temperature. In *air*, $\theta_{3.\mathrm{AIR}}$ is only about 6° and unless pressure is high, the transducers usually need to be placed on *opposite* sides of the pipe. (5) For *liquid* clamp-on, the vee path (shown in (1)) tends to cancel cross-flow as well as double the sensitivity to flow compared to a single traverse. For *gases*, odd numbers of traverses are preferred, to reduce cross talk. This means, if cross-flow is significant, crossed paths are recommended. The velocities measured along the legs of the *X* should be averaged. Best solution: find a long straight run far from disturbances and joints. For *gases* (diagrams (4) & (5)) the flowmeter instrument (6) is Panametrics' GC868 (Ting and Ao (2002)). Diagram (7), drawn for LOX (liquid oxygen) or LN2 (liquid nitrogen), shows the refracted angle $\theta_3 \approx 16°$, nearly midway between water (25°) and air (6°). (Reproduced by permission of Panametrics, Inc.)

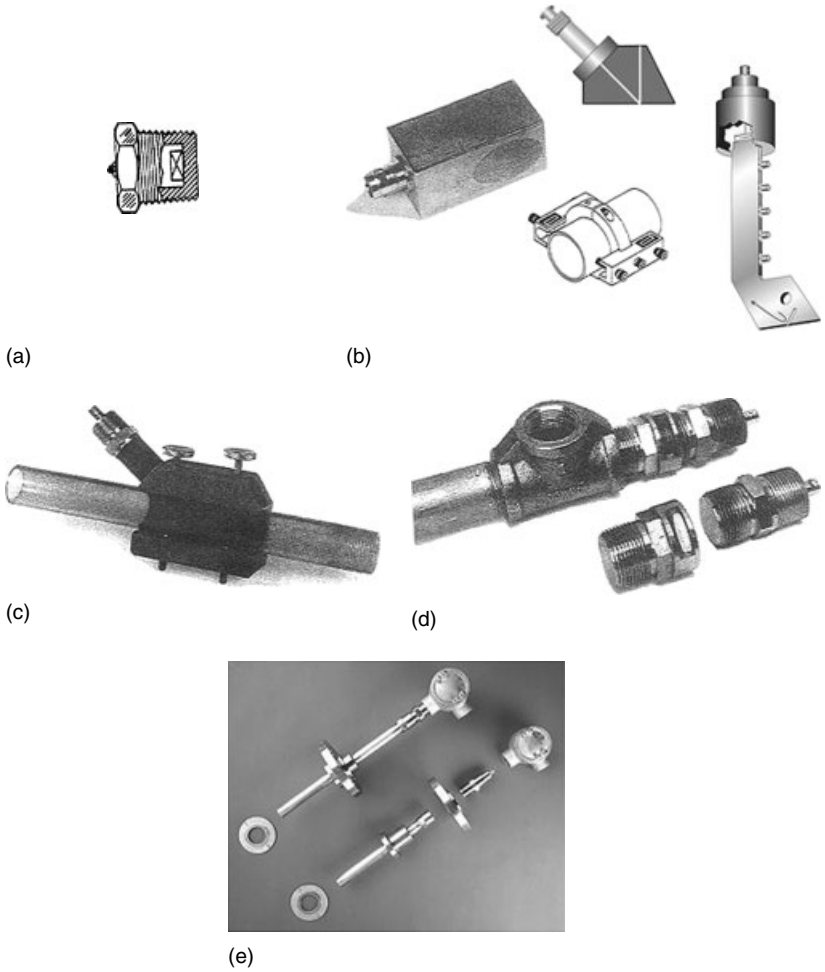


Figure 5. Transducers for liquids: (a) wetted, (b) clamp-on, (c) snap-on, (d) hybrid for ordinary temperature, and (e) hybrid for temperature extremes, using a buffer consisting of a rigid bundle of thin rods within a sealed tube. (Reproduced by permission of Panametrics, Inc.)

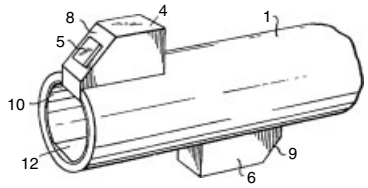
flowmeters in cases in which the fluid does not already contain sufficient scatterers, contrast agents are added. Besides physical contrast agents, investigators sometimes use cavitation bubbles (Takeda *et al.*, 2000). In Lynnworth, Carey and Pedersen (1974), it was suggested that a circumferential external ultrasonic cylindrical transducer, when energized, could generate a hot tag near the axis.

Another way to improve flowmeter performance is to condition the flow to eliminate uncertainty in profile. Flow conditioners (Miller, 1996, Chap. 5; Gallagher, 2002) usually *reduce* cross-flow and swirl, but in principle they could accomplish their purpose of reducing profile uncertainty if they generated a predetermined pattern that includes a known asymmetry and swirl (Hill, 2001). In *small* ducts, flow conditioners have taken the form of static mixers, tube bundles, concentric tubes comprising a Clausius-Mosotti dielectric – constant–based fuel densitometer, or blades (Figure 6).

5 ULTRASONIC TRANSDUCERS FOR NONDESTRUCTIVE TESTING AND EVALUATION OF SOLID MATERIALS

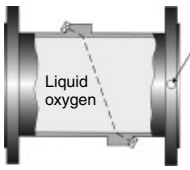
Although the markets and the manufacturers may differ according to one's interpretation of ultrasonic instrumentation (i.e. 'analytic' instrumentation, 'process' instrumentation, 'NDE'(nondestructive evaluation)/NDT instrumentation), one would expect that the physics underlying the wave/measurand interaction would largely determine whether the design of transducers for such applications must differ or could be identical. In practice, many handheld transducers for manual inspection of manufactured parts reflect ergometric design considerations; size, weight, surface contours and finish. Other design considerations include part accessibility, inspectability, transducer wear resistance and longevity. Because of the wide range of materials, geometries, dimensions, microstructures, bonds,

1962



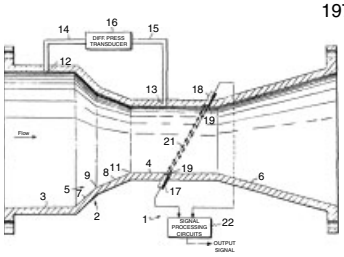
Lake, D.B., *Flowmeters*, U.S. Patent 3,050,997 (Aug. 28, 1962)

1973



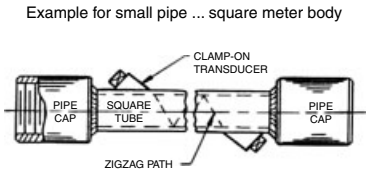
Adapted from Pedersen, N. E., and Lynnworth, L. C., Nonintrusive Dynamic Flowmeter, pp. 178–181 in: *1973 Ultrasonics Symposium Proceedings*, IEEE (1973)

1974



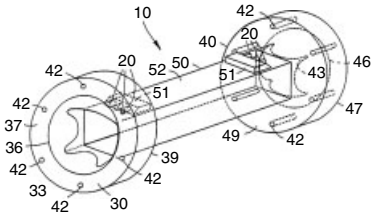
Turtle, Q., *Electroacoustical Flow Metering Apparatus*, U.S. Patent 3,788,140 (Jan. 29, 1974)

1987



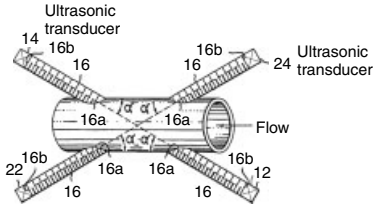
Source: Matson, J. E., Nguyen, T. H. and Lynnworth, L.C., Ultrasonic Measurement of Liquid Flow Using Clamp-On Rayleigh Wave Transducers, 197-206, *IEEE Trans I & M Conference Proceedings* (April 27–29, 1987)

2000



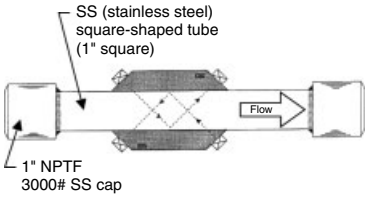
Baumoel; D. S. Baumoel. J., *Pipe Spool Section Having Square or Rectangular Cross-Section for Clamp on Transducer and Method for Flow Measurement*, U.S. Patent 6,026,693 (Feb. 22, 2000)

1977



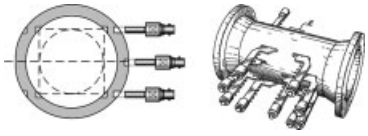
Araki, H., and Yoshihioro, M., *Ultrasonic Flow Meter*, U.S. Patent 4,014,211 (March 29, 1977)

1996



© 1996 Panametrics, Inc.

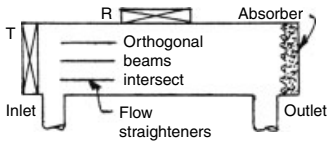
2000



Lynnworth, L. C., Jossinet, G., and Chérifi, E., 300°C Clamp-On Ultrasonic Transducers for Measuring Water Flow and Level, *1996 Ultrasonics Symposium Proceedings*, pp. 407–412, IEEE (1996); Lynnworth, L. C., *Ultrasonic Buffer/Waveguide*, U.S. Patent 6,047,602 (April 11, 2000)

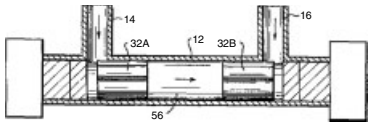
Other flow conditioners

1975



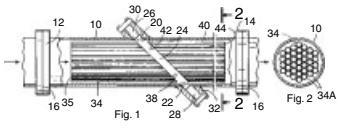
Pedersen, N. E., Lynnworth, L. C., and Bradshaw, J. E., USAAMRDL-TR-75-8, page 178 (June 1975)

1982



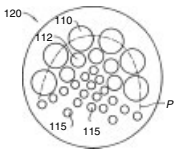
Zacharias, E. M., *Flow Straighteners in Axial Flowmeters*, U.S. Patent 4,365,518 (Dec. 28, 1982)

1985



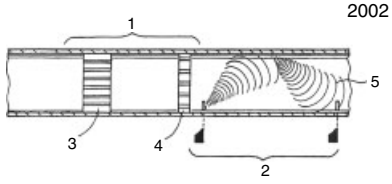
Zacharias, E. M., *Sonic Flow Meter Having Improved Flow Straighteners*, U.S. Patent 4,523,478 (June 18, 1985)

2001



Hill, J. A., *Disturbance Simulating Flow Plate*, U.S. Patent 6,186,179 (Feb. 13, 2001)

2002



Gallagher, J. E., *Method for Determining Flow Velocity in a Channel*, U.S. Patent 6,494,105 (Dec. 17, 2002)

Figure 6. Flowcells and interrogation methods for *liquids*, some of which were later applied or adapted to measuring *gas* flow. Transducers are wetted in some cases and external (not wetted) in others. Flow conditioners were known at least since Turtle (1974); see also, Miller (1996), Chapter 5. A Kenics *static mixer*, not shown, was used as a flow conditioner in 1974 in an R&D (research and development) program and reported by Pedersen, Lynnworth and Bradshaw (1975). (Reproduced by permission of Panametrics, Inc.)

quality levels and a market that has matured over the past half-century, a far greater variety of transducers exists for NDT applications than for any other single process measurand such as flow. Most ultrasonic NDT transducers operate in the 1 to 10-MHz decade. Others are available down to 20 kHz and up to 100 MHz using bondable electroacoustic elements, and up to the gigahertz range using deposited piezoelectrics for acoustic microscopy, see Figure 7.

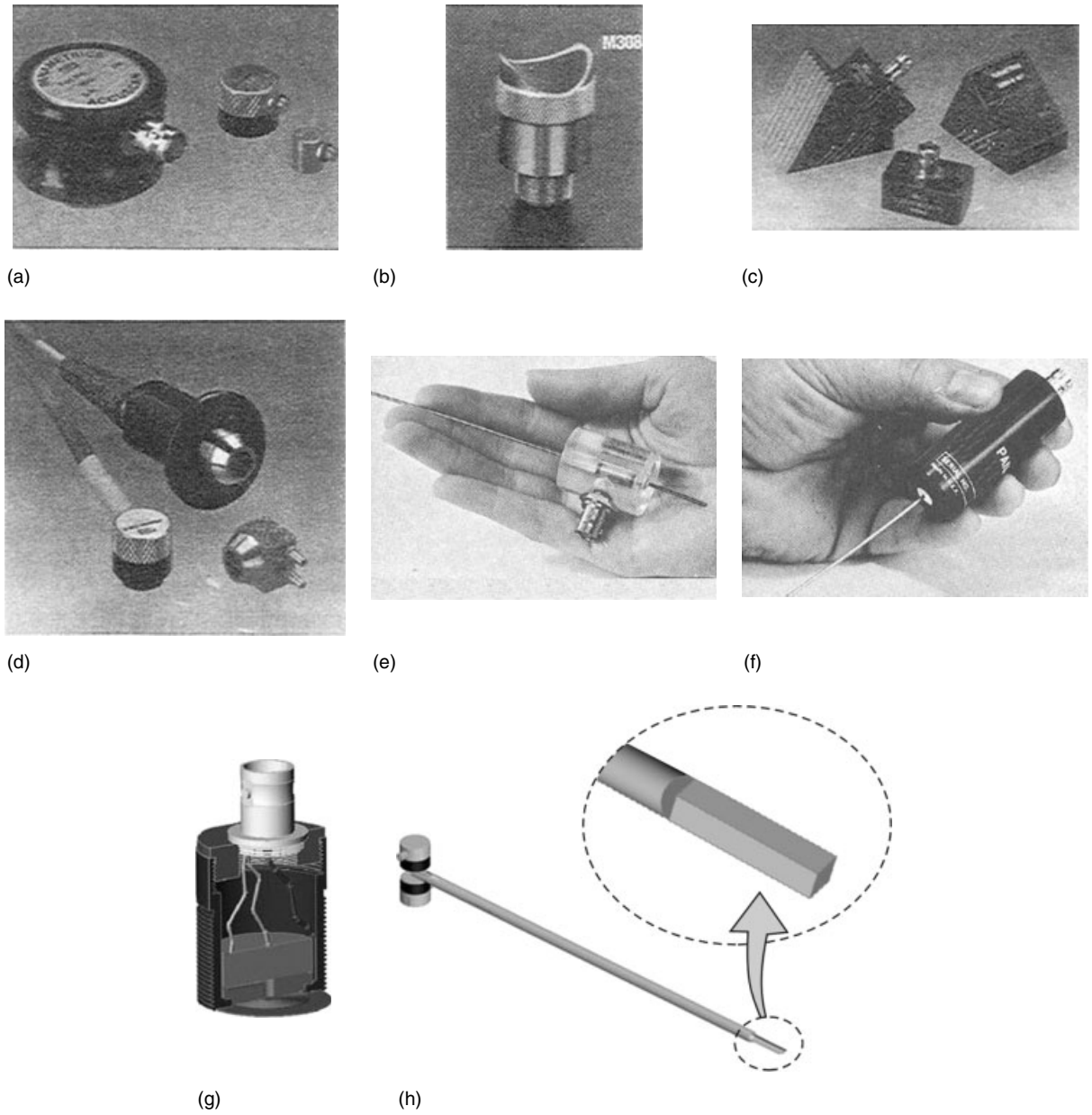


Figure 7. NDT transducers: (a) contact, flat, (b) wetted, focused, (c) angled, (d) dual element for corrosion testing; emat electromagnetic acoustic transducer, (e) slidable magnetostrictive transducer for generating extensional and/or torsional waves, (f) shielded magnetostrictive transducer for generating extensional or torsional waves, (g) emat (electromagnetic acoustic transducer), and (h) shear wave piezo couple used to generate and detect torsional waves by a mode conversion process. The wide range of *sizes and form factors* for 'fixtures' that position NDT transducers is exemplified by comparing (i) a miniature 'clamp-on' transducer that briefly clamps to one's finger; (j) a hand-held transducer in a penlike case versus (k) gantry for robotic ultrasonic scanning system including water squirters for coupling and means to orient transducers normal to an airplane fuselage's contour. Another aspect of *form factor* is the shape or diameter of the specimen. If the specimen is a thin elastic rod, e.g. diameter in the several μm to mm range (l) the methods of Bell (1957) or Fowler (in Lynnworth, Papadakis and Fowler, 1977) may be considered. Post-2000 versions of (l) adapted to moduli measurements and related (derived) characteristics in thin glass fibers are found in Krohn *et al*. (2003). (© 2003 Panametrics, Inc. and reproduced courtesy of GE Panametrics.)

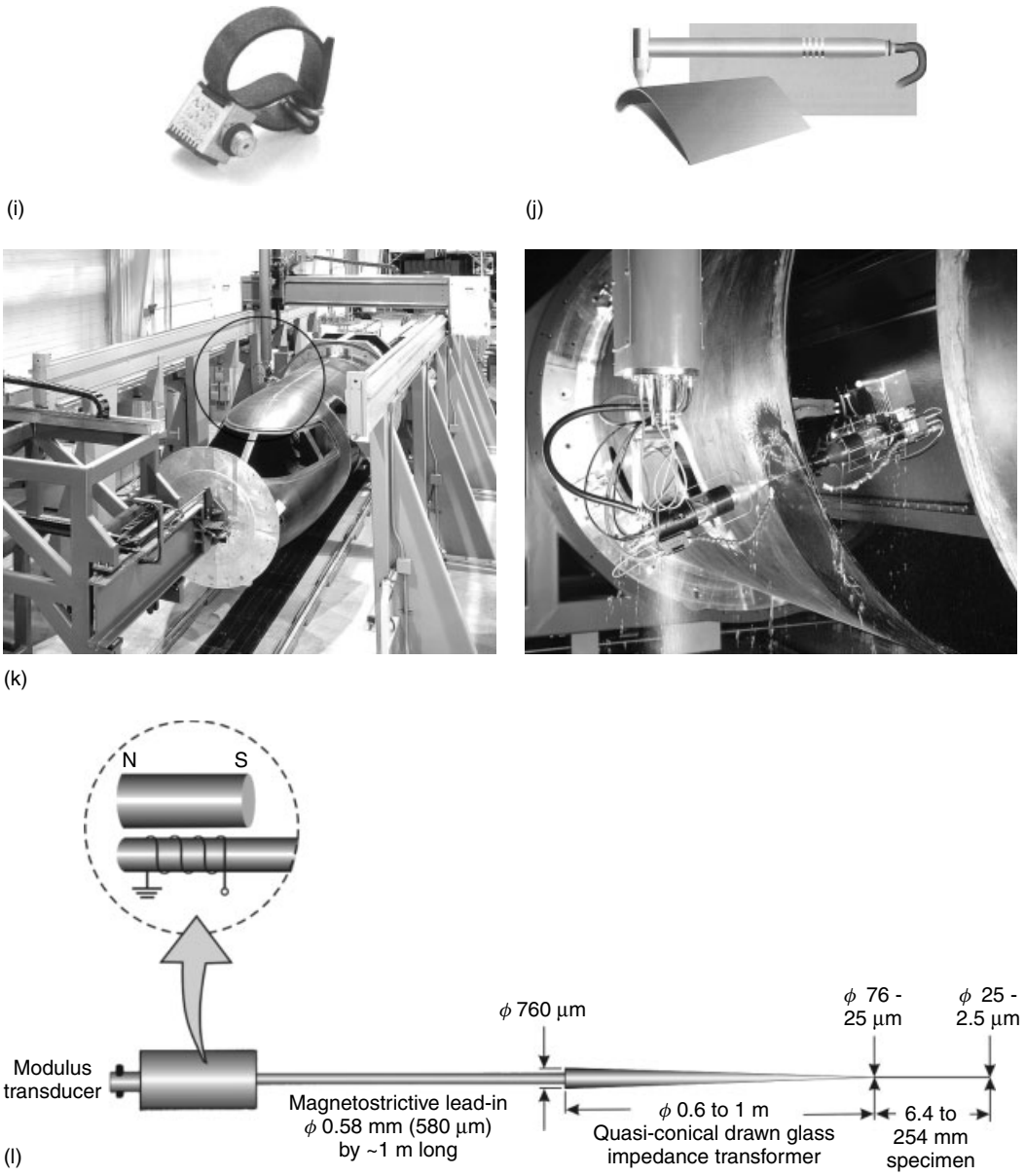


Figure 7. (*Continued*).

REFERENCES

Ao, X. (2003) *Clamp-On Steam/Gas Flow Meter*, U.S. Patent 6,626,049.

Asher, R.C. (1997) *Ultrasonic Sensors for Chemical Process Plant*, Institute of Physics Publishing, London.

Auld, B.A. (1990) *Acoustic Fields and Waves in Solids*, 2nd edn, Krieger Publishing, Malabar Florida.

Baker, R.C. (2000) *Flow Measurement Handbook*, Cambridge University Press.

Bell, J.F.W. (1957) The Velocity of Sound in Metals at High Temperatures. *Philosophical Magazine*, **2**, 1113–1120.

Gallagher, J.E. (2002) *Method for Determining Flow Velocity in a Channel*, U.S. Patent 6,494,105.

Hill, J.A. (2001) *Disturbance Simulating Flow Plate*, U.S. Patent 6,186,179.

Hill, J. and Weber, A. (2002) Qualification of Ultrasonic Flowmeters of Natural Gas Using Atmospheric Air Calibration Facilities, *Proceedings of the 20th North Sea Flow Measurement Workshop*, St. Andrews.

Khuri-Yakub, B.T., Kim, J.H., Chou, C.-H., Parent, P. and Kino, G.S. (1988) A New Design for Air Transducers, *Proceedings of the Ultrasonics Symposium, IEEE* (pp. 503–506).

Lynnworth, L.C. (1989) *Ultrasonic Measurements for Process Control, Theory, Techniques, Applications*, Academic Press.

Lynnworth, L.C. and Mágori, V. (1999) *Industrial Process Control Sensors and Systems*, Chapter 4, in *Ultrasonic Instruments and Devices: Reference for Modern Instrumentation, Techniques, and Technology* (E.P. Papadakis Guest Editor), 23 in the Series *Physical Acoustics*, Academic Press (pp. 275–470).

Lynnworth, L.C., Carey, C.A. and Pedersen, N. (1974) *Nonintrusive (Noninterfering) Ultrasonic Techniques to Measure Gas Mass Flow Rates*, AEDC-TR-74-77, Arnold Engineering Development Center, Tennessee (p. 19).

Lynnworth, L.C., Fowler, K.A. and Patch, D.R. (1981) *Sealed, Matched Piezoelectric Transducer*, U.S. Patent 4,297,607; Lynnworth, L.C., Patch, D.R. and Mellish, W.C. (1984) Impedance-Matched Metallurgically-Sealed Transducers. *IEEE Trans Sonics and Ultras*, **SU-31**(2), 101–104.

Lynnworth, L.C., Papadakis, E.P. and Fowler, K.A. (1977) Ultrasound Propagation Measurements and Applications, in *International Advances in Nondestructive Testing*, Vol. 5 (ed. W.J. McGonnagle), Gordon & Breach (pp. 77–115).

Miller, R.W. (1996) *Flow Measurement Engineering Handbook*, 3rd edn, McGraw-Hill.

Papadakis, E.P. and Fowler, K.A. (1972) Broadband Transducers: Radiation Field and Selected Applications. *Journal of the Acoustical Society of America*, **50**,(3, Part 1), 729–745.

Pedersen, N.E., Lynnworth, L.C. and Bradshaw, J.E. (1975) USAAMRDL-TR-75-8 (p. 178).

Rogers, S.C. (1984) *Tunable Damper for an Acoustic Wave Guide*, U.S. Patent 4,452,334.

Royer, D. and Dieulesaint, E. (2000) *Elastic Waves in Solids I, Free and Guided Propagation*, Springer.

Safari, A., Ebrahimi, M., Toreu, S., Hall, A., Brenan, R. and Hagh, N.M. (2002) Layered Manufacturing for Prototyping of Novel Transducers, *Proceedings of the International Ultrasonic Symposium, IEEE* (pp. 1060–1068).

Takeda, Y., Furuichi, N., Mori, M., Aritomi, M. and Kikura, H. (2000) Development of a New Flow Metering System Using UVP, Preliminary Performance Assessments Using NIST Flow Standards, *Proceedings of the ASME FEDSM 2000*.

Ting, V.C. and Ao, X. (2002) Evaluation of Clamp-On Ultrasonic Gas Flowmeters for Natural Gas Applications, *Proceedings of the 20th North Sea Flow Measurement Workshop*, St. Andrews.

Turtle, Q. (1974) *Electroacoustical Flow Metering Apparatus*, U.S. Patent 3,788,140.

FURTHER READING

Baumoel, D.S. and Baumoel, J. (2000) *Pipe Spool Section Having Square or Rectangular Cross-Section for Clamp On Transducer and Method for Flow Measurement*, U.S. Patent 6,026,693.

Brown, A. and Lynnworth, L.C. (2001) Ultrasonic Flowmeters, Chapter 20, in *Flow Measurement – Practical Guides for Measurement Control*, 2nd edn (ed. D.W. Spitzer), ISA (pp. 515–573).

Cousins, T. and Augenstein, D. (2002) Proving of Multi-Path Liquid Ultrasonic Flowmeters, *Proceedings of the North Sea Flow Measurement Workshop*, St. Andrews, 22–25 October 2002.

Dabirikhah, H. and Turner, C.W. (1994) Leaky Plate Wave Airborne Ultrasonic Transducer. *Electronic Letters*, **30**(18), 1549–1550.

Estrada, H. (2001) *Theory of Ultrasonic Flow Measurement – Gases and Liquids*, Caldon Technical Paper TP58, Presented at ISHM 2001 (International School of Hydrocarbon Measurement), Class 3175.

Guizot, J.-L. (2003) *Ultrasonic Liquid Flow Measurement*, Caldon Technical Paper TP68, Presented at ISHM 2003 (International School of Hydrocarbon Measurement), Class 2405.

Hurd, J.C., Nguyen, T.H. and Lynnworth, L.C. (2004) Shear Wave Polarization Follows Twist of Rectangular Steel Bar. *Materials Evaluation*, **62**(1), 37–42.

Itoh, I., Masanori, H. and Akio, Y. (1996) *Ultrasonic Fluid Vibrating Flowmeter*, U.S. Patent 5,503,035.

Kleppe, J.A. (1989) *Engineering Applications of Acoustics*, Artech House, Boston, MA.

Krohn, M.H., Hellman, J.R., Pantano, C.G., Lower, N. and Brow, R.K. (2003) Effects of Tin on the Physical Properties and Crack Growth in Soda-Lime Silica Float Glass. *Proceedings of the 8th International Symposium on Fracture Mechanics of Ceramics*, Plenum Press, New York [Houston, Texas February 2003].

Lake, D.B. (1962) *Flowmeters*, U.S. Patent 3,050,997.

Lipták, B.G. (1995) Process Control, *Instrument Engineers' Handbook*, 3rd edn, Vol. 2, ISA.

Lynnworth, L.C. (1977) *Ultrasonic Measuring System with Isolation Means*, U.S. Patent 4,004,461.

Lynnworth, L.C. (1992) Ultrasonics in Instrumentation, Chapter 38, in *Handbook of Measurement Science* (eds P.H. Sydenham and R. Thorn), Wiley (pp. 1655–1689).

Lynnworth, L.C. (1994) *Ultrasonic Transducer System with Crosstalk Isolation*, U.S. Patent 5,275,060.

Lynnworth, L.C. (2000) *Ultrasonic Buffer/Waveguide*, U.S. Patent 6,047,602.

Lynnworth, L.C., Jossinet, G. and Chérifi, E. (1996) 300°C Clamp-On Ultrasonic Transducers for Measuring Water Flow and Level, *1996 Ultrasonics Symposium Proceedings, IEEE* (pp. 407–412).

Matson, J.E., Nguyen, T.H. and Lynnworth, L.C. (1987) Ultrasonic Measurement of Liquid Flow Using clamp-On Rayleigh Wave Transducers, *IEEE Transactions I & M Conference Proceedings* (pp. 197–206).

Nguyen, T.H., Smart, C.D. and Lynnworth, L.C. (2004) Transverse Shear to Torsion Mode Converter and Applications. *Materials Evaluation* **62**(6), 690–698.

Oldenziel, D. and Greissmann, M. (1996) *Clamp-On Ultrasonic Volumetric Flowmeter*, U.S. Patent 5,533,408.

Pedersen, N.E. and Lynnworth, L.C. (1973) Nonintrusive Dynamic Flowmeter, *1973 Ultrasonics Symposium Proceedings, IEEE* (pp. 178–181).

Shah, K.C. (1986) *Ultrasonic Temperature Sensor*, U.S. Patent 4,610,551.

Szabo, T.L. (2004) *Diagnostic Ultrasound Imaging – Inside Out*, Academic Press.

Wada, S., Kikura, H., Aritomi, M., Mori, M. and Takeda, Y. (2004) Development of Pulse Ultrasonic Doppler Method for Flow Rate Measurement in Power Plant – Multilines Flow Rate Measurement on Metal Pipe. *Journal of Nuclear Science and Technology* **41**(3), 339–346.

Zacharias Jr., E.M. (1982) *Flow Straighteners in Axial Flowmeters*, U.S. Patent 4,365,518.

Zacharias Jr., E.M. (1985) *Sonic Flow Meter Having Improved Flow Straighteners*, U.S. Patent 4,523,478.

121: Signals in the Presence of Noise

Richard Burdett
Signal Recovery, Wokingham, UK

1 Signal Extraction When Noise is Present 827
2 Basic Methods for Reducing Noise 827
3 Nature of Noise 828
4 White Noise 829
5 Flicker Noise 829
6 Importance of Bandwidth 829
7 Equivalent Noise Bandwidth 829
Related Articles 830
References 830

1 SIGNAL EXTRACTION WHEN NOISE IS PRESENT

A measurement system begins with sensing stages that couple to relevant measurands of the system under study. The power level of the information-bearing signals formed by the sensors is often very low and may be swamped by the unwanted noise signals that are present. Careful attention to sensor and circuit design and assembly, plus use of certain signal processing methods, makes it possible to greatly enhance the original signal-to-noise ratio to usable levels.

This group of articles discusses the various strategies that are available and provides a basis for their use. This is not a widely published topic, the best information probably being available in the literature of companies marketing such products. It is surprising to find that, after well over half a century of signal processing progress, there exist so few full length texts that deal generally with signal recovery and enhancement in general instrumentation applications, see Kester (2002), and the classic work Vainshtein (1985).

It will be found that the techniques, originally conceived for analog methods of signal processing, are often now implemented in digital form, and thus, in many cases provide improved performance at comparable or lower cost. Parts of the systems are, however, necessarily analog and always will be so, since the natural world is not all digital in its structure or form.

2 BASIC METHODS FOR REDUCING NOISE

Recovering or enhancing a signal, or improving a signal-to-noise ratio (SNR) simply means reducing the noise accompanying a signal.

There are two basic ways of doing this:

1. Bandwidth reduction, where the noise is reduced by reducing the system noise bandwidth (B_n). This approach works well if the frequency spectra of the noise and signal do not overlap significantly, so that reducing the noise bandwidth does not affect the signal. With random white noise, the output noise is proportional to $\sqrt{B_n}$. With nonwhite noise, other relationships will apply.
2. Averaging or integrating techniques, where successive samples of the signal are synchronized and added together. The signal will grow as the number (n) of added samples; with random white noise, the noise will grow as $\sqrt{n}$. This is only the case, if the signal characteristics are stationary for the duration of the extraction process.

Sometimes it is useful to combine both techniques. In many applications there is significant overlap between the

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

signal and noise spectra and improving a signal-to-noise ratio must be done at the expense of the response time or measurement time (T); with random white noise interference, the output signal-to-noise ratio is proportional to $\sqrt{T}$. The bandwidth reduction technique is best looked at from a frequency-domain point of view; signal averaging and correlation techniques lend themselves to time-domain analysis.

In this set of articles, – **Part 12, Elements: F – Signals and Noise, Volume 3** – mathematics, and theoretical considerations are kept to a minimum and the models presented are basic. Practical best practice requires considerably more processing to precondition signals than is shown here. For further simplicity, it is assumed that all noise processes are stationary and that both signal and noise are ergodic, analog variables; we will not concern ourselves here with digital signals or discrete-time (sampled) signals except where such signals are involved in the enhancement techniques. They are essential in modern application methods but it is the basic ideas that drive the digital methods.

In addition, only signal recovery techniques will be considered. Further processing, such as least-squares polynomial smoothing of a waveform or Fourier transformation to obtain a frequency spectrum, are not considered here.

Discussion is started by reviewing some basic concepts, then moving on to ways of avoiding adding noise (e.g. hum pickup and preamplifier noise) and finally covering the instrumentation techniques used to reduce the remaining noise content. Finally, special considerations involved in recovering pulse signals from photon (light), ion, or electron beams are covered in this section **Part 12, Elements: F – Signals and Noise, Volume 3**.

3 NATURE OF NOISE

Noise is an undesired signal. All systems have it present to some degree. It usually becomes of interest when it obscures a desired signal. Figure 1 shows the power spectral density (power/unit bandwidth) of the most commonly encountered types of noise.

Deterministic noise can range from simple discrete-frequency components such as power-line hum at harmonics of 50 or 60 Hz, to radio frequency interference, RFI, caused by narrow, high-energy pulses from power-line switching spikes, pulsed lasers, radar transmitters, and the like.

Stochastic or random noise is found in most systems both as white noise, where the power spectral density is independent of frequency, and also as $1/f$ or flicker noise, where the power spectral density decreases as frequency increases. Power spectral density is usually measured in mean-squared-volts/Hz or mean-squared-amperes/Hz; for noise, such specifications are usually referred to as *spot*

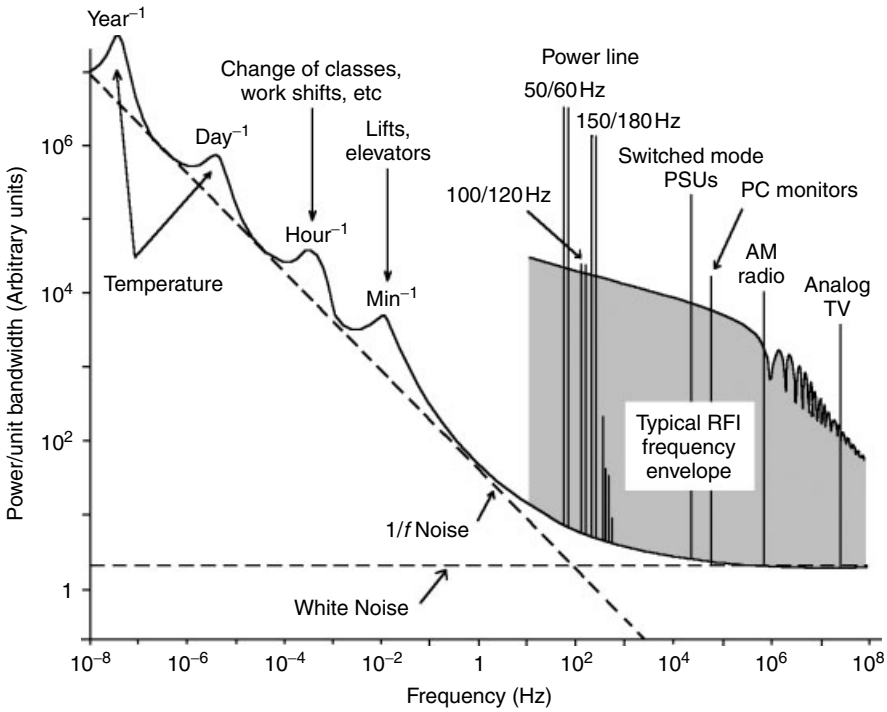


Figure 1. Environmental noise. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

noise data and usually are a function of frequency. Notice that for an rms voltage of v (volts) and a frequency range of Δf (Hz), the power spectral density, S, is given by

$$S = \frac{v^2}{\Delta f} = \left(\frac{v}{\sqrt{(\Delta f)}}\right)^2 \tag{1}$$

The quantity $v/\sqrt{(\Delta f)}$ is usually referred to as *voltage spectral density* and is measured in rms volts/$\sqrt{}$Hz (volts per root hertz). Similarly, we can refer to current spectral density specifications in units of rms amperes/$\sqrt{}$Hz.

4 WHITE NOISE

White noise is usually found in one of two forms: Johnson noise and shot noise. Johnson, or thermal, noise is caused by random motion of thermally agitated electrons in resistive materials, and the mean-square noise voltage is given by

$$v_n^2 = 4kTR\Delta f \tag{2}$$

where k is Boltzmann's constant ($1.381 \times 10^{-23}\,\mathrm{JK}^{-1}$),

T is the absolute temperature (kelvin) and
R is the resistance (ohm).

Alternatively, from Ohm's law, the mean-square noise current is given by

$$i_n^2 = \left(\frac{v_n}{R}\right)^2 = \frac{4kT\Delta f}{R} \tag{3}$$

Shot noise is caused by the random arrival of electrons – *see* **Article 184, Photon Counting, Volume 3** – at, for example, the electrodes of electron tubes or transistor junctions. A DC current, I, will have a noise-current component, i_n, given by

$$i_n^2 = 2AeI\Delta f \tag{4a}$$

where e is the charge of one electron ($\approx 1.6 \times 10^{-19}\,\mathrm{C}$), A is the mean gain experienced by each electron and I is in amperes. In many cases, $A = 1$, so that

$$i_n^2 = 2eI\Delta f \tag{4b}$$

5 FLICKER NOISE

Flicker noise has many different origins and is not clearly understood but exhibits a $1/f^n$ power spectrum with n usually in the range of 0.9 to 1.35. Note that DC drift is a very low frequency form of flicker noise.

What this model represents is that it is increasingly harder to remove noise as the frequency is lowered below 1 Hz or so. This makes practical sense when it is considered that there is a link between noise generation and temperature cycling of mechanical systems. It becomes harder to integrate temperature cycling noise as the cycle time lengthens.

6 IMPORTANCE OF BANDWIDTH

What do we mean by bandwidth? In the simple low-pass filter circuit shown in Figure 2, for example, we usually and somewhat arbitrarily define the signal bandwidth (Figure 3) to be the cutoff frequency, f_c, where $v_o/v_i = 70.7\%$ ($-3\,\mathrm{dB}$) or $v_o^2/v_i^2 = 50\%$ (the half-power point).

7 EQUIVALENT NOISE BANDWIDTH

Notice that frequencies above f_c will obviously pass (although attenuated) through the filter, and therefore are not really cut off. For noise, it is convenient to think in terms of an equivalent noise bandwidth, B_n, defined by the relationship

$$B_n = \frac{1}{G^2}\int_0^\infty |H(j\omega)|^2\,\mathrm{d}f \tag{5}$$

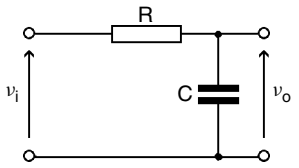


Figure 2. Low-pass filter circuit. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

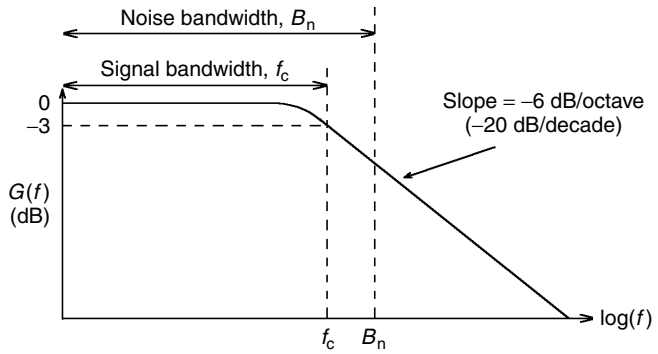


Figure 3. Low-pass filter transfer characteristic. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

where $H(j\omega)$ is the frequency response function of the system and G is a gain parameter suitably chosen to be a measure of the response of the system to some parameter of the signal: for low-pass systems (e.g. Figures 2 and 3), G is usually taken to be the zero-frequency or DC gain. For band-pass responses, G is usually made equal to the maximum gain.

Using the above definition, and taking G to be the zero-frequency gain (i.e. unity), we can readily calculate that for the simple RC filter shown in Figure 2 that

$$B_n = \frac{1}{4}RC \text{ Hz} \tag{6}$$

Noise, of the stochastic form, has been reviewed in relation to instrument systems in a classic paper, Fellgett and Usher (1980).

RELATED ARTICLES

Article 125, Outline of Purpose of Analog Data Filters, Volume 2; **Article 176, Signals and Signal-to-noise Ratio, Volume 3**; **Article 178, Noise Matching and Preamplifier Selection, Volume 3**.

REFERENCES

Fellgett, P.B. and Usher, M.J. (1980) Fluctuation Phenomena in Instrument Science, *Journal of Physics E: Scientific and Instrumentation*, **13**, 104–106.

Kester, W. (2002) *Mixed-signal and DSP Design Techniques*, Engineering Staff of Analog Devices Inc., Newnes, London.

Vainshtein, L.A. (1985) *Extraction of Signals from Noise*, reprinted from Dover Publications, Wokingham, UK, 1970.

This article is based in part on a chapter by D.M. Munro originally published in Handbook of Measurement Science Vol 1, P.H. Sydenham, Wiley, 1982.

122: Operational Amplifiers

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Introduction to Operational Amplifiers (Op-amps) 831
2 Classification of Op-amps 832
3 Types of Semiconductor Amplifier Units used in Op-amps 832
4 Open Loop Voltage Transfer Function 832
5 Op-amp Model 833
6 Inverting Circuits 833
7 Noninverting Circuits 834
8 Frequency Domain Factors and Behavior 836
9 Offset Voltage, Bias, and Noise in Op-amps 837
10 Equivalent Circuit for Real Op-amps 838
Related Articles 839
Further Reading 839

1 INTRODUCTION TO OPERATIONAL AMPLIFIERS (OP-AMPS)

The voltage levels available from passive sensor conditioners are usually too small to be practically useful; it is necessary to amplify the signal. Since the development of the triode valve by Lee de Forest early in the twentieth century, many other significant developments in amplification have been achieved. Notable milestones were the appearance of the point contact transistor, the bipolar junction transistor (BJT), the field-effect transistor (FET) or junction field-effect transistor (JFET) and the metal oxide semiconductor (MOSFET).

From these developments in semiconductor technology, rapid evolution of solid-state amplifiers has resulted in the integrated circuit operational amplifier, abbreviated to IC op-amp or simply op-amp.

It should also be noted that when the expression op-amp is now used, it usually refers to the IC unit as purchased from a manufacturer. Whenever the word *amplifier* is used, it implies a circuit that employs an op-amp and some other components to constitute the overall amplifier. Amplifiers and the op-amps included in them, which are the basic building blocks of almost all data handling measurement systems, have now reached a very high degree of sophistication. The detailed internal design of op-amps will not be considered here; it is very specialized.

Op-amps are directly coupled amplifiers having, for reasons that will be discussed later, a high gain of the order of $>10^4$. In their own right, they are highly reliable electronic systems. In addition, they bring the benefits of compactness, versatility, and ease of use, which make them essential electronic building blocks. This means they can make a substantial contribution to increasing measuring systems sensitivity while exhibiting versatility in the design of the frequency characteristics of the systems in which they are employed.

In many such industrial systems, the measuring points are fairly remote from the data processing location. Op-amps make a considerable contribution to the effective transmission of these remotely acquired signals. They consume comparatively little power. Consequently, it can be concluded that the versatility brought to measuring systems by electronic means is largely due to the impact of the op-amp.

The continuing steadily improving performance of op-amps since their first introduction around 1960 has been accompanied by significant reduction in their cost. Consequently, op-amps now give measuring system designers a

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

versatile building block component for inclusion in numerous applications to realize many diverse and complex functions. The versatility of op-amps lies in their application using associated feedback networks. Thus, op-amps can be configured to act as summers, subtracters, and difference amplifiers. They can also perform the functions of differentiation, integration, and of other types of filters, or as comparators. This concept has eventually resulted in the universal application of the op-amp for a variety of uses, as is obvious from its universality in contemporary measuring systems. As there is a wide diversity of available types of IC, the task of choosing a suitable op-amp for a particular application is not difficult but requires some knowledge of their nature, scope, characteristics, and formal modeling.

2 CLASSIFICATION OF OP-AMPS

No single op-amp type meets the needs of every application. All practically manufactured forms constitute a compromise among the various aspects and features required. Hence, the design of op-amps is an optimizing process. To give a good overall view, the three main op-amp types will be considered by describing their principal features.

Originally, op-amps were designed for the application of external voltage feedback. This keeps the inverting input current small by maintaining high input impedance as well as by keeping the input difference voltage small. Contemporary designs make the inverting input respond to current using current feedback in such a way that the output voltage is proportional to the current. Major benefits of current feedback are the much larger slew rate and the much decreased influence of stray input capacitance on op-amp bandwidth.

3 TYPES OF SEMICONDUCTOR AMPLIFIER UNITS USED IN OP-AMPS

Bipolar transistor differential op-amp: Op-amps using bipolar technology are probably the most widely applied general-purpose units. They are used as differential amplifiers, which are described in **Article 123, Instrumentation Amplifiers, Volume 2**, in low noise applications with low impedance sources and in low performance, and in noncritical computational applications.

They are popular because of their low cost, modularity, packaging, and wide temperature range. Input bias current, which can be in the range 1 nA to 1 μA at 25 °C, has a typical drift in excess of 0.2 nA with offset voltage drift in the range 0.25 to 100 μV/ °C. Open loop gain is in the range from 10^4 to 3×10^6. (Note that these levels of gain are rarely used in open loop operation – see later.)

Field-effect transistor, or FET, differential op-amp: When it is important to minimize input bias current, it is usual to design an op-amp using a complementary pair of input FETs. This type of input also increases the input impedance of the op-amp. Unfortunately, because the bias current essentially depends on the reverse saturation current of the gate-to-source junction, its bias current drift is much larger than for a bipolar input unit. Their potential for high slew rate and fast settling time is accompanied by capacitive pickup of noise signals due to the high input impedance. Some cases are also known where mechanical vibration causes problems if they introduce movement into the wiring. They find use in general-purpose applications, in differential amplifiers, low noise amplifiers from high impedance sources and sample-and-hold amplifiers and integrators, because of their low bias current.

FET op-amps may use either junction FETs or MOSFETs. They require a bias current in the range of 0.01 to 100 pA at 25 °C, which doubles with every 10 °C increase in temperature. Their drift voltage is around 2 to 100 μV/ °C. Open loop gain is 10^4 to 5×10^5. Because their differential input resistance is very high, at around 10^{11} to 10^{12} Ω, they are particularly suitable for creating noninverting amplifiers. MOSFET types require special input circuitry to protect them against damage by electrostatic charges.

Other measuring system amplifiers: Many amplifiers can be constructed by combining the basic op-amp unit with other components. Those considered here are the instrumentation amplifiers in **Article 123, Instrumentation Amplifiers, Volume 2** and the chopper, chopper stabilized/compensated and auto-zero amplifiers, isolation amplifiers, and the charge amplifier discussed in **Article 124, Specialized Amplifiers for Measurement Systems, Volume 2**.

4 OPEN LOOP VOLTAGE TRANSFER FUNCTION

As shown in Figure 1, the voltage transfer relation of an op-amp is essentially nonlinear especially when v_o is close to either of the supply rail voltages. In normal operation, the op-amp is restricted to the linear range of the voltage transfer relation. Note that when $v_{Ig} = v_{Ng}$ then $v_o = 0$. This corresponds to the case of infinite common mode rejection, or CMR. A figure of merit for an op-amp, which specifies its ability to reject common mode signals, is called the *common mode rejection ratio*, or CMRR.

An op-amp will have an infinite CMRR if there is symmetry of all of its parameters. It is never possible to obtain infinite CMRR because production mismatches

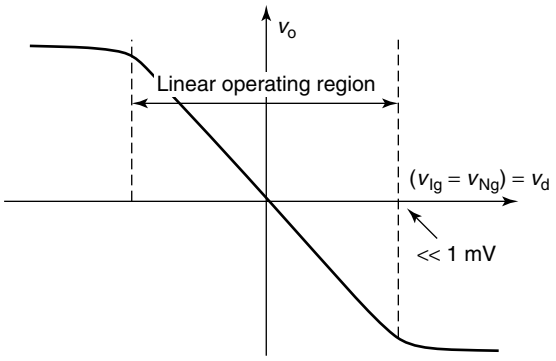


Figure 1. Voltage transfer relationship for op-amp.

always produce asymmetries. This means that the open loop gains of the op-amp to inverting and noninverting inputs are always different. The open loop gain of an op-amp is typically $>2 \times 10^5$, which is sufficiently high for the majority of practical applications.

5 OP-AMP MODEL

Op-amps are usually modeled for their signal handling properties. The simple block diagram and circuit model of an ideal op-amp, which are shown in Figure 2, visualize the op-amp as a voltage controlled voltage source.

Although op-amps require the supply of electrical energy, their connections to the power supply are rarely shown in circuit diagrams.

Various degrees of model complexity exist. Input terminals, referred to as the inverting, I, and noninverting, N, inputs, drive the amplifier as a difference signal to give an output voltage, which is directly proportional to this difference signal. The ideal amplifier has infinite input or output impedance, has an infinite bandwidth of operation without gain change and a gain that is very high, but not quite infinite. This is different from the definition given in many other books where the gain is assumed to be infinite.

Complete symmetry also characterizes the amplifier so there is no input circuit mismatch or gain asymmetry in the signal propagation from I and N to the output.

6 INVERTING CIRCUITS

Basic inverting amplifier with finite gain op-amp: In most practical situations, op-amps are connected with feedback to form a closed loop configuration. One of the two basic closed loop circuits is the single input inverter shown in its general form in Figure 3(a) and for a closed loop gain of 1 in Figure 3(b). A stage gain of $(-R_f/R_1)$ is expected if the op-amp gain approaches infinity. The 5-kΩ resistor in Figure 3(b) is to maintain symmetry for bias purposes. More will be said about this in Section 8.

Assume that the op-amp is ideal in the sense defined before so that $v_o = -A_0 v_{IN}$. Define the frequency independent feedback factor as $\beta_I = R_1/R_1 + R_f$. Also, since the input current $i = 0$ the closed loop gain, A_{CL}, can be shown to be

$$v_o = A_{CL} v_{in} = -\frac{R_f}{R_1} \times \frac{\beta_1 A_0}{1 + \beta_1 A_0} v_{in}$$

$$= -G_0 \frac{\dfrac{1}{1 + G_0} A_0}{1 + \dfrac{1}{1 + G_0} A_0} v_{in} \tag{1}$$

In the case when $\beta_1 A_0 \gg 1$, it can be seen that the closed loop gain has the value $A_{CL} \approx -R_f/R_1 = -G_0$, where G_0 is the low frequency stage gain assuming the op-amp has an infinite open loop gain.

In other words, the closed loop gain at low frequencies is determined purely by the ratio of two resistors. This is an important finding for it means imperfections in manufacture, and the gain actually used, have little effect on the closed loop gain in this feedback configuration. It also shows why basic op-amp units are made with such high values of gain.

Since it is assumed that the op-amp input resistance is infinite, the stage has an equivalent input resistance given by

$$R_{in} = R_1 + \frac{R_f}{1 + A_0} \tag{2}$$

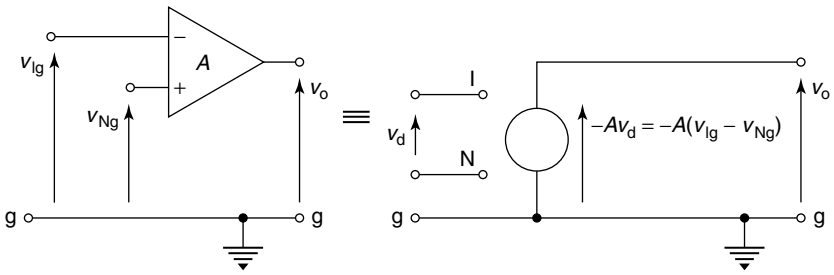


Figure 2. Block diagram and circuit model of an ideal IC op-amp.

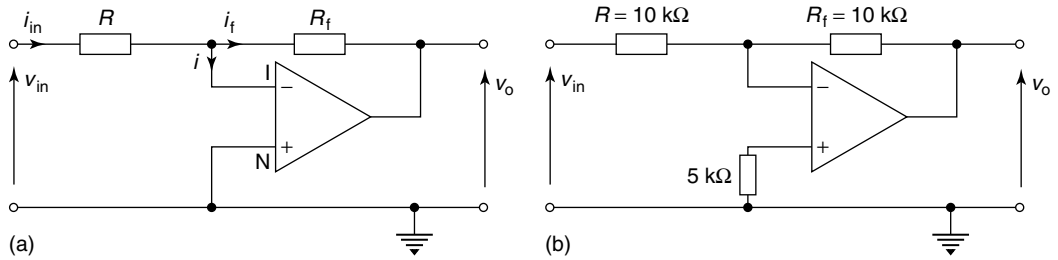


Figure 3. Inverting amplifier.

The term, $R_f/(1 + A_0)$, in (2) is the effect of R_f across the terminals IN of the op-amp in Figure 3 when the Miller effect of R_f is taken into account.

It is obvious that there will be an error in the calculation of the gain if it is assumed that the open loop gain of the op-amp is infinite. To express this error, (1) may be rewritten as

$$v_o = A_{CL}v_{in} = -G_0(1 - \varepsilon_G)v_{in}$$
$$= -G_0\left[1 - \left(1 + \frac{1}{1 + G_0}A_0\right)^{-1}\right]v_{in} \qquad (3)$$

where $\varepsilon_G = (1 + 1/(1 + G_0)A_0)^{-1}$ is the relative error introduced by assuming that the closed loop gain is $G_0 = R_f/R_1$. Usually, but not always so, the variation in gain not being infinite is insignificant.

Summing amplifier: Several inputs may be connected to the summing point as shown in Figure 4.

Assume an infinite gain op-amp so that the inverting input node, I, can be considered as a virtual earth corresponding to $v_i \approx 0$. Assume infinite op-amp input resistance so that i_{IN} may be taken as zero to give

$$i_1 + i_2 + i_3 = -i_f \qquad (4)$$

then

$$\left(\frac{v_1}{R_1}\right) + \left(\frac{v_2}{R_2}\right) + \left(\frac{v_3}{R_3}\right) = \frac{-v_o}{R_f} \qquad (5)$$

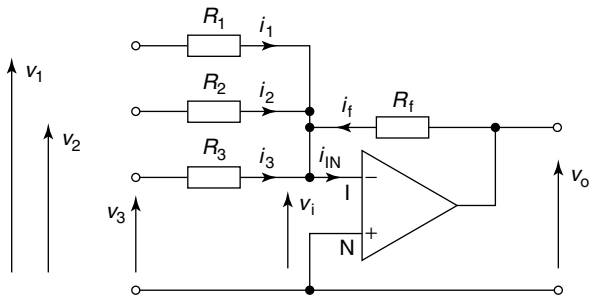


Figure 4. Summing amplifier.

so that

$$v_o = -\left(\frac{R_f}{R_1}\right)v_1 - \left(\frac{R_f}{R_2}\right)v_2 - \left(\frac{R_f}{R_3}\right)v_3 \qquad (6)$$

Hence, also, the channels are noninteracting if the open loop gain is infinite.

Effect of op-amp input resistance: The value of the differential open loop input resistance, R_{indd}, of an op-amp is typically >300 kΩ and is usually neglected.

7 NONINVERTING CIRCUITS

Basic noninverting amplifier: Circuits for noninverting amplifiers are shown in Figure 5. In these circuits the input signal is applied to the noninverting input terminal, N, of the op-amp, which is assumed to have a frequency independent gain of $-A_0$. Taking a voltage sum around the external nodes of the circuit for the noninverting amplifier in Figure 5(a) gives

$$v_{Ig} = v_{IN} + v_{in} = v_{IN} + v_{Ng} \qquad (7)$$

Assume infinite op-amp input resistance so that i_{IN} may be taken as zero. Convert the loop with v_o, R_f, and R_1 into an equivalent voltage source looking left from the terminal I. Also use the frequency independent feedback factor, defined above for (1), to obtain

$$v_{Ig} = \beta_I v_o = \frac{R_1}{R_1 + R_f}v_o = v_{IG} + v_{Ng}$$
$$= -\frac{v_o}{A_0} + v_{in} \quad \text{or} \quad v_o = \frac{A_0}{1 + \beta_I A_0}v_{in} \qquad (8)$$

In the case when $\beta_I A_0 \gg 1$, (5) can be simplified to give

$$v_o = \left(1 + \frac{R_f}{R_1}\right)v_{in} \qquad (9)$$

In the same way as for the inverter, there will be an error in the calculation of the noninverting gain when the

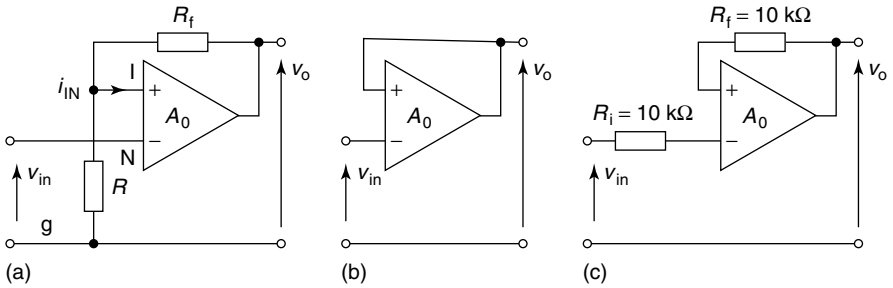


Figure 5. Noninverting amplifiers (a) with gain (b) as a follower and (c) with bias compensation.

open loop gain of the op-amp is assumed to be infinite. To calculate the error, (8) can be rewritten in a similar way as (3) to obtain the expression for the relative error.

Example 1. Deduce an expression for the effective input resistance of the noninverting amplifier whose op-amp has a low frequency open loop gain of A_0 and a differential input resistance of $R_{indd} = 300\,k\Omega$. If $A_0 = 10^6$, $R_f = 100\,k\Omega$ and $R_1 = 10\,k\Omega$, calculate the stage input resistance.

Solution: Represent the loop with the output voltage, v_o, and the resistors R_f and R_1 as an equivalent voltage source. Figure 5(a) can then be redrawn as in Figure 6.

The effective stage input resistance can be deduced immediately as

$$R_{in} = \frac{v_{in}}{i_{IN}} = R_{indd}\beta_I A_0 + \frac{R_1}{R_f} \quad \text{and} \quad \frac{R_1}{R_f} \ll R_{indd}\beta_I A_0$$
$$\approx R_{indd}\left[\frac{A_0}{1+G_0}\right] \text{ with } G_0 = \frac{R_f}{R_1} \tag{10}$$

Completing the calculations gives

$$R_{in} = R_{indd}\left[\frac{A_0}{1+G_0}\right] = \frac{300 \times 10^3 \times 10^6}{1+10} = 27.3\,G\Omega$$

Hence, noninverting amplifiers appear as near open circuits to signals.

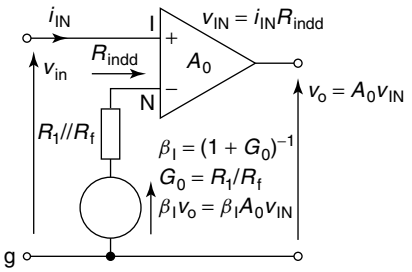


Figure 6. Circuit for evaluating the input resistance of a noninverting amplifier.

Basic voltage follower: The basic voltage follower, shown in two versions in Figure 5(b) and (c), is a special case of the noninverting amplifier in which $R_f = 0$. This circuit is used as a buffer amplifier, allowing a transformation from a high source impedance to a low source impedance. Specific questions of interest are the value of the input resistance and the value of the output resistance. For finite op-amp gain at low frequencies, it is easily shown that

$$v_o = \frac{A_o}{1+A_o} v_{in} \tag{11}$$

The values of R_1 and R_f are not critical. They may be chosen to suit the biasing conditions.

Example 2. An op-amp has an open loop gain of $A_0 = 10^5$. Evaluate the error in the closed loop gain, A_{CL}, of an inverting stage if the feedback resistance is $R_f = 100\,k\Omega$ and the lead-in resistance is $R_1 = 10\,k\Omega$. Also, calculate the input resistance of the stage.

Comment upon the error in calculating the gain in the case of a noninverting amplifier with the same parameter values

Solution: Substituting into (3) gives the gain error as

$$\varepsilon_G = \left(1 + \frac{1}{1+G_0}A_0\right)^{-1} = \left(1 + \frac{1}{11} \times 10^5\right)^{-1}$$
$$\approx \left(\frac{11}{10^5}\right) \approx 110 \text{ parts in 1 million or } 0.011\%$$

The input resistance of the stage is calculated from (4) as

$$R_{in} = R_1 + \frac{R_f}{1+A_0} = 10 + \frac{100}{1+10^5}$$
$$= (10 + 0.001)\,k\Omega \approx R_1$$

The error in calculating the gain of the noninverting stage with the same values is equal to the error for the inverting case.

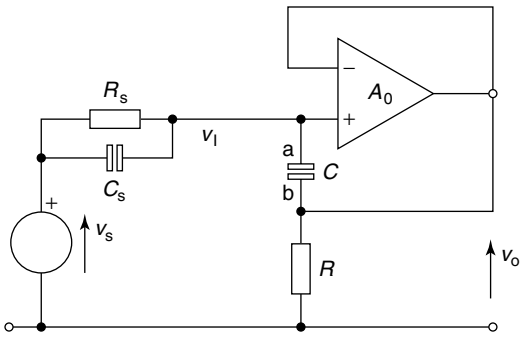


Figure 7. Potential equalization.

Potential equalization: A useful application of the voltage follower is for two nodes of a network to be at the same potential without any physical connection. This can occur on a printed circuit card where it is difficult to isolate items for testing purposes. The circuit points a and b in Figure 7 are at the same potential so there is no current flowing in C. Hence, the source is an open circuit. However, the amplifier drives a current through R. Hence, the source may exhibit an open circuit frequency response.

8 FREQUENCY DOMAIN FACTORS AND BEHAVIOR

Frequency domain parameters of op-amps: Typical values of these parameters can vary widely depending upon op-amp price and application. For example, very low-cost units will have

Unity gain bandwidth = 1 MHz,
Low frequency open loop gain = 2×10^5,
Roll-off rate = −6 dB/octave,
Corner frequency = 5 Hz.

The small signal bandwidth for any closed loop gain $A_{CL} \geq 10$ is given by

$$\text{Bandwidth} \approx \frac{10^6}{A_{CL}}\text{Hz} \tag{12}$$

For ultralow noise and ultralow distortion types these parameters are

Unity gain bandwidth = 100 MHz,
Low frequency open loop gain = 20×10^6,
Roll-off rate = −6 dB/octave,
Corner frequency = 5 Hz.

The small signal bandwidth for any closed loop gain $A_{CL} \geq 10$ is given by

$$\text{Bandwidth} \approx \frac{10^7}{A_{CL}}\text{Hz} \tag{13}$$

Slew rate and full power frequency response: The slew rate of an operational amplifier is the maximum rate of change of output voltage at the rated output.

The slew rate is normally specified for the unity gain follower configuration, as this is often the worst case.

If frequency compensation is optimized, higher closed loop gain often yields higher slew rates. The slew rates for positive and negative swing of the output need not always be the same. For any desired output swing, the slew rate can set the limit for the frequency response obtainable for that particular swing.

An important frequency domain parameter is the full power response frequency, f_p. This is the maximum frequency at which rated output can be obtained without significant distortion. It can be expressed in terms of the slew rate of the amplifier as

$$f_p = \frac{S_r}{2\pi E_o} \tag{14}$$

where S_r is the slew rate and E_o is the rated output voltage.

This relationship is obtained by equating the slew rate to the maximum slope of a sinusoidal signal of peak value, E_o, and frequency, f_p. It indicates that the effective frequency response of an amplifier for a specific output swing is often more limited by the slew rate of the amplifier than by the small signal frequency response predicted by the open loop gain characteristics.

Inverter with frequency dependent gain: Consider the inverting amplifier in Figure 3 with generalized frequency dependent impedances Z_f and Z_1, replacing R_f and R_1 respectively, and an op-amp with an open loop gain $\overline{A_{OL}} = -A_0/[1 + j(\omega/\omega_c)]$. Assume that the op-amp is ideal in every other respect. Define the frequency dependent feedback factor, $\overline{\beta_I} = Z_1/Z_1 + Z_f$. Conduct an analysis similar to that for the basic inverter. To take account of the frequency dependence of all of the circuit parameters, this is done in the frequency domain to obtain the voltage frequency response function as

$$\frac{V_o(j\omega)}{V_{in}(j\omega)} = A_{CL} = -\frac{Z_f}{Z_1}\frac{\overline{\beta_I} \times \overline{A_{OL}}}{1 + \overline{\beta_I} \times \overline{A_{OL}}} \tag{15}$$

Equation (15) provides a general result for frequency domain behavior. Integrators, differentiators, and active filters, which can be analyzed using (13), are described in Section 8, of **Article 124, Specialized Amplifiers for Measurement Systems, Volume 2**.

9 OFFSET VOLTAGE, BIAS, AND NOISE IN OP-AMPS

Asymmetrical fabrication mismatches in the input of op-amps give rise to two undesirable effects. The first is input voltage offset and the second is asymmetry of bias currents. Even when the input voltage of an op-amp is zero, the output is always nonzero.

Input offset voltage is the value of an equivalent input voltage, which would give this output voltage.

The input bias currents are those currents necessary to bias the input transistors of the op-amp for operation. These effects are modeled as shown in Figure 8(a). The offset voltage, v_{os}, and offset current, i_{os}, consist of a nominal component plus components due to its sensitivity to changes in temperature, time, and power supply voltage. It is possible to design the circuit so that the nominal and the temperature components are completely removed.

Voltage offset: Even with the highly sophisticated techniques used in modern IC fabrication, it is virtually impossible to fabricate op-amps without input stage asymmetries. Values of the input offset voltage, v_{os}, can be as high as 2 to 7 mV in low-cost units to as low as 50 to 100 μV in ultralow distortion and ultralow noise units. The polarity of the input offset voltage may differ between units. In practice, the output offset, which is determined by the product of gain and input offset, can be corrected by appropriate offset correction circuits.

The output offset can be calculated in a specific case provided both the input offset voltage, v_{os}, its sensitivity to variations in temperature, power supply voltage and time, as well as the gain of the stage, are known.

Analysis proceeds by treating the input offset voltage as an input at the noninverting terminal of the stage, which is looked upon as a noninverting amplifier like that in Figure 5(a).

Average temperature coefficient of voltage offset: The average temperature coefficient of offset, $\Delta v_{os}/\Delta T$, is typically about 1 to 6 μV/°C. Using this coefficient, the output offset voltage changes can be calculated in specific cases. This approach is also used to calculate temporal drift and power supply changes.

The input bias current: This has a typical value of 0.01 pA to 1.5 μA, with an effective offset current of 80 nA. The bias current is the average of the individual bias currents, I_{B1} and I_{B2}, shown in Figure 8(a). The offset is half the difference. The presence of the bias current introduces an additional offset error, depending on the input and feedback resistors used. Referring to Figure 8(b) the output error, v_o, resulting from the effect of bias current is

$$v_o = \frac{I_{B1}R_f}{R_{i1}} - I_{B2}R_{i2}\left(1 + \frac{R_f}{R_{i1}}\right) \tag{16}$$

The total output offset error will include that due to the input offset voltage. For an amplifier having a typical offset voltage error, V_{os}, and bias current I_B, there is a critical resistance value of $R_i = V_{os}/I_B$. If a value of R_i greater than this is used, the offset error due to bias current will be greater than that due to offset voltage.

Input offset current: The input offset current, i_{os}, is the difference between the bias current I_{B1} and I_{B2}. In many op-amps, the bias currents track each other well with temperature and are very nearly equal in magnitude. Hence, it is possible to minimize the errors due to bias currents by proper choice of input resistors. Note that offset current can exhibit sensitivity to the same circuit variables as the offset voltage. From (16), if

$$R_{i2} = \frac{R_{i1}R_f}{R_{i1} + R_f} \tag{17}$$

then

$$v_o = (I_{B1} - I_{B2})R_f = i_{os}R_f \tag{18}$$

If R_{i2} cannot be selected as in (7), the value of bias current will determine the DC output errors.

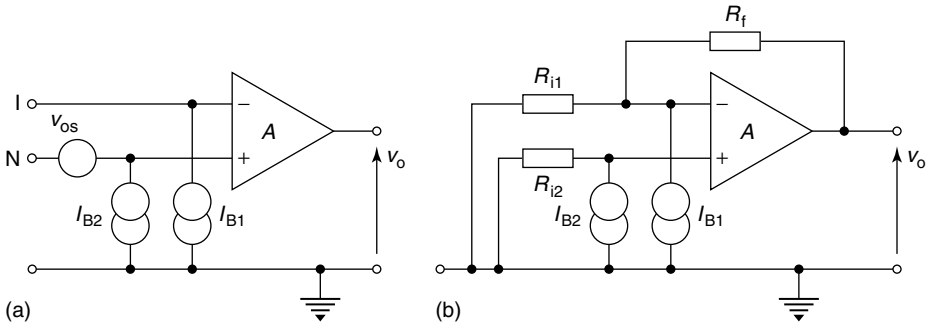


Figure 8. Individual input bias currents in (a) require compensation for resulting offset using (b).

Example 3. Estimate the output offset voltage and its variation from 25 °C to 45 °C for an op-amp in the inverting configuration, as shown in Figure 3(a), having a gain of −10 and a lead-in resistance R_1 of (a) 10 kΩ and (b) 100 kΩ. Assume that op-amp offset parameters are $v_{os} = \pm 2$ mV with a temperature coefficient of $\pm 6\,\mu$V/°C and $i_{os} = 80$ nA with a temperature sensitivity of 1%/°C.

Solution: Consider first (a) $R_1 = 10\,\text{k}\Omega$ at 25 °C. For a gain of −10, $R_f = 100\,\text{k}\Omega$.

The general expression for the output voltage combining the errors due to offset voltage and bias current is

$$v_o = \pm v_{os}\left(1 + \frac{R_f}{R_i}\right) + i_{os}R_f \tag{19}$$

Therefore, taking into account the direction of current flow,

$$[v_o]_{25\,^\circ\text{C}} = \pm 2 \times 10^{-3}(1 + 10) + 80 \times 10^{-9} \times 100 \times 10^3$$
$$= +30 \text{ to } -14\,\text{mV}$$

Knowing the temperature sensitivities, the offset voltage and the bias current, the output offset can be calculated to give

$$[v_o]_{45\,^\circ\text{C}} = +29.3 \text{ to } -17.3\,\text{mV}$$

If the offset voltage correction is arranged at 25 °C, the change in output will be

$$\Delta v_o = (\pm 1.32 - 2)\,\text{mV} = -0.68 \text{ to } -3.32\,\text{mV}$$

Repeating the above calculations for case (b) $R_1 = 100\,\text{k}\Omega$ gives $R_f = 100\,\text{M}\Omega$, $[v_o]_{25\,^\circ\text{C}} = +102$ to $+58$ mV and $[v_o]_{45\,^\circ\text{C}} = +83.3$ to 36.7 mV. Hence,

$$\Delta v_o = -18.68 \text{ to } -21.32\,\text{mV}$$

Note that the errors due to bias current increase as R_1 is increased.

Voltage and current noise: Unfortunately, it is not possible to compensate for the equivalent noise voltage and current in as simple a way as for the offset components. Since the mechanisms that lead to both voltage and current noise are statistical, then, statistical methods, based upon correlation analysis, must be used to calculate effective levels. If the signal is low level, corresponding to small v_d, thus requiring high amplification, usually >100, noise sources will largely determine the fundamental resolution of the amplifier.

Since, op-amps are made from either bipolar or field-effect transistors; the op-amp noise will consist of the same types as occur in these devices. Fabrication technology uses silicon nitride methods, which is a combination of special diffusion techniques and modern surface passivation, to give good surface stability. Bipolar noise is predominantly low frequency voltage noise due to emitter region dislocations, whereas the field-effect MOS device is mostly due to trapping levels within the oxide. Channel carriers jump into these levels and are later expelled back into the channel. Ion-implanted field-effect technology is used in FET input op-amps. These effects are modeled using similar voltage and current sources as for the offset effects.

10 EQUIVALENT CIRCUIT FOR REAL OP-AMPS

Figure 9 illustrates a consolidated op-amp model for assessing the circuit behavior of an op-amp. Assessment is important because the signal propagation properties must be calculable. This can only be done with an effective signal transmission model. Other important aspects are the offset and noise behavior of the op-amp in the actual configuration used.

In the model, the parameters are grouped in a logical manner by offset sources, noise sources, and impedances. Although all of these parameters will generally be asymmetrical as is shown in the model, simplification is possible.

The consolidated equations for offset voltage and current are respectively

$$\begin{aligned} e_{os} &= E_{os} + \frac{\Delta v_{os}}{\Delta T}\Delta T + \frac{\Delta v_{os}}{\Delta V_s}\Delta V_s + \frac{\Delta v_{os}}{\Delta t}\Delta t \\ &= \begin{pmatrix}\text{normal}\\ \text{in } \mu\text{V at } 25\,^\circ\text{C}\end{pmatrix} + \begin{pmatrix}\text{temperature}\\ \text{sensitivity}\\ \mu\text{V}/^\circ\text{C}\end{pmatrix} + \begin{pmatrix}\text{power supply}\\ \text{sensitivity}\\ \mu\text{V}/\%\end{pmatrix} + \begin{pmatrix}\text{temporal drift}\\ \mu\text{V/day}\end{pmatrix} \end{aligned} \tag{20}$$

$$\begin{aligned} i_{os} &= I_{os} + \frac{\Delta i_{os}}{\Delta T}\Delta T + \frac{\Delta i_{os}}{\Delta V_s}\Delta V_s + \frac{\Delta i_{os}}{\Delta t}\Delta t \\ &= \begin{pmatrix}\text{normal}\\ \text{in nA at } 25\,^\circ\text{C}\end{pmatrix} + \begin{pmatrix}\text{temperature}\\ \text{sensitivity}\\ \text{nA}/^\circ\text{C}\end{pmatrix} + \begin{pmatrix}\text{power supply}\\ \text{sensitivity}\\ \text{nA}/\%\end{pmatrix} + \begin{pmatrix}\text{temporal drift}\\ \text{nA/day}\end{pmatrix} \end{aligned} \tag{21}$$

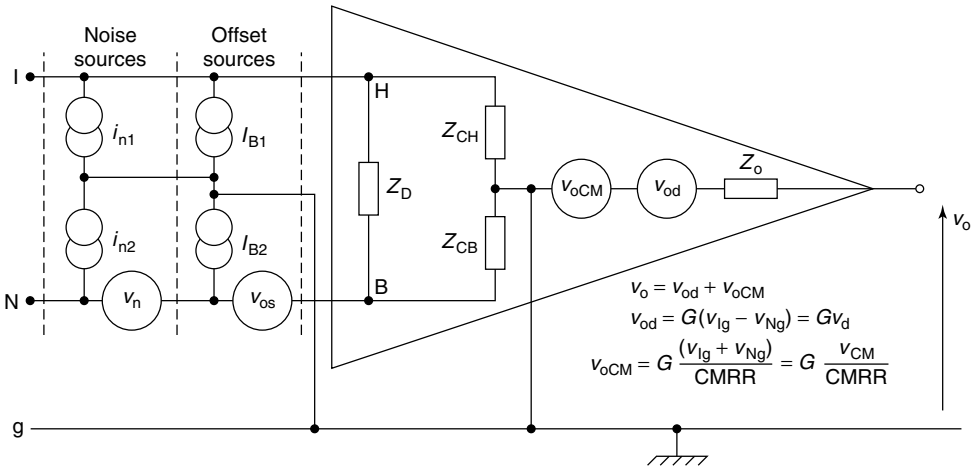


Figure 9. Consolidated signal, offset and noise model for an op-amp.

It can be seen from (20) and (21) that the offset voltage and current can be calculated for any given changes in temperature, power supply voltage and time period from the drift sensitivity coefficients.

It is usually possible to adjust the initial offset voltage and current to zero using a suitable biasing network. The standard practice is always to quote offsets and drift referred to the input in op-amp specifications. This allows the method to be used of multiplying the input referred effects by the noninverting gain to obtain the effect of offsets on the output. The methods used to calculate output offset effects in instrumentation amplifiers, which are different from those used in op-amps are described in Section 6 of **Article 123, Instrumentation Amplifiers, Volume 2**.

Because the noise values are normally considered as energy densities per unit bandwidth, it is more appropriate to perform any noise budgeting analysis in the frequency domain.

RELATED ARTICLES

Article 123, Instrumentation Amplifiers, Volume 2; **Article 124, Specialized Amplifiers for Measurement Systems, Volume 2**; **Article 133, Amplifiers and Filters for DAS, Volume 3**.

FURTHER READING

Clayton, G.B. (1975) *Linear Integrated Circuits*, Macmillan Press, London.

Coughlin, R.F. and Driscoll, F.F. (2000) *Operational Amplifiers and Linear Integrated Circuits*, Prentice Hall, NJ.

Franco, S. (2001) *Design with Operational Amplifiers and Analog Integrated Circuits*, McGraw-Hill, New York.

Huijsing, J.H. (2000) *Operational Amplifiers – Theory and Design*, Kluwer, Amsterdam.

Jacob, J. (2000) *Analog Integrated Circuits Applications*, Prentice Hall, NJ.

Lang, T.T. (1987) *Electronics for Measuring Systems*, Wiley, Chichester.

Mancini, R. (2003) *Op Amps for Everyone*, Newne, London.

Rangan, C.S., Sarma, G.R. and Mani, V.S.V. (1983) *Instrumentation: Devices and Systems*, Tata McGraw Hill, New Delhi.

Rhodes, J.D. (1976) *Theory of Electrical Filters*, Wiley, Chichester.

Stanley, W.D. (2001) *Operational Amplifiers with Linear Integrated Circuits*, Prentice Hall, NJ.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

This article is based in part on a chapter by J. McGhee *et al* originally published in Measurement Data Handling: Hardware Techniques Vol. 2, J. McGhee *et al*, Technical University of Lodz, 2001.

123: Instrumentation Amplifiers

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 The Electronic Instrumentation Amplifier 840
2 Differential Input to Single-ended Output Stage 840
3 Differential-input to Differential-output Stage 842
4 Two op-amp Configuration 843
5 Three op-amp Configuration 843
6 Specifying Instrumentation Amplifiers 844
7 Improving the CMRR – Shielding and Grounding 845
8 Improving the CMRR – Guard Techniques 847
Related Articles 849
Further Reading 849

1 THE ELECTRONIC INSTRUMENTATION AMPLIFIER

An instrumentation amplifier, or IA, is a closed-loop, dedicated circuit differential input gain block. These units can justifiably be considered as one of the most important elements in contemporary measuring systems.

The primary function of an IA is precision amplification of the differential signal applied to its inputs, while rejecting any existing common-mode signals. Ideally, an IA only responds to the difference in voltage between the two signals applied to its input terminals. This difference is often called the *differential input voltage*.

IAs also exhibit extremely high impedances between the I and N input terminals, and from each of these terminals to ground. The output voltage, which is equal to the product of the amplifier gain and the differential input voltage, is also developed as a single-ended output with respect to ground. Note that an IA differs fundamentally from an op-amp. An op-amp is an uncommitted open-loop device, whose closed-loop performance can be designed depending upon the external networks used to close the feedback loop.

To extend the common-mode rejection ratio (CMRR) of the configurations, IAs using discrete realizations of two and three op-amp configurations were widely employed. In contemporary systems where the trend is toward 12-bit, 14-bit, or even higher resolution, the discrete approach is no longer satisfactory. This trend for better resolution also means that tighter specifications for nonlinearity, offset and gain drift, and noise are usually required for instrumentation amplifiers.

In this article, the foundational design principles of flexible high performance IAs are described. These properties are presently obtained by combining additional specially designed circuits with high performance op-amps on the same monolithic chip. Hence, the section commences with a description of the basic difference stage and the differential-input to differential-output stage, which are the two constituent stages of the most common realization of IAs. This is followed by details of the two op-amp and three op-amp forms of IA realization. The importance of dealing with common-mode signals is considered at different places.

2 DIFFERENTIAL INPUT TO SINGLE-ENDED OUTPUT STAGE

The circuit for the general case of an asymmetrical differential stage, sometimes also called a *difference amplifier*, is given in Figure 1. In differential amplifiers, asymmetries

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

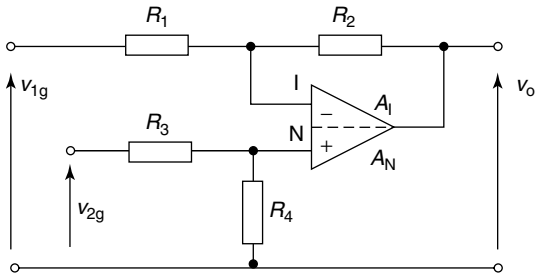


Figure 1. An asymmetrical difference amplifier.

in all of the parameters of the circuit must be considered. Asymmetries will occur because of the inherent asymmetry of the op-amp used. In addition to the offset asymmetries, there will be gain asymmetry in the signal paths between the inverting input and the output and between the noninverting input and the output.

Offset effects may be treated separately on the basis that stage linearity allows the application of superposition. Although the resistors in the stage are high-tolerance components, it is still necessary to take these tolerances into account.

Analyzing the stage gives its output voltage as

$$v_o = -A_I v_{Ig} + A_N v_{Ng} \tag{1}$$

Define

$$\beta_I = \frac{R_1}{R_1 + R_2} \quad \text{and} \quad \beta_N = \frac{R_3}{R_3 + R_4} \tag{2}$$

Hence,

$$v_{Ig} = \frac{R_2}{R_1}\beta_I v_{1\,g} + \beta_I v_o \quad \text{and} \quad v_{Ng} = \frac{R_4}{R_3}\beta_N v_{2g} \tag{3}$$

Substitute (2) into (3) and the result into (1) and simplify to obtain

$$v_o = -\frac{R_2}{R_1}\frac{\beta_I A_I}{1 + \beta_I A_I} v_{1g} + \frac{R_4}{R_3}\frac{\beta_N A_N}{1 + \beta_I A_I} v_{2\,g} \tag{4}$$

Assume that $\beta_I A_I > 1$. Now let $v_{1g} = -v_{2g} = 1/2v_d$ to obtain the differential mode output voltage, v_{od}, as

$$v_{od} = -\frac{1}{2}\left(\frac{R_2}{R_1} + \frac{R_4}{R_3}\frac{\beta_N A_N}{\beta_I A_I}\right) v_d = -A_{vd} v_d \tag{5}$$

where the difference mode–voltage gain, A_{vd}, is

$$A_{vd} = -\frac{1}{2}\left(\frac{R_2}{R_1} + \frac{R_4}{R_3}\frac{\beta_N A_N}{\beta_I A_I}\right) \tag{6}$$

In a similar way, let $v_{1g} = v_{2g} = v_{CM}$ to obtain the common mode–output voltage, v_{oCM}, as

$$v_{oCM} = -\left(\frac{R_2}{R_1} - \frac{R_4}{R_3}\frac{\beta_N A_N}{\beta_I A_I}\right) v_d = -A_{vCM} v_{CM} \tag{7}$$

where the common mode–voltage gain, A_{vCM}, is

$$A_{vCM} = \left(\frac{R_2}{R_1} - \frac{R_4}{R_3}\frac{\beta_N A_N}{\beta_I A_I}\right) \tag{8}$$

Using superposition, the total output voltage, which includes a difference-mode component and a differential mode component, can be written as

$$v_o = -A_{vd} v_d - A_{vCM} v_{CM} = -A_{vd}\left(v_d + \frac{1}{CMRR} v_{CM}\right) \tag{9}$$

where the common-mode rejection ratio, or CMRR, is given by

$$CMRR = \frac{A_{vd}}{A_{vCM}} = \frac{1}{2}\frac{\dfrac{R_2}{R_1} + \dfrac{R_4}{R_3} \times \dfrac{\beta_N}{\beta_I} \times \dfrac{A_N}{A_I}}{\dfrac{R_2}{R_1} - \dfrac{R_4}{R_3}\dfrac{\beta_N}{\beta_I} \times \dfrac{A_N}{A_I}} \tag{10}$$

This rejection ratio is an important figure of merit for a differential amplifier, since it allows a comparison of the ability of one arrangement to reject common-mode signals with the same ability of some other arrangement.

If it is momentarily assumed that the op-amp is ideal and the circuit is symmetrical, then $\beta_I A_I = \beta_N A_N$. This shows that the matching condition for infinite rejection of common-mode inputs corresponds to the denominator of (5) being zero or $(R_2/R_1) = (R_4/R_3)$.

However, all of the circuit resistors have manufacturing tolerances and the op-amp has asymmetrical gain. These inadequacies ensure that there is a practical limit to the common-mode rejection property of the stage. To evaluate the effects, it is possible to express the *CMRR* as

$$CMRR = \frac{1 + G_{0Q}}{\pm\delta_1 \pm \delta_4 \pm \delta_2 \pm \delta_3 \pm (1 + G_{0Q})CMRR^{-1}} \tag{11}$$

where the nominal closed-loop gain to inverting inputs is $G_{0Q} = R_{2Q}/R_{1Q}$ and the factor *cmrr* is the common mode–rejection ratio of the op-amp itself.

Equation (11) is easily understood. It demonstrates that the *CMRR* of a practical differential amplifier depends upon

both the matching of the circuit resistors and the *cmrr* of the op-amp.

It is instructive to consider typical numerical values. Let the stage have $G_{0Q} = 1$ and also that $cmrr_{dB} = 100 \equiv 20\log_{10}\ (A_{vd}/A_{vCM})$ so that $cmrr = (A_{vd}/A_{vCM}) = 10^5$, which is a typical value. This gives the last term in the denominator of (5) a value of 2×10^{-5}.

Compared to typical resistor tolerances of $\pm 1\%$, or 0.01, this figure of 2×10^{-5} can be neglected. It is only when $(1 + G_{0Q})\ cmrr^{-1} \geq 0.004$ that the term due to op-amp *cmrr* needs to be taken into account. This will correspond to $G_{0Q} \approx 4 \times 10^{-3} \times cmrr = 400$ for *cmrr* equal to 10^5. Hence, for good common-mode rejection, use tightly matched resistors, which can track each other well with temperature, select a low value for G_{0Q} and a high value for *cmrr*. Even for values of resistor tolerance of $\pm 1\%$ and $G_{0Q} = 1$, the value of the stage *CMRR* is still only ≈ 50, which is not very impressive. Clearly, steps need to be taken to design a differential amplifier with a much higher *CMRR*.

In the above analysis, it has been assumed that the common mode–input resistance/impedance of the op-amp has been neglected. To give a more complete interpretation, the effects of asymmetry in these resistances need to be taken into account.

It is easily understood that the common-mode resistance between both I and N op-amp inputs and ground act as voltage dividers when combined with their external circuit components. The problem is especially evident if the lead-in resistors are increased in value in an attempt to increase the common mode–input resistance of the stage. This can be counterproductive, since it will lead to increased offset and drift. In the case of FET type op-amps, which can tolerate large values of lead-in resistance because of their significantly lower bias current, such an increase might be counterproductive by introducing an increase in noise and a decrease in bandwidth when combined with the stray circuit capacitance.

Resistor tolerance has been shown to be the critical factor in determining the *CMRR* of the stage, since the tolerance of this component exerts a much greater influence than the *cmrr* of the op-amp used. Resistance asymmetry for whatever reason or from whatever cause, effectively converts common-mode interference into series mode interference. These series mode sources are indistinguishable from the true difference-mode input.

Although trimming the resistors can help improve the *CMRR*, this may be of no benefit if the trimmed resistors do not track each other with changes in temperature. Other sources of drift are due to sensitivity to changes with time, such as temperature drift and aging effects, and power supply voltage.

3 DIFFERENTIAL-INPUT TO DIFFERENTIAL-OUTPUT STAGE

Another important disadvantage of the basic difference amplifier considered before in Section 2 is its comparatively low input resistance. This can be improved using a suitable form of the type of noninverting amplifier previously considered in Section 6 of **Article 122, Operational Amplifiers, Volume 2**.

It will be recalled from (10), in **Article 122, Operational Amplifiers, Volume 2**, that the input resistance of the noninverting stage and the follower stage are very much higher than that of the inverting stage. Also, noninverting stages essentially buffer a high-impedance source to a low-impedance source. The need for high input resistance and high *CMRR* in amplifiers for low-level signals are satisfied by combining differential-input to differential-output amplifiers with the differential-input to single-ended stage in Section 1.

Of the candidate circuits shown in Figure 2, consider the first Figure 2(a) using only follower stages. This combination stage provides a differential output from a differential input with a very high input resistance. Provided the op-amps are well matched, it also has a low common-mode gain. Its only real disadvantage is its unity differential gain.

Consider the second circuit in Figure 2(b), consisting of two cross-connected noninverting stages. The voltages around the circuit are

$$v_w = v_1 \quad \text{and} \quad v_z = v_2 \tag{12}$$

The current, i_0, flowing in the resistor, R_0, is seen to have the value $(v_w - v_z)/R_0$ so that the output voltages, v_{o1} and v_{o2}, are given by

$$v_{o1} = v_1 + \frac{v_1 - v_2}{R_0} R_2 \quad \text{and} \quad v_{o2} = v_2 - \frac{v_1 - v_2}{R_0} R_2 \tag{13}$$

When there is no differential mode input, then $v_1 = v_2$ so that $v_{o1} = v_{o2} = v_{CM}$ giving a common-mode voltage gain

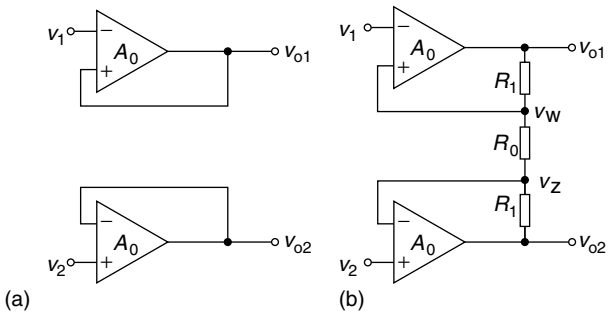


Figure 2. Candidate circuits for differential-input to differential-output stages.

A_{vCM} of unity. For difference-mode inputs corresponding to $v_1 = 1/2v_d$ and $v_2 = -1/2v_d$, the stage has a difference mode–voltage gain, A_{vd}, given by

$$A_{vd} = \frac{v_{o1} - v_{o2}}{v_d} = \left(1 + 2\frac{R_2}{R_0}\right) \tag{14}$$

Because $A_{vCM} = 1$, the *CMRR* of the stage is thus numerically equal to the difference-mode gain, which can be chosen to be very high simply by reducing the value of R_0. Since the stage consists of noninverting amplifiers, the output resistance at each of the output nodes will be very low.

Common-mode signals, which are transmitted with unity gain by both op-amps, can be converted to series mode–interfering components if the op-amps used are not exactly matched. With such asymmetry, any common-mode signal, v_{CM}, will give rise to a differential signal given by

$$v_{CM}\left(\frac{1}{CMRR_1} - \frac{1}{CMRR_2}\right) \tag{15}$$

4 TWO OP-AMP CONFIGURATION

Basic two amplifier configuration: A circuit diagram for a basic two amplifier IA realization is shown in Figure 3.

The voltages around the circuit are

$$v_A = v_1[1 + (1 + R_1/R_2)] = \frac{v_1}{B_{1I}} \quad \text{with} \quad B_{1I} = \frac{R_2}{R_1 + R_2} \tag{16}$$

$$v_o = -\left(\frac{R_4}{R_3}\right)v_A + \left(\frac{v_2}{B_{2I}}\right) \quad \text{with} \quad B_{2I} = \frac{R_3}{R_3 + R_4} \tag{17}$$

$$v_o = -\left(\frac{R_4}{R_3}\right)\left[\frac{R_1 + R_2}{R_2}\right]v_1 + \left[\frac{R_3 + R_4}{R_3}\right]v_2 \tag{18}$$

With $R_4 = R_2$ and $R_1 = R_3$, this becomes

$$v_o = \left[\frac{R_1 + R_2}{R_1}\right](v_1 - v_2) = \left(1 + \frac{R_2}{R_1}\right)v_d \tag{19}$$

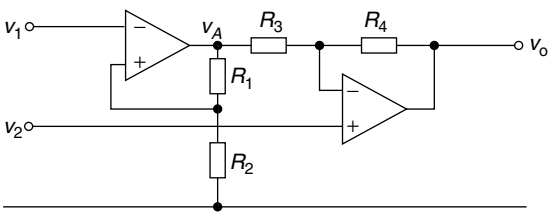


Figure 3. Basic instrumentation amplifier configured using two op-amps.

5 THREE OP-AMP CONFIGURATION

Three op-amp IA configuration: This configuration, illustrated in Figure 4, has a high accuracy and excellent common-mode rejection. Provided that the amplifiers, A_1 and A_2, are well-matched, low-noise, low-offset, and high-gain units then precision difference amplification with high common-mode rejection may be obtained.

The input stages, A_1 and A_2, constitute a differential-input to differential-output stage so that $v_{o1} - v_{o2}$ has an amplified component of v_d and an unamplified component of v_{CM}. The stage A_3 is a standard differential-input to single-ended output stage with unity gain and a high CMRR.

The equations around the circuit are

$$v_{o1} = \left[1 + \left(\frac{2R_1}{R_o}\right)\right]^{\frac{1}{2}} v_d + v_{CM}$$

$$v_{o2} = -\left[1 + \left(\frac{2R_1}{R_o}\right)\right]^{\frac{1}{2}} v_d + v_{CM} \tag{20}$$

Taking into account the finite $CMRR_3$ in A_3 and using (17) gives the output voltage as

$$\begin{aligned} v_o &= -\left(1 + 2\frac{R_1}{R_0}\right)v_d + \frac{1}{CMRR_3}v_{CM} \\ &= -Gv_d + \frac{1}{CMRR_3}v_{CM} \\ &= -G\left[v_d + \frac{1}{G \times CMRR_3}v_{CM}\right] \\ &= -G\left[v_d + \frac{1}{CMRR_o}v_{CM}\right] \end{aligned} \tag{21}$$

where $CMRR_o$ is the overall *CMRR* of the stage.

Note that $CMRR_o = G \times CMRR_3$ shows that the *CMRR* is increased by the difference gain $G = (1 + 2R_1/R_0)$ of the differential-input to differential-output stage. This is

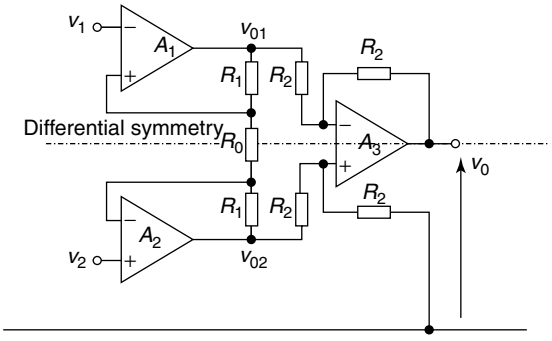


Figure 4. Three op-amp configured instrumentation amplifier.

possible because the common-mode gain of this stage equals unity. The common-mode rejection ratio of A_3, $CMRR_3$, is completely determined with negligible error by the tolerance of the resistors, R_2, in accordance with (11). Since the common-mode rejection ratio of A_3, $CMRR_3 = 1/2\delta_2$, (21) can be written finally as

$$v_o = -G\left[v_d + \frac{2\delta_2}{G} v_{CM}\right] \tag{22}$$

Example 1. A three op-amp IA, which uses resistors with a tolerance of ±1%, has a closed-loop gain of 200. Calculate the common-mode rejection ratio of the stage.

Solution: Since $CMRR_o = G \times CMRR_3$, and $CMRR_3 = (1 + G_3)/4\delta$, the overall $CMRR$ is calculated as $CMRR_o = 200 \times (1 + 1)/(4 \times 0.01) = 10\,000$, which is equivalent to $CMRR_{odB} = 20 \log_{10} 10\,000 = 20 \times 4 = 80$ dB.

6 SPECIFYING INSTRUMENTATION AMPLIFIERS

Closed-loop gain and nonlinearity: The idealized transfer function is

$$v_o = -G(v_1 - v_2) \tag{23}$$

The amplifier gain, G, is usually set by the user to give typical values of 1,200,500, and 1000 with a single external resistor. Temperature coefficient of the gain and gain nonlinearity, shown in Figure 5, can be found on manufacturer's data sheets.

Gain nonlinearity, which is defined as the peak deviation from a best-fit straight line, expressed as a percentage of peak-to-peak full-scale output, is possibly of more importance than the gain accuracy, since the value of the gain can be adjusted to compensate for simple gain errors.

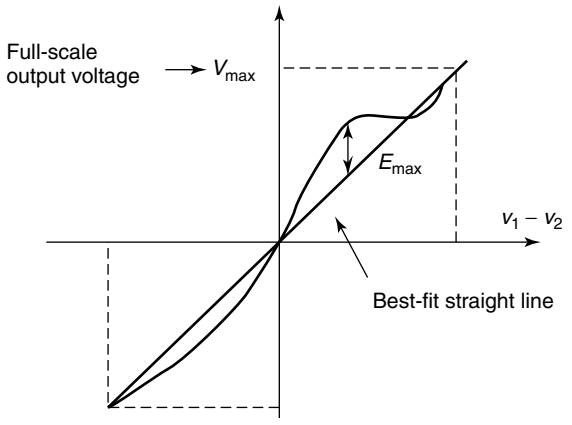


Figure 5. Illustration of nonlinearity error.

Note that for programmable gain IAs, commonly used in computerized data-acquisition systems, the gain control is accomplished through a number of digital inputs.

Offset voltage and input bias currents: Voltage offset and input bias current are often considered as the key figures of merit for IAs.

Offset voltage is defined as the voltage required at the input to drive the output to zero in exactly the same way as for op-amps. While initial offset can be adjusted to zero, shifts in offset voltage in temperature and time introduce errors. They are by far the most important source of error in precision IAs. The temperature coefficients of these parameters, which are always specified by IA manufacturers, have typical values of the order of 1.0 μV/°C in the temperature range $-25\,^{\circ}C \le T_A \le +85\,^{\circ}C$.

The offset at the output of an IA consists of two terms. The first is output offset, V_{os}, sometimes referred to as *unity gain output offset*, while the second is $+GV_{os}$, sometimes referred to as *inverting gain offset*.

The input bias currents, which are currents flowing into or out of the two inputs of the amplifier, correspond to the base currents I_{B1} for BJT and FET leakage currents, depending upon the specific type of op-amp input stage. In the same way as for other op-amp circuits, the bias currents flowing through the source resistance will generate a voltage offset. Although initial bias currents are often adjustable to zero, their drift with temperature, especially for FET type input stages, is very troublesome. Recall from Section 1 of **Article 122, Operational Amplifiers, Volume 2** that the JFET leakage current can double every 10 °C.

There are differences between the methods of calculating offset and drift effects in op-amps and in IAs.

In the case of op-amps, as pointed out in Section 8 of **Article 122, Operational Amplifiers, Volume 2**, offset parameters are always quoted as values referred to as their inputs.

In IAs, extra care needs to be taken because of the effect of the gain-determining nature of the external component. For fixed gain IAs, it does not matter whether the offset parameters are quoted as referred to the input or as referred to the output, because one can be calculated knowing the other, provided the gain is also known.

With adjustable gain IAs, it is essential to remember that each of the offset parameters will give rise to different calculated effects depending upon the gain. In many cases, when the drift sensitivities are quoted at two values of gain, typically 1 and 1000, it is easy to calculate the effective drift at other values of gain. This is achieved by assuming that there is a linear relationship between the drift sensitivity

and the gain. Subsequently, the calculation in specific cases of temperature and power supply ranges and time periods proceeds in the same way as for the op-amp in Section 8 of **Article 122, Operational Amplifiers, Volume 2**.

Frequency response: Specifications such as bandwidth, slew rate, and settling time are defined and specified as for op-amps.

Output impedance: IA output impedance, which is ∼10 to 20 Ω, allows very easy and direct interfacing with other following modules. IAs have much lower output impedances compared with the open-loop output resistance of op-amps.

Consolidated signal, offset, and noise model: This model, shown in Figure 6 may be summarized as possessing the properties of high-input impedance, low offset and drift, low nonlinearity, stable gain, high common-mode rejection, and low effective-output impedance. Hence, it is widely used for applications requiring these advantages. Examples include amplification for various types of transducers such as strain-gauge deflection bridges, load cells, resistance and thermistor thermometer networks, thermocouples, current shunts, and many more, as well as where small differential signals superimposed on high common-mode voltages require preamplification.

Common mode–rejection ratio: An ideal IA must respond only to the difference between the input voltages. The output of the ideal IA will be zero if the input voltages are equal so that $v_1 = v_2 = v_{CM}$, the common-mode voltage.

However, the output voltage of practical IAs has one component proportional to the differential input voltage and a second component proportional to the common-mode input voltage. The common-mode voltage, which appears at the amplifier's input terminal is defined as

$$v_{CM} = \frac{v_1 + v_2}{2} \tag{24}$$

7 IMPROVING THE CMRR – SHIELDING AND GROUNDING

A major source of error in instrumentation amplifier–based systems is the coupling of undesired interfering currents and voltages into systems with low-level signals. This kind of coupled interference, which usually takes the form of common-mode voltages and currents, can occur in many ways through

- differences in potential due to ground loop current flow;
- electrostatic coupling through mutual capacitance between adjacent conductors, such as power supply lines and other high frequency sources;
- electromagnetic coupling through mutual inductive coupling in closed conductive loops;
- asymmetrical effects such as loading of the signal leads.

Handling common-mode signals is crucial in precision low-level measurements, where it is especially essential to improve the *CMRR*. Figures 7 and 8 illustrate two examples of common-mode interference.

In Figure 7, the emf, e, that develops across a thermocouple, which is bonded to a metal plate, has to be measured. This metal plate is at a potential v_{CM}.

In Figure 8, mechanical stress is to be measured with a strain-gauge deflection bridge. When the applied stress is zero, the bridge is at balance and the emf, e, is equal to zero. However, the common-mode voltage, which is applied at the amplifier's input terminal, is equal to

$$V_{CM} = \frac{V_A + V_B}{2} \approx \frac{E}{2} \tag{25}$$

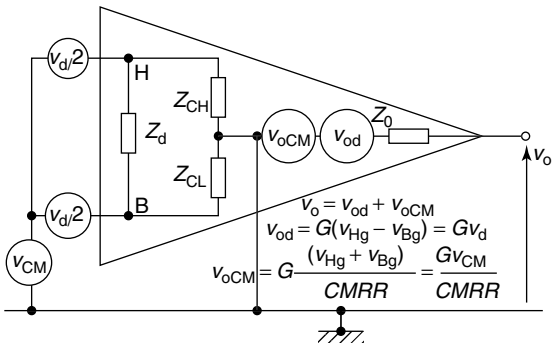


Figure 6. A simplified consolidated model of an instrumentation amplifier.

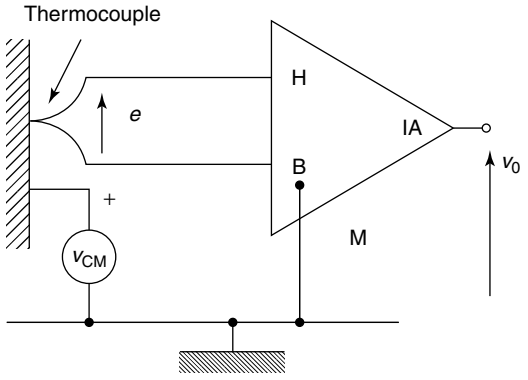


Figure 7. Thermocouple bonded to a metal plate which is at a potential v_{CM}.

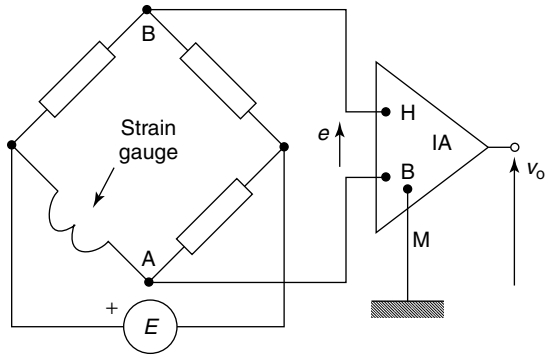


Figure 8. Circuit for measurement of mechanical strain with a strain gauge.

Common-mode interference can exert a critical, sometimes even determining, influence in such systems as industrial-process control systems whose remote transducers are often spread over large distances from their measurement instruments. Analog data must be transmitted over these long distances. Since the input lead resistances are no longer negligible, the undesirable voltages, which can be developed, are added both to any common-mode signal and potential difference between two ground connections, which may influence many transducer circuits.

In the equivalent circuit of a measuring scheme shown in Figure 9(a), R_1 and R_2 represent the input lead resistances, Z_d the IA input impedance, and Z_{CH} and Z_{CB} the leakage impedances from H to ground and B to ground respectively.

Considering only the common-mode voltage, v_{CM}, this circuit may be drawn as shown in Figure 9(b). It is clear that the undesirable voltage, v_P, produced by v_{CM} is zero only when the bridge is perfectly balanced.

When specifying the error caused by v_{CM}, it is usual to consider the worst-case off-balance. Assume that the IA used in this measurement system has infinite *CMRR*. The IA measures the potential difference, v_P which exists across the H and B input terminals. The *CMRR* of the measurement system, expressed in dB, is

$$CMRR_{dB} = 20 \log_{10} \frac{v_{CM}}{v_P} \tag{26}$$

The worst-case off-balance specified by measuring instrument manufacturers is that R_1 is zero, while R_2 is some finite resistance such as, say, 1 kΩ. The equivalent circuit is now shown in Figure 9(c) so that the

$CMRR_{dB}$ is

$$CMRR_{dB} = 20 \log_{10} \left[\frac{Z_{CB} + R_2 \| Z_d}{R_2 \| Z_d} \right] \tag{27}$$

Normally, $Z_d \gg R_2$ and $Z_{CB} \gg R_2$, so (27) simplifies to

$$CMRR_{dB} = 20 \log_{10} \frac{Z_{CB}}{R_2} \tag{28}$$

Example 2. In a certain measuring system, it is known that there is a direct common-mode voltage, V_{CM}. The

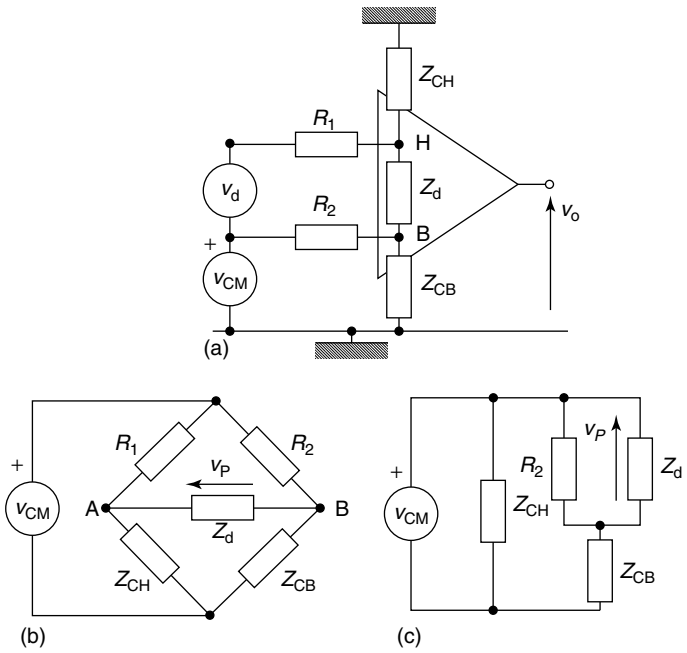


Figure 9. Measuring system in (a) has the parasitic voltage, v_P, in (b) produced by common-mode voltage, v_{CM}, may be analyzed by the equivalent circuit in the worst-case unbalance ($R_1 = 0$) in (c).

leakage impedance, Z_{CB}, between the B terminal of the IA and ground is purely resistive with a value of $R_{CB} = 100\,\text{M}\Omega$ and the off-balance resistance is $R_2 = 1000\,\Omega$. Calculate the $CMRR_{dB}$.

If the common-mode voltage is now alternating, and the leakage impedance is the parallel combination of $R_{CB} = 100\,\text{M}\Omega$ and a capacitance of 100 pF, calculate the $CMRR_{dB}$ if the frequency of v_{CM} is 50 Hz.

Solution: Immediately applying (28) gives

$$CMRR_{dB} = 20\,\log_{10}\frac{10^8}{10^3} = 100\,\text{dB} \tag{29}$$

Now calculate to get

$$|Z_{CL}| = \left|\frac{R_{CB}}{1 + j\omega R_{CB}C}\right| = \left|\frac{10^8}{1 + j2\pi 50 \times 10^8 \times 10^{-7}}\right| \approx \frac{10^8}{2\pi 50 \times 10^8 \times 10^{-7}} \approx 320 \times 10^3 = 320\,\text{k}\Omega$$

Substituting in (28) gives

$$CMRR_{dB} = 20\,\log_{10}\frac{320 \times 10^3}{10^3} \approx 50\,\text{dB} \tag{30}$$

8 IMPROVING THE CMRR – GUARD TECHNIQUES

Guard techniques allow effective shunting of the common-mode bridge and at the same time they provide high common-mode impedances, Z_{CH} and Z_{CB}.

To improve the *CMRR*, the input circuits of the IA are mounted inside a guard shield, which must be well isolated from the ground and connected directly to the common-mode source, v_{CM}, as shown in Figure 10(a) and its equivalent circuit of Figure 10(b).

Note that R_G must be as low as possible and Z_L represents the leakage impedance from the guard shield to ground.

Referring to Figure 10 allows the unwanted voltage, v_P, to be written as

$$v_P = v_{CM}\frac{R_G}{R_G + Z_L} \times \frac{R_2}{Z_{CB}} \approx v_{CM}\frac{R_G}{Z_L} \times \frac{R_2}{Z_{CB}} \tag{31}$$

Hence, using (26), the $CMRR_{dB}$ can be expressed as

$$CMRR_{dB} = 20\,\log_{10}\frac{Z_{CB}}{R_2} \times \frac{Z_L}{R_G} \tag{32}$$

Example 3. An IA mounted inside a guard shield has $Z_{CB} = 1000\,\text{M}\Omega$. If the guard shield leakage impedance is $Z_L = 1000\,\text{M}\Omega$, $R_2 = 10^3\,\Omega$, and $R_G = 100\,\Omega$ calculate the $CMRR_{dB}$ if v_{CM} is direct.

Comment upon the $CMRR_{dB}$ if v_{CM} is alternating.

Solution: Using (32) gives $CMRR_{dB} = 20\log_{10}\ 10^6 + 20\log_{10}\ 10^7 = 260\,\text{dB}$.

When v_{CM} is alternating, the effects of stray capacitance must be taken into account.

Consider the connections illustrated in Figure 11. The instrumentation amplifier often has a very high common mode–rejection ratio. This ensures that the common mode–voltage gain of the amplifier is very low compared to its gain for differential mode signals. In cases in which a cable with a length of 1 m or more runs from the transducer to the amplifier, the common-mode signals presented to the amplifier will not be the same as those at the transducer.

This arises because of mismatches in the resistance and capacitance of the cable. For example, in Figure 11(a)

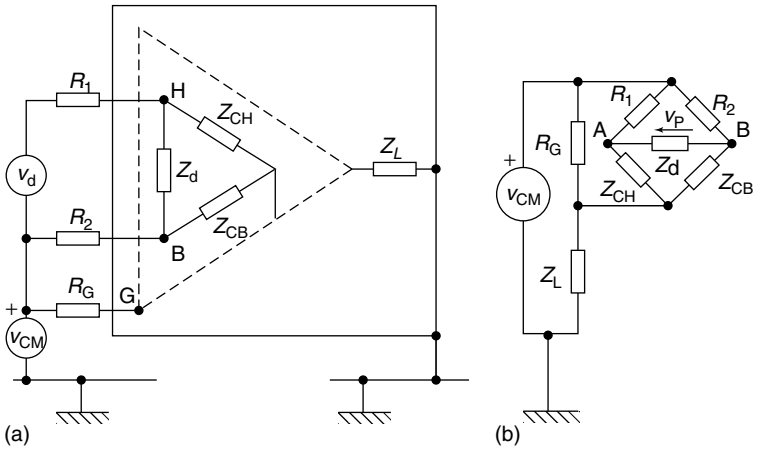


Figure 10. IA mounted inside a guard shield in (a) has the equivalent circuit in (b).

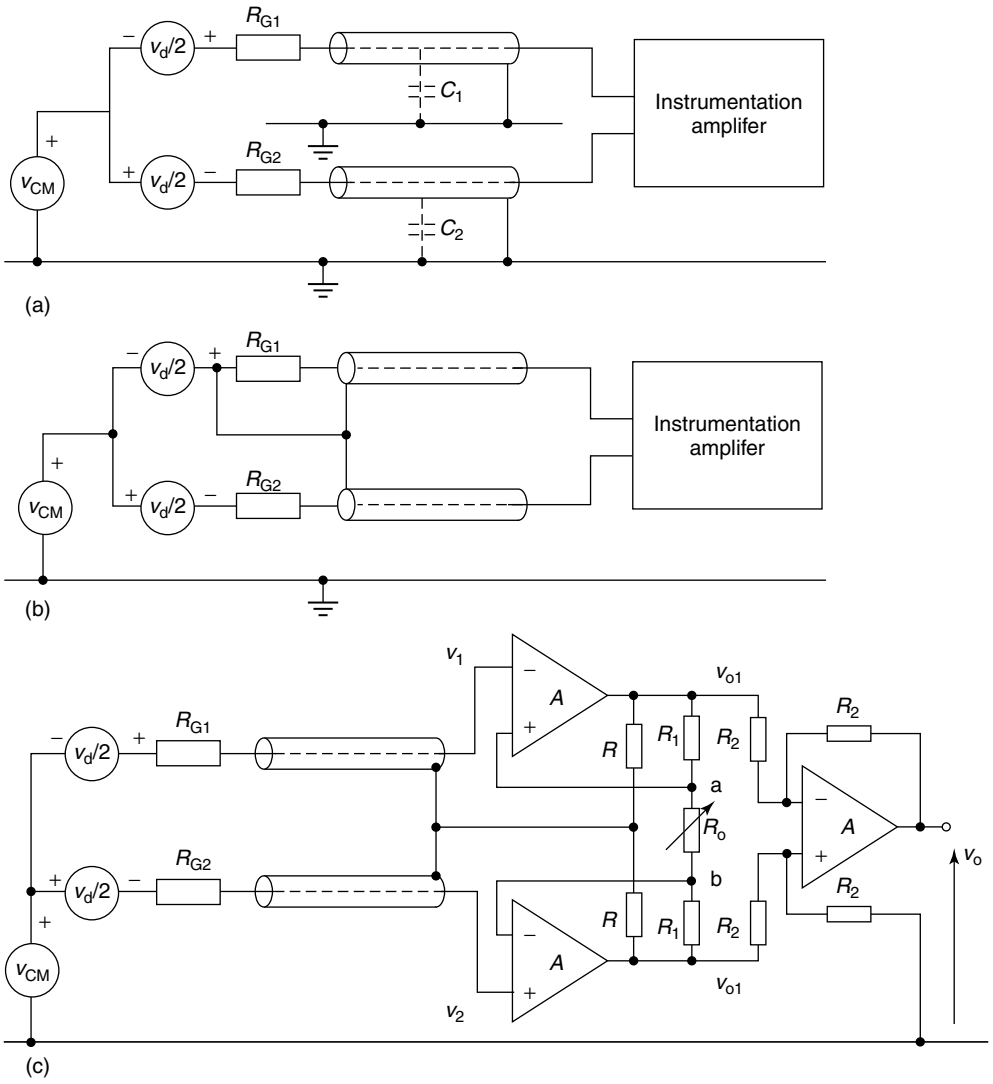


Figure 11. Input guarding showing the effect of cable capacitance in (a), driving the shield from one of the inputs to enhance *CMRR* in (b) and deriving the common-mode signal for driving the shielded cable in (c).

suppose that $R_{G1}C_1$ is not equal to $R_{G2}C_2$, where R_{G1} and R_{G2} are the source resistances and C_1 and C_2 are the cable capacitances. In that case, the common-mode voltage, which is converted to a series mode interfering signal, is then amplified by the differential mode gain. Consequently, the common mode–rejection ratio of the system deteriorates significantly. This problem can be tackled by connecting the cables preferably to the common-mode voltage, instead of to the ground, so that the voltage across the cable capacitance then reduces to the difference-mode signal. In this manner, the common-mode voltage at the input of the amplifier is now unattenuated by the cable capacitance.

When the common-mode voltage, v_{CM}, is high, the cables can be driven by one of the input signals itself, as shown in Figure 11(b).

If the common-mode voltage, v_{CM}, is not large compared to the difference-mode signal, v_d, it is necessary to derive the common-mode voltage as the average of the two input signals. In the schematic shown in Figure 11(c), this is achieved by the resistive summing of the outputs of amplifiers, A_1 and A_2, to get $v_{CM} = 1/2\ (v_1 + v_2)$ where $v_{01} = v_{CM} + Gv_d$ and $v_{02} = v_{CM} - Gv_d$, where G is the difference-mode gain of the first stage.

In situations in which large shield capacitances are to be driven at high frequencies, the divider resistances, R, have to be low. If this causes excessive output loading on the differential-input to differential-output stage, a unity gain buffer amplifier should be added between the summing-point output and the shield connection.

RELATED ARTICLES

Article 122, Operational Amplifiers, Volume 2; **Article 124, Specialized Amplifiers for Measurement Systems, Volume 2**; **Article 133, Amplifiers and Filters for DAS, Volume 3**; **Article 176, Signals and Signal-to-noise Ratio, Volume 3**.

FURTHER READING

Coughlin, R.F. and Driscoll, F.F. (2000) *Operational Amplifiers and Linear Integrated Circuits*, Prentice Hall, NJ.

Franco, S. (2001) *Design with Operational Amplifiers and Analog Integrated Circuits*, McGraw-Hill, New York.

Huijsing, J.H. (2000) *Operational Amplifiers – Theory and Design*, Kluwer, Amsterdam.

Lang, T.T. (1987) *Electronics for Measuring Systems*, Wiley, Chichester.

Stanley, W.D. (2001) *Operational Amplifiers with Linear Integrated Circuits*, Prentice Hall, NJ.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

This article is based in part on a chapter by J. McGhee *et al* originally published in Measurement Data Handling: Hardware Techniques Vol. 2, J. McGhee *et al*, Technical University of Lodz, 2001.

124: Specialized Amplifiers for Measurement Systems

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Amplifiers for Low-level Signals 850
2 Chopper Amplifier 850
3 Chopper Stabilized Amplifier 852
4 Auto-zero Amplifier 852
5 Magnetic and Optical Isolation Amplifiers 853
6 Necessity for the Charge Amplifier 856
7 Principles and Practicalities of a Charge Amplifier 856
8 Integrators 857
9 Differentiators 858
Related Articles 859
Further Reading 859

1 AMPLIFIERS FOR LOW-LEVEL SIGNALS

Instrumentation amplifiers (IA) covered in **Article 123, Instrumentation Amplifiers, Volume 2** will solve many measuring system amplification and conditioning problems but not all. There exist many situations in which low offset and low noise preamplification are required. Additionally, in many other applications, it is essential to ensure that the sensing input sections of the measuring system are galvanically isolated from the rest of the system.

Chopper amplifiers and others based upon the principles of signal chopping, or more precisely modulation, are perhaps the most common form of low-level signal amplifiers based upon the op-amp. All of them can help reduce undesirable variations at low frequencies such as voltage and current offset, drift effects, and flicker noise effects. The three chopper type amplifiers to be described in this article are the chopper amplifier, the chopper stabilized amplifier, and the auto-zero amplifier.

In the case in which isolation is required, applying a generic type of amplifier known as an *isolation amplifier* is usually appropriate. The magnetic isolation barrier amplifier and the optical isolation barrier amplifier are also described.

2 CHOPPER AMPLIFIER

A chopper amplifier, which operates in the manner outlined in Figure 1, uses an important method to reduce to a great extent the effects of low-frequency drift and noise in low-level amplification. In measuring systems, the frequency ranges of most of the interference noise and inherent noise sources at least overlap the bandwidths of the sensor signals if they do not totally coincide with them. Interference-type noise is picked up from outside the circuit under consideration. Inherent noise, which arises inside the circuit under consideration, can be due to the equipartition of energy, the randomness in flowing current or in the characteristics, and fabrication of the materials.

Interference noise, which may be periodic, irregularly occurring, or essentially random, can be minimized, if not very substantially reduced or even eliminated, using suitable precautions to avoid electrostatic and electromagnetic pickup from nearby radiators of electrical energy.

Inherent noise is usually totally random in nature although a nonrandom inherent interference such as chopper noise in chopper amplifiers can occur. Random inherent noise usually takes the form of:

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

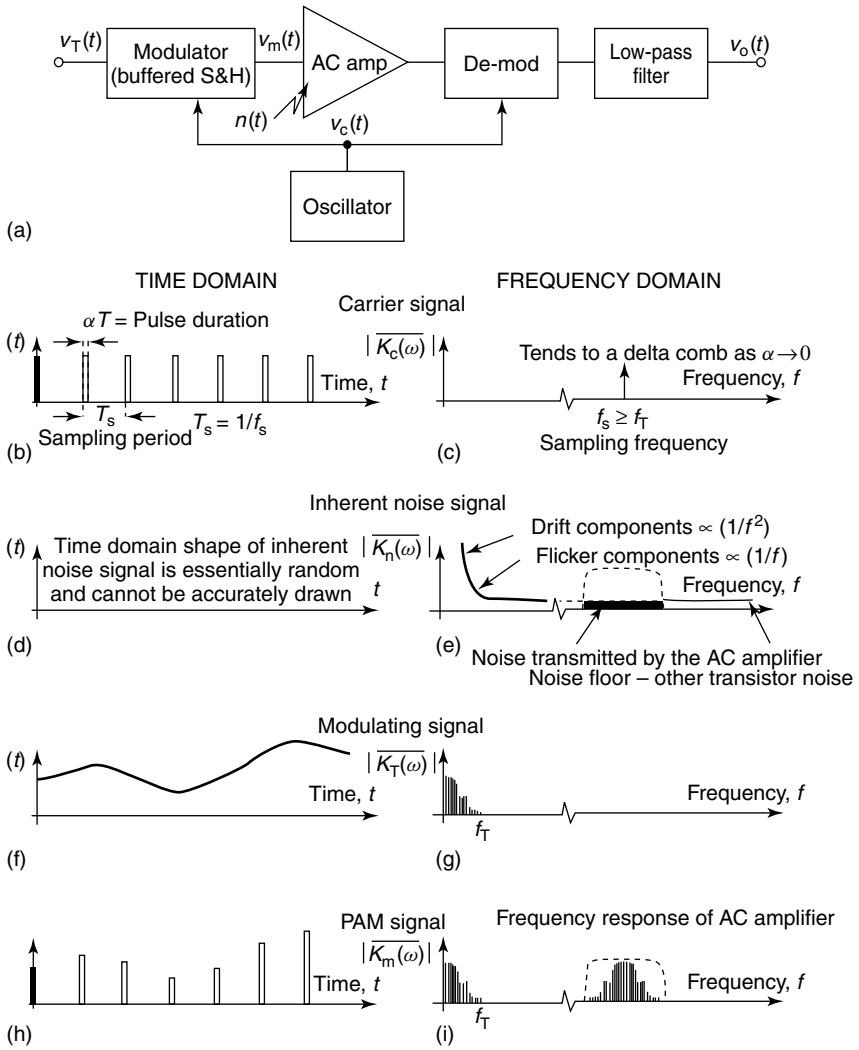


Figure 1. Block diagram and operational waveforms of a chopper amplifier.

- Johnson noise, which occurs in resistors;
- Schottky noise, which results when current carriers flow through a region depleted of carriers, such as the junctions of semiconductors;
- flicker noise, or $(1/f)$ noise, which is mainly caused by surface effects and crystalline dislocation, which are particularly troublesome deficiencies in semiconductor devices;
- popcorn noise, which is caused by erratic jittering of the forward current gain of bipolar transistors.

The only really effective way to minimize the effects of random inherent noise is to use modulation to shift the spectrum of the sensor signal into a band in which flicker effect is much less. Hence, these op-amp-based structures include some form of modulation and demodulation. In effect, they allow the use of very narrow bandwidth filtering in a stable manner.

Consider Figure 1. An oscillator, whose output signal, $v_c(t)$, has the time domain waveform with the frequency, f_s, in Figure 1(b), also has the frequency domain spectrum, $K_c(\omega)$, shown in Figure 1(c).

This carrier signal is pulse amplitude modulated by the transducer signal, $v_T(t)$, shown in Figure 1(f), whose typical spectrum, $K_T(\omega)$, is given in Figure 1(g). For the purposes of comparison, the internal drift and noise in the AC amplifier is shown in Figure 1(e). Note that the drift gives rise to the $(1/f^2)$ asymptote, flicker noise gives the $(1/f)$ asymptote, and other amplifier noise gives rise to the base, or floor, level noise. The resulting pulse amplitude modulated signal, $v_m(t)$, shown in Figure 1(h), which has the spectrum, $K_m(\omega)$, in Figure 1(i), is applied to the AC band-pass amplifier whose frequency response is shown in Figure 1(i).

If the AC amplifier has internal drifts and noise, their spectrum being as shown in Figure 1(e). Comparison of the

region where this interference spectrum is concentrated with Figure 1(i) shows that only the noise inside the bandwidth of the band-pass amplifier, with the comparatively low-level noise floor spectrum, will be transmitted. If the bandwidth adequately covers that of the transducer spectrum around f_s, then the net mean square value of the transmitted noise appearing at the amplifier output will be very low. The chopper has removed most of the amplifier interference.

3 CHOPPER STABILIZED AMPLIFIER

Chopper amplifiers can be employed with other high-performance op-amps to compensate for the already small inadequacies of these high-quality op-amps. The resulting combination, shown in Figure 2, gives a circuit called a *chopper stabilized amplifier*, which was first patented more than 50 years ago. Its designed aim was to combat the drift of directly coupled amplifiers by converting the DC voltage to an AC voltage.

The most important disadvantage of this inverting chopper stabilized stage is the relatively low input resistance, which is the same order as the lead-in resistance. A non-inverting, high input resistance version of the chopper stabilized type first appeared about 30 years ago. By causing interference from digital switching noise, the bandwidth is limited. Intermodulation distortion (IMD) between the clock and the input signal, which is similar to aliasing, produces errors at the sum and difference frequencies. Although both of these types have now been ostensibly replaced by the auto-zero amplifier, which is described in Section 4, they are still worth examining.

The chopper amplifier in Figure 2 measures the voltage of the inverting input of a conventional high-performance op-amp and then applies its amplified output to the non-inverting input of the same op-amp. In this way, any disturbances occurring inside the feedback loop are reduced in accordance with the well-known theory of negative feedback. To illustrate this reduction, consider that a disturbance $\overline{K_d}$, shown in Figure 2, occurs inside the feedback loop.

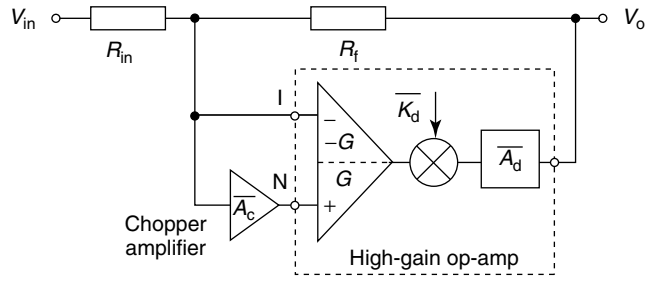


Figure 2. Chopper stabilized amplifier.

Analyzing in the frequency domain, it can be seen that the relations in the circuit are

$$V_o = -\overline{A_d} \times \overline{K_d} G V_{Ig} + \overline{A_d} \times G \times \overline{A_c} V_{Ig} \qquad (1)$$

Assuming very large input impedance at I gives

$$V_{Ig} = \frac{R_1}{R_f + R_1} V_o + \frac{R_f}{R_f + R_1} V_{in} = \beta_I V_o + G_{ex} \times \beta_I V_{in} \qquad (2)$$

where the feedback factor β_I is defined as $\beta_I = R_1/R_f + R_1 = 1/1 + G_{ex}$ and the external ideal gain G_{ex}, which is defined as $G_{ex} = R_f/R_1$, is the gain, assuming an ideal op-amp.

Inserting (2) into (1) and simplifying gives

$$\begin{aligned} V_o &= -\frac{R_f}{R_1} \frac{\beta_I \overline{A_d} \times G(1 - \overline{A_c})}{1 + \beta_I A_d \times G(1 - A_c)} V_{in} \\ &\quad + \frac{\overline{A_d}}{1 + \beta_I \overline{A_d} \times G(1 - A_c)} \overline{K_d} \\ &= \text{Desired output signal} \\ &\quad + \text{Disturbance caused output signal} \end{aligned} \qquad (3)$$

Assuming that $\beta_I \overline{A_d} \times G(1 - \overline{A_c}) \gg 1$ allows (3) to be written as

$$\begin{aligned} V_o &= -\frac{R_f}{R_1} V_{in} + \frac{1}{\beta_I G(1 - \overline{A_c})} \overline{K_d} \\ &= -\frac{R_f}{R_1} V_{in} + \frac{1 + (R_f/R_1)}{G(1 - \overline{A_c})} \overline{K_d} \end{aligned} \qquad (4)$$

When $\overline{A_c} \equiv 0$, which corresponds to a conventional inverting amplifier circuit, (4) becomes

$$V_o = -\frac{R_f}{R_1} V_{in} + \frac{1 + (R_f/R_1)}{G} \overline{K_d} \qquad (5)$$

Hence, with chopper stabilization, the disturbance input, $\overline{K_d}$, is reduced by the factor $(1 - \overline{A_c})^{-1}$. Note that the chopper amplifier gain $\overline{A_c}$ should be inverting at low frequencies, corresponding to $\overline{A_c(\omega \to 0)} = -A_{c0}$.

4 AUTO-ZERO AMPLIFIER

Operating principle: Auto-zero chopper stabilized amplifiers, with the block diagram in Figure 1, are now the most widely available form of chopper stabilized amplifiers. Their operation consists of two phases. During Phase 1, as in Figure 3(a), the input is applied only to the main amplifier, A_1, whose nulling input voltage, V_{N1}, is supplied by the voltage on capacitor, C_{M1}.

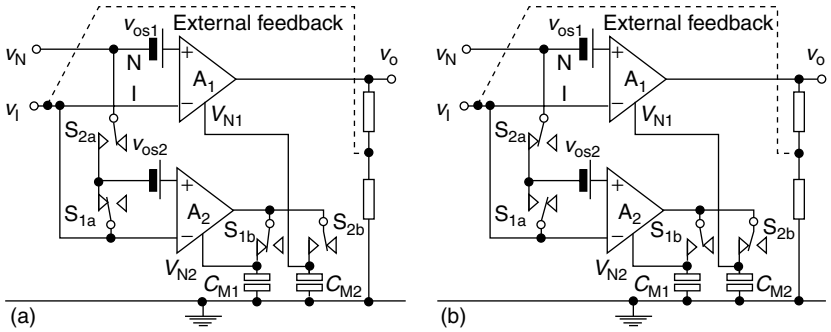


Figure 3. Phase 1 of an auto-zero amplifier in (a) is followed in operation by Phase 2 in (b).

Meantime, the I and N inputs of amplifier, A_2, are connected by S_{2a} to allow this amplifier to measure its auto-zero voltage that is stored on capacitor, C_{M1}. At the start of Phase 2, the make-before-break switches, S_1 and S_2, each change over. This allows the previously measured auto-zero values for A_2 to be applied at V_{N2}. When Phase 2 is complete, the whole sequence of operations is repeated.

Applications and characteristics of auto-zero amplifiers: Auto-zero amplifiers possess a number of highly desirable characteristics. They are eminently suitable where very high input impedance buffering is necessary. For this reason, they are mostly applied in integrators, electrode buffers, ionization chamber buffers, and electrometer-type applications for femto-ampere current measurement.

Although the clock frequency is in the range from 100 Hz to a few kHz, the gain bandwidth product can be about 1 to 3 MHz or even more. The open-loop, low-frequency gain is $>10^7$ while at the same time offering very low offset voltages in the region of 1 μV. Offset drift, which is a negligible 50 nV per month, gives the auto-zero amplifier the lowest drift of all contemporary types of amplifiers.

The same comment can be made about offset temperature sensitivity, which is ≈5 to 10 nV/°C. The CMRR and the power supply rejection ratio, or PSRR, which are also strongly affected by these low levels of offset, can have values in the order of 140 dB or more. They are also characterized by low voltage and low current noise.

The high-speed chopper action means that the low-frequency noise is nearly uniform down to zero frequency. This should be compared with the flicker noise problems of other noncompensated amplifiers, whose flicker noise corner frequency can be in the range from ≈5 to 10 Hz. Auto-zero amplifiers use MOS input devices, which ensures that the bias currents required are in the range 10 to 20 pA at 25 °C while exhibiting drifts of about 0.5 pA/°C. This is a considerable advantage if large value source resistances are likely. Chopper radiation can sometimes cause problems due to intermodulation distortion. This type of distortion, which is very similar to aliasing in sampled data theory, can be reduced by a careful layout design.

5 MAGNETIC AND OPTICAL ISOLATION AMPLIFIERS

The isolation amplifier is particularly appropriate in data acquisition systems where galvanic isolation (i.e. no DC electrical connection path) between the input and the output circuits is important such as in the following examples:

- amplification of low-level measurement signals in high common-mode environments;
- breaking of ground loops;
- isolation protection of personnel and equipment due to hazardous common-mode situations such as in biological monitoring.

They also allow effective utilization of gain by inserting the gain at the signal source where the bandwidth is likely to be lowest. The isolation barrier may be a screened transformer or an optocoupler.

Isolation amplifiers, which are represented in Figure 4(b), appear inside the isolation units of the general system architecture shown in Figure 4(a). A generalized data acquisition system in an industrial process environment, requiring multiple signal isolation solutions, is given in this system block diagram.

A diversity of signals is acquired from a range of resistance type transducers, voltage sources, current sources, and perhaps other types of transducers such as thermocouples.

Each of these signal sources is associated with an individual conditioning card, whose function is to perform any necessary energizing power supply, screening, guarding, and grounding, and perhaps preamplification. Power supply and control lines connect the conditioning card to the isolation units.

Isolation amplifiers consist of three parts. The first part is a power supply, which is necessary for correct

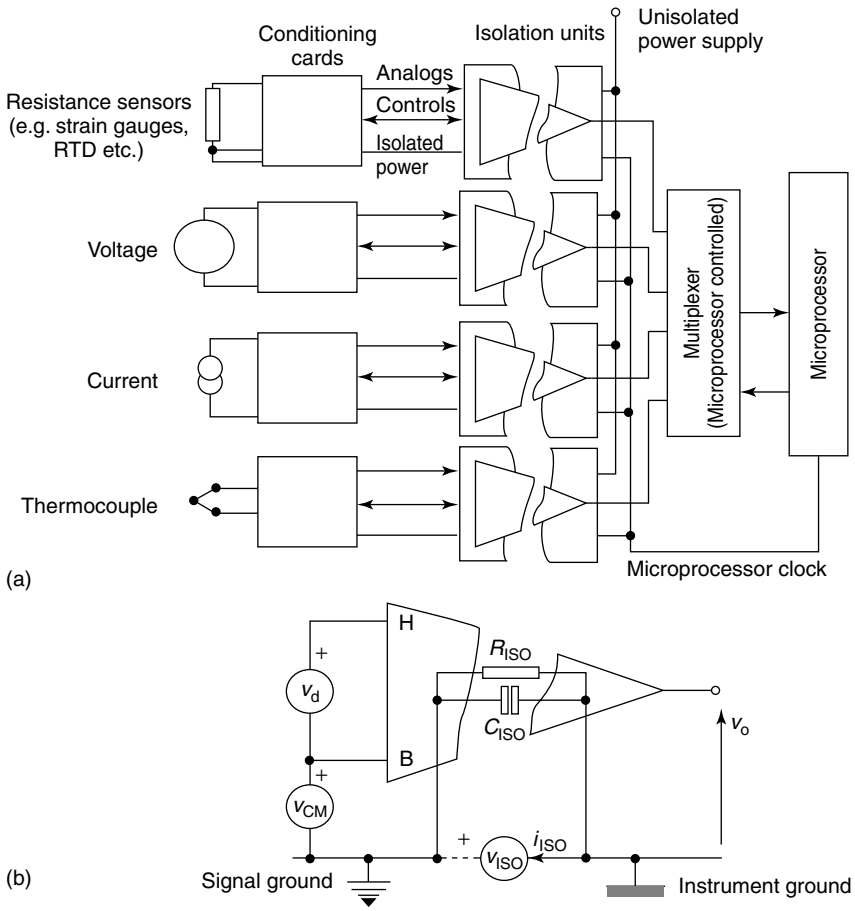


Figure 4. A data acquisition system using isolation units containing isolation amplifiers in (a) has a model for isolation amplifiers given in (b).

operation. It is arranged using some kind of isolating and chopping DC-to-DC converter. An input stage with floating input connections and an output stage with output connections are the other two subunits. All of the subunits are electrically isolated from each other. Residual and stray impedances will always be associated with practical isolation amplifiers. As in Figure 4(b), isolation resistance and isolation capacitance, which are the most important parameters influencing the effectiveness of the isolation between the signal input terminals and the output terminals, have respective values of 10^{10} to $10^{12}\,\Omega$ and 10 to 20 pF, so ensuring isolation potentials up to about 5 kV and leakage currents $i_{ISO} \approx 0.5$ to $2\,\mu$A.

It can be seen from Figure 4(b) that the effect of v_{ISO} is very similar to v_{CM}. For this reason, it can be taken into account in numerical evaluations by using a figure of merit called the *isolation mode rejection ratio*, or *IMRR*, which is defined to give

$$v_o = G\left(v_d + \frac{v_{CM}}{CMRR}\right) + \frac{v_{ISO}}{IMRR} \qquad (6)$$

Where bandwidth and response speed are important, optical isolation amplifiers are preferable to the magnetic type, which are best suited for gain accuracy and linearity.

Magnetic barrier isolation amplifier: The most commonly encountered form of isolation amplifier, which has a block diagram similar to that in Figure 5, uses magnetic isolation barriers in the form of miniature toroidal transformers.

A high-power oscillator is magnetically coupled through a DC-to-DC converter to provide the power supply for the input stages. It is necessary to provide two stages of isolating modulation and demodulation units. One of these units allows transmission of the signal from the input across the magnetic isolation barrier to the output. A second modulator and demodulator allows transfer of the output signal back into the isolation side for feedback purposes.

Magnetic barrier isolation amplifiers can cope with typical continuous DC isolation voltages of 2 kV or a value of 1.4 kV rms for alternating isolation voltages. The

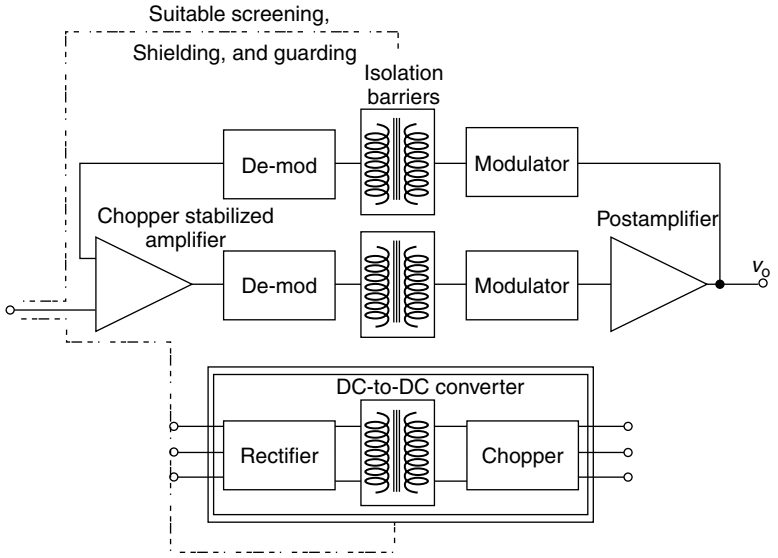


Figure 5. Block diagram of a magnetic barrier isolation amplifier.

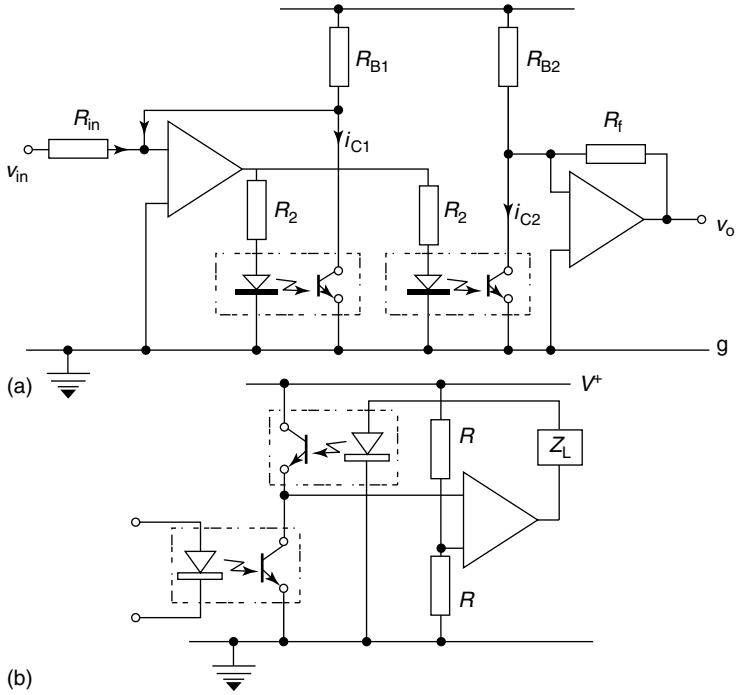


Figure 6. Optical barrier isolation amplifier.

IMRR is about 160 dB for direct voltages and 130 dB for alternating forms. Leakage current can be about 0.5 μA.

Optical barrier isolation amplifier: Figure 6 illustrates two versions of optical isolation amplifiers for two different possibilities. When optoisolation of analog voltages is required, Figure 6(a) shows how feedback in the preamplifier can be used to linearize the response of the optoisolator.

The current in the input LED, which requires a standing bias current, I_B, to ensure full range input swing, is controlled by the input op-amp. The standing bias is $I_B = V^+/R_B$. Hence, the collector current of the transistor in the input optoisolator is

$$i_{C1} = \frac{v_{in}}{R_{in}} + \frac{V^+}{R_{B1}} \tag{7}$$

If the two optocouplers are closely matched, the currents, i_{C1} and i_{C2}, in the collectors of the two transistors will be equal. It can be seen that the output op-amp is configured as a current to voltage converter so that its output voltage is

$$v_o = \frac{R_f}{R_{in}} v_{in} + \frac{V^+}{R_{B1}} - \frac{V^+}{R_{B2}} = \frac{R_f}{R_{in}} v_{in} \quad \text{for} \quad R_{B1} = R_{B2} \tag{8}$$

In the case when the measured variable has been converted to a current, a circuit like that in Figure 6(b) can be used. The input current, i_{in}, which consists of a standing bias current, I_B, and the signal current, i_s, means that the collector current of the input transistor $i_{C1} = i_{in}$. The output current $i_o = v_o/Z_L$, which is fed back to constrain the potential at the inverting input, I, of the op-amp to consist of a component due to the standing bias current at the input and another component proportional to the signal current. The component due to the standing bias current is backed off with the potential at the noninverting input, N. The output current is therefore directly proportional to the input signal current, i_{in}.

Optoisolators have DC isolation voltages up to about 2 kV, which becomes about 5 kV for a pulse test. Their rejection ratio is usually better than 140 dB for DC conditions or 120 dB at 50 to 60 Hz. Their leakage current is typically 0.25 μA and isolation leakage capacitance ≈2 pF.

6 NECESSITY FOR THE CHARGE AMPLIFIER

Quartz and barium titanite are asymmetrical crystalline materials that produce an electrical charge when weights are placed on them. Piezoelectric transducers (PZT), which are used for force measurement, particularly in accelerometers, utilize this property where the produced electrical charge, Q, is proportional to the applied force F. As a consequence, PZT can be modeled as a current source $i = dQ/dt$ in parallel with a capacitor, C_c.

Transducers are often remote from measurement amplifiers in practical measurement systems. The connection between the transducer and an amplifier, as in Figure 7, is realized with coaxial cables. These present a capacitance C, depending on the cable length, appearing in parallel with the amplifier input resistance R_{IN}.

Referring to Figure 7, it can be seen that

$$\begin{aligned} V_c(s) &= \frac{sR_{IN}}{1 + sR_{IN}(C_c + C)} Q(s) \\ &= \frac{sR_{IN}(C_c + C)}{1 + sR_{IN}(C_c + C)} \left[\frac{Q(s)}{C_c + C} \right] \\ &= \frac{(s/\omega_{ci})}{1 + (s/\omega_{ci})} \left[\frac{Q(s)}{C_c + C} \right] \end{aligned} \tag{9}$$

The stage acts like an active high-pass filter, whose Bode frequency response in Figure 7 has a cutoff frequency, f_{ci}, given by

$$f_{ci} = \frac{\omega_{ci}}{2\pi} = \frac{1}{2\pi R_{IN}(C_c + C)} \tag{10}$$

As the frequency response at low frequencies is restricted, slowly varying phenomena cannot be measured. A charge amplifier is needed to interface the PZT so that the high-pass corner frequency can be reduced.

Since the output voltage of the transducer depends on C and thus on the cable length, a new calibration has to be carried out every time the cable length is changed. This is an unfortunate drawback of this measurement system.

7 PRINCIPLES AND PRACTICALITIES OF A CHARGE AMPLIFIER

In the simplified charge amplifier of Figure 8, assume that the amplifier, A, is ideal as described previously in Section 1 of **Article 122, Operational Amplifiers, Volume 2**.

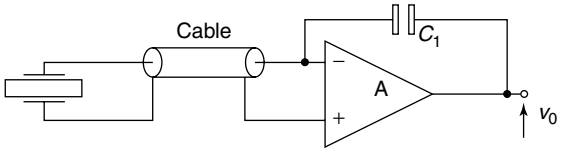


Figure 8. Functional diagram of a charge amplifier.

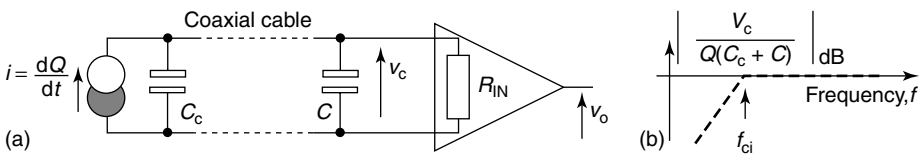


Figure 7. A piezoelectric transducer in (a) connected to an amplifier with a coaxial cable with a capacitance depending on its length has the Bode frequency response in (b).

The charge, Q, produced by the transducer will be stored in the capacitor, C_1. Therefore the amplifier output voltage will be equal to

$$v_o = \frac{Q}{C_1} \tag{11}$$

As this is no longer dependent on C and R, recalibration is not necessary when the cable length is changed.

Charge amplifier with a real op-amp: Suppose the amplifier is nonideal with finite values of gain, A_0, and input resistance, R_{IN}. Using the relations in Section 7 of **Article 122, Operational Amplifiers, Volume 2** with appropriate modifications, the amplifier output signal can be expressed as

$$v_o(s) = -Q(s)\frac{A_0 s R_{IN}}{1 + sR_{IN}[C_c + C + C_1(1 + A_0)]} \tag{12}$$

Owing to the Miller effect transformation, the feedback capacitance appears as a high-value capacitor $C_1(1 + A_0)$ in parallel with the cable and the transducer capacitance. Since $C_1(1 + A_0) \gg (C_c + C)$, the output signal is effectively independent of the cable length.

By contrast with (10), the cutoff, or corner frequency, is now

$$f_c = \frac{1}{2\pi R_{IN} C_1(1 + A_0)} \ll f_{ci} \tag{13}$$

The high input resistance, R_{IN}, implies the need for an op-amp with a low input bias current. This can be met by using either op-amps with FET input stages or BJT input stages, specifically designed with low input bias currents. Although FET input op-amps have well under 1-nA input bias current at 25 °C, this current rises with increasing temperature to values above that of bipolar input types. This small bias current, i_b, causes an output voltage that increases with time in the manner

$$v_o = -\frac{i_b}{C_1}t \tag{14}$$

Hence, care must be taken to choose an adequate op-amp in the design of charge amplifiers. It is possible to avoid this integration effect by placing a large resistance, R_1, with a value of the order of 1 to 10 MΩ in parallel with C_1. Unfortunately, this also causes an output offset voltage with a value of

$$\Delta v_o = -R_1 i_b \tag{15}$$

8 INTEGRATORS

The need for integrators and differentiators often arises in measuring systems. Integrators are employed when the acceleration measured has to be converted to velocity or when velocity has to be converted to displacement. Conversely, to obtain velocity from displacement, differentiators are required.

The basic integrator circuit is shown in Figure 9(a). It has limited practical use because the nonzero input offset errors produce high output offset errors when they are integrated at very low frequencies.

These errors can be minimized by reducing the gain at very low frequencies by the low-frequency feedback in the 15-MΩ resistor in Figure 9(b). Though this results in nonideal integrator characteristics, the final circuit is useful for many applications over a limited frequency range. For the component values indicated, the phase lag of this circuit is $\pm 270^\circ$ $\pm 0.5^\circ$ compared with $+270^\circ$ for an ideal inverting integrator working over a frequency range of 10 Hz to 10 kHz. At very low frequencies, the gain is limited to -1000, resulting in output drift of less then a few millivolts.

The gain, which is approximately -10 at 10 Hz, reduces to about 0.01 at 10 kHz. To ensure that the output at 10 kHz does not become comparable with the drift v_{in} at this frequency, this should be of the order of a few volts. The phase error at the high-frequency end is the direct result of the limited unity gain bandwidth of the op-amp used. Low-frequency phase errors are related to the basic

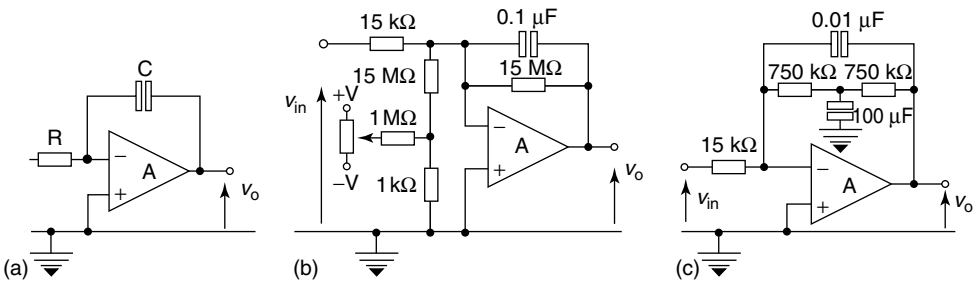


Figure 9. Integrator with the basic configuration of (a) is practically realized using (b) whose performance with frequency is improved with (c).

low-pass characteristics of the configuration used with a −3 dB cutoff, or corner, frequency of approximately 0.1 Hz, relative to DC gain. The useful operating range of this integrator can be shifted to a frequency range of 1 Hz to 1 kHz by increasing the feedback capacitance to 1 μF.

Further reduction of DC errors can be obtained by providing higher feedback for DC and making this feedback ineffective for AC use in the desired frequency range. The arrangement shown in Figure 9(c) has a DC gain of 100, reducing the output DC errors by a factor of 10, as compared to the performance of the integrator shown in Figure 9(b).

9 DIFFERENTIATORS

The basic differentiator shown in Figure 10(a) has limitations. Owing to the 6 dB/octave rolloff of the op-amp gain at high frequencies and the differentiating characteristics of the configuration at the operating frequency, the closed-loop gain shows a peak in its frequency response, similar to that in Figure 10(b). Using a low cost op-amp with the parameters in Section 1 in **Article 122, Operational Amplifiers, Volume 2** and a value of $RC = 0.16$ s, the gain is unity at 1 Hz with the peak at a frequency of approximately 1 kHz. The damping factor, ζ, is obtained from the equation

$$\frac{V_o}{V_i} = \frac{-RCs}{1 + \frac{1}{A_0}(\tau_A + RC)s + \left(\frac{RC\tau_A}{A_0}\right)s^2} \tag{16}$$

as

$$\zeta = \sqrt{\frac{(\tau_A + RC)^2}{RC\tau_A}}\frac{1}{A_0} \tag{17}$$

where A_0 is the low-frequency open-loop gain and τ_A is the open-loop time constant of the op-amp. If $A_0 = 10^5$ and $\tau_A = 0.016$ s, then $\zeta = 1/1000$ for $RC = 0.16$ s. This corresponds to a unity gain bandwidth product of 1 MHz.

The peaking of the frequency response means that the differentiator, however, gives a very noisy output since it amplifies its own internal noise in the 1-kHz range by as much as 10 times. If the amplifier has a response rolling off at more than 6 dB/octave, the circuit may even oscillate.

The modified differentiator shown in Figure 10(c) can be used in practical situations since the closed-loop response is designed to be well damped. With the values shown in Figure 10(c), the response is typically as shown in Figure 10(d). The resistance in series with the input circuit capacitance flattens the response at about 30 kHz. At 100 kHz, the response starts to roll off because of the influence of the open-loop gain of the op-amp. For this to happen at 100 kHz, the unity gain bandwidth of the op-amp should be ≈10 MHz. This can be achieved readily with many op-amps using an external compensating capacitor with a typical value of 3 pF. The phase shift is 90° lead, in addition to the 'inversion' of the amplifier, with less than 2° error for all frequencies less than 1 kHz. At frequencies of

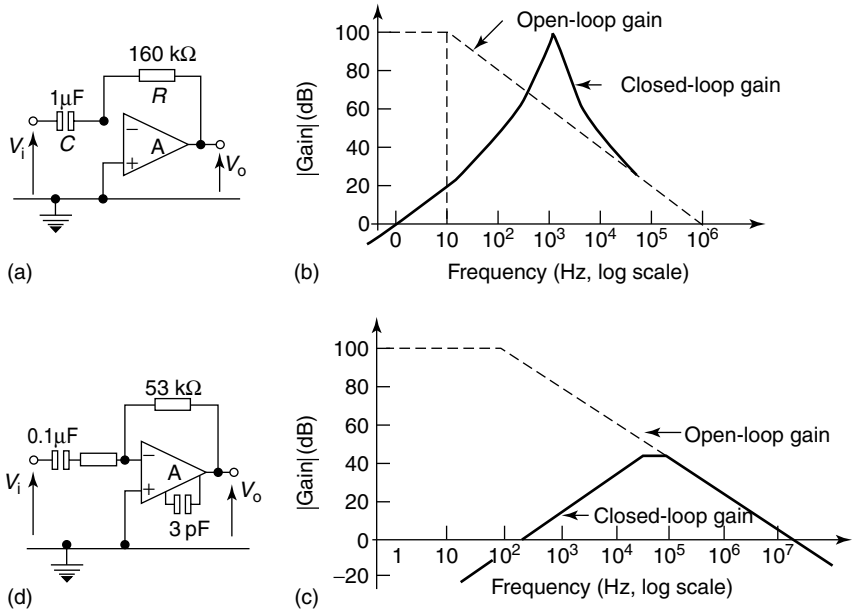


Figure 10. A basic differentiator in (a) has the frequency response in (b), while the modified differentiator of (c) has the frequency response of (d).

0.1 Hz and below, the reduced gain for the input signal can affect the performance. A differentiator for the frequency range of 0.01 to 100 Hz is achieved by using the op-amp with 30 pF compensation and by changing the input circuit capacitance to 1 μF.

RELATED ARTICLES

Article 122, Operational Amplifiers, Volume 2; **Article 123, Instrumentation Amplifiers, Volume 2**; **Article 133, Amplifiers and Filters for DAS, Volume 3**; **Article 181, Amplitude Modulated Signals: The Lock-in Amplifier, Volume 3**.

FURTHER READING

Huijsing, J.H. (2000) *Operational Amplifiers – Theory and Design*, Kluwer, Amsterdam.

Stanley, W.D. (2001) *Operational Amplifiers with Linear Integrated Circuits*, Prentice Hall, NJ.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

This article is based in part on a chapter by J. McGhee *et al* originally published in Measurement Data Handling: Hardware Techniques Vol. 2, J. McGhee *et al*, Technical University of Lodz, 2001.

125: Outline of Purpose of Analog Data Filters

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Role of a Signal Filter 860
2 Features That Need Selecting or Filtering Out 861
3 Classification of Active Filter Types 862
4 Application of Filters 862
5 Comparison of Analog and Digital Filters 862
6 Overview of Active Analog Filter Design Methodology 863
7 Proprietary Filter Modules 863
Related Articles 863
References 863

1 ROLE OF A SIGNAL FILTER

It is always the case that the signal acquired from a sensor or transducer requires conditioning of some kind or another. Even after passive conditioning with constant voltage, constant current or bridge circuits, or active conditioning with various kinds of amplifiers, there will still be the need to introduce some form of filtering. Filters, which usually are regarded as modifiers in the terminology of **Article 109, Structure and Energy in Sensor Systems, Volume 2**, are specially designed to modify the range of their transmitted frequencies. Filtering operations may treat analog or digital signals. Although the filters used in measurement and instrumentation use electric circuit realizations, filters exist in other energy forms. Electric filters are also called *electric wave filters*.

The frequency ranges of filters can be used as a means of defining their different types. For example, filters using passive *R-L-C* elements, frequently referred to as *passive filters*, are used in the range from about 100 Hz to >100 MHz. *Active filters*, which are widely applied in the frequency range from <0.01 Hz to about 100 kHz, are normally realized using individual resistors, capacitors, and op-amps. Microwave filters, also called *cavity resonators*, as their name implies, operate in the microwave frequency range above about 10 to 50 MHz. Crystal filters using piezoelectric techniques cover the range from about 100 Hz to the low microwave region. Monolithic filters, which use large-scale integration for their realization, cover the same basic range as active filters but have the possibility for higher frequency operation because of their smaller general physical dimensions. This is because the *R-C* elements used are normally included on a chip.

Analog filters are classified as passive if they use only passive *R-L-C* elements. More recently, active filters using combinations of passive *R-C* elements and active op-amps have been widely developed. Digital filters, which process discrete interval sampled data signals, are realized using dedicated digital hardware structures or software programs. The digital hardware used takes the form of shift registers, multipliers, and accumulators.

The concepts of filtering provide the measurement system designer with additional tools to tackle the problems of minimizing error, signal-to-noise ratio, response time, and bandwidth. At the same time, analog filters allow modification of the frequency content of the acquired signal so that the problems of *foldover* associated with digital signal processing are at least minimized by the use of *antialiasing* precautions. Before the era of ICs, filters were exclusively realized using passive *R-L-C* elements. Presently, the op-amp has exerted a profound influence upon the design of data filters for analog signals. This

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

Table 1. Advantages and disadvantages of active filters.

Advantages	Disadvantages
1. Small in size and weight because there are no inductors	1. Limitations on voltage/current levels makes them unsuitable in electrical power systems
2. Fairly easily designed	2. Inadequacies due to imperfections in the static and dynamic characteristics of the active elements
3. Easily 'tuned' to give the desired response	3. Parasitic oscillations may occur when feedback is used
4. Isolation between cascaded stages	4. High sensitivity to tolerances of passive components
5. Large available gain so amplification is possible	5. A power supply is required

means that for frequencies up to about 100 kHz there is no need for inductors.

The design techniques for active analog filters are largely the same as for passive filters. However, as active networks contain op-amps employing feedback, the stability of the filters needs to be considered. Also, as inductors are bulky, it is preferred that only R and C elements are used, even though inductors may be realized using gyrators. Gyrators will not be considered further. A general treatment is quite difficult as there is such a large variety of possible methods. Table 1 gives advantages and disadvantages of active filters.

2 FEATURES THAT NEED SELECTING OR FILTERING OUT

Filtering is essentially a signal-processing operation to obtain optimal detection of some specified parameter or parameters of a signal immersed in interference or noise. Problems of filter design are concerned with optimizing the preservation of the important signal parameters and removing or at least minimizing the effect of the noise.

The first item of interest is the signal and the manner in which it carries the measurement information. Information can be carried in the amplitude, the phase, or the frequency of an analog measurement signal, or perhaps in the shape in the case of a pulse modulated signal. When the amplitude carries the information, the design of the filter concentrates upon uniform amplification of the frequencies within the band of interest so that the relative magnitudes of each frequency component are maintained after the filtering operation. Ideally, all signal components outside the band of interest should be attenuated by as large an amount as possible. Using the same reasoning, filter design concentrates on optimizing the phase characteristic of the filter if the information is carried in the phase angles of the signal components. In this case, optimal retention of the relative phases of the frequency components within the band of interest after filtering is the aim. This is especially important if the shape of the waveform needs to be maintained. In some cases, information is carried in both the amplitude and the phase of the measurement signal. Simultaneous preservation of relative amplitude and relative phase difference are conflicting requirements. In these cases, it is necessary to design some filter that compromises between the two requirements.

Interference, more widely known as noise, occurs in all measurement systems. It is shown in **Article 68, Setting the System Boundaries, Volume 2** that noise may penetrate a measurement system from the outside, in which case it may be called *interference* or *interfering noise*. Noise may also arise inside a measurement system owing to imperfections in system design or of the materials used in its construction. All of these noise sources contribute to the overall error. Errors should be minimized to realize the overall aim of measurement, which Solopchenko (1994) asserts as being the discovery of errors and their minimization.

One of the most commonly occurring problems in measurement and instrumentation systems is due to interfering pickup, or 'hum', from the electric supply systems used in the electrical elements of the overall system. Thus, undesirable inductive and capacitive coupling into the more sensitive parts of the system from nearby electrical sources needs to be guarded against. Removal of this type of interference may also be accomplished using a band-reject filter, sometimes also called a *notch filter*. Other kinds of interference may be due to cosmic effects.

Naturally occurring fundamental fluctuations are endemic throughout nature. In mechanical systems, they are manifested in Brownian motion, while in electrical systems, they are referred to as thermal, Johnson, or Nyquist noise. These fluctuations, which are due to the equipartitioning of energy, have a uniform power spectral density up to about 10 GHz. Quantum corrections must be applied beyond this frequency. Since most measurement systems use electrical/electronic building blocks, other types of noise such as flicker, or low frequency noise, shot noise, and other types of noise characteristic of semiconductor components require precautions to be taken to minimize their effects.

In some cases, the measurement information is associated with a narrow range of frequencies. To ensure that unwanted components outside this range are not transmitted in the measuring system, band-pass filters are used.

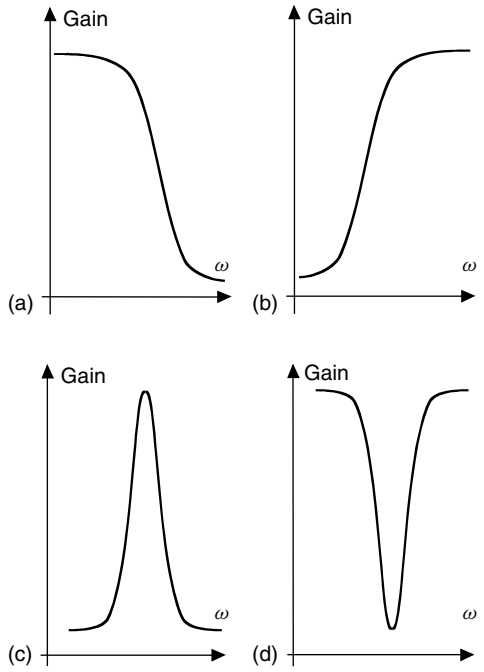


Figure 1. Classes of filters: (a) low pass, (b) high pass, (c) band pass, and (d) band stop/notch.

3 CLASSIFICATION OF ACTIVE FILTER TYPES

There are four main types of frequency responses realized using active filters as in Figure 1. Low-pass sections, which widely occur in instrumentation systems, are used to overcome some limitations of a sensor signal by removing unwanted higher frequency noise and providing some antialiasing protection. The other three are the high-pass section, the band-pass section and the band-reject or notch-filter section. All-pass sections, which are not shown in Figure 1, are used as phase equalizers. They possess a flat amplitude response with a phase response, which can be designed to a given value of phase delay or advance.

4 APPLICATION OF FILTERS

The rapid development of large-scale communication systems provided a major impetus in the need for and development of both passive and active filters. In this field, an very important application is the pulse-code modulation coder/decoder, or PCM-CODEC, in digital telephony. In this application, the filters require very tight specifications of both their cutoff sharpness and pass band ripple. Other communication system examples of filtering occur in signal preemphasis, equalization, impedance matching in repeaters, and for harmonic suppression in some low-frequency applications.

A major application area of filters is in measurement and instrumentation systems. When measuring voltages with digital voltmeters, it is important to reject interference from noise sources and from power line harmonics. Similar requirements for noise rejection occur in the study of vibration and shock. Many physiological signals such as those occuring in heart monitoring by electrocardiogram, or ECG, or in brain wave monitoring by electroencephalogram, or EEG, have the information content of their spectrum in a three-decade band below about 20 Hz. Special filters to remove ultralow-frequency interference below about 0.002 Hz, and noise penetration above 20 Hz are required.

Electric wave filters, and especially active filters, are also widely used in control systems, such as servo-systems and motor drive controllers, as well as in process control equipment, consumer and entertainment systems such as stereo amplifiers, videocassette recorders, or VCRs, digital versatile disc, or DVD, systems and in all types of public broadcast receivers.

Applications also occur in sonar systems, telemetry systems, prefilters for analog-to-digital converters, spectrum analyzer systems and many others.

5 COMPARISON OF ANALOG AND DIGITAL FILTERS

Some advantages and disadvantages of analog filters have been given above. Digital filters take the form of recursive filters, also called *infinite impulse response filters*, or *IIR filters*, and nonrecursive filters, sometimes referred to as *finite impulse response filters*, or *FIR filters*.

Digital filters have some significant advantages over analog filters. In the first place, digital types possess guaranteed accuracy, which is only limited by the number of bits, and guaranteed reproducibility since there are no variations due to component tolerances or aging or temperature drift. In contrast, the accuracy and reproducibility of analog filters cannot be guaranteed because they can depend heavily upon the tolerance of the R and C elements, upon the characteristics of the op-amps used, and upon the aging and temperature drift effects, which they are all subjected to. Thus, digital filters have superior performance over analog types. The flexibility of digital filters arises from their programmable nature, which allows a variety of functions to be performed using a single hardware structure. It is impossible to provide this flexibility with analog filters.

Nevertheless, digital filters possess three main disadvantages. Designing DSP can be time consuming unless the

development resources and knowledge are available. Realizing DSP can be quite expensive, especially with wideband signals. Such signals need fast, and thus expensive, analog-to-digital and digital-to-analog converters. Because real-time applications are implemented using only a limited number of bits, there is a real possibility for errors due to quantization effects.

6 OVERVIEW OF ACTIVE ANALOG FILTER DESIGN METHODOLOGY

Only the design of active analog filters will be described. Many authors have described the methodology (Bowron and Stephenson, 1979; Chen, 1986; Horowitz and Hill, 1980; Huelsman, 1993; Lang, 1987; Ludeman, 1987; McGhee *et al*., 2001; Rangan, Sarama and Mani, 1983; Schaumann and Van Valkenburg, 2000; Temes, 1985; Winder, 2002).

In all cases, the design commences with a specification of the filter frequency characteristics based upon the information bearing parameters of the signal, its frequency range, and the likelihood and types of any noise and their sources. This specification will most likely lead to an ideal filter, also called a *brick-wall filter*. If the specification points in the direction of high-pass, band-pass or notch sections, they are transposed to an equivalent low-pass region for a low-pass design. This design is then retransposed using suitable frequency domain transformations to gain the desired design. Since ideal filters cannot be realized in practice, a suitable approximation must be designed on the basis of the four main possibilities. These are

- the maximally flat amplitude filters, or *Butterworth filters*;
- the maximally flat delay filters, also called *Thomson filters* or *Bessel filters*;
- the equiripple pass-band filter, or Chebyshev type 1 filter;
- the elliptic filter, or Chebyshev type 2 filter.

Both kinds of Chebyshev filters are considered in the references given at the end of this article.

In this way, the three main frequency regions of a low-pass filter may be described. If a band-pass, high-pass, or band-notch filter is required, these may be achieved using suitable low-pass-to-band-pass, low-pass-to-high-pass, and low-pass-to-notch transformations (Ludeman, 1987; Rangan, Sarama and Mani, 1983) features.

7 PROPRIETARY FILTER MODULES

There are many IC manufacturers such as Burr-Brown, now a division of Texas Instruments, Maxim (http://www.miximic.com), Linear Technology (http://www.linear.com) Mixed Signal Integration, or MSI (http://www.mix-sig.com), and many others.

RELATED ARTICLES

Article 75, Electronic System Building Blocks, Volume 2; **Article 121, Signals in the Presence of Noise, Volume 2**; **Article 133, Amplifiers and Filters for DAS, Volume 3**.

REFERENCES

Bowron, P. and Stephenson, F.W. (1979) *Active Filters for Communications and Instrumentation*, McGraw-Hill, London.

Chen, W.K. (1986) *Passive and Active Filters: Theory and Implementations*, John Wiley & Sons, New York.

Horowitz, P. and Hill, W. (1980) *The Art of Electronics*, Cambridge University Press, Cambridge, MA.

Huelsman, L.P. (1993) *Active and Passive Analog Filter Design: An Introduction*, McGraw-Hill Education.

Lang, T.T. (1987) *Electronics of Measuring Systems*, John Wiley & Sons, Chichester.

Ludeman, L.C. (1987) *Fundamentals of Digital Signal Processing*, John Wiley & Sons.

McGhee, J., Henderson, I.A., Kulesza, W. and Korczynski, J. (2001) *Measurement Data Handling: Hardware Technique*, ISBN 83 7283 008 8, A.C.G.M. Lodart S. A., Lodz.

Rangan, C.S., Sarama, G.R. and Mani, V.S. (1983) *Instrumentation Devices and Systems*, Tata McGraw-Hill, New Delhi.

Schaumann, R. and Van Valkenburg, M. (2000) *Design of Analog Filters*, Oxford University Press.

Solopchenko, G.N. (1994) Formal Metrological Components of Measuring Systems. *Measurement*, **13**, 1–12.

Temes, G.C. (ed.) (1985) *Selected Papers on Integrated Analog Filters: Advances in Circuits and Systems*, ISBN: 0879422157, IEEE Publications.

Winder, S. (2002) *Analog and Digital Filter Design*, Butterworth-Heinemann.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

This article is based in part on a chapter by J. McGhee *et al* originally published in Measurement Data Handling: Hardware Techniques Vol. 2, J. McGhee *et al*, Technical University of Lodz, 2001.

126: Electrical Bridge Circuits – Basic Information

Zygmunt L. Warsza
Polish Metrological Society, Warsaw, Poland

1 Balance in Measurement Circuits – General Remarks 867
2 Basic Bridge Definitions 868
3 Wheatstone Bridge and History of Bridge Developments 869
4 Bridges in Measurement of Nonelectrical Quantities 871
5 Bridge Circuits Classification 871
6 Z Matrix Equations of Two Port 872
7 Impedance Bridge as Two Port and its Z Matrix 873
8 Basic Equations of Unbalanced Bridges 875
Related Articles 876
References 877
Further Reading 877

1 BALANCE IN MEASUREMENT CIRCUITS – GENERAL REMARKS

In the measurement process, special circuit states, such as equality of two nodes' potentials or of two branch currents connected, for example, in the common node as shown in Figure 1(a), are of special significance. Such states are known as balanced states. Hence, existing relations between circuit parameters are called *balance conditions*.

A circuit could have more than one state of balance. If the potentials of two nodes are equal, it is possible to connect them together directly, or by some impedance, without changing the current and voltage distribution in all the circuit branches.

Similarly, after disconnecting two such linked branches from the node, no changes will appear because the currents are opposite to each other – see Figure 1(b).

The equations describing balance conditions are usually much simpler than those describing the circuit in general conditions, and, therefore, are often used for measurements.

An unknown parameter, or combination of a few of them, is measurable if it is possible to calculate them from balance equations as a function of other known values of parameters obtained after balancing the circuit.

Opportunities given by balanced states are used both in a single bridge, and also in the more sophisticated measurement circuits including a bridge as a part of them.

In the circuits with time-variable signals, the variety of other possible balance states are also applied for measurements, for example, the equality of

- node potentials or branch currents as above, but only in some instants of time, for example, as in successive approximations balance circuits, and A/D converters;
- integrals of two signals: in the different types of integral A/D converters;
- both, module and phase of two pure sinusoidal signals of the same frequency, or of harmonic components of the differently disturbed periodic signals;
- only amplitudes, or only phases called the *quasi-balanced state*;
- maximum, mean, or RMS values of two signals of different shapes.

Other possible balance states also exist.

External and internal voltage and current sources, of stationary magnitude or that are adjustable, are used in the measuring circuit.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

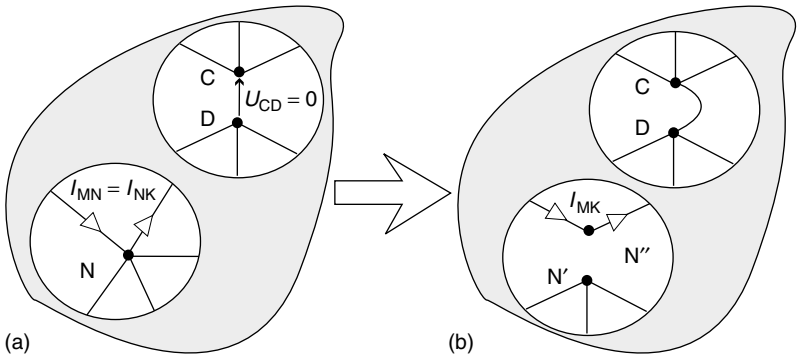


Figure 1. (a) Examples of the voltage and current circuit balance states and (b) possible modifications of the circuit.

Adjustable parameters of circuit elements may be changed or modulated, either manually or automatically, inside or outside the main circuit, such as in a bridge.

If, in the measurement process, the relationships between circuit parameters in the balance state are only taken for obtaining measuring results then such a method of measurement is called a *balanced* or *null-balance* method.

If changes of parameters result in a different state from the null-balance one, such a method is a differential one and the circuit is called *deflection*, *off-balance*, or *unbalanced* one.

Both methods offer high accuracy and are widely used not only in electrical measurements.

2 BASIC BRIDGE DEFINITIONS

An electrical measurement circuit including at least one quadrilateral topology network having all possible branches present is called a *bridge circuit*, or in short, the *bridge*.

Usually, such a circuit is excited (driven, powered supplied) across one of its diagonals and has the output on the other one (see Figure 2 as an example of the bridge).

Bridges could work as balanced or unbalanced measurement circuits.

Discovery of the various bridge circuits resulted originally from the measurement needs of electricity and magnetism science for an explanation of its historical emergence – *see also* Sydenham (1979).

Later, their development became necessary for the research and practice of a rapidly growing electrical technology, telecommunication, and other disciplines.

More accurate measurements of the resistance R and conductance G in the direct current (DC) circuits were initially needed. Later, alternating current circuit parameters such as resistance R, inductance L, capacitance C, mutual inductance M, components of impedance Z or admittance Y, frequency, and other auxiliary parameters were measured in the continuously rising range of frequencies that

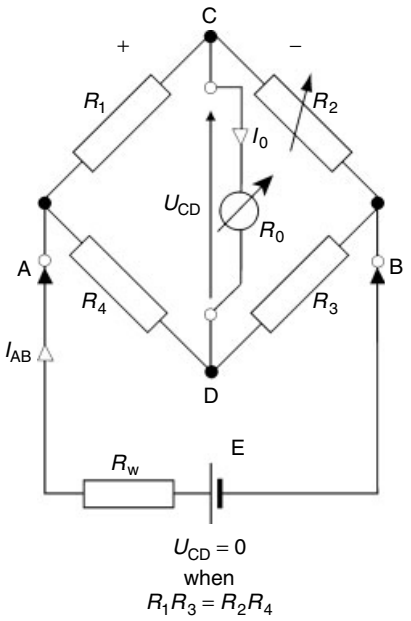


Figure 2. Wheatstone bridge and its balance condition $R_1, \ldots R_4$ – resistance of bridge arms; R_0 – resistance of galvanometer (bridge load resistance R_L); R_w – source internal resistance (series resistance to the primary bridge circuit R_G).

find application. They not only included measurements of parameters of the two terminal components but also their values in circuits with different structures of lumped and distributed elements.

Presented here is the basic unchanging foundational knowledge useful for setting up and understanding practical bridge application.

The term *bridge* is used with two different meanings: for a circuit and for a measuring device utilizing this circuit as its main part. During recent years, the term 'bridge' has also been used sometimes by some instrument producers, for a device measuring the same parameters as, for example, RLC bridges, but working on a completely different measuring principle.

Basic properties of different bridge-like circuits and the background of their applications in measurements are presented in detail in the literature given in the end and in **Article 127, Unbalanced DC Bridges, Volume 2**.

Detailed technical data of bridge instruments, converters, and computer measurement systems working with them that are offered on the market are given in the manufacturer's publications, including the Internet.

3 WHEATSTONE BRIDGE AND HISTORY OF BRIDGE DEVELOPMENTS

The history of bridges and their development is some 170 years old now. The DC bridge principle was first introduced by Christy in 1833, but it was unnoticed during the next ten years. In 1843, Charles Wheatstone applied the bridge arrangement for the comparison of resistors in his 'differential analyzer' gaining it a place as a basic electrical circuit. It is, thus, well known since as the Wheatstone bridge.

Together with its balance condition for $U_{CD} = 0$ this bridge is shown in Figure 2. Resistance R_1 is the one to be measured, R_2 is the adjustable resistance and auxiliary resistances R_3 and R_4 are equal and used to set the potential of point D. A battery is connected across A B diagonal. A galvanometer (this was originally an electromechanical detector but is now usually an electronic amplifier) 'bridges' the C D points and is used to detect the balance of their potentials. No current will flow through the galvanometer R_0 if R_2 is adjusted to be equal to R_1. At that state, the value of R_1 follows directly from the obtained value of R_2.

By analogy to the process of mass weighing, resistances R_3 and R_4 were called by Wheatstone as the *arms of the balance*. This term, shortened to *bridge arms* has been used up to now for all the bridge side branches.

Five years later, in 1848, Werner von Siemens developed a simple method of changing the range of R_1 measurements. He applied unequal arms R_3, R_4 with any desired ratio. The balance condition of the Wheatstone bridge given in Figure 2 was obtained from application of Kirchoff and Ohm's laws for linear resistances. It is independent of the voltage value of source E and of resistances connected in both diagonals. Siemen's idea made it possible to use a slide-wire or decade adjustable resistor R_2 for different ranges of R_1. The bridge ratio R_4/R_3 then is changed in set sequences, for example, 10^n.

The range, accuracy, and sensitivity of bridge measurements are limited by

- accuracy, values, and permissible power (as balance currents still flow in all resistors!) loads of R_2, R_3, R_4;
- resolution of adjustable resistor ΔR_2;
- value of applied voltage E;
- sensitivity of the galvanometer now more usually called the *detector* – for details refer to Dally, Riley and McConnell (1993) and Karandeev (1966).

The influence of the thermoelectric voltages formed in these circuits should be avoided when two measurements are made for opposite directions of supply. The correct result is then found as a mean value of the two readings that average out the effects.

The structure of the Wheatstone bridge was later applied to the design of several bridge versions for making R, L, C, and M measurements.

The Wheatstone bridge is also commonly used in off-balance measurements. In such a case, the supplying source E and resistances connected to both diagonals should be set to appropriate values and kept stable during the measurements. Then no additional adjustments are needed to obtain the result.

The output signal level depends on changes of all arm resistances from their values in the balance state. It can be measured by a direct reading instrument or conditioned, converted, and transmitted to remote display devices, or to the computer system.

Measurement of the ratio of the unbalanced bridge output signal and of its supply current or voltage is also used. Not long ago, a quadrilateral structure was also applied in active bridges with computer controlled voltage sources, for example, in the Cutkosky bridge.

The Christy–Wheatstone bridge principle opened the 170-year-long history of bridge circuit developments and their numerous applications in measurement, control, testing, monitoring, and diagnostics. The first 15 years of this history are covered already in Section 3.

In 1862, Sir Thompson Kelvin developed a 'double' – six-arm DC bridge for measurements of small values of resistances, lower than 1 Ω. This bridge provided the means for avoiding the influence of the resistance of connecting lead wires and of their changes during measurements. With this arrangement, small value, high precision four-terminal standard resistances could be easily measured and compared with high accuracy.

The next invention was the application of the bridge circuit to measure and to compare passive parameters such as inductance L, mutual inductance M and capacitance C.

The steady and transient states of the bridge circuit were used together. Firstly, the bridge is balanced in the steady state by adjusting one of the resistances for the DC supply. It is then also balanced in the unsteady state by L or C adjustments of its arms to avoid pulse (AC) movements of the galvanometer working in the ballistic regime. The pulses are generated just after repeated commutations of the supplying voltage or current DC (they can be switched off,

switched on, or reversed) and are imposed. Such 'ballistic bridges' were developed, one after another, by de Sauty in 1871 – for C measurements; by Maxwell in 1873 – for the L, M, LM measurements; by Carry-Foster in 1887 – for the C, M measurements; and also used and improved by Perry, Ayrton, and Brulion. After the invention of the telephone by Bell in 1875, Lord Rayleigh, Hevisade, Hughes, Kohlrausch, and others tried to use it in the above bridges supplied by an induction coil generator, but its voltage was of unknown shape and not of one frequency.

The principle of the first modern AC bridge was introduced by Max Wien in 1891 – for C measurements. He first used an alternating current supply of defined steady frequency and a 'vibrating' galvanometer mechanically adjusted to resonance with this frequency. Its very light moving part was equipped with a small mirror, and a light beam pointer.

Wien's AC bridge, together with such a galvanometer, was subsequently used for several decades, mainly at frequencies up to a few kilo hertz. For higher frequencies, high sensitivity electronic tube (valve) resonance amplifiers were developed.

Presently, for all frequency ranges, integrated, solid state, very high sensitivity, adjustable resonance amplifiers, and also phase detectors and shifters with operational amplifiers are commonly applied in AC bridges. Use of the electromechanical indicator has all but become a museum piece but its performance will still sometimes be useful in standard work.

The balance state: $U_{\mathrm{CD}}(t) = 0$, in the four-arm AC bridge demands that for every frequency of this voltage there is equality of products of the complex impedances of the opposite arms. This results in two real equations for equality of the polar or rectangular components of those two products. Here it is not always possible to obtain the solution.

Furthermore, only a few AC bridges have their balance conditions independent of the signal frequency. In the balancing process, two, but not any two, circuit parameters should be adjusted for a successively smaller and smaller readout from the detector.

Thus, in all AC bridges, the process of balancing is much more complicated than for DC bridges.

In the early days of development, dozens of AC bridge variation were developed for measurement of AC passive parameters under different conditions. Between them in 1920, Schering introduced a high voltage bridge to measure the capacity C of a condenser and its dielectric losses (expressed by tg δ) measurements, used up to now. In 1924, Robinson presented a bridge for frequency measurements.

More about the history of the first period of developments of AC bridges and deep studies of many circuits and their properties can be found in the references at the end, revised by T.R. Foord, edition of Hague and Foord (1971) famous book.

Parallel to this, classical DC bridges have also been continuously improved. In the beginning of the second part of the twentieth century, analog immittance measurements, provided by DC and AC bridges and by DC potentiometers, obtained levels of accuracy, sensitivity, and measurement ranges not very far from the present day ones. However, they were not as 'user friendly' as the modern instruments, because they needed many manual lengthy adjustments to obtain balance and much skilled experimental experience.

During the last fifty years, new revolutionary technological possibilities have been coming up for use in measurements. This has been made possible by the avalanche of inventions and developments in physics and electronics, such as transistors; new almost perfect ferromagnetic materials; Zener diodes; solid-state operational amplifiers (OPAMP)s; analog to digital A/D and digital to analog D/A signal converters; integrated digital and analog solid-state circuits; microprocessors; application specific integrated circuits (ASIC); huge software possibilities; programmable digital signal processors (DSP); and so on.

An important stage in high accuracy AC bridges was that connected with the development of high precision, adjustable induction dividers. They were made on cores of new ferromagnetic materials of very high permeability, very narrow hysteresis loop, and very low losses. These offered highly accurate, up to seven decades, AC voltage division dependent nearly exactly on the ratio of the coils' winding turns.

The accuracy of earlier developed precision resistance decade dividers is limited, because it depends on the stability of the resistor material and of their specialized technology process. The principle of the transformer bridge was shown quite early by Walsh in 1930, but in practice it was only first applied in 1949. Developments for high precision measurements have been due to Hill, Miller, and Decon in the 1962–1968 period.

A new idea in DC bridges was introduced by Milianic and others in 1962, and by Kuster and Moore in 1964–1966. They made use of the high precision adjustable induction current comparator. In this device, the zero of the DC magnetic flux in a ferromagnetic core is precisely detected due to additional AC coils, as in transductors. They gave the possibility of accurate comparing DC ampere turns of a few core windings.

In the book edited by Schnell and Sydenham (1993), Osvad presented the details of the development of the transformer bridges and also of active circuits based on operational amplifiers, both dedicated to immittance measurements.

Some basic information about bridges are in instrumentation handbooks, for example, in one edited by Nolting (1995) and revised by Boyes *et al*. (2003), and also in the latest one by Dyer (2001). Some valuable information about bridges can also be found in several academic textbooks and engineering books, including those given below in the References and Further Reading at the end of this article.

During recent years, research on AC impedance measurement has focused on the development of virtual measurement circuits. High accuracy, high speed, and precise DSP processors are used for that; they make it possible to build virtual bridges. These circuits could include only two hardware arms: the measured and the reference impedance. Results obtained in the laboratory are comparable with the best technical data of hardware automatic digital bridges. New algorithms for such measurements have also been found, but it seems that the stability of the best R or C and M reference elements is still better than that of the A/D converters used as DSP inputs.

4 BRIDGES IN MEASUREMENT OF NONELECTRICAL QUANTITIES

Applications of the bridges in the measurement of nonelectrical quantities by indirect means based on electrical detection began after immittance sensors of various quantities were developed.

Siemens, in 1871, tried to design platinum resistance sensors for temperature measurement, but without success. Callendar, in 1891, described such a sensor and also formulated the relation of its resistance to temperature.

Finally, in 1925, van Dusen developed corrections to this curve, which has now been standardized. According to these contributions, it became possible to build specialized, laboratory and portable temperature bridges. In recent decades, electronic analog and digital intelligent industrial transmitters have been made with bridges at their inputs.

For recording temperature, Leeds, and Northrop Company, about 1940, designed the first automatic bridge recorder. A robust electromechanical galvanometer was used with periodically operated special scissors closing on its deflected pointer. They actuated the clutch of a precision slide-wire resistor balancing the bridge step by step. Such an instrument had enough power to move the ink pen on a slowly moved circular, or strip recording paper. This instrument was still being used in dangerous areas even up to 1975. Their new versions with electronic amplifier and servomechanism were applied later and it is still possible to meet them in industry or in laboratories.

Thompson discovered the piezoresistive effect in 1856, but it was applied firstly in measurements by Simmons (CallTech) and by AC Rugy (MIT) in 1937. Very soon Baldwin Southwork Corp. started the production of metal strain gauges glued on paper background, ready to be stuck to a measured stressed object with a special adhesive. For their applications in static and in dynamic measurements, special AC resistance bridges were developed.

Some other immittance sensors, such as photoresistors and magnetoresistors, can work in the DC or AC bridges. Self- and mutual-inductance sensors and capacitance sensors have to use AC bridges.

In all industrial and laboratory intelligent transmitters, and in measuring cards of computer systems working with immittance sensors, unbalanced bridge circuits of both passive and active forms, are now commonly used. They are the input part of the signal conditioning circuits and still offer many properties not given by other circuits.

Bridge circuits and their theoretical backgrounds are still very useful in measurements, mainly for the following reasons:

- Most physical quantities can be measured using analog immittance sensors connected in the bridge circuit.
- Influence on sensor parameters of additional, not measured quantities, can be easily eliminated by using a differential sensor connected to the half-bridge or two such sensors in the full bridge circuit.
- If the measured changes of a sensors parameters are very small (e.g. lower than 0.5%, as it is in metallic strain gauges), the bridge circuits give accurate output signals.
- Bridges can also easily operate with multiparameter sensors for simultaneous measurements of a few variables.
- The equivalent circuit of four or more terminal sensors that is integrated within a bridge circuit is accessible only from external terminals.
- Bridge circuits are also used in the measurement of tensor components of material properties and of the distribution of multiparameter fields; and in nondestructive testing, impedance tomography, and diagnostics of circuit parameters.

Knowledge of bridge circuits and their properties are very important not only for designers but also for users of instrumentation. Thus, a unified description of bridge circuits is needed.

5 BRIDGE CIRCUITS CLASSIFICATION

Bridge circuits are commonly called by the names of their inventors; bridge instruments – by the names given by their manufacturers; or according to the field of their

applications, or parameters to be measured. A large diversity of bridge circuits, especially AC ones, demands that there is a clear criteria of their classifications. Some of them are presented below. Bridges could be classified according to

- static characteristic of the parameters (validity of the use of the superposition law in the circuit): linear, nonlinear (for at least one parameter), being quasi-linear around the working for relatively small excursions;
- type or number of supplied sources:
  - DC and AC of different frequency ranges,
  - current or voltage sources,
  - low or high voltage,
  - one, or several sources,
  - stationary and controlled sources;
- type of the bridge circuit:
  - passive, active;
- number of branches:
  - four, five, or six arms, and so on.
- number of external terminals:
  - four, six terminal, and so on.
- number of output:
- number of measured parameters;
- number of variable arms;
- variability of the circuit structure:
  - invariable, switched;
- type and range of measured parameters:
  R, RC, RLC bridge, low impedance, high impedance, and so on;
- type of application: high precision, laboratory, portable;
- accuracy of measurement of different parameters for different ranges in percents;
- field of application:
  - temperature bridges, strain gauge bridges, conductometric bridges, and so on.

AC bridges can also be classified according to the phase relations of circuit arms.

6 Z MATRIX EQUATIONS OF TWO PORT

In the general case, a bridge circuit can be considered as multinode, multiloop, and multiterminal [n-T] network in which, for obtaining measurement result, its one, or several branch parameters are changed or modulated externally.

The two Kirchoff Laws are valid in any case, at any time and any part of the circuit having negligible electromagnetic propagation. The circuit is linear if the superposition theorem is valid for it. It can then be described by a set of linear equations: algebraic ones for DC; complex form for AC of one frequency; Fourier transforms for few frequencies; Laplace transforms for any shape of signals in time; and partial differential equations for the distributed parameter circuits.

For a single supply and single output set up, any bridge could be transformed to a two port (two-terminal-pair) circuit – see Figure 3.

A set of equations of any linear circuit working as a two-port device could be reduced to a set of only two linear relations between currents and voltages at their ports.

It is thus possible to write these relations in six different ways; and for simplicity in matrix form. Then the one of the following matrices expresses the two-port parameters: **Z**, **Y**, **A**, **B**, **H**, **H′**. If it is needed, anyone of them could be easily transformed to another one.

For a particular circuit, some of the above matrices will not exist at all or their parameters are infinite in some states, for example, parameters of the transmission matrix **A** of the balanced bridge – see Hague and Foord (1971). More valuable information about the two-port circuit theory can be found in academic textbooks on electrical circuits, for example, such as the classic one by Carter and Richardson (1972).

Impedances are used commonly to describe the bridge arms. Therefore, the **Z** matrix will be taken for further consideration as the main one. Also, it is very important that this matrix exists in the balanced states of the bridge, but matrix **A** does not.

The linear two-port device with all typical symbols is given in Figure 3. An additional assumption is taken that there are no stationary sources inside a two-port circuit. If it is not so, the influence of such sources is eliminated from measurement considerations by short-circuiting the voltage ideal sources and disconnecting the current ones. They can also be taken into account in the next steps of the bridge circuit analysis, for example, by the superposition method.

The **Z** type matrix equation for a two-port device is the following:

$$\mathbf{U} = \mathbf{ZI} \tag{1}$$

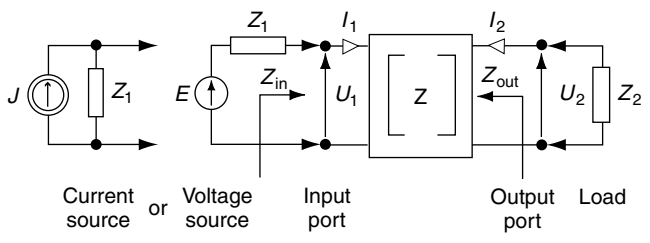


Figure 3. Two port network: standard conventions of currents, voltages, and impedances.

or in expanded form

$$\begin{vmatrix} U_1 \\ U_2 \end{vmatrix} = \begin{vmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{vmatrix} \begin{vmatrix} I_1 \\ I_2 \end{vmatrix} \tag{1a}$$

In the algebraic form it is:

$$\begin{aligned} U_1 &= z_{11}I_1 + z_{12}I_2 \\ U_2 &= z_{21}I_1 + z_{22}I_2 \end{aligned} \tag{1b}$$

In the general case, all impedance coefficients z_{ij} are different. A linear two-port circuit could include different, but only linear, two- and more terminal passive elements, for example, R L C single or coupled impedances, transformers, and so on, and only sources controlled proportionally by the circuit's voltages or currents.

If the two-port is nonreciprocal, then $z_{12} \neq z_{21}$ holds. For only one source supplying input or output port, the two-port has two different balance states on its terminals: $z_{21} = 0$ and $z_{12} = 0$.

If $z_{12} = z_{21}$, then the **Z** matrix has only three independent impedance elements.

If external circuits connected to both ports are also linear, then according to the Thevenin (after Carter and Richardson (1972) this theorem first developed by H. Hemholtz 30 years earlier) or Norton theorems, they have linear equivalent circuits. Such a one is shown for the primary side of the two-port of Figure 3.

Input and output equivalent impedances, terminal currents, and voltages under different load Z_2 and source Z_1 impedances are given in Table 1.

These relations are valid for all linear two-ports including bridge circuits. If elements of circuits connected to any side of the two-port are of known nonlinearities, it is possible also to find the terminal currents and voltages of this two-port by using its equations for each value of supply voltage E or current J.

The practical meaning of the **Z** matrix elements is apparent from the equations of Table 1. It is seen that element z_{11} is the input impedance Z_{in} of the two-port if the output circuit is open, and z_{22} is the output impedance Z_{out} when $Z_1 \to \infty$.

Element z_{12} and z_{21} of the two-port matrix **Z** are current to voltage transmittances, respectively in the reverse and forward direction of transmission when the output port is open.

It is possible to find all elements of the **Z** matrix from two-port terminal parameters measured with an open or short-circuited output port that is, $Z_2 = \infty$, $Z_2 = 0$ or $Z_1 \to \infty$, $Z_1 = 0$.

Equations of Table 1 could also be presented in different forms. For example, after inserting equations for Z_{in} and Z_{out} to the equation of voltage U_2 we obtain

$$U_2 = J\frac{Z_1 Z_2 z_{21}}{(Z_1 + z_{11})(Z_2 + z_{22}) - z_{12}z_{21}} \tag{1c}$$

Then, if z_{ij} elements were known from calculations or terminal measurements for a particular bridge working as the two-port, equations of Table 1 would be a useful tool to analyze it as a passive or active electrical linear circuit with variable parameters.

7 IMPEDANCE BRIDGE AS TWO PORT AND ITS Z MATRIX

As it was mentioned earlier, four-terminal bridges including only linear elements, could operate as linear two-ports. If the bridge is a part of the larger circuit, it can be considered as a two-port set up, only when it has two pairs of terminals with currents opposite to each other in each pair. It means that currents in connections around the bridge that is, between two terminals, each of a different pair, are negligible, or their influence additionally balanced. If it is not so, the bridge should be treated as a four-terminal (4-T) or generally an n-terminal (n-T) circuit.

Any four arms being used in a linear impedance bridge is the X-type two-port. It is shown in Figure 4. Designations of the two-port in Figure 3 are here changed to the ones used in bridges, similar to the Wheatstone bridge in Figure 2.

The impedance matrix **Z** of the linear passive bridge is given below

$$\mathbf{Z} = \begin{vmatrix} \dfrac{(Z_1 + Z_2)(Z_3 + Z_4)}{\sum Z_i} & \dfrac{Z_1 Z_3 - Z_2 Z_4}{\sum Z_i} \\ \dfrac{Z_1 Z_3 - Z_2 Z_4}{\sum Z_i} & \dfrac{(Z_1 + Z_4)(Z_2 + Z_3)}{\sum Z_i} \end{vmatrix} \tag{2}$$

where $\sum Z_i \equiv Z_1 + Z_2 + Z_3 + Z_4$.

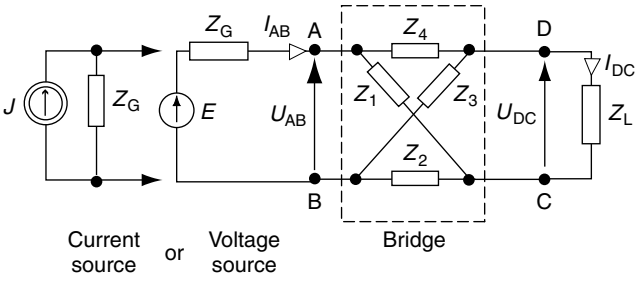


Figure 4. Notations of four-arm impedance bridge as X type two port.

Table 1. Definitions and basic equations of terminal impedances, currents and voltages of linear two port under work of linear load and supply source.

Parameters and basic relations			General case	Particular cases			
			$\infty > Z_1 > 0, \infty > Z_2 > 0$	$Z_1 = \infty$	$Z_1 = 0$	$Z_2 = \infty$	$Z_2 = 0$
Port impedance	Input	$Z_{\text{in}} \equiv \frac{U_1}{I_1}$	$z_{11} - \frac{z_{12}\, z_{21}}{z_{22} + Z_2} = \frac{\det \mathbf{Z} + Z_2\, z_{11}}{z_{22} + Z_2} = Z_{\text{in}}^{\infty} \frac{Z_{\text{out}}^{0} + Z_2}{Z_{\text{out}}^{\infty} + Z_2}$			$Z_{\text{in}}^{\infty} = z_{11}$	$Z_{\text{in}}^{0} = \frac{\det \mathbf{Z}}{z_{22}}$
	Output	$Z_{\text{out}} \equiv \frac{U_2}{I_2}$	$z_{22} - \frac{z_{12}\, z_{21}}{z_{11} + Z_1} = Z_{\text{out}}^{\infty} \frac{Z_{\text{in}}^{0} + Z_1}{Z_{\text{in}}^{\infty} + Z_1}$	$Z_{\text{out}}^{\infty} = z_{22}$	$Z_{\text{out}}^{0} = \frac{\det \mathbf{Z}}{z_{11}}$	$z_{22} - \frac{z_{12}\, z_{21}}{z_{11} + Z_1} = \frac{\det \mathbf{Z} + Z_1\, z_{22}}{z_{11} + Z_1}$	
Supply source E or J in primary port	Output variables	U_2	$\frac{E}{Z_{\text{in}} + Z_1} z_{21} \frac{Z_2}{Z_{\text{out}} + Z_2}$	0	$\frac{E}{Z_{\text{in}}} \frac{z_{21}\, Z_2}{Z_{\text{out}}^{0} + Z_2}$	$\frac{E\, z_{21}}{z_{11} + Z_1}$	0
			$J \frac{Z_1}{Z_{\text{in}} + Z_1} z_{21} \frac{Z_2}{Z_{\text{out}} + Z_2}$	$J \frac{z_{21}\, Z_2}{z_{22} + Z_2}$	0	$J \frac{Z_1\, z_{21}}{z_{11} + Z_1}$	0
		$-I_2$	$\frac{E}{Z_{\text{in}} + Z_1} \frac{z_{21}}{Z_{\text{out}} + Z_2}$	0	$\frac{E}{Z_{\text{in}}} \frac{z_{21}}{Z_{\text{out}}^{0} + Z_2}$	0	$\frac{E}{Z_{\text{in}}^{0} + Z_1} \frac{z_{21}}{Z_{\text{out}}}$
			$J \frac{Z_1}{Z_{\text{in}} + Z_1} \frac{z_{21}}{Z_{\text{out}} + Z_2}$	$J \frac{z_{21}}{z_{22} + Z_2}$	0	0	$J \frac{Z_1}{Z_{\text{in}}^{0} + Z_1} \frac{z_{21}}{Z_{\text{out}}}$
Supply source E or J in primary port	Input variables	U_1	$\frac{E\, Z_{\text{in}}}{Z_{\text{in}} + Z_1}$	0	E	$E \frac{z_{11}}{z_{11} + Z_1}$	$E \frac{\det \mathbf{Z}}{\det \mathbf{Z} + Z_1\, z_{22}}$
			$J \frac{Z_{\text{in}}\, Z_1}{Z_{\text{in}} + Z_1}$	$J\, Z_{\text{in}}$	0	$J \frac{z_{11}\, Z_1}{z_{11} + Z_1}$	$J \frac{Z_1 \det \mathbf{Z}}{\det \mathbf{Z} + Z_1 z_{22}}$
		I_1	$\frac{E}{Z_{\text{in}} + Z_1}$	0	$\frac{E}{Z_{\text{in}}}$	$\frac{E}{Z_1 + z_{11}}$	$\frac{E\, z_{22}}{\det \mathbf{Z} + Z_1 z_{22}}$
			$J \frac{Z_{\text{in}}\, Z_1}{Z_{\text{in}} + Z_1}$	J	0	$J \frac{Z_1}{z_{11} + Z_1}$	$J \frac{Z_1\, z_{22}}{\det \mathbf{Z} + Z_1\, z_{22}}$

Note: Where $\det \mathbf{Z} = z_{11}\, z_{22} - z_{12}\, z_{21}$.

It is a symmetric matrix, that is, $z_{12} = z_{21}$, because the circuit is reciprocal. This means that open circuit output voltages U_{DC} and U_{AB} are equal when the bridge is supplied by the same value current $I_{\text{AB}} = I_{\text{DC}}$ at primary AB or secondary DC port. Then the balance conditions for $U_{\text{DC}} = 0$ or $U_{\text{AB}} = 0$ for both ways of supply are equal and the following applies:

$$Z_1 Z_3 - Z_2 Z_4 = 0 \tag{3}$$

Or for arm admittances $Y_i = 1/Z_i$

$$Y_1 Y_3 - Y_2 Y_4 = 0 \tag{3a}$$

Balance condition could be also expressed in ratio forms for example,

$$\frac{Z_1}{Z_2} = \frac{Z_4}{Z_3} \text{ or } \frac{Z_1}{Z_4} = \frac{Z_2}{Z_3} \tag{3b) or (3c)}$$

Input and output bridge diagonal impedances are, in the general case, different. They depend mainly on the bridge arm impedances and to a smaller degree also on the impedance connected to the opposite diagonal. Only in the balanced state they are independent of the last ones, as it results from Figure 1(a).

If the supply source E or J, as well as Z_L and Z_G are given, the output voltage U_{DC} of the loaded bridge as two-port (see equation (1c)) is

$$U_{DC} = I\frac{Z_G Z_L z_{21}}{(Z_G + z_{11})(Z_L + z_{22}) - z_{21}^2} \tag{4}$$

Inserting in (4) the values of bridge matrix elements given in (2) we obtain:

$$U_{DC} = J\frac{Z_G Z_L (Z_1 Z_3 - Z_2 Z_4)}{Z_G Z_L \sum Z_i + Z_L (Z_1 + Z_2)(Z_3 + Z_4) + Z_G (Z_1 + Z_4)(Z_2 + Z_3) + A} \tag{5}$$

Where:

$$A = \sum Z_i Z_j Z_k = Z_1 Z_2 (Z_3 + Z_4) + Z_3 Z_4 (Z_1 + Z_2) = \cdots$$

Moving the product of $Z_G Z_L$ to the denominator we get:

$$U_{DC} = J\frac{Z_1 Z_3 - Z_2 Z_4}{\sum Z_i + \frac{1}{Z_G}(Z_1 + Z_2)(Z_3 + Z_4) + \frac{1}{Z_L}(Z_1 + Z_4)(Z_2 + Z_3) + \frac{A}{Z_G Z_L}} \tag{6}$$

From the equations of lines 1 and 2 in Table 1 it is also possible to find Z_{in} and Z_{out} impedances of the analyzed bridge for different values of source impedance Z_G and of the load one Z_L – *see* **Article 127, Unbalanced DC Bridges, Volume 2**.

8 BASIC EQUATIONS OF UNBALANCED BRIDGES

In the analysis of bridges in off-balance conditions and properties of unbalanced (deflection) bridges, it is convenient to express matrix $\mathbf{Z}$ as a sum of two components

$$\mathbf{Z} = \mathbf{Z_0} + \mathbf{\Delta Z} = \mathbf{Z_0} + \mathbf{\Delta Z}(\varepsilon) \tag{7}$$

or in the expanded forms:

$$\mathbf{Z} = \begin{vmatrix} Z_{AB0} & 0 \\ 0 & Z_{CD0} \end{vmatrix} + \begin{vmatrix} \Delta Z_{AB} & z_{21} \\ z_{21} & \Delta Z_{CD} \end{vmatrix} = \begin{vmatrix} Z_{AB0} & 0 \\ 0 & Z_{CD0} \end{vmatrix} + \begin{vmatrix} \varepsilon_{AB} Z_{AB0} & t_0 f(\varepsilon_i) \\ t_0 f(\varepsilon_i) & \varepsilon_{CD} Z_{CD0} \end{vmatrix} \tag{8}$$

Where $\mathbf{Z}_0 - \mathbf{Z}$ matrix in the bridge balanced state; $\mathbf{\Delta Z}$ – matrix of absolute increments Δz_{ij} from this state; $\mathbf{\Delta Z}(\varepsilon)$ – above matrix expressed by relative increments of impedances and by the initial sensitivity t_0 and unbalance function $f(\varepsilon_i) = \Delta L(\varepsilon_i)/M(\varepsilon_i)$ of the output voltage given below.

All bridge parameters corresponding to the balance state have now additional subscript 0. Then the arm impedances are

$$Z_i \equiv Z_{i0} + \Delta Z_i = Z_{i0}(1 + \varepsilon_i) \tag{9}$$

$\Delta Z_i;\ \varepsilon_i \equiv \dfrac{\Delta Z_i}{Z_{i0}}$ – absolute and relative increments of impedances Z_i, from their values Z_{i0} in the balance state;

$t_0 \equiv \dfrac{Z_{10} Z_{30}}{\sum Z_{i0}}$ – initial voltage sensitivity per one unit of current I_{AB} of the open output of the bridge;

$f(\varepsilon_i) = \dfrac{\Delta L(\varepsilon_i)}{M(\varepsilon_i)}$ – voltage unbalance relative function of bridge impedance increments and

$\Delta L(\varepsilon_i) = \varepsilon_1 - \varepsilon_2 + \varepsilon_3 - \varepsilon_4 + \varepsilon_1\varepsilon_3 - \varepsilon_2\varepsilon_4$ – numerator and

$M(\varepsilon_i) = 1 + \dfrac{\sum \Delta Z_i}{\sum Z_{i0}} = 1 + \dfrac{\sum \varepsilon_i Z_{i0}}{\sum Z_{i0}}$ – denominator of the above function.

The following additional relations of arm initial impedances to the value of Z_{10} are defined as

$$Z_{20} \equiv mZ_{10};\ Z_{40} \equiv nZ_{10};\ Z_{30} \equiv mnZ_{10} \tag{10}$$

Two-port matrix $\mathbf{Z}$ parameters of the bridge, in the above new notation, are presented in Table 2.

The output voltage of the loaded bridge, given by (6) also could be written now as

$$U_{DC} \equiv T_0 \frac{\Delta L(\varepsilon_i)}{M_L(\varepsilon_i)} \tag{11}$$

$T_0 \equiv JZ_{10}Z_{30}/M_{UJ0}$ – initial voltage sensitivity of loaded bridge;

$M_{UJ} \equiv M_{UJ0} + \Delta M_{UI}(\Delta Z_i)$ – denominator in (6) of U_{DC}.

$\Delta L(\varepsilon_i) = \varepsilon_1 - \varepsilon_2 + \varepsilon_3 - \varepsilon_4 + \varepsilon_1\varepsilon_3 - \varepsilon_2\varepsilon_4$ – numerator as before,

$M_L(\varepsilon_i) \equiv 1 + \Delta M_{UJ}(\Delta Z_i)/M_{UJ0}$ – new denominator of the voltage relative unbalance function of the loaded bridge with a nonideal supply source.

Increment $\Delta M_{UJ}(\Delta Z_i)$ of the denominator is a complicated function of the balance impedances and of their increments for general cases of value Z_G and Z_L, because:

Table 2. Parameters of matrices **Z, Z**$_0$, **Δ Z** (ε_i), **ΔZ** of the four-arm linear impedance bridge as two port.

No	Matrix	Bridge state	z_{11}		$z_{12} = z_{21}$	z_{22}	
1	$\mathbf{Z}_0$	In balance	Z_{AB0}	$\dfrac{Z_{40}(Z_{10}+Z_{20})}{Z_{10}+Z_{40}} = Z_{10}\dfrac{n(1+m)}{1+n}$	$\dfrac{Z_{10}Z_{30}-Z_{20}Z_{40}}{\sum Z_{i0}} = 0$	Z_{CD0}	$Z_{20}\dfrac{Z_{10}+Z_{40}}{Z_{10}+Z_{20}} = Z_{10}\dfrac{m(1+n)}{1+m}$
2	**Z**	Out of balance	Z_{AB} ($Z_L = \infty$)	$\dfrac{(Z_1+Z_2)(Z_3+Z_4)}{\sum Z_i}$ or	$\dfrac{Z_1Z_3-Z_2Z_4}{\sum Z_i}$	Z_{CD} ($Z_G = \infty$)	$\dfrac{(Z_1+Z_4)(Z_2+Z_3)}{\sum Z_i}$ or
3				$Z_{AB0}[1+\varepsilon_{AB}]$	or		$Z_{CD0}[1+\varepsilon_{CD}]$
4	ε		ε_{AB} ($Z_L = \infty$)	$\dfrac{1}{M(\varepsilon_i)}\left(\dfrac{n\varepsilon_{12}}{1+n}+\dfrac{\varepsilon_{43}}{1+n}+\varepsilon_{12}\varepsilon_{43}\right)$		ε_{CD} ($Z_G = \infty$)	$\dfrac{1}{M(\varepsilon_i)}\left(\dfrac{m\varepsilon_{14}}{1+m}+\dfrac{\varepsilon_{23}}{1+m}+\varepsilon_{14}\varepsilon_{23}\right)$
			ε_{12}	$\varepsilon_{12} = \dfrac{\varepsilon_1+m\varepsilon_2}{1+m}$	$t_0\dfrac{\Delta L(\varepsilon_i)}{M(\varepsilon_i)}$	ε_{14}	$\varepsilon_{14} = \dfrac{\varepsilon_1+n\varepsilon_4}{1+n}$
			ε_{43}	$\varepsilon_{43} = \dfrac{\varepsilon_4+m\varepsilon_3}{1+m}$		ε_{23}	$\varepsilon_{23} = \dfrac{\varepsilon_2+n\varepsilon_3}{1+n}$
5	$\Delta\mathbf{Z}$		ΔZ_{AB} ($Z_L = \infty$)	$\dfrac{Z_{AB0}}{\sum Z_{i0}+\sum \Delta Z_i} \times \left[\begin{array}{c} n(\Delta Z_i+\Delta Z_2)+\dfrac{1}{n}(\Delta Z_3+\Delta Z_4) \\ +\dfrac{(\Delta Z_1+\Delta Z_2)(\Delta Z_3+\Delta Z_4)}{Z_{AB0}} \end{array}\right]$		ΔZ_{CD} ($Z_G = \infty$)	$\dfrac{Z_{CD0}}{\sum Z_{i0}+\sum \Delta Z_i} \times \left[\begin{array}{c} m(\Delta Z_1+\Delta Z_4)+\dfrac{1}{m}(\Delta Z_2+\Delta Z_3) \\ +\dfrac{(\Delta Z_1+\Delta Z_4)(\Delta Z_2+\Delta Z_3)}{Z_{AB0}} \end{array}\right]$

Note: Where initial bridge output voltage to current sensitivity – $t_0 \equiv \dfrac{T_0}{J} = \dfrac{Z_{20}Z_{40}}{\sum Z_{i0}} = \dfrac{Z_{10}mn}{(1+m)(1+n)}$,

numerator of unbalance function – $\Delta L(\varepsilon_i) = \varepsilon_1 - \varepsilon_2 + \varepsilon_3 - \varepsilon_4 + \varepsilon_1\varepsilon_3 - \varepsilon_2\varepsilon_4$,

denominator of unbalance function – $M(\varepsilon_i) = 1 + \dfrac{\sum \Delta Z_i}{\sum Z_{i0}} = 1 + \dfrac{\varepsilon_{12}+n\varepsilon_{43}}{1+n} = 1 + \dfrac{\varepsilon_{14}+m\varepsilon_{23}}{1+m} = 1 + \dfrac{\varepsilon_1+m\varepsilon_2+n(\varepsilon_4+m\varepsilon_3)}{(1+m)(1+n)}$.

$$\Delta M_{UJ}(\Delta Z_i) = \sum \frac{\partial M_{UJ}}{\partial Z_i}\Delta Z_i + \sum_{i\neq j}\frac{\partial^2 M_{UJ}}{\partial Z_i \partial Z_j}\Delta Z_i \Delta Z_j + \sum_{\substack{i\neq j\neq k \\ i\neq k}}\frac{\partial^3 M_{UJ}}{\partial Z_i \partial Z_j \partial Z_k}\Delta Z_i \Delta Z_j \Delta Z_k \qquad (12)$$

There are 14 elements to add together $(4+6+4)$. If modules of related resistance increments $|\varepsilon_i|$ are not too high, the last sum in the above equation, dependant on the products of the three of them, can be neglected; and also for $\varepsilon_i \ll 1$, the second sum is negligible.

Further simplification of (6) and (11) is obtained assuming ideal J or E supply sources $|Z_G| \gg |Z_{AB}|$ or $|Z_G| \ll |Z_{AB}|$ respectively, for ideal voltage output $|Z_L| \gg |Z_{CD}|$ or ideal current output $|Z_L| \ll |Z_{CD}|$.

It is also the case if some initial bridge arm impedances are equal, or not changing and if pairs of the increments, or all of them, are of the same absolute value but of the opposite signs in the neighboring arms. Such cases are very common in practice. Detailed discussion of these cases are in books by Bentley (1995), Dally, Riley and McConnell (1993), Holman (1994), Tattamangalam (2000), Schnell and Sydenham (1993), and also in **Article 127, Unbalanced DC Bridges, Volume 2**.

RELATED ARTICLES

Article 56, Transfer Characteristics of Instrument Stages, Volume 1; **Article 112, Systematic Description of Sensors, Volume 2**; **Article 127, Unbalanced DC**

Bridges, Volume 2; **Article 202, Current Measurement, Volume 3**; **Article 204, Capacitance and Inductance Measurement, Volume 3**

REFERENCES

This list extends the cited references given in the text above. Bridges were of strong interest to publishers until the 1950s after which, despite, their still continued use, titles on the topic twindled. Some of the best literature in bridges is to be found in texts now considered as too old to be relevant by many!

Bentley, J.P (1995) *Principles of Measurement Systems*, 3rd edn, Longman Scientific & Technical, Longman Groupe UK Ltd, Harlow, Essex (pp. 178–187).

Boyes, W., Parr, M. and Noltingk, B.E. (eds) (2003) *Instrumentation Reference Book*, 3rd edn, Butterworth-Heinemann, Elsevier Science, Boston (pp. 463–479).

Carter, G.W. and Richardson, A. (1972) *Technique of Circuit Analysis*, Cambridge University Press, London (pp. 17–22, 248–281).

Dally, J.W., Riley, W.F. and McConnell, K.G. (1993) *Instrumentation for Engineering Measurements*, 2nd edn, John Wiley & Sons, NewYork (pp. 170–176, 220–236).

Dyer, S.A. (ed) (2001) *Survey of Instrumentation and Measurement*, Wiley-Interscience, John Wiley & Sons, New York (Nicolas, J., Section: *Bridge Instruments*, pp. 309–326; Kumar, V.J., Murti, V.G.K. and Sankaran, P., Sections: *Methods of Measurement of Self-inductance* and *... of Mutual Inductance*, pp. 352–357; Eren, H., Section: *Capacitance Measurements*, pp. 369–371).

Hague, B. and Foord, T.R. (1971) *Alternating Current Bridge Methods*, 6th edn, Pitman Publishing, London.

Holman, J.P. (1994) *Experimental Methods for Engineers*, 6th edn, McGraw-Hill International, New York (pp. 136–144).

Schnell, L. and Sydenham, P.H. (eds) (1993) *Technology of Electrical Measurements*, *Series on Measurement Science*, John Wiley & Sons (Chapter 4 by P. Osvath, pp. 162–218).

Sydenham, P.H. (1979) *Measuring Instruments; Tools of Knowledge and Control*, Peter Perigrinus.

Tattamangalam, R.P. (2000) *Industrial Instrumentation Principles and Design*, Springer-Verlag, London (pp. 29–51).

FURTHER READING

Gregory, B.A. (1973) *An Introduction to Electrical Instruments*, Macmilan Press Ltd, London (pp. 88–116).

Hai Hung, C. (1984) *Electrical and Electronic Instrumentation*, John Wiley & Sons, New York (pp. 48–92).

Karandeyev, K.B. (1966) *Bridge and potentiometer methods in electrical measurements*, Energia Moscow (1966) and Peace Publisher MIR (1967) (Eng. transl. from Russian edition 1960).

McGhee, J., Korczyński, M.J., Henderson, I.A. and Kulesza W. (1998) *Scientific Metrology*, 2nd edn Lodard S.A., Lodz (pp. 205–217, 289–297).

Morris, A.S. (1993) *Principles of Measurement and Instrumentation*, 2nd edn, Prentice Hall, New York (pp. 106–127).

Noltingk, B.E. (ed) (1995) *Instrumentation Reference Book*, Part 3, Butterworth-Heinemann, London (pp. 40–57).

Oliver, B.M. and Cage, J.M. (1971) *Electronic Measurements and Instrumentation*, McGraw-Hill, New York (Chapter 9 by H.P. Hall, pp. 264–319).

Sydenham, P.H. (ed) (1983) *Handbook of Measurement Science*, Vol. 2, Wiley-Interscience Publications, John Wiley (Chapter 20.6 by L. Schnell).

Webster, J.G. (ed) (1996, 2000) *The Measurement, Instrumentation and Sensors Handbook*, CRC Press London, Springer, IEEE Press New York; Eren, H. and Goh J., Chapter 45: *Capacitance and capacitance measurements*, pp. 45-23–45-27; Misra D.K., Chapter 46: *Permitivity measurement*, pp. 46-6–46-7; Szyper M., Chapter 50: *Inductance measurement*, pp. 50-8–50-13; Dreher A., Chapter 51: *Immitance Measurement*, pp. 51-11–51-15.

127: Unbalanced DC Bridges

Zygmunt L. Warsza
Polish Metrological Society, Warsaw, Poland

1 Introduction 878
2 Four-terminal (4T) DC Circuits 878
3 Basic Equations of the Four-arm Resistance Bridge (4R) as Twoport 879
4 Simplified Bridge Equivalent Circuits 884
5 Relations Between Open Circuit Terminal Parameters of the 4R Bridge 884
6 Bridge Open Circuit Terminal Parameters' Linearization 885
7 Terminal Parameters of the DC Bridge of any Supply Source and Load 886
Related Articles 889
References 889
Further Reading 889

1 INTRODUCTION

Bridge circuits are still commonly used in the input of autonomic immitance measurement instruments and transmitters of industrial systems. It is so for the following reasons:

- They are either directly sensitive to changes of immitance components or else they are indirectly sensitive to other variables actuating the immitance of analog sensors.
- They perform some simple signal-conditioning functions, such as dependence of output signal on changes of parameters of several sensors easily and accurately.

During the last decades, many new input signal-conditioning circuits have been developed (see – Anderson (1997), Kester (1999), Zakrzewski (2000), printed catalogs and Internet information of National Instruments™ (2004) and of Analog Devices Co.). They use various differential operational amplifiers, feedback loops, single and double stabilized supply sources, many methods of switching (e.g. integrating or successive approximation AD converters). Ratio measurements are also applied when analog circuit and AD converter are supplied by a common source. However, classic unbalanced DC bridges are still most commonly used either on their own or in above mentioned circuits for measurement of resistance and conductance for their merits and simplicity. They work most often with nearly unloaded (open circuit) voltage output, sometimes with current output loaded with resistance or short circuited, and they deliver measurement signals for amplifying and converting them to the digital form.

Furthermore, measured objects as well as multiterminal solid state and integrated semiconductor sensors can also include inseparable bridge structures in their equivalent circuits. Parameters of this structure are measurable indirectly only on terminals of the circuit or on electrodes used to check material properties.

Below, properties of the four terminal (4T) unbalanced resistance bridge networks are analyzed in detail. The relevant knowledge in design and understanding of measurement bridge circuits and in programming digital circuits connected with them is required. Some new developments have also been introduced. A short history of measuring bridges and their basic equations have already been given in **Article 126, Electrical Bridge Circuits – Basic Information, Volume 2**.

2 FOUR-TERMINAL (4T) DC CIRCUITS

Four-terminal circuits (4T) could have arbitrary currents of three terminals and the fourth one determined by Kirchoff's Law. If there are two pairs of opposite terminal

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

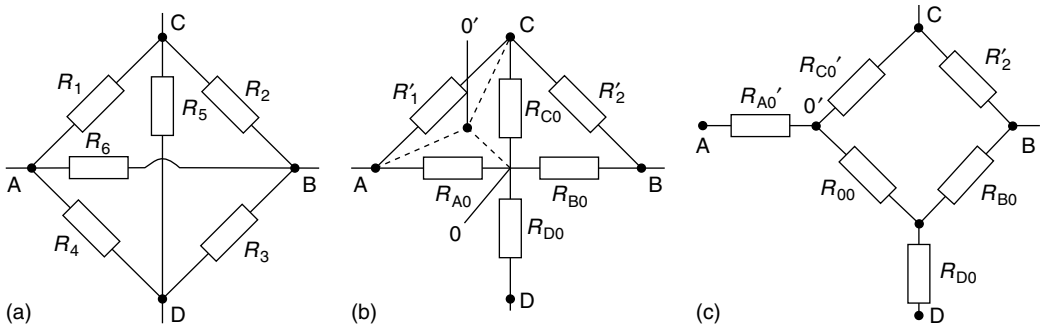


Figure 1. Examples of 4T circuit, every one equivalent to each other: (a) without nodes inside – four arms bridge with diagonals; (b) with one node inside; (c) with two nodes inside.

currents, then the 4T circuit behaves as the twoport – see, for example, Carter and Richardson (1971). Only passive and reciprocal 4T circuits are taken below for analysis. Their basic equivalent circuit has six immitance branches, which are connected to the quadrilateral network given for DC on Figure 1(a).

Terminals are also the only nodes in this network, and it has the structure of the four-arm bridge with two diagonals. There is a direct relation of the admittance of each branch and the value of the off-diagonal element of the circuit admittance matrix with the sign 'plus'. Any of four-terminal circuits, which are more complex inside, can always be transformed to the above network or to some of its simpler structures.

In Figure 1(b) and (c), there are shown two other networks of six branches equal to the first one. They had originally been developed by the author for the reciprocal part of the Hall devices' equivalent 4T network to find zero correction circuits that were stabile in a wide range of temperatures – see Warsza (1980). The network in Figure 1(b) has one node inside. It was obtained from circuit in Figure 1(a) by circuit substituting two smaller resistances, each one of two smaller opposite resistances' products, for example, R_1 and R_2 by two parallel ones respectively R_1', R_1'' and R_2', R_2''. So, that is $R_1'' R_3 = R_2'' R_4 = R_5 R_6$ and to the network of these six resistances the Rosen's theorem of the four-arm star-delta inverse transformation is applicable. The last network, given in Figure 1(c) has two nodes inside. It is obtained from the previous one by transformation of elements of the triangular ACO to the equivalent star. There are six different networks of two serial immitances connected to neighboring nodes of the inside bridge and they are equivalent to each other. All bridges of networks of Figure 1 are in balance together.

The four-arm bridge (4R) is a particular case of the above networks of six immitances. It has no other four branch circuits that are equivalent to it. Then, in the circuit theory it plays an analogous role to the prime number in mathematics.

Resistance or conductance 4T circuits could be applied in measurements in different ways. In addition to measurements of branch immitances, they could be used for indirect measurements of one and theoretically up to six variables differently influencing these immitances.

In measurements, a bridge could be supplied from voltage or current sources, ideal ones, or of the particular internal resistance. The classic type supply is connected to one of the bridge diagonals and the output is taken from the second one. Two sources could be connected in different ways, for example, voltage sources, one in the diagonal and the second one in the bridge arm or in the output diagonal. Both the ones in the bridge branches, also have been used. Some years ago, this author proposed to use unconventional power supply for the bridge by providing two equal and ideal current sources connected in parallel to opposite bridge arms, or by only one such source switched between these arms and averaging of two measurement results obtained. Such a bridge has two outputs from both diagonals and offers original metrological possibilities. It could be used simultaneously for two and more variable measurements – see Warsza (2002), (2004) and the literature list in it.

3 BASIC EQUATIONS OF THE FOUR-ARM RESISTANCE BRIDGE (4R) AS TWOPORT

In Figure 2, the four-arm resistance bridge circuit (4R) classically powered from current or from voltage source is given. This bridge appears under the name of Wheatstone bridge. It is drawn here as the twoport.

In the general case, the bridge input current I_{AB} depends on supply source parameters J and R_G, or E and R_G and also on the equivalent input bridge resistance R_{AB}. The last one is a function of actual values of all bridge resistances R_i and of its load R_L, if any. Then, for a full description of the bridge from Figure 2 as a linear

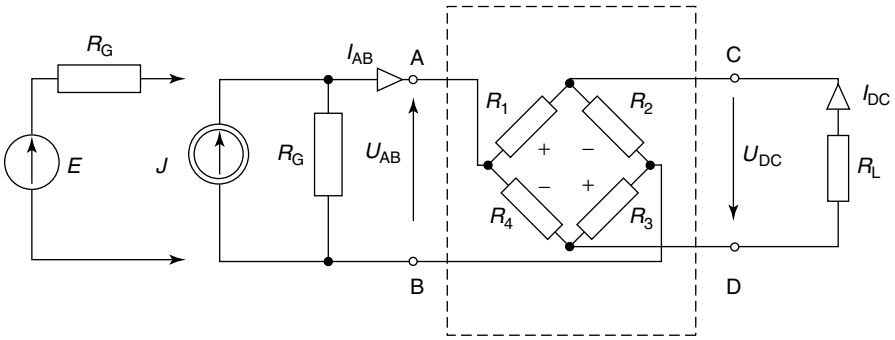


Figure 2. The four-arm bridge as twoport, together with the single supply source J or E of the internal resistance R_G and the load R_L, both connected to its diagonals.

circuit of variable parameters, twoport equations could be used – see such bridge equations of the impedance type given in **Article 126, Electrical Bridge Circuits – Basic Information, Volume 2** and of the particular case when all parameters are real, and matrix $\mathbf{Z} = \mathrm{Re}(\mathbf{Z}) \equiv \mathbf{Z_R}$. In generalized form, it is

$$\mathbf{U} = \mathbf{Z_R I}$$

and after expansion of matrices:

$$\begin{vmatrix} U_{AB} \\ U_{DC} \end{vmatrix} = \begin{vmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{vmatrix} \begin{vmatrix} I_{AB} \\ I_{CD} \end{vmatrix} \equiv \begin{vmatrix} R^{\infty}_{AB} & r_{12} \\ r_{21} & R^{\infty}_{CD} \end{vmatrix} \begin{vmatrix} I_{AB} \\ I_{CD} \end{vmatrix} \quad (1)$$

where current $I_{CD} = -I_{DC}$,

$r_{11} \equiv R^{\infty}_{AB}, r_{22} \equiv R^{\infty}_{CD}$ – input and output resistances of twoport when the opposite side port is opencircuited.
r_{12}, r_{21} – current to voltage both direction transmittances (transfer resistances) when the opposite side port is open circuited ($R_G \to \infty$, or $R_L \to \infty$). (Superscript ∞ or 0 is used here for open or shorted circuit of the opposite side of the bridge and subscript 0 – for the balance state).

Formula (1) in algebraic form and with the opposite direction output current I_{DC} as shown in Figure 2, is

$$\begin{aligned} U_{AB} &= R^{\infty}_{AB} I_{AB} - r_{12} I_{DC} \\ U_{DC} &= r_{21} I_{AB} - R^{\infty}_{CD} I_{DC} \end{aligned} \quad (1a)$$

Matrix $\mathbf{Z_R}$ of the four-arms resistance bridge is:

$$\mathbf{Z}_R \equiv \begin{vmatrix} \dfrac{(R_1+R_2)(R_3+R_4)}{\sum R_i} & \dfrac{R_1R_3 - R_2R_4}{\sum R_i} \\ \dfrac{R_1R_3 - R_2R_4}{\sum R_i} & \dfrac{(R_1+R_4)(R_2+R_3)}{\sum R_i} \end{vmatrix} \quad (2)$$

where $\sum R_i = R_1 + R_2 + R_3 + R_4$ – sum of resistances of bridge main loop A B C D;

Because of the equivalence of both directions' bridge transmittances (or transfer resistances) $r_{21} = r_{12}$, only three elements of the matrix $\mathbf{Z_R}$ are different, but they depend on four bridge circuit resistances. Values of these elements are always finite, including balance state when $r_{21} = 0$ and

$$R_1 R_3 = R_2 R_4 \quad (2a)$$

Let us consider a general case of the bridge when variable resistances are in all arms. As it was introduced in **Article 126, Electrical Bridge Circuits – Basic Information, Volume 2**, analysis of the circuit of Figure 2 should be more general and easier if resistances are referenced to one of them, for example, to R_{10}, and their changes – to initial values of resistances in the balanced state. Then,

$$R_i \equiv R_{i0} + \Delta R_i \equiv R_{i0}(1 + \varepsilon_i) \equiv r_{i0} R_{10}(1 + \varepsilon_i) \quad (3)$$

where R_{i0} – initial (in balance) value of the R_i resistance;
$\Delta R_i, \varepsilon_i$ – are absolute and relative increments;
r_{i0} – relative values of arms' resistances in balance conditions.

The open circuit output voltage U^{∞}_{DC} is given by the equation

$$U^{\infty}_{DC} = I_{AB} r_{12} = \frac{I_{AB}(R_1R_3 - R_2R_4)}{\sum R_i} \quad (4)$$

where $\Sigma R_i = R_1 + R_2 + R_3 + R_4$

With the notations given by formula (3) and after transformation of (4), one gets

$$\begin{aligned} U^{\infty}_{DC} &= I_{AB} R_{10} \frac{r_{20} r_{40}}{\sum r_{i0}} \\ &\quad \times \frac{(\varepsilon_1 - \varepsilon_2 + \varepsilon_3 - \varepsilon_4 + \varepsilon_1\varepsilon_3 - \varepsilon_2\varepsilon_4)}{1 + \dfrac{\sum r_{i0}\varepsilon_i}{\sum r_{i0}}} \\ &\equiv T'_0 \times f'(\varepsilon_i) \end{aligned} \quad (5)$$

where $T_0' \equiv I_{AB} t_0'$ – initial output voltage sensitivity to the unbalance function $f'(\varepsilon_i)$.

When the bridge is in balance, output voltage $U_{DC}^{\infty} = 0$ and balance condition (2a) could be now written as:

$$R_{30} = r_{20} r_{40} R_{10}; \tag{5a}$$

Balance state of the bridge is the property of this particular network. Output signal is equal to zero only if the bridge is supplied from the single source connected to one of its diagonals.

If, for more simplicity of the notation, we put down additionally: $R_{20} \equiv mR_{10}$, $R_{40} \equiv nR_{10}$, then it is $R_{30} = mnR_{10}$ and initial open circuit sensitivity of transmittance r_{21} is

$$t_0' \equiv \frac{R_{10} R_{30}}{\sum R_{i0}} = R_{10} \frac{mn}{(1+m)(1+n)} \tag{5b}$$

and unbalance function of r_{21}:

$$f'(\varepsilon_i) = \frac{(\varepsilon_1 - \varepsilon_2 + \varepsilon_3 - \varepsilon_4 + \varepsilon_1\varepsilon_3 - \varepsilon_2\varepsilon_4)}{1 + \dfrac{\varepsilon_1 + m\varepsilon_2 + n(m\varepsilon_3 - \varepsilon_4)}{(1+m)(1+n)}} \equiv \frac{\Delta L \varepsilon_i}{1 + \varepsilon_{\Sigma R}(\varepsilon_i)} \tag{5c}$$

where $\Delta L(\varepsilon_i); \varepsilon_{\Sigma R}$. – increments of its numerator and denominator.

$\sum R_{i0}$ – value of ΣR_i of the balanced bridge.

From (5) and (5c), it is clear how output open-circuited voltage U_{DC}^{∞} and transmittance r_{21} depend on signs and values of relative increments ε_i of resistances of particular bridge arms. These increments of opposite signs in neighboring arms, given in Figure 2, unbalance the bridge in the same direction. If absolute values of these increments are the same, the output voltage is proportional to the number of variable arms (e.g. multiplied by 2 or by 4).

Transmittance r_{21} and its unbalance function $f'(\varepsilon_i)$ theoretically could take values from the range $(-\infty, +\infty)$. In practice, there are some limitations due to existing or permitted extreme values of ε_i, maximum dissipated powers of arms and maximum voltage of the current source or maximum current of the voltage source. Transmittance $r_{21} = 0$, if $f'(\varepsilon_i) = 0$. It could happen for many different combinations of ε_i. The basic balance state is defined in such a one as all $\varepsilon_i = 0$.

Open circuit bridge terminal resistances are

$$R_{AB}^{\infty} = R_{AB0}(1 + \varepsilon_{AB}^{\infty}) = R_{AB0} \frac{(1+\varepsilon_{12})(1+\varepsilon_{34})}{(1+\varepsilon_{\Sigma R})} \tag{6}$$

$$R_{CD}^{\infty} = R_{CD0}(1 + \varepsilon_{DC}^{\infty}) = R_{CD0} \frac{(1+\varepsilon_{14})(1+\varepsilon_{23})}{(1+\varepsilon_{\Sigma R})} \tag{7}$$

where

– input and output bridge resistances in the balance state:

$$R_{AB0} = R_{10} \frac{n(1+m)}{1+n} \tag{6a}$$

and

$$R_{CD0} = R_{10} \frac{m(1+n)}{1+m} \tag{7a}$$

– relative increments of R_{AB} and of R_{CD} from the balance state:

$$\varepsilon_{AB}^{\infty} = \frac{1}{1 + \varepsilon_{\Sigma R}(\varepsilon_i)} \left(\frac{n\varepsilon_{12} + \varepsilon_{43}}{1+n} + \varepsilon_{12}\varepsilon_{43} \right) \tag{6b}$$

and

$$\varepsilon_{CD}^{\infty} = \frac{1}{1 + \varepsilon_{\Sigma R}(\varepsilon_i)} \left(\frac{m\varepsilon_{14} + \varepsilon_{23}}{1+m} + \varepsilon_{14}\varepsilon_{23} \right) \tag{7b}$$

– relative increment of the sum of $R_i + R_j$ resistances of i, j arms;

$$\varepsilon_{ij} \equiv \frac{\Delta R_i + \Delta R_j}{R_{i0} + R_{j0}} \tag{8}$$

– relative increment of the sum $\sum R_i$ of bridge resistances:

$$\varepsilon_{\Sigma R} \equiv \frac{\varepsilon_{12} + n\varepsilon_{43}}{1+n} = \frac{\varepsilon_{14} + m\varepsilon_{23}}{1+m} \tag{9}$$

Matrix $\mathbf{Z_R}$ from (2) could now also be written in a different form as

$$\mathbf{Z}_R = \begin{vmatrix} R_{AB0}^{\infty}(1 + \varepsilon_{AB}^{\infty}) & t_0' f'(\varepsilon_i) \\ t_0' f'(\varepsilon_i) & R_{CD0}^{\infty}(1 + \varepsilon_{CD}^{\infty}) \end{vmatrix} \tag{10}$$

In Tables 1(a) and (b), are given elements of the matrix $\mathbf{Z_R}$ of the resistance bridge 4R operating as the twoport X. In subsequent lines, there are open circuit terminal parameters r_{21}, R_{AB}^{∞}, R_{DC}^{∞} and their components. The first column of Table 1(a) corresponds to the general case of the bridge. In columns a to d of this table and in columns e to j of Table 1(b), there are few particular cases depending on

- the number of variable arms,
- the relations between their increments, and
- the relations between initial arm resistances (in balance).

On the basis of these tables, it is possible to find terminal parameters of the four-arm bridge in any kind of its operation as twoport. Some particular cases of unbalanced 4R bridge are discussed in detail by Dall (1992), Karandeev (1966), Morris (1993), Padmanabhan (2000), Sanderson in Bentley (1995) handbook, Warsza (2004) and by others.

Table 1a. Matrix $\mathbf{Z_R}$ elements of the resistance bridge as the linear twoport X of four variable arms – general case and its four particular cases – only two arms or single arm variable.

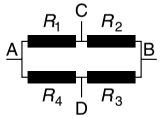
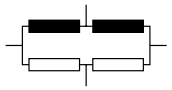
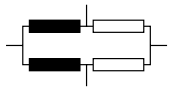
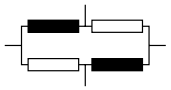
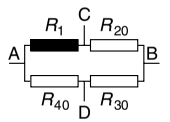
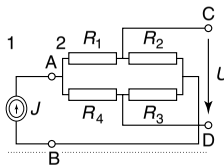
No	Bridge matrix $\mathbf{Z_R}$ elements (terminal parameters)	Symbol	General case	Particular cases			
			variable: $R_1 = R_{10} \quad (1+\varepsilon_1)$ $R_2 = mR_{20} \quad (1+\varepsilon_2)$ $R_3 = mnR_{10} \quad (1+\varepsilon_3)$ $R_4 = nR_{10} \quad (1+\varepsilon_4)$ 	(a) variable: R_1, R_2 	(b) variable: R_1, R_4 	(c) variable: R_1, R_3 	(d) variable: R_1 
1	**Transmittance:** $r_{21} = \frac{U'_{DC}}{J} = \frac{R_1R_3 - R_2R_4}{\sum R_i}$ $\equiv t_0 f(\varepsilon_i) \equiv t_0 \frac{\Delta L(\varepsilon_i)}{1+\Delta M(\varepsilon_i)}$	r_{21}	$t_0 \frac{\varepsilon_1 - \varepsilon_2 + \varepsilon_3 - \varepsilon_4 + \varepsilon_1\varepsilon_3 - \varepsilon_2\varepsilon_4}{1 + \frac{\varepsilon_1 + m\varepsilon_2 + n(\varepsilon_4 + m\varepsilon_3)}{(1+n)(1+m)}}$	$t_0 \frac{\varepsilon_1 - \varepsilon_2}{1 + \frac{\varepsilon_1 + m\varepsilon_2}{(1+n)(1+m)}}$	$t_0 \frac{\varepsilon_1 - \varepsilon_4}{1 + \frac{\varepsilon_1 + n\varepsilon_4}{(1+n)(1+m)}}$	$t_0 \frac{\varepsilon_1 + \varepsilon_3}{1 + \frac{\varepsilon_1 + mn\varepsilon_3}{(1+n)(1+m)}}$	$t_0 \frac{\varepsilon_1}{1 + \frac{\varepsilon_1}{(1+n)(1+m)}}$
	Initial sensitivity of r_{21} when : $f(\varepsilon_i) \to 0$	t_0	$t_0 = R_{10} \frac{mn}{(1+m)(1+n)}$				
	Increment of function $f(\varepsilon_i)$ denominator	ΔM	$\Delta M = \frac{\varepsilon_1 + m\varepsilon_2 + n(\varepsilon_4 + m\varepsilon_3)}{(1+n)(1+m)}$	$\Delta M_{12} = \frac{\varepsilon_1 + m\varepsilon_2}{(1+n)(1+m)}$	$\Delta M_{14} = \frac{\varepsilon_1 + n\varepsilon_4}{(1+n)(1+m)}$	$\Delta M_{13} = \frac{\varepsilon_1 + mn\varepsilon_3}{(1+n)(1+m)}$	$\Delta M_1 = \frac{\varepsilon_1}{(1+n)(1+m)}$
2	**Input resistance:** $r_{11} = \frac{(R_1+R_2)(R_4+R_3)}{\sum R_i}$ $\equiv R_{AB0}(1+\varepsilon^{\infty}_{AB})$	R^{∞}_{AB}	$\frac{R_{10}(1+m+\varepsilon_1+m\varepsilon_2)n(1+m+\varepsilon_4+m\varepsilon_3)}{(1+m)(1+n)+\varepsilon_1+m\varepsilon_2+n(\varepsilon_4+m\varepsilon_3)}$ $= R_{AB0}\frac{(1+\varepsilon_{12})(1+\varepsilon_{43})}{1+\Delta M} \quad \langle R_L = \infty \rangle$	$R_{AB0}\frac{1+\frac{\varepsilon_1+m\varepsilon_2}{1+m}}{1+\Delta M_{12}}$	$R_{AB0}\frac{\left(1+\frac{\varepsilon_1}{1+m}\right)\left(1+\frac{\varepsilon_4}{1+m}\right)}{1+\Delta M_{14}}$	$R_{AB0}\frac{\left(1+\frac{\varepsilon_1}{1+m}\right)\left(1+\frac{m\varepsilon_3}{1+m}\right)}{1+\Delta M_{13}}$	$R_{AB0}\frac{1+\frac{\varepsilon_1}{1+m}}{1+\Delta M_1}$
	In balance	R_{AB0}	$R_{AB0} = R_{10}\frac{n(1+m)}{(1+n)}$				
	Its relative increment	$\varepsilon^{\infty}_{AB}$	$\frac{1}{1+\Delta M(\varepsilon_i)}\left(\frac{n\varepsilon_{12}+\varepsilon_{43}}{1+n}+\varepsilon_{12}\varepsilon_{43}\right)$ where $\varepsilon_{12} \equiv \frac{\varepsilon_1+m\varepsilon_2}{1+m}$; $\varepsilon_{43} \equiv \frac{\varepsilon_4+m\varepsilon_3}{1+m}$	$\frac{n(\varepsilon_1+m\varepsilon_2)}{(1+n)(1+m)+\varepsilon_1+m\varepsilon_2}$	$\frac{n\varepsilon_1+\varepsilon_4+\varepsilon_1\varepsilon_4\left(\frac{1+n}{1+m}\right)}{(1+n)(1+m)+\varepsilon_1+n\varepsilon_4}$	$\frac{n\varepsilon_1+m\varepsilon_3+m\varepsilon_1\varepsilon_3\left(\frac{1+n}{1+m}\right)}{(1+n)(1+m)+\varepsilon_1+mn\varepsilon_3}$	$\frac{n\varepsilon_1}{(1+n)(1+m)+\varepsilon_1}$
3	**Output resistance:** $r_{22} = \frac{(R_1+R_4)(R_2+R_3)}{\sum R_i}$ $\equiv R_{CD0}(1+\varepsilon^{\infty}_{CD})$	R^{∞}_{CD}	$\frac{R_{10}(1+n+\varepsilon_1+n\varepsilon_4)m(1+n+\varepsilon_2+n\varepsilon_3)}{(1+m)(1+n)+\varepsilon_1+m\varepsilon_2+n(\varepsilon_4+m\varepsilon_3)}$ $= R_{CD0}\frac{(1+\varepsilon_{14})(1+\varepsilon_{23})}{1+\Delta M} \quad \langle R_G = \infty \rangle$	$R_{CD0}\frac{\left(1+\frac{\varepsilon_1}{1+n}\right)\left(1+\frac{\varepsilon_2}{1+n}\right)}{1+\Delta M_{12}}$	$R_{CD0}\frac{1+\frac{\varepsilon_1+n\varepsilon_4}{1+n}}{1+\Delta M_{14}}$	$R_{CD0}\frac{\left(1+\frac{\varepsilon_1}{1+n}\right)\left(1+\frac{n\varepsilon_3}{1+n}\right)}{1+\Delta M_{13}}$	$R_{CD0}\frac{1+\frac{\varepsilon_1}{1+n}}{1+\Delta M_1}$
	In balance	R_{CD0}	$R_{CD0} = R_{10}\frac{m(1+n)}{(1+m)}$				
	Its relative increment	$\varepsilon^{\infty}_{CD}$	$\frac{1}{1+\Delta M(\varepsilon_i)}\left(\frac{m\varepsilon_{14}+\varepsilon_{23}}{1+m}+\varepsilon_{14}\varepsilon_{23}\right)$ where $\varepsilon_{14} \equiv \frac{\varepsilon_1+n\varepsilon_4}{1+n}$; $\varepsilon_{23} \equiv \frac{\varepsilon_2+n\varepsilon_3}{1+n}$	$\frac{m\varepsilon_1+\varepsilon_2+\varepsilon_1\varepsilon_2\left(\frac{1+m}{1+n}\right)}{(1+n)(1+m)+\varepsilon_1+m\varepsilon_2}$	$\frac{m(\varepsilon_1+n\varepsilon_4)}{(1+n)(1+m)+\varepsilon_1+n\varepsilon_4}$	$\frac{m\varepsilon_1+n\varepsilon_3+n\varepsilon_1\varepsilon_3\left(\frac{1+m}{1+n}\right)}{(1+n)(1+m)+\varepsilon_1+mn\varepsilon_3}$	$\frac{m\varepsilon_1}{(1+n)(1+m)+\varepsilon_1}$

Table 1b. Matrix $\mathbf{Z_R}$ elements of particular cases: if increments of two arm resistances R_1, R_2 are opposite or small or if all arm initial resistances are equal.

No	Bridge open circuit terminal parameters (matrix $\mathbf{Z_R}$ elements)	Symbol	Bridge of arbitrary resistance ratio m and n		Initially double symmetric bridge $M = N = 1$		
			Variable: only two resistances: R_1, R_2		Variable: all resistances: R_1, R_2, R_3, R_4		
			(e) Large value opposite increments	(f) Small increments[1)]	(g) Large increments	(h) Very small increments[2)]	(i) $\lvert\varepsilon_i\rvert = \varepsilon$
			 $R_1 = R_{10}\ (1+\varepsilon)$; $R_2 = mR_{10}(1-\varepsilon)$; $R_3 = mnR_{10}$; $R_4 = nR_{10}$ $U'_{DC} = J\ r_{21}$	[1)] $\lvert\Delta R_1 + \Delta R_2\rvert \ll \sum R_{10}$ equal to: $\lvert\varepsilon_1 + m\varepsilon_2\rvert \ll (1+m)(1+n)$	$R_1 = R_{10}(1+\varepsilon_1)$; $R_2 = R_{10}(1+\varepsilon_2)$; $R_3 = R_{10}(1+\varepsilon_3)$; $R_4 = R_{10}(1+\varepsilon_4)$ [2)] $\left\lvert\sum \Delta R_i\right\rvert \ll 4R_{10}$ equal to: $\left\lvert\sum \varepsilon_i\right\rvert \ll 4$		$R_i = R_{10}(1\pm\varepsilon)$ $\varepsilon_1 = \varepsilon_3 = \varepsilon$ $\varepsilon_2 = \varepsilon_4 = -\varepsilon$
1	Current to Voltage transmittance	r_{21}	$t'_0\dfrac{\varepsilon_1-\varepsilon_2}{1+\Delta M_{12}} = t'_0\dfrac{2\varepsilon}{1+\dfrac{\varepsilon(1-m)}{(1+n)(1+m)}}$	$\approx t'_0(\varepsilon_1-\varepsilon_2)$	$t'_0\dfrac{\varepsilon_1-\varepsilon_2+\varepsilon_3-\varepsilon_4+\varepsilon_1\varepsilon_3-\varepsilon_2\varepsilon_4}{1+\dfrac{\varepsilon_1+\varepsilon_2+\varepsilon_3+\varepsilon_4}{4}}$	$\approx t'_0(\varepsilon_1-\varepsilon_2+\varepsilon_3-\varepsilon_4)$	$t'_0 4\varepsilon = \varepsilon R_{10}$
	Its initial sensitivity	t'_0	$t'_0 = R_{10}\dfrac{mn}{(1+m)(1+n)}$		$t'_0 = \dfrac{R_{10}}{4}$		
	Its denominator increment	ΔM	$\Delta M_{12} = \dfrac{\varepsilon(1-m)}{(1+n)(1+m)}$	$\Delta M_{12} \approx 0$	$\Delta M = \dfrac{\varepsilon_1+\varepsilon_2+\varepsilon_3+\varepsilon_4}{4}$	$\Delta M \approx 0$	$\Delta M = 0$
2	**Input resistance:** $R^{\infty}_{AB} \equiv R_{AB0}(1+\varepsilon^{\infty}_{AB})$	R^{∞}_{AB}	$R_{AB0}\dfrac{1+\dfrac{\varepsilon(1-m)}{(1+m)}}{1+\Delta M_{12}} = n\dfrac{(1+m)^2+\varepsilon(1-m^2)}{(1+m)(1+n)+\varepsilon(1-m)}R_{10}$	$\approx R_{AB0}\left(1+n\dfrac{\varepsilon_1+m\varepsilon_2}{(1+n)(1+m)}\right)$	$R_{10}\dfrac{\left(1+\dfrac{\varepsilon_1+\varepsilon_2}{2}\right)\left(1+\dfrac{\varepsilon_3+\varepsilon_4}{2}\right)}{1+0,25(\varepsilon_1+\varepsilon_2+\varepsilon_3+\varepsilon_4)}$	$\approx R_{10}\left(1+\dfrac{\varepsilon_1+\varepsilon_2+\varepsilon_3+\varepsilon_4}{4}\right)$	R_{10}
	In balance	R_{AB0}	$R_{AB0} = R_{10}\dfrac{n(1+m)}{(1+n)}$		R_{10}		
	Its related increment	$\varepsilon^{\infty}_{AB}$	$\dfrac{n\varepsilon(1-m)}{(1+n)(1+m)+\varepsilon(1-m)}$	$\approx \dfrac{n(\varepsilon_1+m\varepsilon_2)}{(1+n)(1+m)}$	$0,25\dfrac{\varepsilon_1+\varepsilon_2+\varepsilon_3+\varepsilon_4+(\varepsilon_1+\varepsilon_2)(\varepsilon_3+\varepsilon_4)}{1+0,25(\varepsilon_1+\varepsilon_2+\varepsilon_3+\varepsilon_4)}$	$\approx \dfrac{\varepsilon_1+\varepsilon_2+\varepsilon_3+\varepsilon_4}{4}$	0
3	**Output resistance:** $R^{\infty}_{CD} \equiv R_{CD0}(1+\varepsilon^{\infty}_{CD})$	R^{∞}_{CD}	$R_{CD0}\dfrac{1-\left(\dfrac{\varepsilon}{1+n}\right)^2}{1+\Delta M_{12}} = m\dfrac{(1+n)^2-\varepsilon^2}{(1+n)(1+m)+\varepsilon(1-m)}R_{10}$	$\approx R_{CD0}\left(1+\dfrac{\varepsilon_1+\varepsilon_2}{1+n}\right)$	$R_{10}\dfrac{\left(1+\dfrac{\varepsilon_1+\varepsilon_4}{2}\right)\left(1+\dfrac{\varepsilon_2+\varepsilon_3}{2}\right)}{1+0,25(\varepsilon_1+\varepsilon_2+\varepsilon_3+\varepsilon_4)}$	$\approx R_{10}\left(1+\dfrac{\varepsilon_1+\varepsilon_2+\varepsilon_3+\varepsilon_4}{4}\right)$	R_{10}
	In balance	R_{CD0}	$R_{CD0} = R_{10}\dfrac{m(1+n)}{1+m}$		R_{10}		
	Its related increment	$\varepsilon^{\infty}_{CD}$	$\dfrac{-\varepsilon(1-m)+\varepsilon^2\left(\dfrac{1+m}{1+n}\right)}{(1+n)(1+m)+\varepsilon(1-m)}$	$\approx \dfrac{m\varepsilon_1+\varepsilon_2}{(1+n)(1+m)}$	$0,25\dfrac{\varepsilon_1+\varepsilon_2+\varepsilon_3+\varepsilon_4+(\varepsilon_1+\varepsilon_4)(\varepsilon_2+\varepsilon_3)}{1+0,25(\varepsilon_1+\varepsilon_2+\varepsilon_3+\varepsilon_4)}$	$\approx \dfrac{\varepsilon_1+\varepsilon_2+\varepsilon_3+\varepsilon_4}{4}$	0

4 SIMPLIFIED BRIDGE EQUIVALENT CIRCUITS

In particular cases of the operation of the bridge as the twoport type **X**, some of its equivalent circuits could also be applied. Two of them are given in Figures 3(b) and 3(c).

The circuit in Figure 3(b) is twoport of type T. It has only the same values of input–output terminal parameters as the bridge (elements of its matrix $\mathbf{Z_R}$) and could be used when outside the bridge there is no other connection between circuits of both sides. In this twoport, two terminals, B and C, are connected together but in the bridge they are not. The equivalent circuit type **Π** of twoport X also exists and could be sometimes more proper, for example, for describing the short circuit current of the conductance bridge.

The other circuit, given in Figure 3(c), is the type 2T symmetric twoport. Two horizontally drawn resistances of circuit in Figure 3(b) are divided here to two parts each. They depend on bridge resistances, each one differently as follows

$$R_{\mathrm{A}0'} = \frac{R_4(R_1 + R_2)}{\sum R_{\mathrm{i}}} \tag{11a}$$

$$R_{\mathrm{B}0''} = \frac{R_2(R_3 + R_4)}{\sum R_{\mathrm{i}}} \tag{11b}$$

$$R_{\mathrm{C}0''} = \frac{R_4(R_2 + R_3)}{\sum R_{\mathrm{i}}} \tag{11c}$$

$$R_{\mathrm{D}0'} = \frac{R_2(R_1 + R_4)}{\sum R_{\mathrm{i}}} \tag{11d}$$

This circuit not only keeps the twoport matrix parameters equal but voltages of all terminals as well. Only input resistances from the arms' sides are here different from ones of 4R bridge, for example, input resistance from terminals AD. This equivalent circuit of the bridge has not been found in literature yet but could be very useful in the analysis of the bridge and different circuits connected to it, including operational amplifiers, in the case in which they are supplied from the common source.

In both the equivalent circuits given above, the vertical resistance simulates the transmittance r_{21}. If r_{21} is negative, this resistance should also be negative or wires of output terminals should be reversed.

In conclusion, it is clear that in the general case it is impossible to describe the operation of any four-terminal circuit, working as the twoport in full ranges of its internal resistance increment changes, only by three-terminal circuit

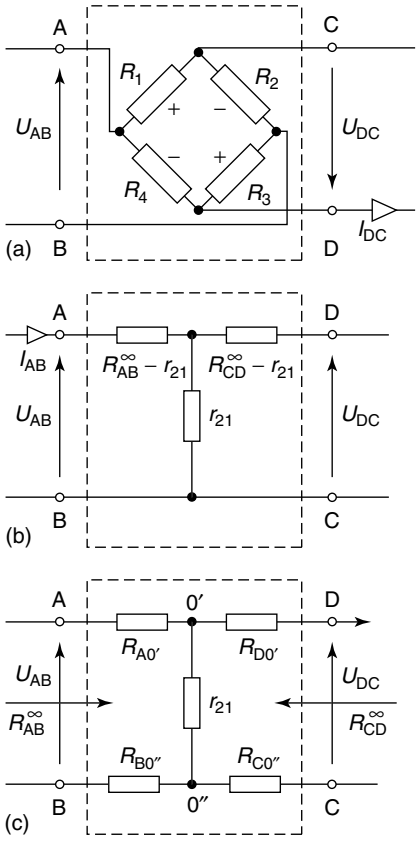


Figure 3. Bridge 4R as twoport and its equivalent T and 2T twoport circuits: (a) bridge as differently drawn twoport **X**, (b) twoport T circuit equivalent only for one sign of U_{DC} and of input and of output parameters of the bridge (the same elements of matrix $\mathbf{Z}_R$); (c) 2T symmetric twoport equivalent as circuit b) and of all terminal voltages.

of **T** or of **Π** type structure. Reversing of output lines is needed when output signal is crossing zero. Then the only proper circuit for simulating in such a case is the twoport X, that is, the four-arm bridge itself.

5 RELATIONS BETWEEN OPEN CIRCUIT TERMINAL PARAMETERS OF THE 4R BRIDGE

From (6a) and (7a) of terminal resistances $R_{\mathrm{AB}0}$, $R_{\mathrm{CD}0}$, of the balanced 4R bridge, it resulted that their product was also equal to the product of opposite arms' resistances as in the balance condition (2a) or (5a), then

$$R_{\mathrm{AB}0}R_{\mathrm{CD}0} = mn\ R_{10}^2 = R_{10}R_{30} = R_{20}R_{40} \tag{12}$$

The above expression could be called *extended balance condition of the four-arm bridge*.

Additionally, from (5b) results also the simple relationship for initial parameters of the balanced 4R bridge

$$\frac{R_{AB0}R_{CD0}}{\sum R_{i0}} = t_0' \tag{13}$$

Then they are related to each other and an arbitrary value cannot be assigned to all of them. Two above relationships could be applied to the syntheses of elements of the 4R bridge equivalent to the unknown network's inside structure, based only on the results of measurements performed on its terminals.

From the values of resistances R_{AB0}, R_{CD0}, ratios m and n of balance state resistances of the four-arm bridge could be directly obtained:

$$m = \frac{r_{CD0}\left(r_{AB0} - 1\right)}{r_{CD0} - 1} \tag{14a}$$

and

$$n = \frac{r_{AB0}\left(r_{CD0} - 1\right)}{r_{AB0} - 1} \tag{14b}$$

where: $r_{AB0} \equiv R_{AB0}/R_{10}$; $r_{CD0} \equiv R_{CD0}/R_{10}$.

When resistance R_{10} is given then single solutions $m > 0$, $n > 0$ are obtain when $r_{AB0} > 1$ and $r_{CD0} > 1$ or $r_{AB0} < 1$ and $r_{CD0} < 1$. If $r_{AB0} = 1$ and $r_{CD0} = 1$ then there is infinite number of solutions $n = m^{-1}$.

These values of m and n and condition (13) should be taken also as some kind of criterion to choose the structure of the equivalent circuit of the two ports 'black box'. Could it be considered only as a four arms resistance bridge or if some additional resistances in series to bridge nodes are needed.

It follows from (6) and (7) that relative increments $\varepsilon_{AB}^{\infty}$ and $\varepsilon_{CD}^{\infty}$ of the bridge resistances R_{AB}^{∞}, R_{CD}^{∞} are related to each other through the common expression $1 + \varepsilon_{\Sigma R}$, as follows

$$\begin{aligned}\frac{1+\varepsilon_{CD}^{\infty}}{1+\varepsilon_{AB}^{\infty}} &= \frac{\left(1+\varepsilon_{14}\right)\left(1+\varepsilon_{23}\right)}{\left(1+\varepsilon_{12}\right)\left(1+\varepsilon_{34}\right)} \\ &= \left(\frac{1+m}{1+n}\right)^2 \frac{\begin{matrix}\left(1+n+\varepsilon_1+n\varepsilon_4\right) \\ \times\left(1+n+\varepsilon_2+n\varepsilon_3\right)\end{matrix}}{\begin{matrix}\left(1+m+\varepsilon_1+m\varepsilon_2\right) \\ \times\left(1+m+\varepsilon_4+m\varepsilon_3\right)\end{matrix}}\end{aligned} \tag{14c}$$

Actual arm resistances or their relative increments of the unbalanced bridge are not easy to reveal from measurements of input–output terminal parameters. In the general case, the set of nonlinear equations should be solved.

6 BRIDGE OPEN CIRCUIT TERMINAL PARAMETERS' LINEARIZATION

Bridge open circuit terminal parameters are related practically linearly only for small values of independent resistance increments ε_i. It is possible to obtain the linear unbalance function $f'(\varepsilon_i)$ for large values of ε_i too. In this case, increments should simultaneously be at least in two bridge arms and they are not independent. It happens when increments of the sum of resistances of neighboring arms are equal to each other, that is, $\varepsilon_{12} = \varepsilon_{34}$, or $\varepsilon_{14} = \varepsilon_{23}$. In these cases, from (6b) and (9) or from (7b) and (9) the outcomes respectively are $\varepsilon_{AB}^{\infty} = \varepsilon_{12} = \varepsilon_{\Sigma R}$ or $\varepsilon_{CD}^{\infty} = \varepsilon_{14} = \varepsilon_{\Sigma R}$ and according to the definition (8), *generalized bridge linearity conditions* could be formulated as

$$\varepsilon_1 + m\varepsilon_2 = \varepsilon_4 + m\varepsilon_3 \tag{15a}$$

or

$$\varepsilon_1 + n\varepsilon_4 = \varepsilon_2 + n\varepsilon_3 \tag{15b}$$

With every one of them, formulas of bridge terminal parameters become simpler. If condition (15a) is valid, then it is possible to find from (5) that

$$r_{21} = R_{10}\frac{n}{1+n}(\varepsilon_1 - \varepsilon_4) \tag{16a}$$

and from (6a), that

$$\varepsilon_{AB}^{\infty} = \frac{\varepsilon_1 + m\varepsilon_2}{1+m} \tag{16b}$$

In this case, output resistance R_{CD}^{∞} depends nonlinearly on increments ε_i of arm resistances.

If condition (15b) is satisfied then

$$r_{21} = R_{10}\frac{m}{1+m}(\varepsilon_1 - \varepsilon_2) \tag{17a}$$

and

$$\varepsilon_{DC}^{\infty} = \frac{\varepsilon_1 + n\varepsilon_4}{1+n} \tag{17b}$$

and now open circuit input resistance R_{AB}^{∞} nonlinearly depends on increments ε_i.

In the literature, only some particular cases of general linearity conditions (15a) and (15b) can be found, for example, for $m = 1$, $n = 1$. If each of them is to be separately satisfied, at most three increments can be independent, and the fourth one results from others according to the above formulas. It is difficult to implement such cases in practice. Furthermore, in any pair of terminal parameter formulas (16a,b) or (17a,b) three increments ε_i are present and an additional equation is needed to find all of them separately from the terminal measurements. It is not so easy because the third terminal parameter (R_{CD}^{∞} or R_{AB}^{∞}) depends on ε_i

nonlinearly. It is much easier to obtain two pairs of increments related to each other within each pair. It can be done by respectively forming the relations to common measured variables. For example, stationary resistance connected in series to the arm resistance (or parallel to arm conductance) proportionally decreases its increment. In force or pressure sensors, it is also possible to place two strain gauges in points stressed differently but of known ratio. For condition (15a), it could be

$$\varepsilon_1 = m\varepsilon_3, \quad \varepsilon_2 = m\varepsilon_4 \tag{18a}$$

or

$$\varepsilon_1 = -m\varepsilon_2, \quad \varepsilon_4 = -m\varepsilon_3 \tag{18b}$$

It is obvious that the output signal is higher if increments of neighboring arms are of the opposite signs, that is, $sign\ \varepsilon_4 = -sign\ \varepsilon_1$ and $sign\ \varepsilon_2 = -sign\ \varepsilon_1$.

In (18a) conditions, positive values of increments ε_i are limited only by the permissible dissipate power of arm resistances or by maximum voltage of the current supply source. Increment $\varepsilon_{AB}^{\infty}$ of the input resistance R_{AB}^{∞} of the unloaded bridge is now given by the formula

$$\varepsilon_{AB}^{\infty} = \frac{\varepsilon_1 + \varepsilon_4}{1 + m} \tag{19}$$

From formulas (16a), (18a), and (19) it is now possible to find increments of any arm.

As an example of the (18b) case is the bridge built by connecting two resistance potentiometers with output on their slides. Both ends of these resistances limit extreme increments and are as follows: $|\varepsilon_i| \leq 1 + m$. In this bridge, the input resistance R_{AB}^{∞} of the open output is constant and the output resistance R_{CD} nonlinearly depends on out-of-balance increments, but it should be measured instead.

Linear relationships of the bridge transmittance r_{21} and one of its terminal resistances could be used in one variable measurement to correct the influence of another variable on the sensor set or to simultaneously obtain two variables measured indirectly on bridge terminals.

As an example of the first case, it is the use of the strain gauge bridge input resistance changes to compensate temperature influences on it.

7 TERMINAL PARAMETERS OF THE DC BRIDGE OF ANY SUPPLY SOURCE AND LOAD

If in the circuit there are branches that can be considered as internal or external diagonals connected to the 4R bridge diagonals of finite equivalent resistances of values comparable to terminal open circuit resistances, then bridge terminal parameters are different from the ones in matrix $\mathbf{Z_R}$ of 4R bridge alone. Here they are called *working terminal parameters*. Any four-terminal (4T) circuit operating as the X type twoport could be substituted by the equivalent 4R bridge of arm resistances and its terminal parameters dependent now on diagonal resistances. Current to voltage transmittance (or transfer resistance) r_{21} of unloaded bridge is replaced now by *working transmittance* $T \equiv U_{DC}/J$ and open circuit terminal resistances R_{AB}^{∞}, R_{CD}^{∞} – by working terminal resistances R_{AB}, R_{CD}.

From (6) of **Article 126, Electrical Bridge Circuits – Basic Information, Volume 2**, if $Z_i = R_i$, output voltage $U_{DC} = R_L I_{DC}$ is

$$U_{DC} = J\frac{R_1R_3 - R_2R_4}{M_{UJ}\left(R_i, R_G, R_L\right)} \equiv JT \tag{20}$$

where

$$T = \frac{R_1R_3 - R_2R_4}{M_{UJ}\left(R_i, R_G, R_L\right)} \tag{20a}$$

$$M_{UJ} = \Sigma R_i + \frac{1}{R_G}(R_1 + R_2)(R_3 + R_4) + \frac{1}{R_L}(R_1 + R_4)(R_2 + R_3) + \frac{A}{R_G R_L}$$

$$A = \Sigma R_i R_j R_k = R_1R_2(R_3 + R_4) + R_3R_4(R_1 + R_2) = \ldots$$

Numerators of the sensitivity T in (20) and then the balance condition are in all cases the same as before for the bridge without diagonals. The denominator is the function of bridge resistances R_i and of diagonal resistances R_G and R_L. When $R_G \to \infty$, the second and the last addend (component of the sum) in it become negligible. It is also the same with the third one and the last one of the addends if $R_L \to \infty$. For J = *constant*, and if also together $R_G \to \infty$ and $R_L \to \infty$ then output voltage U_{DC} is the highest one and given in (5) as U_{CD}^{∞}.

Similarly it is possible to find equivalent working terminal resistances R_{AB}, R_{CD}, of the bridge. After Karandeyev (1966) they are

$$R_{AB} = \frac{R_L(R_1 + R_2)(R_3 + R_4) + A}{R_L\Sigma R_i + (R_1 + R_4)(R_2 + R_3)} \tag{21}$$

and

$$R_{DC} = \frac{R_G(R_1 + R_4)(R_2 + R_3) + A}{R_G\Sigma R_i + (R_1 + R_2)(R_3 + R_4)} \tag{22}$$

Formulas (20) to (22) are quite complex. Then, in many cases it should be more useful in practice to apply equivalent terminal working parameters of the bridge only as functions of its matrix $\mathbf{Z_R}$ elements.

From (4) in **Article 126, Electrical Bridge Circuits – Basic Information, Volume 2**, formulas of the output voltage of the bridge loaded by R_L and supplied by the current or voltage source of equivalent internal resistance R_G could be obtained. They are

$$U_{DC} = J\frac{R_G R_L\, r_{21}}{\left(R_G + R_{AB}^{\infty}\right)\left(R_L + R_{CD}^{\infty}\right) - r_{21}^2} \tag{23a}$$

or after putting $J = E R_G^{-1}$:

$$U_{DC} = E\frac{R_L\, r_{21}}{\left(R_G + R_{AB}^{\infty}\right)\left(R_L + R_{CD}^{\infty}\right) - r_{21}^2} \tag{23b}$$

Formula (23a) simplified when $R_G \to \infty$ or/and $R_L \to \infty$ and formula (23b) when $R_G = 0$ or/and $R_L \to \infty$.

The equivalent source resistance R_G and load R_L are generally constant in measuring circuits. They could be variable in equivalent circuits of four-terminal (4T) tested objects and sensors, for example, Hall devices – see Chapters 6 and 7 by Warsza (1980) in the monograph by Kobus *et al.*

After applying formulas of resistances R_{AB}^{∞}, R_{CD}^{∞} and putting $R_G \equiv gR_{10} \equiv k_G R_{AB0}$ and $R_L \equiv lR_{10} \equiv k_L R_{CD0}$ it is

$$U_{DC} = J\frac{r_{21}}{\left(1 + \dfrac{1+\varepsilon_{AB}^{\infty}}{k_G}\right)\left(1 + \dfrac{1+\varepsilon_{CD}^{\infty}}{k_L}\right) - \dfrac{r_{21}^2}{R_G R_L}} \tag{24a}$$

and

$$U_{DC} = \frac{E}{R_G}\frac{r_{21}}{\left(1 + \dfrac{1+\varepsilon_{AB}^{\infty}}{k_G}\right)\left(1 + \dfrac{1+\varepsilon_{CD}^{\infty}}{k_L}\right) - \dfrac{r_{21}^2}{R_G R_L}} \tag{24b}$$

Then the output voltage depends on resistances R_G, R_L, and also on their ratios to terminal resistances of the balanced bridge given by coefficients k_G, k_L.

Working terminal resistances of the bridge within both constant diagonal resistances are given by formulas:

$$R_{AB} = R_{AB}^{\infty} - \frac{r_{21}^2}{R_{CD}^{\infty} + R_L} = R_{AB0}\left[1 + \varepsilon_{AB}^{\infty} - \frac{r_{12}^2}{mn\ R_{10}^2\left(1 + \varepsilon_{CD}^{\infty} + k_L\right)}\right] \tag{25a,b}$$

and

$$R_{CD} = R_{CD}^{\infty} - \frac{r_{21}^2}{R_{AB}^{\infty} + R_G} = R_{CD0}\left[1 + \varepsilon_{CD}^{\infty} - \frac{r_{12}^2}{mnR_{10}^2\left(1 + \varepsilon_{AB}^{\infty} + k_G\right)}\right] \tag{26a,b}$$

These resistances in balance state are equal to R_{AB0} and to R_{CD0}. Out of the balance both of them decrease with the square of transmittance r_{21}, that is, independently of its sign. Each of them also decreases on the decreasing of the bridge opposite diagonal resistances R_L or R_G.

If $R_L = 0$ or $R_G = 0$, short circuit resistances obtained are R_{AB}^0, R_{CD}^0. After comparison of (25a) and (26a) for any value of r_{21} it is

$$\frac{R_{CD}^0}{R_{CD}^{\infty}} = \frac{R_{AB}^0}{R_{AB}^{\infty}} = 1 - \frac{r_{21}^2}{R_{CD}^{\infty} R_{AB}^{\infty}} \tag{27a}$$

or

$$R_{CD}^0 R_{AB}^{\infty} = R_{AB}^0 R_{CD}^{\infty} \tag{27b}$$

Then ratios of working open circuit and short circuit terminal resistances of both ports are equal and decrease with r_{21}^2.

When $R_L = 0$, from (23a) or (24a), and (25a) two forms of current I_{DC}^0 of the bridge shortened output could be found:

$$I_{DC}^0 = \frac{U_{DC}^{\infty}}{R_{CD}} = J\frac{R_G r_{21}}{\left(R_G + R_{AB}^{\infty}\right)R_{CD}^{\infty} - r_{21}^2} = \frac{J\ r_{21}}{R_{CD0}}\frac{1}{\left(1 + \dfrac{1+\varepsilon_{AB}^{\infty}}{k_G}\right)\left(1+\varepsilon_{CD}^{\infty}\right) - \dfrac{r_{21}^2}{k_G\ mn\ R_{10}^2}} \tag{28a,b}$$

All the above formulas became simpler, in particular, cases of ideal current or voltage supply sources and of the load resistance $R_L = 0$ or $R_L = \infty$. All such cases are given in Table 2.

The output voltage of any case could be expressed similarly as U_{DC}^{∞} in formula (5), that is:

$$U_{DC} \equiv T_0 f_{UJ}(\varepsilon_i) \tag{29}$$

where

$T_0 \equiv J(R_{10}R_{30})/M_{UJ0} = E/(R_G)(R_{10}R_{30})/M_{UJ0}$ – initial sensitivity of the working output voltage of the bridge supplied by source of internal resistance $R_G > 0$ and loaded by any $R_L > 0$,

$f_{UJ}(\varepsilon_i) \equiv \Delta L(\varepsilon_i)/1 + \varepsilon_{UJ}(\varepsilon_i)$ – unbalance function of the working output voltage of the bridge.

$\Delta L(\varepsilon_i) = \varepsilon_1 + \varepsilon_2 + \varepsilon_3 - \varepsilon_4 + \varepsilon_1\varepsilon_3 - \varepsilon_2\varepsilon_4$ – numerator of above function (as for 4R bridge alone),

$\varepsilon_{UJ}(\varepsilon_i) \equiv \Delta M_{UJ}(\varepsilon_i)/M_{UJ0}$ – relative increment of its denominator,

Table 2. Bridge 4R output signals and terminal resistances of any mode of DC supply and load as functions of its $\mathbf{Z_R}$ matrix elements.

No	Bridge output signals and terminal resistances	Symbol	General case	Particular cases		
			$U_{AB} = R^{\infty}_{AB} I_{AB} - r_{21} I_{DC}$ $U_{DC} = r_{21} I_{AB} - R^{\infty}_{CD} I_{DC}$	(a) Current supply $R_G \to \infty,\ I_{AB} = J$	(b) Voltage supply $R_G \to 0,\ U_{AB} = E$	(c) Source with resistance R_G, voltage U^{∞}_{CD} as output signal $R_L \to \infty,\ I_{DC} = 0$
1	**Output voltage:** $U_{DC} = J\dfrac{R_1R_3 - R_2R_4}{M_{UJ}} \equiv T_0\dfrac{\Delta L(\varepsilon_i)}{1+\Delta M_L(\varepsilon_i)}$	U_{DC}	$J\dfrac{R_G R_L r_{21}}{(R_G + R^{\infty}_{AB})(R_L + R^{\infty}_{CD}) - r^2_{21}} = \dfrac{Jr_{21}}{\left(1+\dfrac{1+\varepsilon^{\infty}_{AB}}{k_G}\right)\left(1+\dfrac{1+\varepsilon^{\infty}_{CD}}{k_L}\right) - \dfrac{r^2_{21}}{R_G R_L}}$	$J\dfrac{R_L r_{21}}{R_L + R_{CD0}(1+\varepsilon^{\infty}_{CD})} = J\dfrac{r_{21}}{1+\dfrac{1+\varepsilon^{\infty}_{CD}}{k_L}}$	$E\dfrac{R_L r_{21}}{R_{AB0}(1+\varepsilon^{\infty}_{AB})[R_L + R_{CD0}(1+\varepsilon^{\infty}_{CD})] - r^2_{21}} = \dfrac{Er_{21}}{R_{AB0}}\dfrac{1}{(1+\varepsilon^{\infty}_{AB})\left(1+\dfrac{1+\varepsilon^{\infty}_{CD}}{k_L}\right) - \dfrac{r^2_{21}}{R_{AB0}R_L}}$	$J\dfrac{R_G r_{21}}{R_G + R_{AB0}(1+\varepsilon^{\infty}_{AB})} = J\dfrac{r_{21}}{1+\dfrac{1+\varepsilon^{\infty}_{AB}}{k_G}}$
	Initial sensitivity $T_0 \equiv J\dfrac{R_{10}R_{30}}{M_{UJ0}} = \dfrac{E}{R_G}\dfrac{R_{10}R_{30}}{M_{UJ0}}$	T_0	$J\dfrac{R_{10}mn}{(1+m)(1+n)}\dfrac{k_G k_L}{(1+k_G)(1+k_L)} = E\dfrac{m}{(1+m)^2}\dfrac{1}{(1+k_G)}\dfrac{k_L}{(1+k_L)}$	$J\dfrac{R_{10}mn}{(1+m)(1+n)}\dfrac{k_L}{1+k_L}$	$E\dfrac{m}{(1+m)^2}\dfrac{k_L}{(1+k_L)}$	$T_0 = J\dfrac{R_{10}mn}{(1+m)(1+n)}\dfrac{k_G}{1+k_G} \times E\dfrac{m}{(1+m)^2}\dfrac{1}{(1+k_G)}$
	Denominator of U_{DC} in balance	M_{UJ0}	$R_{10}\left[1+n+\dfrac{n}{g}(1+m)\right]\left[1+m+\dfrac{m}{l}(1+n)\right]$	$R_{10}(1+n)\left[1+m+\dfrac{m}{l}(1+n)\right]$	$R_{10}[g(1+n)+n(1+m)] \times\left[1+m+\dfrac{m}{l}(1+n)\right]$	$R_{10}\left[1+n+\dfrac{n}{g}(1+m)\right](1+m)$
2	**Input resistance:** In balance: $R_{AB0} = R_{10}\dfrac{n(1+m)}{(1+n)}$	R_{AB}	$R^{\infty}_{AB}\left[1 - \dfrac{r^2_{21}}{R^{\infty}_{AB}(R^{\infty}_{CD} + R_L)}\right] = R_{AB0}(1+\varepsilon^{\infty}_{AB}) - \dfrac{r^2_{21}}{R_{CD0}(1+\varepsilon^{\infty}_{CD}+k_L)}$			$R_{AB0}(1+\varepsilon^{\infty}_{AB})$
3	**Output resistance:** In balance: $R_{CD0} = R_{10}\dfrac{m(1+n)}{(1+m)}$	R_{CD}	$R^{\infty}_{CD}\left[1 - \dfrac{r^2_{21}}{R^{\infty}_{CD}(R^{\infty}_{AB} + R_G)}\right] = R_{CD0}(1+\varepsilon^{\infty}_{CD}) - \dfrac{r^2_{21}}{R_{AB0}(1+\varepsilon^{\infty}_{AB}+k_G)}$	$R_{CD0}(1+\varepsilon^{\infty}_{CD})$	$R^{\infty}_{CD} - \dfrac{r^2_{21}}{R^{\infty}_{AB}} = R^{\infty}_{CD}\left(1 - \dfrac{r^2_{21}}{R^{\infty}_{CD}R^{\infty}_{AB}}\right)$	$R_{CD0}(1+\varepsilon^{\infty}_{CD}) - \dfrac{r^2_{21}}{R_{AB0}(1+\varepsilon^{\infty}_{AB}+k_G)}$
4	**Shorted output current:** $I^0_{DC} = \dfrac{U^{\infty}_{DC}}{R_{CD}}$	I^0_{DC}	$\dfrac{Jr_{21}}{R_{CD0}}\dfrac{1}{\left(1+\dfrac{1+\varepsilon^{\infty}_{AB}}{k_G}\right)(1+\varepsilon^{\infty}_{CD}) - \dfrac{r^2_{21}}{R_G R_{CD0}}}$	$J\dfrac{r_{21}}{R_{CD0}(1+\varepsilon^{\infty}_{CD})}$	$\dfrac{Er_{21}}{mnR_{10}}\dfrac{1}{(1+\varepsilon^{\infty}_{AB})(1+\varepsilon^{\infty}_{CD}) - \dfrac{r^2_{21}}{mnR^2_{10}}}$	$J\dfrac{R_G r_{21}}{(R_G + R^{\infty}_{AB})R^{\infty}_{CD} - r^2_{21}}$

Note: Some useful relations: $E = JR_G$; $R_G \equiv gR_{10}$; $R_L \equiv lR_{10}$; $\dfrac{R_G}{R_{AB0}} \equiv k_G$; $\dfrac{R_L}{R_{CD0}} \equiv k_L$; $r_{12} = \dfrac{R_{10}mn}{(1+m)(1+n)}$; $\Delta L(\varepsilon_i) = \varepsilon_1 - \varepsilon_2 + \varepsilon_3 - \varepsilon_4 + \varepsilon_1\varepsilon_3 - \varepsilon_2\varepsilon_4$; $\varepsilon^{\infty}_{AB}, \varepsilon^{\infty}_{CD}$ – see Table 1a and b.

$M_{UJ} \equiv M_{UJ0} + \Delta M_{UI}(\varepsilon_i)$ – denominator of (20) expressed as a sum of its balance value and the resultant increment (sum of 14 components dependent differently on products of 6 resistances increments of bridge with diagonals – 4T circuit) – see (8) and (9) of **Article 126, Electrical Bridge Circuits – Basic Information, Volume 2**.

The working output voltage initial sensitivity T_0 is

$$T_0 = JR_{10}\frac{mn}{(1+n)(1+m)}\frac{k_G}{1+k_G}\frac{k_L}{1+k_L} \tag{30a}$$

or

$$T_0 = E\frac{m}{(1+m)^2}\frac{k_L}{(1+k_G)(1+k_L)} \tag{30b}$$

If supplied current $J = constant$, initial sensitivity $T_0 \to T_0' = JR_{10}(mn/(1+m)(1+n))$ when: $k_G \to \infty$ and $k_L \to \infty$. It is the highest one: $T_{0\ max}' = JR_{10}$ is when $m \to \infty, n \to \infty$.

For $E = constant$ $T_0 \to T_0'' = E(m)/(1+m)^2$ when $k_G \to 0, k_L \to \infty$ and is independent of n. Maximal sensitivity $T_{0\ max}'' = 1/4E$ is now when $m = 1$. It is independent of n. Particular cases are in Table 2.

After applying formula (5b) to (30a) is obtained

$$T_0 = Jt_0'\frac{k_G}{1+k_G}\frac{k_L}{1+k_L} \tag{31}$$

where: $t_0' = (R_{AB0}R_{CD0}/\sum R_{i0})$ – initial sensitivity of transmittance r_{21}.

This is the extension of formula (13) to the case of the resistance bridge with additional diagonal resistances. It relates the bridge initial working sensitivity T_0 to the product of the current J, the initial open-circuit sensitivity t_0' and coefficients of diagonal branches – on the right side.

If changes of arm immitances are given as conductance increments, then it is possible to obtain linear relations of two bridge terminal parameters in dual circuit, that is, bridge supplied from the ideal voltage source E and the output signal as the short circuited output current I_{DC}^0.

More information about unbalanced DC bridges can be found in the literature given below. Measurements of two to four variables provided by the resistance bridge, supplied classically from one current source and unconventionally by two of them connected in parallel to opposite bridge arms, are discussed in detail in Warsza (2002).

RELATED ARTICLES

Article 56, Transfer Characteristics of Instrument Stages, Volume 1; **Article 112, Systematic Description of Sensors, Volume 2**; **Article 126, Electrical Bridge Circuits – Basic Information, Volume 2**; **Article 202, Current Measurement, Volume 3**; **Article 204, Capacitance and Inductance Measurement, Volume 3**.

REFERENCES

Anderson, K.F. (1997) The New Current Loop, An Instrumentation and Measurement Circuit Topology. *IEEE Transactions on Instrumentation and Measurement*, October 1997, http://www.vm-usa.com/links.html.

Bentley, J.P. (1995) *Principles of Measurement Systems*, 3rd edn, Longman Scientific & Technical UK Ltd., Harlow, Essex, (pp. 178–185).

Carter, G.W., Richardson, A. (1971) *Technique of Circuit Analysis*, Cambridge University Press, London UK, (pp. 17–22, 248–281).

Dally, J.W. (ed.) (1993) *Instrumentation for Engineering Measurements*, 2nd edn, John Wiley & Sons, New York (pp. 170–176).

Karandeev, K.B. (1966) *Bridge and Potentiometer Methods in Electrical Measurements*, Energia, Moscow (1966) and Peace Publisher MIR, Moscow (1967) (English translation from Russian edition 1960), Chapter 5.

Kester W. (1999) *Practical Design Techniques for Sensor Signal Conditioning*, Analog Devices Co., Norwood, MA.

Morris A.S. (1993) *Principles of Measurement and Instrumentation*, 2nd edn, Prentice Hall (pp. 106–118).

Padmanabhan T.R. (2000) *Industrial Instrumentation*, Springer Verlag, London (pp. 29–42).

Warsza, Z.L. (1980) Chapters: 1, 6 i 7 in monography: Kobus, A., Tuszynski J., Warsza Z.L., *Technika Hallotronowa (Hall Devices Technology)*, WNT Warszawa, in Polish.

Warsza, Z.L. (2002) Bridges Supplied by Two Current Sources – New Tool for Impedance Measurements and Signal Conditioning, in *Proceedings of IMEKO-TC 7 Symposium*, Cracow University of Technology, Cracow Poland, (pp. 231–236).

Warsza Z.L. (2004) *Resistance Bridges as Signal Conditioning Circuits of Few Variable Measurements*, Prace Naukowe (Reasearch Works) 'Elektryka' **1**(7), 2004, Pulaski Technical University in Radom, Poland, pp. 78–157 (in Polish).

Zakrzewski, J. (2000) The Analogue Signal Processing in Measuring Systems with a Single Passive Sensors, in *Proceedings of XVI IMEKO World Congress*, Vol. IX, Viena (pp. 185–190).

FURTHER READING

Boyes, W. (ed.) *et al*. (2003) *Instrumentation Reference Book*, 3rd edn, Butterworth Heinemann, Elsevier Science, Boston, *DC bridge measurements*, by M.L. Sanderson (pp. 463–468).

Dyer, S.A. (ed.) (2001) *Survey of Instrumentation and Measurements*, Wiley- Interscience, New York, Section *Bridge Instruments* by Nicolas J. (pp. 309–317).

Holman J.P. (1994) *Experimental Methods for Engineers*, 6th edn, McGraw-Hill International Editions, New York, (pp. 141–144).

McGhee, J., *et al*. (1998) *Scientific Metrology*, 2nd edn, Lodard S.A., Lodz, (pp. 205–217, 289–297).

The Measurement and Automation Catalogue, Condensed Version (2003) *modules SCXI*, SCC, National Instruments™, Austin, TX (pp. 22–24). info@ni.com.

Signal Conditioning and Data Acquisition Solutions Catalogue; Analog Devices Co., Norwood, MA, http://www.analog.com/IOS.

128: Nature and Scope of AI Techniques

Ajith Abraham
Oklahoma State University, Stillwater, OK, USA

1 Introduction to Computational Intelligence 893
2 Artificial Neural Networks 894
3 Neural Network Architectures 895
4 Fuzzy Logic 895
5 If-then Rules and Fuzzy Inference Systems 896
6 Evolutionary Algorithms 897
7 Intelligent Paradigms 898
8 Hybrid Intelligent Systems 898
9 Models of Hybrid Systems 899
10 Summary 900
References 900

1 INTRODUCTION TO COMPUTATIONAL INTELLIGENCE

Machine intelligence dates back to 1936, when Alan Turing proposed the idea of a universal mathematics machine, a theoretical concept in the mathematical theory of computability. Turing and Emil Post independently proved that determining the decidability of mathematical propositions is equivalent to asking what sort of sequences of a finite number of symbols can be recognized by an abstract machine with a finite set of instructions.

Such a mechanism is now known as a Turing machine (Turing Machine, 2004). Turing's research paper addressed the question of machine intelligence, assessing the arguments against the possibility of creating an intelligent computing machine and suggesting answers to those arguments; it proposed the Turing test as an empirical test of intelligence (Turing, 1950).

The Turing test, called the *imitation game* by Alan Turing, measures the performance of a machine against that of a human being. The machine and a human (A) are placed in two rooms. A third person, designated the interrogator, is in a room apart from both the machine and the human. The interrogator cannot see or speak directly to either A or the machine, communicating with them solely through some text messages or even a chat window. The task of the interrogator is to distinguish between the human and the computer on the basis of questions he or she may put to both of them over the terminals. If the interrogator cannot distinguish the machine from the human, then, Turing argues, the machine may be assumed to be intelligent. In the 1960s, computers failed to pass the Turing test because of the low processing speed of the computers.

The last few decades have seen a new era of artificial intelligence (AI) focusing on the principles, theoretical aspects, and design methodology of algorithms gleaned from nature. Examples are artificial neural networks inspired by mammalian neural systems, evolutionary computation inspired by natural selection in biology, simulated annealing inspired by thermodynamics principles, and swarm intelligence inspired by the collective behavior of insects or microorganisms, and so on, interacting locally with their environment, therein causing coherent functional global patterns to emerge. These techniques have found their way into solving real-world problems in science, business, technology, commerce, and also to a great extent in measuring systems.

Computational intelligence is a well-established paradigm, where new theories with a sound biological understanding have been evolving. The current experimental systems have many of the characteristics of biological computers (brains, in other words) and are beginning to be built

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

to perform a variety of tasks that are difficult or impossible to do with conventional computers.

To name a few, we have microwave ovens, washing machines, and digital cameras that can figure out on their own what settings to use to perform their tasks optimally; they have a reasoning capability, make intelligent decisions, and learn from experience.

As usual, defining computational intelligence is not an easy task. In a nutshell, which becomes quite apparent in light of the current research pursuits, the area is heterogeneous with a combination of such technologies as neural networks, fuzzy systems, evolutionary computation, swarm intelligence, and probabilistic reasoning.

The recent trend is to integrate different components to take advantage of complementary features and to develop a synergistic system. Hybrid architectures like neuro-fuzzy systems, evolutionary-fuzzy systems, evolutionary-neural networks, evolutionary-neuro-fuzzy systems, and so on, are widely applied for real-world problem solving. In the following sections, the main functional components of computational intelligence are introduced along with their key advantages and application domains.

2 ARTIFICIAL NEURAL NETWORKS

Artificial neural networks (ANN) have been developed as generalizations of mathematical models of biological nervous systems.

In a simplified mathematical model of the neuron, the effects of the synapses are represented by connection weights that modulate the effect of the associated input signals, and the nonlinear characteristic exhibited by neurons is represented by a transfer function, which is usually the sigmoid, Gaussian, trigonometric function, and so on.

The neuron impulse is then computed as the weighted sum of the input signals, being transformed by the transfer function.

The learning capability of an artificial neuron is achieved by adjusting the weights in accordance to the chosen learning algorithm. Most applications of neural networks fall into the following categories:

- Prediction: Use input values to predict some output
- Classification: Use input values to determine the classification
- Data Association: Like classification, but it also recognizes data that contains errors
- Data conceptualization: Analyze the inputs so that grouping relationships can be inferred.

A typical multilayered neural network and an artificial neuron are illustrated in Figure 1. Each neuron is characterized by an activity level (representing the state of polarization of a neuron), an output value (representing the firing rate of the neuron), a set of input connections (representing synapses on the cell and its dendrite), a bias value (representing an internal resting level of the neuron), and a set of output connections (representing a neuron's axonal projections). Each of these aspects of the unit is represented mathematically by real numbers. Thus, each connection has an associated weight (synaptic strength), which determines the effect of the incoming input on the activation level of the unit. The weights may be positive or negative. Referring to Figure 1, the signal flow from inputs $x_1, \ldots, x_n$ is considered to be unidirectional, indicated by arrows, as is a neuron's output signal flow (O). The neuron output signal O is given by the following relationship:

$$O = f(net) = f\left(\sum_{j=1}^{n} w_j x_j\right) \tag{1}$$

where w_j is the weight vector and the function *f(net) i*s referred to as an *activation (transfer) function*. The variable net is defined as a scalar product of the weight and input vectors

$$net = w^{\mathrm{T}} x = w_1 x_1 + \cdots + w_n x_n \tag{2}$$

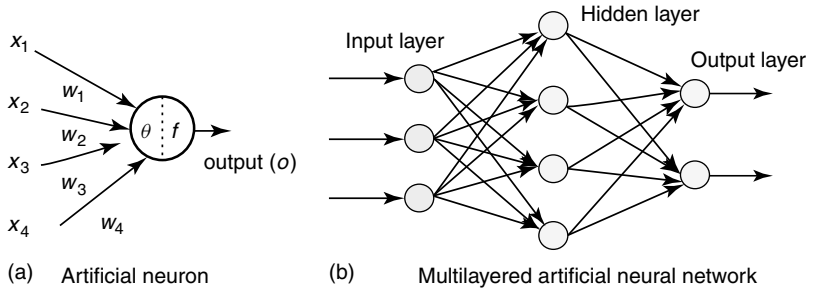


Figure 1. Architecture of an artificial neuron and a multilayered neural network.

where T is the transpose of a matrix and in the simplest case the output value O is computed as

$$O = f(net) = \begin{cases} 1 & if\ w^{\mathrm{T}}x \geq \theta \\ 0 & otherwise \end{cases} \quad (3)$$

where θ is called the *threshold level*, and this type of node is called a *linear threshold unit*.

3 NEURAL NETWORK ARCHITECTURES

The behavior of the neural network depends largely on the interaction between the different neurons. The basic architecture consists of three types of neuron layers:

1. Input
2. Hidden
3. Output.

In feed-forward networks, the signal flow is from input to output units, strictly in a feed-forward direction. The data processing can extend over multiple (layers of) units, but no feedback connections are present, that is, connections extending from outputs of units to inputs of units in the same layer or previous layers.

Recurrent networks contain feedback connections. Contrary to feed-forward networks, the dynamical properties of the network are important. In some cases, the activation values of the units undergo a relaxation process such that the network will evolve into a stable state in which these activations do not change anymore. In other applications, the changes of the activation values of the output neurons are significant, such that the dynamical behavior constitutes the output of the network.

There are several other neural network architectures (Elman network, adaptive resonance theory maps, competitive networks, etc.) depending on the properties and requirement of the application. The reader may refer to Bishop (1995) for an extensive overview of the different neural network architectures and learning algorithms.

A neural network has to be configured such that the application of a set of inputs produces the desired set of outputs. Various methods to set the strengths of the connections exist. One way is to set the weights explicitly, using a priori knowledge.

Another way is to train the neural network by feeding it, teaching patterns and letting it change its weights according to some learning rule.

The learning situations in neural networks may be classified into three distinct sorts of learning:

1. Supervised
2. Unsupervised
3. Reinforcement.

In supervised learning, an input vector is presented at the inputs together with a set of desired responses, one for each node, at the output layer. A forward pass is done and the errors or discrepancies, between the desired and actual response for each node in the output layer, are found. These are then used to determine weight changes in the net according to the prevailing learning rule.

The term 'supervised' originates from the fact that the desired signals on individual output nodes are provided by an external teacher. The best-known examples of this technique occur in the backpropagation algorithm, the delta rule, and perceptron rule.

In unsupervised learning (or self-organization), a (output) unit is trained to respond to clusters of pattern within the input. In this paradigm, the system is supposed to discover statistically salient features of the input population (Kohonen, 1988). Unlike the supervised learning paradigm, there is no a priori set of categories into which the patterns are to be classified; rather, the system must develop its own representation of the input stimuli.

Reinforcement learning is learning what to do – how to map situations to actions – so as to maximize a numerical reward signal. The learner is not told which actions to take, as in most forms of machine learning, but instead must discover which actions yield the most reward by trying them. In the most interesting and challenging cases, actions may affect not only the immediate reward but also the next situation and, through that, all subsequent rewards. These two characteristics, trial-and-error search and delayed reward, are the two most important distinguishing features of reinforcement learning.

4 FUZZY LOGIC

Zadeh (1965) introduced the concept of fuzzy logic to represent vagueness in linguistics and to further implement and express human knowledge and inference capability in a natural way. Fuzzy logic starts with the concept of a fuzzy set.

A fuzzy set is a set without a crisp, clearly defined boundary. It can contain elements with only a partial degree of membership.

A Membership Function (MF) is a curve that defines how each point in the input space is mapped to a membership value (or degree of membership) between 0 and 1. The input space is sometimes referred to as the *universe of discourse*.

Let X be the universe of discourse and x be a generic element of X. A classical set A is defined as a collection of elements or objects $x \in X$, such that each x can either belong to or not belong to the set A.

By defining a characteristic function (or membership function) on each element x in X, a classical set A can

be represented by a set of ordered pairs $(x, 0)$ or $(x, 1)$, where 1 indicates membership and 0, nonmembership.

Unlike the conventional set mentioned above, the fuzzy set expresses the degree to which an element belongs to a set. Hence, the characteristic function of a fuzzy set is allowed to have a value between 0 and 1, denoting the degree of membership of an element in a given set. If X is a collection of objects denoted generically by x, then a fuzzy set A in X is defined as a set of ordered pairs:

$$A = \{(x, \mu_A(x)) | x \in X\} \tag{4}$$

$\mu_A(x)$ is called the *membership function* of linguistic variable x in A, which maps X to the membership space M, $M = [0,1]$, where M contains only two points 0 and 1, A is crisp and μ_A is identical to the characteristic function of a crisp set.

Triangular and trapezoidal membership functions are the simplest membership functions, formed using straight lines. Some of the other shapes are Gaussian, generalized bell, sigmoidal, and polynomial-based curves. Figure 2 illustrates the shapes of two commonly used MFs. The most important thing to realize about fuzzy logical reasoning is the fact that it is a superset of standard Boolean logic.

4.1 Fuzzy logic operators

It is interesting to note the correspondence between two-valued and multivalued logic operations for the **AND, OR**, and **NOT** logical operators.

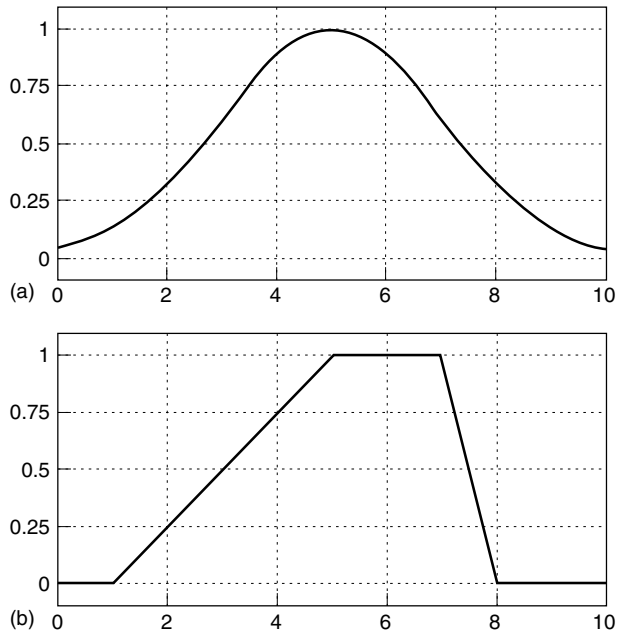


Figure 2. Membership functions; (a) Gaussian and (b) trapezoidal.

It is possible to resolve the statement A **AND** B, where A and B are limited to the range (0, 1) by using the operator minimum (A, B). Using the same reasoning, we can replace the **OR** operation with the maximum operator, so that A **OR** B becomes equivalent to maximum (A, B). Finally, the operation **NOT** A becomes equivalent to the operation 1 – A.

In fuzzy logic terms, these are popularly known as fuzzy intersection or conjunction (AND), fuzzy union or disjunction (OR), and fuzzy complement (NOT). The intersection of two fuzzy sets A and B is specified, in general, by a binary mapping T, which aggregates two membership functions as follows.

$$\mu_{A \cap B}(x) = T(\mu_A(x), \mu_B(x)) \tag{5}$$

The fuzzy intersection operator is usually referred to as *T-norm (Triangular norm) operator*. The fuzzy union operator is specified in general by a binary mapping S.

$$\mu_{A \cup B}(x) = S(\mu_A(x), \mu_B(x)) \tag{6}$$

This class of fuzzy union operators are often referred to as *T-conorm (S-norm) operators*.

5 IF-THEN RULES AND FUZZY INFERENCE SYSTEMS

The fuzzy rule base is characterized in the form of *if-then* rules in which preconditions and consequents involve linguistic variables. The collection of these fuzzy rules forms the rule base for the fuzzy logic system. Owing to their concise form, fuzzy *if-then* rules are often employed to capture the imprecise modes of reasoning that play an essential role in the human ability to make decisions in an environment of uncertainty and imprecision. A single fuzzy *if-then* rule assumes the form

$$if\ x \text{ is } A\ then\ y \text{ is } B$$

where A and B are linguistic values defined by fuzzy sets in the ranges (universes of discourse) X and Y, respectively.

The *if* part of the rule 'x is A' is called the *antecedent (precondition)* or premise, while the *then* part of the rule 'y is B' is called the *consequent* or conclusion. Interpreting an *if-then* rule involves evaluating the antecedent (*fuzzification of* the input and applying any necessary *fuzzy operators*) and then applying that result to the consequent (known as *implication*). For rules with multiple antecedents, all parts of the antecedent are calculated simultaneously and resolved to a single value using the logical operators. Similarly, all the consequents (rules with multiple consequents)

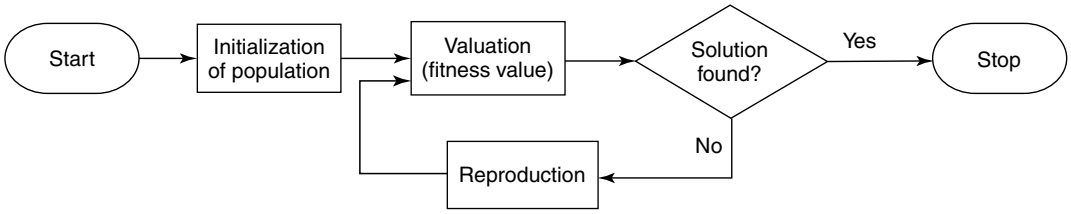


Figure 3. Flowchart of genetic algorithm iteration.

are affected equally by the result of the antecedent. The consequent specifies a fuzzy set be assigned to the output. The implication function then modifies that fuzzy set to the degree specified by the antecedent. For multiple rules, the output of each rule is a fuzzy set. The output fuzzy sets for each rule are then aggregated into a single output fuzzy set. Finally, the resulting set is defuzzified, or resolved, to a single number.

The defuzzification interface is a mapping from a space of fuzzy actions defined over an output universe of discourse into a space of nonfuzzy actions, because the output from the inference engine is usually a fuzzy set, while for most practical applications, crisp values are often required.

The three commonly applied defuzzification techniques are max-criterion, center-of-gravity, and the mean-of-maxima. The max-criterion is the simplest of these three to implement. It produces the point at which the possibility distribution of the action reaches a maximum value.

The reader can refer to Nguyen and Walker (1999) for more information related to fuzzy systems. It is typically advantageous if the fuzzy rule base is adaptive to a certain application. The fuzzy rule base is usually constructed manually or by automatic adaptation by some learning techniques using evolutionary algorithms and/or neural network learning methods (Abraham, 2001).

6 EVOLUTIONARY ALGORITHMS

Evolutionary algorithms (EA) are adaptive methods, which may be used to solve search and optimization problems, based on the genetic processes of biological organisms. Over many generations, natural populations evolve according to the principles of natural selection and 'survival of the fittest', first clearly stated by Charles Darwin in his work '*The Origin of Species*'. By mimicking this process, evolutionary algorithms are able to 'evolve' solutions to real-world problems, provided they have been suitably encoded (Fogel, 1999).

Usually grouped under the term evolutionary algorithms or evolutionary computation, we find the domains of genetic algorithms (GA) (Holland, 1975), (Goldberg, 1989); evolution strategies (Rechenberg, 1973), (Schwefel, 1977); evolutionary programming (Fogel, Owens and Walsh, 1967); genetic programming (Koza, 1992); and learning classifier systems. They all share a common conceptual base of simulating the evolution of *individual* structures via processes of *selection, mutation*, and *reproduction*. The processes depend on the perceived performance of the individual structures as defined by the environment (problem).

EA's deal with parameters of finite length, which are coded using a finite alphabet, rather than directly manipulating the parameters themselves. This means that the search is unconstrained neither by the continuity of the function under investigation nor by the existence of a derivative function. Figure 3 depicts the functional block diagram of a genetic algorithm, and the various aspects are discussed below. It is assumed that a potential solution to a problem may be represented as a set of parameters. These parameters (known as *genes*) are joined together to form a string of values (known as a *chromosome*). A gene (also referred to as a *feature, character*, or *detector*) refers to a specific attribute that is encoded in the chromosome. The particular values that the genes can take are called its *alleles*.

Encoding issues deal with representing a solution in a chromosome and, unfortunately, no one technique works best for all problems. A fitness function must be devised for each problem to be solved.

Given a particular chromosome, the fitness function returns a single numerical fitness or figure of merit, which will determine the ability of the individual that the chromosome represents.

Reproduction is the second critical attribute of GAs, where two individuals selected from the population are allowed to mate to produce an offspring, which will comprise the next generation. Having selected the parents, the offsprings are generated, typically using the mechanisms of crossover and mutation.

Selection is the survival of the fittest within GAs. It determines which individuals are to survive to the next generation. The selection phase consists of three parts.

The first part involves determination of the individual's fitness by the fitness function. A fitness function must be devised for each problem; given a particular chromosome, the fitness function returns a single numerical fitness value,

which is proportional to the ability, or utility, of the individual represented by that chromosome.

The second part involves converting the fitness function into an expected value, followed by the last part where the expected value is then converted to a discrete number of offspring.

Some of the commonly used selection techniques are roulette wheel and stochastic universal sampling. If the GA has been correctly implemented, the population will evolve over successive generations so that the fitness of the best and the average individual in each generation increases toward the global optimum.

7 INTELLIGENT PARADIGMS

7.1 Probabilistic computing

Probabilistic models are viewed as being similar to that of a game; actions are based on expected outcomes. The center of interest moves from the deterministic to probabilistic models using statistical estimations and predictions.

In the probabilistic modeling process, risk means uncertainty for which the probability distribution is known. Therefore, risk assessment means a study to determine the outcomes of decisions along with their probabilities. Decision makers often face a severe lack of definite information. Probability assessment quantifies the information gap between what is known and what needs to be known for an optimal decision. Probabilistic models are used for protection against adverse uncertainty and exploitation of propitious uncertainty (Pearl, 1997).

A good example is the probabilistic neural network (Bayesian learning) in which probability is used to represent uncertainty about the relationship being learned. Before we have seen any data, our *prior* opinions about what the true relationship might be can be expressed in a probability distribution over the network weights that define this relationship.

After we look at the data, our revised opinions are captured by a *posterior* distribution over network weights. Network weights that seemed plausible before, but which do not match the data very well, will now be seen as being much less likely, while the probability for values of the weights that do fit the data well will have increased. Typically, the purpose of training is to make predictions for future cases in which only the inputs to the network are known. The result of conventional network training is a single set of weights that can be used to make such predictions.

7.2 Swarm intelligence

Swarm intelligence is aimed at collective behavior of intelligent agents in decentralized systems. Although there is typically no centralized control dictating the behavior of the agents, local interactions among the agents often cause a global pattern to emerge (Eberhart, Shi and Kennedy, 2001).

Most of the basic ideas are derived from the real swarms in nature, which includes ant colonies, bird flocking, honeybees, bacteria and microorganisms, and so on. Ant Colony Optimization (ACO) has already been applied successfully to solve several engineering optimization problems.

Swarm models are population based and the population is initialized with a population of potential solutions. These individuals are then manipulated (optimized) over several iterations using several heuristics inspired from the social behavior of insects in an effort to find the optimal solution.

Ant colony algorithms are inspired by the behavior of natural ant colonies, in the sense that they solve their problems by multiagent cooperation using indirect communication through modifications in the environment. Ants release a certain amount of pheromone (hormone) while walking, and each ant prefers (probabilistically) to follow a direction that is rich in pheromone. This simple behavior explains why ants are able to adjust to changes in the environment, such as optimizing the shortest path to a food source or a nest. In ACO, ants use information collected during past simulations to direct their search and this information is available and modified through the environment. Recently, ACO algorithms have also been used for clustering data sets.

8 HYBRID INTELLIGENT SYSTEMS

Several adaptive hybrid intelligent systems have, in recent years, been developed for model expertise, image, and video segmentation techniques, process control, mechatronics, robotics, complicated automation tasks and so on.

Many of these approaches use the combination of different knowledge representation schemes, decision-making models, and learning strategies to solve a computational task. This integration aims at overcoming limitations of individual techniques through hybridization or fusion of various techniques.

These ideas have led to the emergence of several different kinds of intelligent system architectures. Most of the current Hybrid Intelligent Systems (HIS) consist of three essential paradigms: artificial neural networks, fuzzy inference systems, and global optimization algorithms (example, evolutionary algorithms). Nevertheless, HIS is an

Table 1. Hybrid intelligent system basic ingredients.

Methodology	Advantage
Artificial neural networks	Adaptation, learning, and approximation
Fuzzy logic	Approximate reasoning
Global optimization algorithms	Derivative free optimization

open, instead of conservative, concept, that is, it is evolving those relevant techniques together with the important advances in other new computing methods. Table 1 lists the three principal ingredients together with their advantages (Abraham, 2002).

Experience has shown that it is crucial for the design of HIS to primarily focus on the integration and interaction of different techniques rather than merge different methods to create ever-new techniques. Techniques already well understood should be applied to solve specific domain problems within the system. Their weakness must be addressed by combining them with complementary methods.

Neural networks offer a highly structured architecture with learning and generalization capabilities. The generalization ability for new inputs is then based on the inherent algebraic structure of the neural network. However, it is very hard to incorporate human a priori knowledge into a neural network. This is mainly due to the fact that the connectionist paradigm gains most of its strength from a distributed knowledge representation.

In contrast, fuzzy inference systems exhibit complementary characteristics, offering a very powerful framework for approximate reasoning as it attempts to model the human reasoning process at a cognitive level. Fuzzy systems acquire knowledge from domain experts and this is encoded within the algorithm in terms of the set of *if-then* rules. Fuzzy systems employ this rule-based approach and interpolative reasoning to respond to new inputs. The incorporation and interpretation of knowledge is straightforward, whereas learning and adaptation constitute major problems.

Global optimization is the task of finding the absolutely best set of parameters to optimize an objective function. In general, it may be possible to have solutions that are locally optimal but not globally optimal. Evolutionary computing (EC) works by simulating evolution on a computer. Such techniques could be easily used to optimize neural networks, fuzzy inference systems, and other problems.

Owing to the complementary features and strengths of different systems, the trend in the design of hybrid system is to merge different techniques into a more powerful integrated system to overcome their individual weaknesses.

9 MODELS OF HYBRID SYSTEMS

The various HIS architectures could be broadly classified into four different categories based on the systems overall architecture:

1. Stand alone
2. Transformational
3. Hierarchical hybrid
4. Integrated hybrid.

9.1 Stand-alone architecture

Stand-alone models of HIS applications consist of independent software components, which do not interact in any way. Developing stand-alone systems can have several purposes. First, they provide direct means of comparing the problem-solving capabilities of different techniques with reference to a certain application. Running different techniques in a parallel environment permits a loose approximation of integration. Stand-alone models are often used to develop a quick initial prototype while a more time-consuming application is developed. Some of the benefits are simplicity and ease of development using commercially available software packages.

9.2 Transformational hybrid architecture

In a transformational hybrid model, the system begins as one type and ends up as the other. Determining which technique is used for development and which is used for delivery is based on the desirable features that the technique offers. Expert systems and neural networks have proven to be useful transformational models. Variously, either the expert system is incapable of adequately solving the problem, or it requires the speed, adaptability, and robustness of neural network. Knowledge from the expert system is used to set the initial conditions and training set for neural network. Transformational hybrid models are often quick to develop and ultimately require maintenance on only one system. Most of the developed models are just application oriented.

9.3 Hierarchical hybrid architectures

The architecture is built in a hierarchical fashion, associating a different functionality with each layer. The overall functioning of the model will depend on the correct functioning of all the layers. A possible error in one of the layers will directly affect the desired output.

9.4 Integrated hybrid architectures

These models include systems, which combine different techniques into one single computational model. They share data structures and knowledge representations. Another approach is to put the various techniques on a side-by-side basis and focus on their interaction in the problem-solving task. This method might allow integrating alternative techniques and exploiting their mutuality. The benefits of fused architecture include robustness, improved performance, and increased problem-solving capabilities. Finally, fully integrated models can provide a full range of capabilities such as adaptation, generalization, noise tolerance, and justification. Fused systems have limitations caused by the increased complexity of the intermodule interactions, and specifying, designing, and building fully integrated models is complex.

10 SUMMARY

Artificial intelligence is the study of intelligent behavior. Its ultimate goal is a theory of intelligence that accounts for the behavior of naturally occurring intelligent entities, and this guides the creation of artificial entities capable of intelligent behavior. The stagnation of artificial intelligence during the 1970s and 1980s does not have much bearing on the likelihood of artificial intelligence to succeed in the future, since we know that the cause responsible for stagnation (mainly due to insufficient hardware resources) is no longer present. More detail of the various methods introduced here is found in **Article 129, Artificial Neural Networks, Volume 2**; **Article 130, Rule-based Expert Systems, Volume 2**; and **Article 131, Evolutionary Computation, Volume 2**.

REFERENCES

Abraham, A. (2001) in *Neuro-Fuzzy Systems: State-of-the-art Modeling Techniques, Connectionist Models of Neurons, Learning Processes, and Artificial Intelligence, Lecture Notes in Computer Science*, Vol. 2084, (eds J. Mira and A. Prieto), Springer Verlag, Germany, (pp. 269–276).

Abraham, A. (2002) Intelligent Systems: Architectures and Perspectives, Recent Advances in Intelligent Paradigms and Applications, in *Studies in Fuzziness and Soft Computing*, Chapter 1, (eds A. Abraham, L. Jain and J. Kacprzyk), Springer Verlag Germany, (pp. 1–35).

Bishop, C.M. (1995) *Neural Networks for Pattern Recognition*, Oxford University Press, UK.

Eberhart, R., Shi, Y. and Kennedy, J. (2001) *Swarm Intelligence*, Morgan Kaufmann, San Francisco, CA.

Fogel, D.B. (1999) *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*, 2nd edn, IEEE Press, Piscataway, NJ.

Fogel, L.J., Owens, A.J. and Walsh, M.J. (1967) *Artificial Intelligence Through Simulated Evolution*, John Wiley & Sons, New York.

Goldberg, D.E. (1989) *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley Publishing Corporation, Inc, Reading, MA.

Holland, J. (1975) *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Harbor, MI.

Kohonen, T. (1988) *Self-organization and Associative Memory*, Springer-Verlag, New York.

Koza, J.R. (1992) *Genetic Programming*, MIT Press, Cambridge, MA.

Nguyen, H.T. and Walker, E.A. (1999) *A First Course in Fuzzy Logic*, CRC Press, USA.

Pearl, J. (1997) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers, San Francisco, CA.

Rechenberg, I. (1973) *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*, Fromman-Holzboog, Stuttgart.

Schwefel, H.P. (1977) *Numerische Optimierung von Computermodellen Mittels der Evolutionsstrategie*, Birkhaeuser, Basel.

Turing, A.M. (1950) Computing Machinery and Intelligence http://abelard.org/turpap/turpap.htm.

Turing Machine. (2004) http://www.turing.org.uk/turing/.

Zadeh, L.A. (1965) Fuzzy Sets. *Journal of Information and Control*, **8**, 338–353.

129: Artificial Neural Networks

Ajith Abraham
Oklahoma State University, Stillwater, OK, USA

1 Introduction to Artificial Neural Networks 901
2 Neural Network Architectures 902
3 Neural Network Learning 903
4 Backpropagation Learning 903
5 Training and Testing Neural Networks 904
6 Higher Order Learning Algorithms 905
7 Designing Artificial Neural Networks 905
8 Self-organizing Feature Map and Radial Basis Function Network 906
9 Recurrent Neural Networks and Adaptive Resonance Theory 907
10 Summary 908
References 908

1 INTRODUCTION TO ARTIFICIAL NEURAL NETWORKS

A general introduction to artificial intelligence methods of measuring signal processing is given in **Article 128, Nature and Scope of AI Techniques, Volume 2**.

The human brain provides proof of the existence of massive neural networks that can succeed at those cognitive, perceptual, and control tasks in which humans are successful. The brain is capable of computationally demanding perceptual acts (e.g. recognition of faces, speech) and control activities (e.g. body movements and body functions). The advantage of the brain is its effective use of massive parallelism, the highly parallel computing structure, and the imprecise information-processing capability. The human brain is a collection of more than 10 billion interconnected neurons. Each neuron is a cell (Figure 1) that uses biochemical reactions to receive, process, and transmit information.

Treelike networks of nerve fibers called *dendrites* are connected to the cell body or soma, where the cell nucleus is located. Extending from the cell body is a single long fiber called the *axon*, which eventually branches into strands and substrands, and are connected to other neurons through synaptic terminals or synapses.

The transmission of signals from one neuron to another at synapses is a complex chemical process in which specific transmitter substances are released from the sending end of the junction. The effect is to raise or lower the electrical potential inside the body of the receiving cell. If the potential reaches a threshold, a pulse is sent down the axon and the cell is 'fired'.

Artificial neural networks (ANN) have been developed as generalizations of mathematical models of biological nervous systems. A first wave of interest in neural networks (also known as *connectionist models* or *parallel distributed processing*) emerged after the introduction of simplified neurons by McCulloch and Pitts (1943).

The basic processing elements of neural networks are called *artificial neurons*, or *simply neurons* or *nodes*. In a simplified mathematical model of the neuron, the effects of the synapses are represented by connection weights that modulate the effect of the associated input signals, and the nonlinear characteristic exhibited by neurons is represented by a transfer function. The neuron impulse is then computed as the weighted sum of the input signals, transformed by the transfer function. The learning capability of an artificial neuron is achieved by adjusting the weights in accordance to the chosen learning algorithm.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

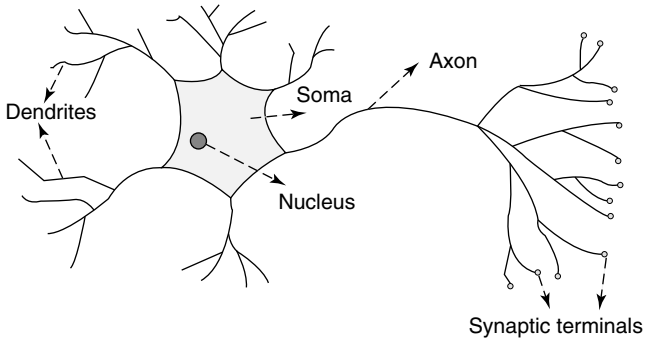


Figure 1. Mammalian neuron.

A typical artificial neuron and the modeling of a multilayered neural network are illustrated in Figure 2. Referring to Figure 2, the signal flow from inputs $x_1, \ldots, x_n$ is considered to be unidirectional, which are indicated by arrows, as is a neuron's output signal flow (O). The neuron output signal O is given by the following relationship:

$$O = f(net) = f\left(\sum_{j=1}^{n} w_j x_j\right) \tag{1}$$

where w_j is the weight vector, and the function *f(net)* is referred to as an activation (transfer) function. The variable *net* is defined as a scalar product of the weight and input vectors,

$$net = w^{\mathrm{T}} x = w_1 x_1 + \cdots + w_n x_n \tag{2}$$

where T is the transpose of a matrix, and, in the simplest case, the output value O is computed as

$$O = f(net) = \begin{cases} 1 & if\ w^{\mathrm{T}} x \geqslant \theta \\ 0 & otherwise \end{cases} \tag{3}$$

where θ is called the threshold level; and this type of node is called a *linear threshold unit*.

2 NEURAL NETWORK ARCHITECTURES

The basic architecture consists of three types of neuron layers: input, hidden, and output layers. In feed-forward networks, the signal flow is from input to output units, strictly in a feed-forward direction. The data processing can extend over multiple (layers of) units, but no feedback connections are present. Recurrent networks contain feedback connections. Contrary to feed-forward networks, the dynamical properties of the network are important. In some cases, the activation values of the units undergo a relaxation process such that the network will evolve to a stable state in which these activations do not change anymore. In other applications, the changes of the activation values of the output neurons are significant, such that the dynamical behavior constitutes the output of the network. There are several other neural network architectures (Elman network, adaptive resonance theory maps, competitive networks, etc.), depending on the properties and requirement of the application. The reader can refer to Bishop (1995) for an extensive overview of the different neural network architectures and learning algorithms.

A neural network has to be configured such that the application of a set of inputs produces the desired set of outputs. Various methods to set the strengths of the connections exist. One way is to set the weights explicitly, using a priori knowledge. Another way is to train the neural network by feeding it teaching patterns and letting it change its weights according to some learning rule. The learning situations in neural networks may be classified into three distinct sorts. These are supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, an input vector is presented at the inputs together with a set of desired responses, one for each node, at the output layer. A forward pass is done, and the errors or discrepancies between the desired and actual response for each node in the output layer are found. These are then used to determine weight changes in the net according to the prevailing learning rule. The term *supervised* originates from the fact that the desired signals on individual output nodes are provided by an external teacher.

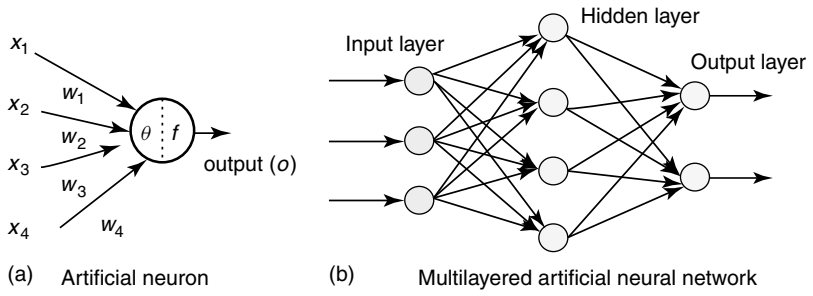


Figure 2. Architecture of an artificial neuron and a multilayered neural network.

The best-known examples of this technique occur in the backpropagation algorithm, the delta rule, and the perceptron rule. In unsupervised learning (or self-organization), a (output) unit is trained to respond to clusters of pattern within the input. In this paradigm, the system is supposed to discover statistically salient features of the input population. Unlike the supervised learning paradigm, there is no a priori set of categories into which the patterns are to be classified; rather, the system must develop its own representation of the input stimuli. Reinforcement learning is learning what to do – how to map situations to actions – so as to maximize a numerical reward signal. The learner is not told which actions to take, as in most forms of machine learning, but instead must discover which actions yield the most reward by trying them. In the most interesting and challenging cases, actions may affect not only the immediate reward, but also the next situation and, through that, all subsequent rewards. These two characteristics, trial-and-error search and delayed reward are the two most important distinguishing features of reinforcement learning.

3 NEURAL NETWORK LEARNING

3.1 Hebbian learning

The learning paradigms discussed above result in an adjustment of the weights of the connections between units, according to some modification rule. Perhaps the most influential work in connectionism's history is the contribution of Hebb (1949), where he presented a theory of behavior based, as much as possible, on the physiology of the nervous system.

The most important concept to emerge from Hebb's work was his formal statement (known as *Hebb's postulate*) of how learning could occur. Learning was based on the modification of synaptic connections between neurons. Specifically, when an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased. The principles underlying this statement have become known as *Hebbian Learning*. Virtually, most of the neural network learning techniques can be considered as a variant of the Hebbian learning rule. The basic idea is that if two neurons are active simultaneously, their interconnection must be strengthened. If we consider a single layer net, one of the interconnected neurons will be an input unit and one an output unit. If the data are represented in bipolar form, it is easy to express the desired weight update as

$$w_i(new) = w_i(old) + x_i o,$$

where o is the desired output for

$$i = 1 \text{ to } n(\text{inputs}).$$

Unfortunately, plain Hebbian learning continually strengthens its weights without bound (unless the input data is properly normalized).

3.2 Perceptron learning rule

The perceptron is a single layer neural network whose weights and biases could be trained to produce a correct target vector when presented with the corresponding input vector. The training technique used is called the *perceptron-learning rule*. Perceptrons are especially suited for simple problems in pattern classification.

Suppose we have a set of learning samples consisting of an input vector x and a desired output $d(k)$. For a classification task, the $d(k)$ is usually $+1$ or -1. The perceptron-learning rule is very simple and can be stated as follows:

1. Start with random weights for the connections.
2. Select an input vector x from the set of training samples.
3. If output $y_k \neq d(k)$ (the perceptron gives an incorrect response), modify all connections w_i according to: $\delta w_i = \eta(d_k - y_k)x_i$; ($\eta =$ learning rate).
4. Go back to step 2.

Note that the procedure is very similar to the Hebb rule; the only difference is that when the network responds correctly, no connection weights are modified.

4 BACKPROPAGATION LEARNING

The simple perceptron is just able to handle linearly separable or linearly independent problems. By taking the partial derivative of the error of the network with respect to each weight, we will learn a little about the direction the error of the network is moving.

In fact, if we take the negative of this derivative (i.e. the rate change of the error as the value of the weight increases) and then proceed to add it to the weight, the error will decrease until it reaches a local minima. This makes sense because if the derivative is positive, this tells us that the error is increasing when the weight is increasing. The obvious thing to do then is to add a negative value to the weight and vice versa if the derivative is negative. Because the taking of these partial derivatives and then applying them to each of the weights takes place, starting from the output layer to hidden layer weights, then the hidden layer to input layer weights (as it turns out, this is necessary since

changing these set of weights requires that we know the partial derivatives calculated in the layer downstream), this algorithm has been called the *backpropagation algorithm*.

A neural network can be trained in two different modes: online and batch modes. The number of weight updates of the two methods for the same number of data presentations is very different.

The online method weight updates are computed for each input data sample, and the weights are modified after each sample.

An alternative solution is to compute the weight update for each input sample, but store these values during one pass through the training set which is called an *epoch*.

At the end of the epoch, all the contributions are added, and only then the weights will be updated with the composite value. This method adapts the weights with a cumulative weight update, so it will follow the gradient more closely. It is called the *batch-training mode*.

Training basically involves feeding training samples as input vectors through a neural network, calculating the error of the output layer, and then adjusting the weights of the network to minimize the error.

The average of all the squared errors (E) for the outputs is computed to make the derivative easier. Once the error is computed, the weights can be updated one by one. In the batched mode variant, the descent is based on the gradient ∇E for the total training set

$$\Delta w_{ij}(n) = -\eta^* \frac{\delta E}{\delta w_{ij}} + \alpha^* \Delta w_{ij}(n-1) \qquad (4)$$

where η and α are the learning rate and momentum respectively.

The momentum term determines the effect of past weight changes on the current direction of movement in the weight space. A good choice of both η and α are required for the training success and the speed of the neural-network learning.

It has been proven that backpropagation learning with sufficient hidden layers can approximate any nonlinear function to arbitrary accuracy. This makes backpropagation learning neural network a good candidate for signal prediction and system modeling.

5 TRAINING AND TESTING NEURAL NETWORKS

The best training procedure is to compile a wide range of examples (for more complex problems, more examples are required), which exhibit all the different characteristics of the problem.

To create a robust and reliable network, in some cases, some noise or other randomness is added to the training data to get the network familiarized with noise and natural variability in real data.

Poor training data inevitably leads to an unreliable and unpredictable network. Usually, the network is trained for a prefixed number of epochs or when the output error decreases below a particular error threshold.

Special care is to be taken not to overtrain the network. By overtraining, the network may become too adapted in learning the samples from the training set, and thus may be unable to accurately classify samples outside of the training set.

Figure 3 illustrates the classification results of an overtrained network. The task is to correctly classify two patterns X and Y. Training patterns are shown by '●' and test patterns by '♣'. The test patterns were not shown during the training phase.

As shown in Figure 3 (left side), each class of test data has been classified correctly, even though they were not seen during training. The trained network is said to have good generalization performance. Figure 3 (right side) illustrates some misclassification of the test data. The network initially learns to detect the global features of the input and, as a consequence, generalizes very well. But after prolonged training, the network starts to recognize individual input/output pairs rather than settling for weights that generally describe the mapping for the whole training set (Fausett, 1994).

5.1 Choosing the number of neurons

The number of hidden neurons affects how well the network is able to separate the data. A large number of hidden neurons will ensure correct learning, and the network is able to correctly predict the data it has been trained on, but its performance on new data, its ability to generalize, is compromised. With too few hidden neurons, the network may be unable to learn the relationships amongst the data and the error will fail to fall below an acceptable level. Thus, selection of the number of hidden neurons is a crucial decision.

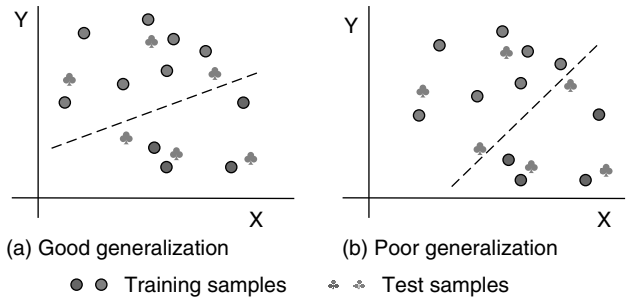


Figure 3. Illustration of generalization performance.

5.2 Choosing the initial weights

The learning algorithm uses a steepest descent technique, which rolls straight downhill in weight space until the first valley is reached. This makes the choice of initial starting point in the multidimensional weight space critical. However, there are no recommended rules for this selection except trying several different starting weight values to see if the network results are improved.

5.3 Choosing the learning rate

Learning rate effectively controls the size of the step that is taken in multidimensional weight space when each weight is modified. If the selected learning rate is too large, then the local minimum may be overstepped constantly, resulting in oscillations and slow convergence to the lower error state. If the learning rate is too low, the number of iterations required may be too large, resulting in slow performance.

6 HIGHER ORDER LEARNING ALGORITHMS

Backpropagation (BP) often gets stuck at a local minimum mainly because of the random initialization of weights. For some initial weight settings, BP may not be able to reach a global minimum of weight space, while for other initializations the same network is able to reach an optimal minimum.

A long recognized bane of analysis of the error surface and the performance of training algorithms is the presence of multiple stationary points, including multiple minima.

Empirical experience with training algorithms show that different initialization of weights yield different resulting networks. Hence, multiple minima not only exist, but there may be huge numbers of them.

In practice, there are four types of optimization algorithms that are used to optimize the weights. The first three methods, gradient descent, conjugate gradients, and quasi-Newton, are general optimization methods whose operation can be understood in the context of minimization of a quadratic error function.

Although the error surface is surely not quadratic, for differentiable node functions, it will be so in a sufficiently small neighborhood of a local minimum, and such an analysis provides information about the behavior of the training algorithm over the span of a few iterations and also as it approaches its goal.

The fourth method of Levenberg and Marquardt is specifically adapted to the minimization of an error function that arises from a squared error criterion of the form we are assuming. A common feature of these training algorithms is the requirement of repeated efficient calculation of gradients. The reader can refer to Bishop (1995) for an extensive coverage of higher-order learning algorithms.

Even though artificial neural networks are capable of performing a wide variety of tasks, in practice, sometimes, they deliver only marginal performance. Inappropriate topology selection and learning algorithm are frequently blamed. There is little reason to expect that one can find a uniformly best algorithm for selecting the weights in a feedforward artificial neural network. This is in accordance with the no free lunch theorem, which explains that for any algorithm, any elevated performance over one class of problems is exactly paid for in performance over another class (Macready and Wolpert, 1997).

The design of artificial neural networks using evolutionary algorithms has been widely explored. Evolutionary algorithms are used to adapt the connection weights, network architecture, and so on, according to the problem environment.

A distinct feature of evolutionary neural networks is their adaptability to a dynamic environment. In other words, such neural networks can adapt to an environment as well as changes in the environment. The two forms of adaptation, evolution and learning in evolutionary artificial neural networks, make their adaptation to a dynamic environment much more effective and efficient than the conventional learning approach. Refer to Abraham (2004) for more technical information related to evolutionary design of neural networks.

7 DESIGNING ARTIFICIAL NEURAL NETWORKS

To illustrate the design of artificial neural networks, the Mackey-Glass chaotic time series (Box and Jenkins, 1970) benchmark is used. The performance of the designed neural network is evaluated for different architectures and activation functions. The Mackey-Glass differential equation is a chaotic time series for some values of the parameters $x(0)$ and τ.

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = \frac{0.2x(t-\tau)}{1+x^{10}(t-\tau)} - 0.1\ x(t). \tag{5}$$

We used the value $x(t-18)$, $x(t-12)$, $x(t-6)$, $x(t)$ to predict $x(t+6)$. Fourth order Runge-Kutta method was used to generate 1000 data series. The time step used in the method is 0.1 and initial condition were $x(0) = 1.2$, $\tau =$

Table 1. Training and test performance for Mackey-Glass Series for different architectures.

Hidden neurons	Root mean-squared error	
	Training data	Test data
14	0.0890	0.0880
16	0.0824	0.0860
18	0.0764	0.0750
20	0.0452	0.0442
24	0.0439	0.0437

17, $x(t) = 0$ for $t < 0$. The first 500 data sets were used for training and remaining data for testing.

7.1 Network architecture

A feed-forward neural network with four input neurons, one hidden layer and one output neuron is used. Weights were randomly initialized and the learning rate and momentum are set at 0.05 and 0.1 respectively. The numbers of hidden neurons are varied (14, 16, 18, 20, 24) and the generalization performance is reported in Table 1. All networks were trained for an identical number of stochastic updates (2500 epochs).

7.2 Role of activation functions

The effect of two different node activation functions in the hidden layer, log-sigmoidal activation function LSAF and tanh-sigmoidal activation function TSAF), keeping 24 hidden neurons for the backpropagation learning algorithm, is illustrated in Figure 4. Table 2 summarizes the empirical results for training and generalization for the two node transfer functions. The generalization looks better with TSAF.

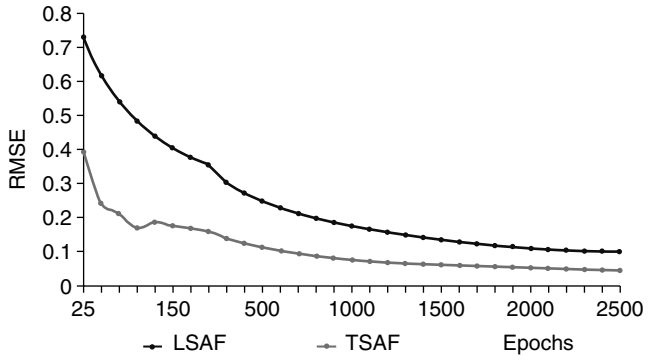


Figure 4. Convergence of training for different node transfer function.

Table 2. Mackey-Glass time series: training and generalization performance for different activation functions.

Activation function	Root mean-squared error	
	Training	Test
TSAF	0.0439	0.0437
LSAF	0.0970	0.0950

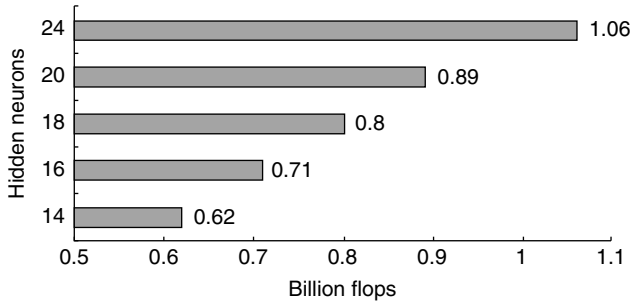


Figure 5. Computational complexity for different architectures.

Figure 5 illustrates the computational complexity in billion flops for different numbers of hidden neurons. At present, neural network design relies heavily on human experts who have sufficient knowledge about the different aspects of the network and the problem domain. As the complexity of the problem domain increases, manual design becomes more difficult.

8 SELF-ORGANIZING FEATURE MAP AND RADIAL BASIS FUNCTION NETWORK

8.1 Self-organizing feature map

Self-organizing Feature Maps SOFM is a data visualization technique proposed by Kohonen (1988), which reduces the dimensions of data through the use of self-organizing neural networks.

A SOFM learns the categorization, topology, and distribution of input vectors. SOFM allocate more neurons to recognize parts of the input space where many input vectors occur and allocate fewer neurons to parts of the input space where few input vectors occur. Neurons next to each other in the network learn to respond to similar vectors.

SOFM can learn to detect regularities and correlations in their input and adapt their future responses to that input accordingly. An important feature of the SOFM learning

algorithm is that it allows neurons that are neighbors to the winning neuron to be output values. Thus, the transition of output vectors is much smoother than that obtained with competitive layers, where only one neuron has an output at a time.

The problem that data visualization attempts to solve is that humans simply cannot visualize high-dimensional data. The way SOFM goes about reducing dimensions is by producing a map of usually 1 or 2 dimensions, which plot the similarities of the data by grouping similar data items together (data clustering). In this process, SOFM accomplish two things, they reduce dimensions and display similarities.

It is important to note that while a self-organizing map does not take long to organize itself so that neighboring neurons recognize similar inputs, it can take a long time for the map to finally arrange itself according to the distribution of input vectors.

8.2 Radial basis function network

The Radial Basis Function (RBF) network is a three-layer feed-forward network that uses a linear transfer function for the output units and a nonlinear transfer function (normally the Gaussian) for the hidden layer neurons (Chen, Cowan and Grant, 1991). Radial basis networks may require more neurons than standard feed-forward backpropagation networks, but often they can be designed with lesser time. They perform well when many training data are available.

Much of the inspiration for RBF networks has come from traditional statistical pattern classification techniques. The input layer is simply a fan-out layer and does no processing. The second or hidden layer performs a nonlinear mapping from the input space into a (usually) higher dimensional space whose activation function is selected from a class of functions called basis functions.

The final layer performs a simple weighted sum with a linear output. Contrary to BP networks, the weights of the hidden layer basis units (input to hidden layer) are set using some clustering techniques. The idea is that the patterns in the input space form clusters. If the centers of these clusters are known, then the Euclidean distance from the cluster center can be measured. As the input data moves away from the connection weights, the activation value reduces. This distance measure is made nonlinear in such a way that for input data close to a cluster center gets a value close to 1. Once the hidden layer weights are set, a second phase of training (usually backpropagation) is used to adjust the output weights.

9 RECURRENT NEURAL NETWORKS AND ADAPTIVE RESONANCE THEORY

9.1 Recurrent neural networks

Recurrent networks are the state of the art in nonlinear time series prediction, system identification, and temporal pattern classification. As the output of the network at time t is used along with a new input to compute the output of the network at time $t + 1$, the response of the network is dynamic (Mandic and Chambers, 2001).

Time Lag Recurrent Networks (TLRN) are multilayered perceptrons extended with short-term memory structures that have local recurrent connections. The recurrent neural network is a very appropriate model for processing temporal (time-varying) information.

Examples of temporal problems include time-series prediction, system identification, and temporal pattern recognition. A simple recurrent neural network could be constructed by a modification of the multilayered feed-forward network with the addition of a 'context layer'. The context layer is added to the structure, which retains information between observations. At each time step, new inputs are fed to the network. The previous contents of the hidden layer are passed into the context layer. These then feed back into the hidden layer in the next time step. Initially, the context layer contains nothing, so the output from the hidden layer after the first input to the network will be the same as if there is no context layer. Weights are calculated in the same way for the new connections from and to the context layer from the hidden layer.

The training algorithm used in TLRN (backpropagation through time) is more advanced than standard backpropagation algorithm. Very often, TLRN requires a smaller network to learn temporal problems when compared to MLP that use extra inputs to represent the past samples. TLRN is biologically more plausible and computationally more powerful than other adaptive models such as the hidden Markov model.

Some popular recurrent network architectures are the Elman recurrent network in which the hidden unit activation values are fed back to an extra set of input units and the Jordan recurrent network in which output values are fed back into hidden units.

9.2 Adaptive resonance theory

Adaptive Resonance Theory (ART) was initially introduced by Grossberg (1976) as a theory of human information processing. ART neural networks are extensively used for

supervised and unsupervised classification tasks and function approximation.

There exist many different variations of ART networks today (Carpenter and Grossberg, 1998). For example, ART1 performs unsupervised learning for binary input patterns, ART2 is modified to handle both analog and binary input patterns, and ART3 performs parallel searches of distributed recognition codes in a multilevel network hierarchy. Fuzzy ARTMAP represents a synthesis of elements from neural networks, expert systems, and fuzzy logic.

10 SUMMARY

This section presented the biological motivation and fundamental aspects of modeling artificial neural networks. Performance of feed-forward artificial neural networks for a function approximation problem is demonstrated. Advantages of some specific neural network architectures and learning algorithms are also discussed.

REFERENCES

Abraham, A. (2004) Meta-Learning Evolutionary Artificial Neural Networks, *Neurocomputing Journal*, Vol. 56c, Elsevier Science, Netherlands, (1–38).

Bishop, C.M. (1995) *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford, UK.

Box, G.E.P. and Jenkins, G.M. (1970) *Time Series Analysis, Forecasting and Control*, Holden Day, San Francisco, CA.

Carpenter, G. and Grossberg, S. (1998) in *Adaptive Resonance Theory (ART), The Handbook of Brain Theory and Neural Networks*, (ed. M.A. Arbib), MIT Press, Cambridge, MA, (pp. 79–82).

Chen, S., Cowan, C.F.N. and Grant, P.M. (1991) Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks. *IEEE Transactions on Neural Networks*, **2**(2), 302–309.

Fausett, L. (1994) *Fundamentals of Neural Networks*, Prentice Hall, USA.

Grossberg, S. (1976) Adaptive Pattern Classification and Universal Recoding: Parallel Development and Coding of Neural Feature Detectors. *Biological Cybernetics*, **23**, 121–134.

Hebb, D.O. (1949) *The Organization of Behavior*, John Wiley, New York.

Kohonen, T. (1988) *Self-Organization and Associative Memory*, Springer-Verlag, New York.

Macready, W.G. and Wolpert, D.H. (1997) The No Free Lunch Theorems. *IEEE Transactions on Evolutionary Computing*, **1**(1), 67–82.

Mandic, D. and Chambers, J. (2001) *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability*, John Wiley & Sons, New York.

McCulloch, W.S. and Pitts, W.H. (1943) A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*, **5**, 115–133.

130: Rule-based Expert Systems

Ajith Abraham

Oklahoma State University, Stillwater, OK, USA

1 Problem Solving Using Heuristics 909
2 What are Rule-based Systems? 910
3 Inference Engine in Rule-based Systems 911
4 Expert System Development 911
5 Fuzzy Expert Systems 912
6 Modeling Fuzzy Expert Systems 914
7 Illustration of Fuzzy Expert System Design 914
8 Adaptation of Fuzzy Inference Systems 918
9 Summary 918
References 919

1 PROBLEM SOLVING USING HEURISTICS

A general introduction to artificial intelligence methods of measurement signal processing is given in **Article 128, Nature and Scope of AI Techniques, Volume 2**.

Problem solving is the process of finding a solution when the path leading to that solution is uncertain. Even though we are familiar with several problem-solving techniques, in the real world, sometimes many problems cannot be solved by a technique we are familiar with.

Surprisingly, for some complicated problems, no straightforward solution technique is known at all. For these problems, heuristic solution techniques may be the only alternative. A heuristic can be simplified as a strategy that is powerful and general, but not absolutely guaranteed to provide best solutions.

Heuristic methods are very problem specific. Previous experience and some general rules – often called *rules of thumb* – could help find good heuristics easier. Humans use heuristics a great deal in their problem solving. Of course, if the heuristic does fail, it is necessary for the problem solver to either pick another heuristic, or know that it is appropriate to give up.

Choosing random solutions, adopting greedy approaches, evolving the basic heuristics for finding better heuristics are just some of the popular approaches used in heuristic problem solving (Michalewicz and Fogel, 1999).

Heuristic problem solving involves finding a set of rules, or a procedure, that finds satisfactory solutions to a specific problem. A good example is finding one's way through a maze. To make the way toward the final goal, a step-by-step movement is necessary. Very often false moves are made but in most cases we solve the problem without much difficulty. For the maze problem, a simple heuristic rule could be '*choose the direction that seems to make progress*'. Another good example is the job shop scheduling problem wherein the task is to schedule J_n independent jobs, where $n = \{1, 2, \ldots .N\}$ on R_m heterogeneous resources and $m = \{1, 2, \ldots ., M\}$, with an objective of minimizing the completion time of all the jobs and utilizing all the resources effectively.

Each job J_n has processing requirement P_j cycles and resource R_m has speed of S_i cycles/unit time. Any job J_n has to be processed in resource R_m, until completion. If C_j is the completion time and the last job j finishes processing, then we define $C_{\max} = \max\{C_j, j = 1, \ldots, N\}$, the make-span and ΣC_j, as the flow-time.

The task is to find an optimal schedule that optimizes the flow-time and make-span. Some simple heuristic rules to achieve this are by scheduling the Shortest Job on the Fastest Resource (SJFR), which would minimize ΣC_j or by

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.

© 2005 John Wiley & Sons, Ltd.

ISBN: 0-470-02143-8.

scheduling the Longest Job on the Fastest Resource (LJFR), which would minimize $C_{\max}$.

Minimizing ΣC_j asks that the average job finishes quickly, at the expense of the largest job taking a long time, whereas minimizing $C_{\max}$, asks that no job takes too long, at the expense of most jobs taking a long time.

In summary, minimization of $C_{\max}$ will result in maximization of ΣC_j, which makes the problem more interesting.

By contrast, algorithms are straightforward procedures that are guaranteed to work every time for they are fully determinate and time invariant. For example, certain daily routine tasks could be formulated in a strict algorithm format (example, starting up an automobile). However, for a 'problem solver' to be more adaptive, novel elements or new circumstances must be introduced. Many real-world problems cannot be reduced to algorithms, which leads us to the quest to find more powerful techniques.

2 WHAT ARE RULE-BASED SYSTEMS?

Conventional problem-solving computer programs make use of well-structured algorithms, data structures, and crisp reasoning strategies to find solutions. For the difficult problems with which expert systems are concerned, it may be more useful to employ heuristics: strategies that often lead to the correct solution, but that also sometimes fail.

Conventional rule-based expert systems, use human expert knowledge to solve real-world problems that normally would require human intelligence. Expert knowledge is often represented in the form of *rules* or as *data* within the computer.

Depending upon the problem requirement, these rules and data can be recalled to solve problems. Rule-based expert systems have played an important role in modern intelligent systems and their applications in strategic goal setting, planning, design, scheduling, fault monitoring, diagnosis and so on.

With the technological advances made in the last decade, today's users can choose from dozens of commercial software packages having friendly graphic user interfaces (Ignizio, 1991). Conventional computer programs perform tasks using a decision-making logic containing very little knowledge other than the basic algorithm for solving that specific problem. The basic knowledge is often embedded as part of the programming code, so that as the knowledge changes, the program has to be rebuilt. Knowledge-based expert systems collect the small fragments of human know-how into a knowledge base, which is used to reason through a problem, using the knowledge that is appropriate. An important advantage here is that within the domain of the knowledge base, a different problem can be solved using the same program without reprogramming efforts. Moreover, expert systems could explain the reasoning process and handle levels of confidence and uncertainty, which conventional algorithms do not handle (Giarratano and Riley, 1989). Some of the important advantages of expert systems are as follows:

- ability to capture and preserve irreplaceable human experience;
- ability to develop a system more consistent than human experts;
- minimize human expertise needed at a number of locations at the same time (especially in a hostile environment that is dangerous to human health);
- solutions can be developed faster than human experts.

The basic components of an expert system are illustrated in Figure 1. The knowledge base stores all relevant information, data, rules, cases, and relationships used by the expert system. A knowledge base can combine the knowledge of multiple human experts. A rule is a conditional statement that links given conditions to actions or outcomes. A frame is another approach used to capture and store knowledge in a knowledge base. It relates an object or item to various facts or values. A frame-based representation is ideally suited for object-oriented programming techniques. Expert systems making use of frames to store knowledge are also called *frame-based expert systems*.

The purpose of the inference engine is to seek information and relationships from the knowledge base and to provide answers, predictions, and suggestions in the way a human expert would. The inference engine must find the right facts, interpretations, and rules and assemble them correctly. Two types of inference methods are commonly used – Backward chaining is the process of starting with conclusions and working backward to the supporting facts. Forward chaining starts with the facts and works forward to the conclusions.

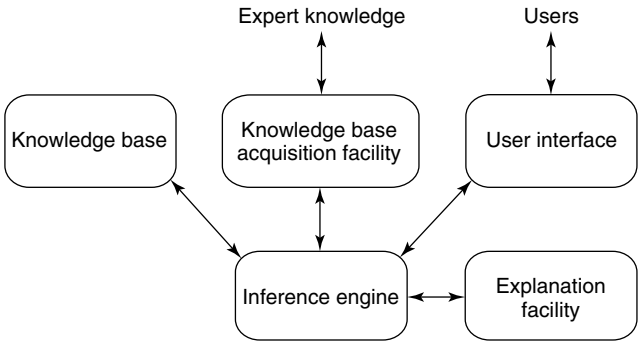


Figure 1. Architecture of a simple expert system.

The explanation facility allows a user to understand how the expert system arrived at certain results. The overall purpose of the knowledge acquisition facility is to provide a convenient and efficient means for capturing and storing all components of the knowledge base.

Very often specialized user interface software is used for designing, updating, and using expert systems. The purpose of the user interface is to ease use of the expert system for developers, users, and administrators.

3 INFERENCE ENGINE IN RULE-BASED SYSTEMS

A rule-based system consists of *if-then rules*, a bunch of *facts*, and an *interpreter* controlling the application of the rules, given the facts.

These *if-then* rule statements are used to formulate the conditional statements that comprise the complete knowledge base. A single *if-then* rule assumes the form 'if x is A then y is B' and the if-part of the rule 'x is A' is called the *antecedent* or *premise*, while the then-part of the rule 'y is B' is called the *consequent* or *conclusion*. There are two broad kinds of inference engines used in rule-based systems: *forward chaining* and *backward chaining* systems.

In a *forward chaining* system, the initial facts are processed first, and keep using the rules to draw new conclusions given those facts. In a *backward chaining* system, the hypothesis (or solution/goal) we are trying to reach is processed first, and keep looking for rules that would allow to conclude that hypothesis. As the processing progresses, new subgoals are also set for validation. Forward chaining systems are primarily data-driven, while backward chaining systems are goal-driven. Consider an example with the following set of *if-then* rules

Rule 1: *If A and C then Y*
Rule 2: *If A and X then Z*
Rule 3: *If B then X*
Rule 4: *If Z then D*

If the task is to prove that D is true, given A and B are true. According to *forward chaining*, start with Rule 1 and go on downward till a rule that fires is found. Rule 3 is the only one that fires in the first iteration. After the first iteration, it can be concluded that A, B, and X are true. The second iteration uses this valuable information. After the second iteration, Rule 2 fires adding Z is true, which in turn helps Rule 4 to fire, proving that D is true. Forward chaining strategy is especially appropriate in situations where data are expensive to collect, but few in quantity. However, special care is to be taken when these rules are constructed, with the preconditions specifying as precisely as possible when different rules should fire.

In the backward chaining method, processing starts with the desired goal, and then attempts to find evidence for proving the goal. Returning to the same example, the task to prove that D is true would be initiated by first finding a rule that proves D. Rule 4 does so, which also provides a subgoal to prove that Z is true. Now Rule 2 comes into play, and as it is already known that A is true, the new subgoal is to show that X is true. Rule 3 provides the next subgoal of proving that B is true. But that B is true is one of the given assertions. Therefore, it could be concluded that X is true, which implies that Z is true, which in turn also implies that D is true. Backward chaining is useful in situations where the quantity of data is potentially very large and where some specific characteristic of the system under consideration is of interest. If there is not much knowledge what the conclusion might be, or there is some specific hypothesis to test, forward chaining systems may be inefficient. In principle, we can use the same set of rules for both forward and backward chaining. In the case of backward chaining, since the main concern is with matching the conclusion of a rule against some goal that is to be proved, the 'then' (consequent) part of the rule is usually not expressed as an action to take but merely as a state, which will be true if the antecedent part(s) are true (Donald, 1986).

4 EXPERT SYSTEM DEVELOPMENT

Steps in the expert systems development process include determining the actual requirements, knowledge acquisition, constructing expert system components, implementing results, and formulating a procedure for maintenance and review.

Knowledge acquisition is the most important element in the development of expert system (Niwa, Sasaki and Ihara, 1988). Knowledge could be obtained by interviewing domain experts and/or learning by experience.

Very often people express knowledge as natural language (spoken language), or using letters or symbolic terms. There exist several methods to extract human knowledge. Cognitive Work Analysis (CWA) and the Cognitive Task Analysis (CTA) provide frameworks to extract knowledge.

The CWA is a technique to analyze, design, and evaluate the human computer interactive systems (Vicente, 1999).

The CTA is a method to identify cognitive skill, mental demands, and needs to perform task proficiency (Militallo and Hutton, 1998). This focuses on describing the representation of the cognitive elements that defines goal generation and decision-making. It is a reliable method for extracting

human knowledge because it is based on the observations or an interview.

Most expert systems are developed using specialized software tools called *shells*. These shells come equipped with an inference mechanism (backward chaining, forward chaining, or both), and require knowledge to be entered according to a specified format.

One of the most popular shells widely used throughout the government, industry, and academia is the CLIPS (CLIPS, 2004). CLIPS is an expert system tool that provides a complete environment for the construction of rule- and/or object-based expert systems. CLIPS provides a cohesive tool for handling a wide variety of knowledge with support for three different programming paradigms: rule-based, object-oriented, and procedural. CLIPS is written in C for portability and speed and has been installed on many different operating systems without code changes.

5 FUZZY EXPERT SYSTEMS

The world of information is surrounded by uncertainty and imprecision. The human reasoning process can handle inexact, uncertain, and vague concepts in an appropriate manner. Usually, the human thinking, reasoning, and perception process cannot be expressed precisely. These types of experiences can rarely be expressed or measured using statistical or probability theory. Fuzzy logic provides a framework to model uncertainty, the human way of thinking, reasoning, and the perception process. Fuzzy systems were first introduced by Zadeh (1965).

A fuzzy expert system is simply an expert system that uses a collection of fuzzy membership functions and rules, instead of Boolean logic, to reason about data (Schneider *et al.,* 1996). The rules in a fuzzy expert system are usually of a form similar to the following:

If A is *low* and B is *high* then $X = medium$

where A and B are input variables, X is an output variable. Here low, high, and medium are fuzzy sets defined on A, B, and X respectively. The antecedent (the rule's premise) describes to what degree the rule applies, while the rule's consequent assigns a membership function to each of one or more output variables.

Let X be a space of objects and x be a generic element of X. A classical set A, $A \subseteq X$, is defined as a collection of elements or objects $x \in X$, such that x can either belong or not belong to the set A. A fuzzy set A in X is defined as a set of ordered pairs

$$A = \{(x, \mu_A(x)) | x \in X\} \tag{1}$$

where $\mu_A(x)$ is called the *membership function* (MF) for the fuzzy set A. The MF maps each element of X to a membership grade (or membership value) between zero and one. Obviously (1) is a simple extension of the definition of a classical set in which the characteristic function is permitted to have any values between zero and one.

The intersection of two fuzzy sets A and B is specified in general by a function T: [0,1] × [0,1] → [0,1], which aggregates two membership grades as follows:

$$\mu_{A \cap B}(x) = T(\mu_A(x), \mu_B(x)) = \mu_A(x) \bar{*} \mu_B(x) \tag{2}$$

where $\bar{*}$ is a binary operator for the function T. This class of fuzzy intersection operators are usually referred to as *T-norm operators* (Jang, Sun and Mizutani, 1997). Four of the most frequently used T-norm operators are

$$\text{Minimum: } T_{\min}(a, b) = \min(a, b) = a \wedge b \tag{3}$$

$$\text{Algebraic product: } T_{\text{ap}}(a, b) = ab \tag{4}$$

$$\text{Bounded product: } T_{\text{bp}}(a, b) = 0 \vee (a + b - 1) \tag{5}$$

$$\text{Drastic product: } T_{\text{dp}}(a, b) = \begin{cases} a, & if\ b = 1 \\ b, & if\ a = 1 \\ 0, & if\ a, b < 1 \end{cases} \tag{6}$$

Like intersection, the fuzzy union operator is specified in general by a function S: [0,1] × [0,1] → [0,1], which aggregates two membership grades as follows:

$$\mu_{A \cup B}(x) = S(\mu_A(x), \mu_B(x)) = \mu_A(x) \mp \mu_B(x) \tag{7}$$

where $\mp$ is the binary operator for the function S. This class of fuzzy union operators are often referred to as *T-conorm* (or *S-norm*) operators (Jang, Sun and Mizutani, 1997). Four of the most frequently used T-conorm operators are

$$\text{Maximum: } S_{\max}(a, b) = \max(a, b) = a \vee b \tag{8}$$

$$\text{Algebraic sum: } S_{\text{as}}(a, b) = a + b - ab \tag{9}$$

$$\text{Bounded sum: } S_{\text{bs}}(a, b) = 1 \wedge (a + b) \tag{10}$$

$$\text{Drastic sum: } S_{\text{ds}}(a, b) = \begin{cases} a, & if\ b = 0 \\ b, & if\ a = 0 \\ 1, & if\ a, b > 0 \end{cases} \tag{11}$$

Both the intersection and union operators retain some properties of the classical set operation. In particular, they are associative and commutative.

Figure 2 illustrates the basic architecture of a fuzzy expert system. The main components are a fuzzification interface, a fuzzy rule base (knowledge base), an inference engine (decision-making logic), and a defuzzification interface. The input variables are fuzzified whereby the membership functions defined on the input variables are applied to their actual values, to determine the degree of truth for each rule antecedent. Fuzzy *if-then* rules and fuzzy reasoning are the backbone of fuzzy expert systems, which are the most

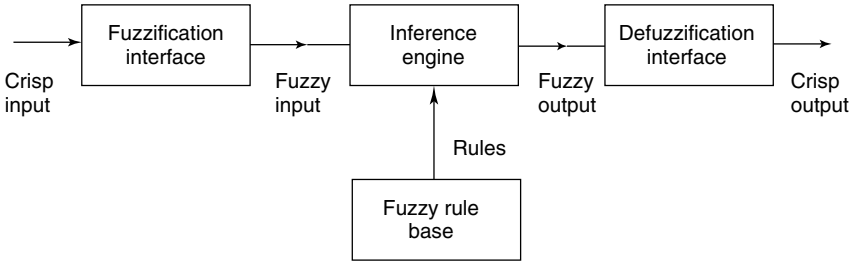


Figure 2. Basic architecture of a fuzzy expert system.

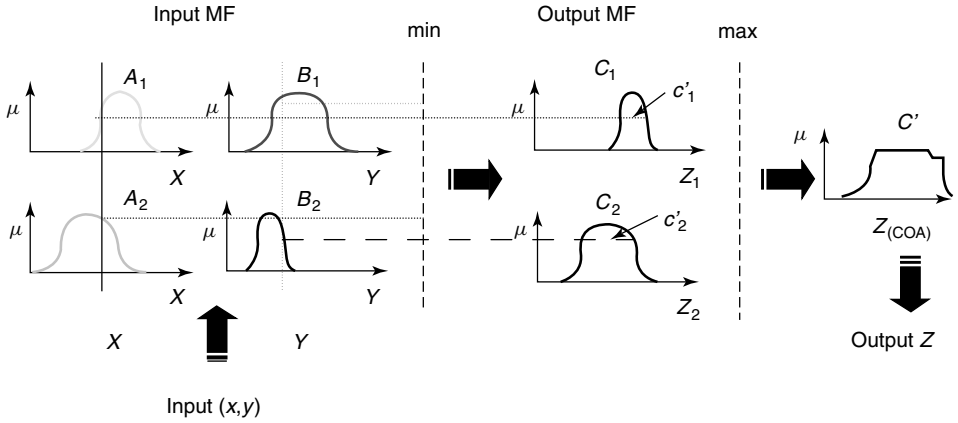


Figure 3. Mamdani fuzzy inference system using min and max for T-norm and T-conorm operators.

important modeling tools based on fuzzy set theory. The fuzzy rule base is characterized in the form of *if-then* rules in which the antecedents and consequents involve linguistic variables. The collection of these fuzzy rules forms the rule base for the fuzzy logic system. Using suitable inference procedure, the truth value for the antecedent of each rule is computed, and applied to the consequent part of each rule. This results in one fuzzy subset to be assigned to each output variable for each rule. Again, by using suitable composition procedure, all the fuzzy subsets assigned to each output variable are combined together to form a single fuzzy subset for each output variable. Finally, defuzzification is applied to convert the fuzzy output set to a crisp output.

The basic fuzzy inference system can take either fuzzy inputs or crisp inputs, but the outputs it produces are always fuzzy sets. The defuzzification task extracts the crisp output that best represents the fuzzy set. With crisp inputs and outputs, a fuzzy inference system implements a nonlinear mapping from its input space to output space through a number of fuzzy *if-then* rules.

In what follows, the two most popular fuzzy inference systems are introduced that have been widely deployed in various applications. The differences between these two fuzzy inference systems lie in the consequents of their fuzzy rules, and thus their aggregation and defuzzification procedures differ accordingly.

According to Mamdani, fuzzy inference system (Mamdani and Assilian, 1975) – see Figure 3 – the rule antecedents and consequents are defined by fuzzy sets and has the following structure:

$$if\ x\ is\ A_1\ and\ y\ is\ B_1\ then\ z_1 = C_1^{'} \tag{12}$$

There are several defuzzification techniques. The most widely used defuzzification technique uses the centroid of area method as follows

$$\text{Centroid of area } Z_{\text{COA}} = \frac{\int_Z \mu_A(z)\, z\, \mathrm{d}z}{\int_Z \mu_A(z)\, \mathrm{d}z} \tag{13}$$

where $\mu_A(z)$ is the aggregated output MF.

Takagi and Sugeno (1985) proposed an inference scheme in which the conclusion of a fuzzy rule is constituted by a weighted linear combination of the crisp inputs rather than a fuzzy set. A basic Takagi–Sugeno fuzzy inference system is illustrated in Figure 4 and the rule has the following structure

$$if\ x\ is\ A_1\ and\ y\ is\ B_1,\ then\ z_1 = p_1 x + q_1 y + r_1 \tag{14}$$

where p_1, q_1, and r_1 are linear parameters. TSK Takagi–Sugeno Kang fuzzy controller usually needs a smaller

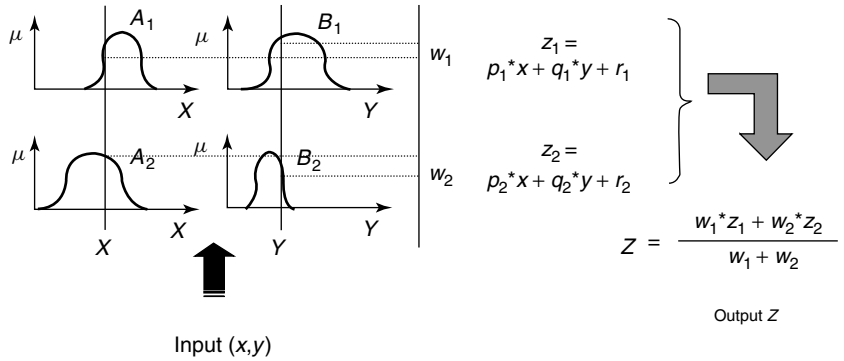


Figure 4. Takagi–Sugeno fuzzy inference system using a min or product as T-norm operator.

number of rules, because their output is already a linear function of the inputs rather than a constant fuzzy set.

6 MODELING FUZZY EXPERT SYSTEMS

Fuzzy expert system modeling can be pursued using the following steps.

- Select relevant input and output variables. Determine the number of linguistic terms associated with each input/output variable. Also, choose the appropriate family of membership functions, fuzzy operators, reasoning mechanism, and so on.
- Choose a specific type of fuzzy inference system (for example, Mamdani, Takagi–Sugeno etc.). In most cases, the inference of the fuzzy rules is carried out using the 'min' and 'max' operators for fuzzy intersection and union.
- Design a collection of fuzzy *if-then* rules (knowledge base). To formulate the initial rule base, the input space is divided into multidimensional partitions and then actions are assigned to each of the partitions.

In most applications, the partitioning is achieved using one-dimensional membership functions using fuzzy *if-then* rules as illustrated in Figure 5. The consequent parts of the rule represent the actions associated with each partition. It is evident that the MFs and the number of rules are tightly related to the partitioning.

7 ILLUSTRATION OF FUZZY EXPERT SYSTEM DESIGN

This section illustrates the development of a reactive power prediction model using Mamdani and Takagi–Sugeno fuzzy inference expert systems. The MatLab™ fuzzy logic toolbox was used to simulate the various experiments (Fuzzy Logic Tool Box, 2004).

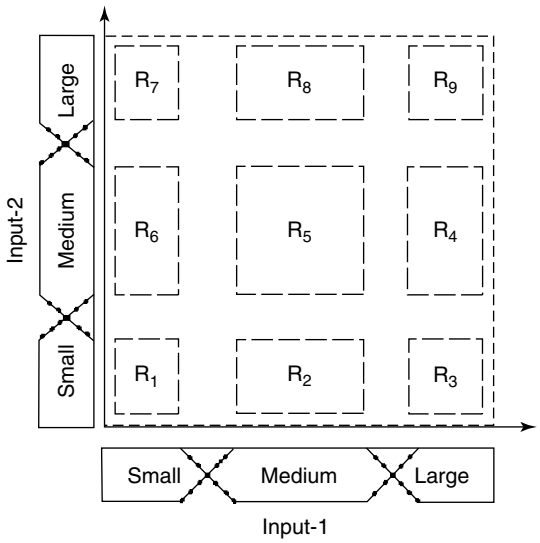


Figure 5. Example showing how the two-dimensional spaces are partitioned using three trapezoidal membership functions per input dimension. A simple if-then rule will appear as *If input-1 is medium and input 2 is large, then rule* R_8 *is fired*.

The task is to develop a fuzzy expert system to forecast the reactive power (P) at time $t + 1$ by knowing the load current (I) and voltage (V) at time t. The experiment system consists of two stages: developing the fuzzy expert system, and performance evaluation using the test data. The model has two in–out variables (V and I) and one output variable (P). Training and testing data sets were extracted randomly from the master dataset. Sixty percent of data was used for training and the remaining 40% for testing (Abraham and Khan, 2003).

7.1 Design and experiments: fuzzy expert systems

First, the effects of (a) shape and quantity of membership functions (b) T-norm and T-conorm operators (c) defuzzification methods and (d) inference method for

designing the fuzzy expert system is analyzed. Experiments were carried out using four different settings using the same rule base.

Experiment 1 (To evaluate the effect on the number of membership functions) The following settings were used for designing the expert system

1. Two triangular membership functions (MFs) for each input variable and four triangular MFs for the output variable (power). Using the grid partitioning method (Figure 5), four *if-then* rules were developed.
2. Three triangular MFs for each input variable and nine triangular MFs for the output variable (power). The rule base consisted of nine *if-then* rules.

'min' and 'max' were used as T-norm and T-conorm operators and the centroid method of defuzzification for Mamdani inference method and weighted average defuzzification method for Takagi–Sugeno Fuzzy Inference System (FIS). The developed fuzzy inference systems using Mamdani and Takagi–Sugeno models are depicted in Figures 6 to 9. Table 1 summarizes the training and testing Root Mean Squared Error (RMSE) values.

Experiment 2 (To evaluate the effect of shape of membership functions) For the Mamdani FIS, three Gaussian MFs for each input variable and nine Gaussian MFs for the output variable were used. The rule base consisted of nine *if-then* rules. 'min' and 'max' as T-norm and T-conorm operators, and the centroid method of defuzzification for Mamdani FIS and the weighted average defuzzification method for Takagi–Sugeno FIS were also used. The developed fuzzy inference systems using Mamdani and Takagi–Sugeno models are depicted in Figures 10 and 11. Table 2 summarizes the training and testing RMSE values.

Experiment 3 (To evaluate the effect of fuzzy operators) For Mamdani FIS, three Gaussian MFs for each input variable and nine Gaussian MFs for the output variable were used. The rule base consisted of nine *if-then* rules. T-norm and T-conorm operators were 'product' and 'sum' and the centroid method of defuzzification for Mamdani FIS, and weighted average defuzzification method for Takagi–Sugeno FIS were used. Table 3 summarizes the training and testing RMSE values.

Experiment 4 (To evaluate the effect of defuzzification operators) For the Mamdani FIS, three Gaussian MFs for each input variable and nine Gaussian MFs for the output variable were used. The rule base consisted of nine *if-then* rules. T-norm and T-conorm operators were 'product' and 'sum' and the following defuzzification operators were tested for Mamdani FIS.

Table 1. Empirical comparison of fuzzy inference systems and quantity of Membership Functions (MFs).

No. of MFs	Mamdani FIS		Takagi – Sugeno FIS	
	Root mean squared error			
	Training	Test	Training	Test
2	0.401	0.397	0.024	0.023
3	0.348	0.334	0.017	0.016

Table 2. Empirical comparison of fuzzy inference systems using Gaussian MFs.

Mamdani FIS		Takagi – Sugeno FIS	
Root mean squared error			
Training	Test	Training	Test
0.243	0.240	0.021	0.019

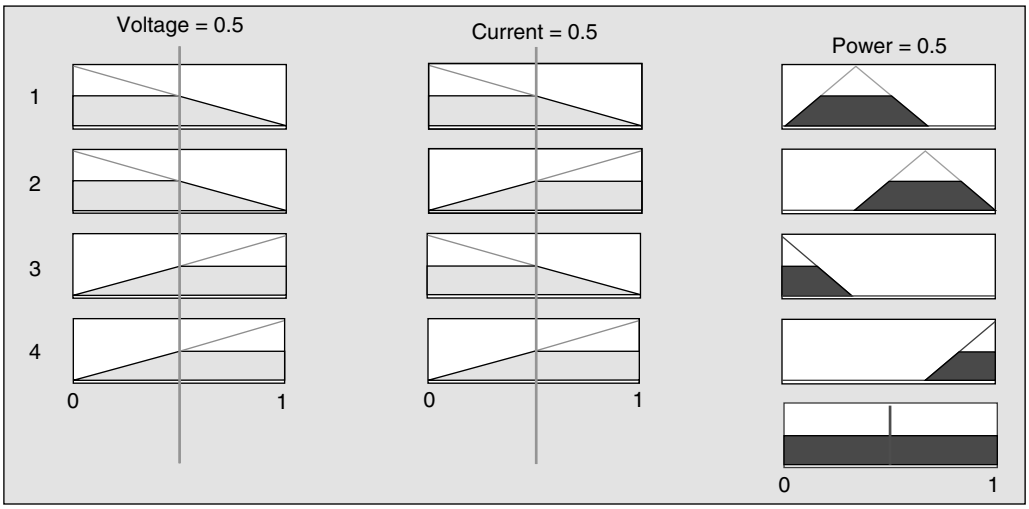


Figure 6. Mamdani fuzzy inference system using two triangular MFs for input variables.

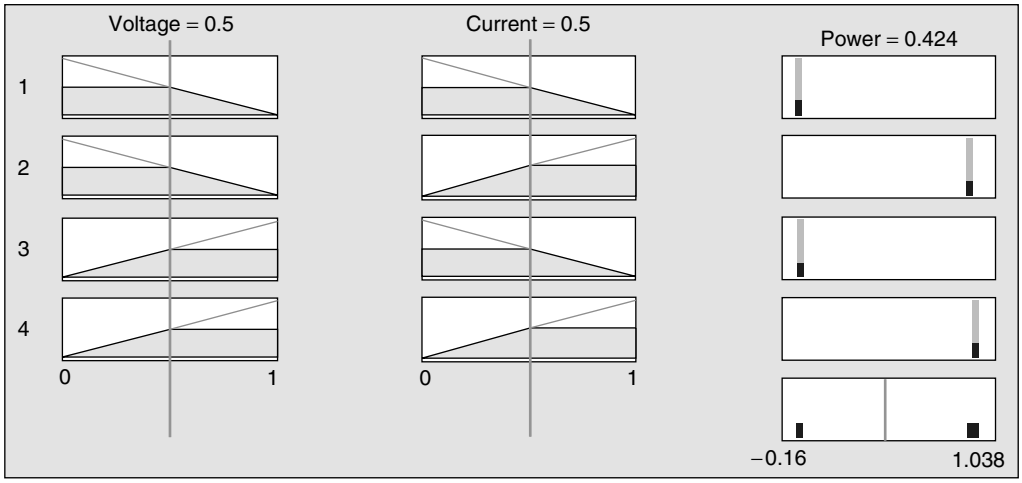


Figure 7. Takagi–Sugeno fuzzy inference system using two triangular MFs for input variables.

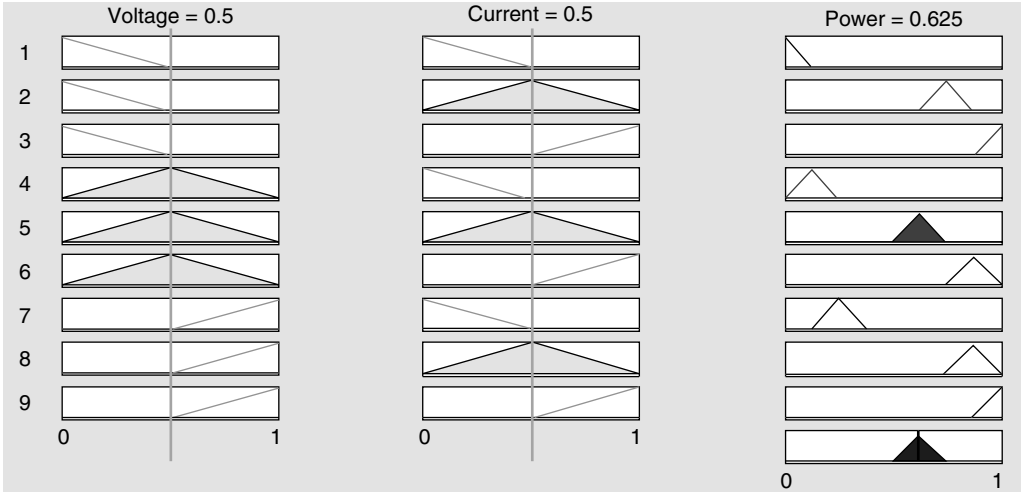


Figure 8. Mamdani fuzzy inference system using three triangular MFs for input variables.

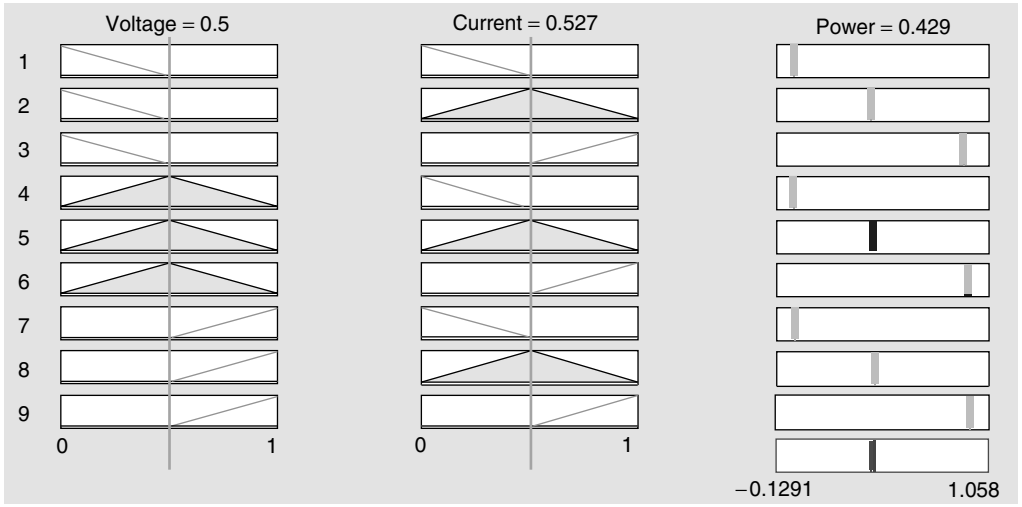


Figure 9. Takagi–Sugeno fuzzy inference system using three triangular MFs for input variables.

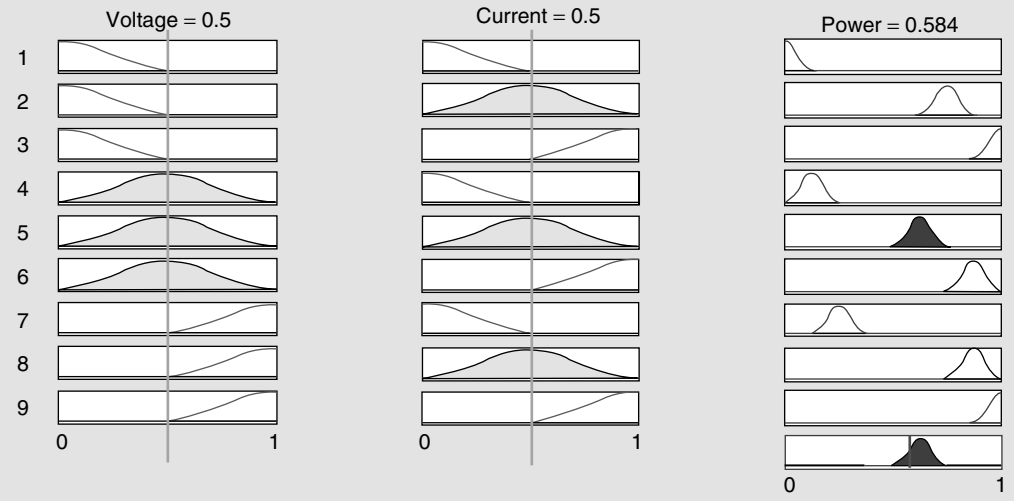


Figure 10. Mamdani fuzzy inference system using three Gaussian MFs for input variables.

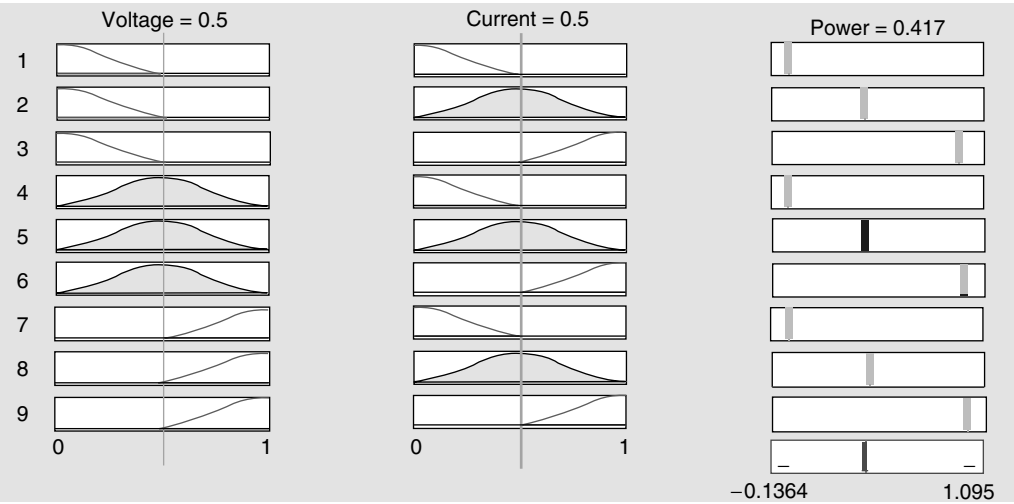


Figure 11. Takagi–Sugeno fuzzy inference system using three Gaussian MFs for input variables.

Table 3. Empirical comparison of fuzzy inference systems for different fuzzy operators.

Mamdani FIS		Takagi – Sugeno FIS	
Root mean squared error			
Training	Test	Training	Test
0.221	0.219	0.019	0.018

- Centroid
- Bisector of Area (BOA)
- Mean of Maximum (MOM)
- Smallest of Maximum (SOM).

For the Takagi–Sugeno FIS, the weighted sum and weighted average defuzzification methods were used. Table 4 summarizes the training and testing of RMSE values.

Discussions of Results and Problem Solution As depicted in Table 1, when the number of input MFs were increased from two to three, the RMSE values reduced regardless of the inference system used. However, when the shape of the MF was changed to Gaussian, RMSE values for Mamdani FIS decreased but the RMSE values for Takagi–Sugeno FIS increased (Table 2). Using Gaussian MFs, when the T-norm and T-conorm operators were changed to 'product' and 'sum' (instead of 'min' and 'max') both the inference methods performed better (Table 3). Finally, the selection of an ideal defuzzification operator also has a direct influence in the performance of FIS as shown in Table 4.

Table 4. Empirical comparison of fuzzy inference systems for different defuzzification operators.

Mamdani FIS			Takagi–Sugeno FIS		
Defuzzification	RMSE		Defuzzification	RMSE	
	Training	Test		Training	Test
Centroid	0.221	0.0219	Weighted sum	0.019	0.018
MOM	0.230	0.232	Weighted average	0.085	0.084
BOA	0.218	0.216			
SOM	0.229	0.232			

The design of the rule base (number of rules and how the inputs and outputs are related) is also very important for the good performance of FIS. The role of weighting factors emphasizing the importance of certain rules also bears a prominent role for the overall performance. When the input/output dimensions become larger, manual design becomes tedious and sometimes could even lead to poor design and implementation.

8 ADAPTATION OF FUZZY INFERENCE SYSTEMS

Expert knowledge is often the main source to design the fuzzy expert systems. Figure 12 illustrates the various parameters and components that need to be adapted for controlling a process. According to the performance measure of the problem environment, the membership functions, rule bases, and the inference mechanism are to be adapted (Abraham, 2002).

Neural network learning, self-organizing maps and clustering methods could be used to generate rules. Gradient descent and its variants could be applied to fine-tune the parameters of parameterized input/output membership functions and fuzzy operators (Abraham, 2001). Adaptation of fuzzy inference systems using evolutionary computation techniques has been widely explored. Evolutionary Computation (EC) is a population based adaptive method, which may be used to solve optimization problems, based on the genetic processes of biological organisms (Michalewicz and Fogel, 1999).

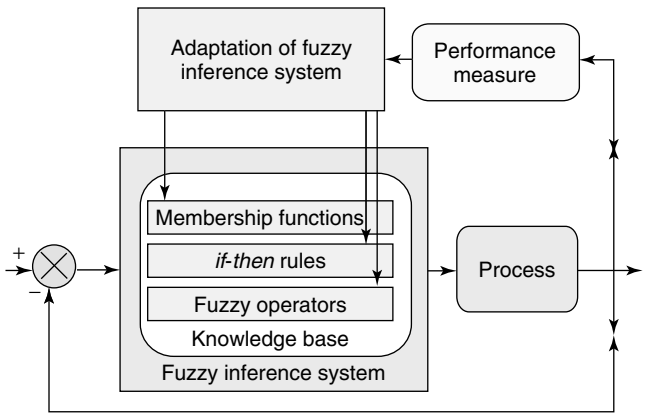


Figure 12. Adaptation of fuzzy inference systems.

Over many generations, natural populations evolve according to the principles of natural selection and 'survival of the fittest', first clearly stated by Charles Darwin in '*On the Origin of Species*'. By mimicking this process, EC could 'evolve' solutions to real-world problems, if they have been suitably encoded (problem representation is called *chromosome*). Automatic adaptation of membership functions is popularly known as *self tuning* and the chromosome encodes parameters of trapezoidal, triangle, logistic, hyperbolic-tangent, Gaussian membership functions, and so on. Evolutionary search of fuzzy rules can be carried out using three approaches. In the first method (Michigan approach), the fuzzy knowledge base is adapted as a result of antagonistic roles of competition and cooperation of fuzzy rules.

The second method (Pittsburgh approach), evolves a population of knowledge bases rather than individual fuzzy rules. Reproduction operators serve to provide a new combination of rules and new rules.

The third method (iterative rule learning approach), is very much similar to the first method with each chromosome representing a single rule, but contrary to the Michigan approach, only the best individual is considered to form part of the solution, discarding the remaining chromosomes of the population. The evolutionary learning process builds up the complete rule base through an iterative learning process (Cordón *et al*., 2001).

9 SUMMARY

Rule-based expert systems have been applied in a vast number of application areas. An important advantage of the fuzzy expert system is that the knowledge is expressed as easy-to-understand linguistic rules. If we have data, the fuzzy expert system can be taught using neural network

learning, EC, or other adaptation techniques. It is to be expected that the number of applications will grow considerably in the future now that success is clearly proven for these methods.

REFERENCES

Abraham, A. (2001) *Neuro-Fuzzy Systems: State-of-the-Art Modeling Techniques, Connectionist Models of Neurons, Learning Processes, and Artificial Intelligence*, in Lecture Notes in Computer Science, Vol. 2084, (eds. Mira., Jose and Prieto., Alberto) Springer Verlag, Germany (pp. 269–276).

Abraham, A. (2002) Intelligent Systems: Architectures and Perspectives, Recent Advances in Intelligent Paradigms and Applications, in *Studies in Fuzziness and Soft Computing*, Chapter 1, (eds A., Abraham, L., Jain and J., Kacprzyk), Springer Verlag, Germany (pp. 1–35).

Abraham, A. and Khan, M.R. (2003) Neuro-Fuzzy Paradigms for Intelligent Energy Management, Innovations in Intelligent Systems: Design, Management and Applications, in *Studies in Fuzziness and Soft Computing*, Chapter 12, (eds A., Abraham, L., Jain and B., Jan van der Zwaag), Springer Verlag, Germany (pp. 285–314).

CLIPS (2004) *Expert System Shell* <http://www.ghg.net/clips/CLIPS.html>.

Cordón, O., Herrera, F., Hoffmann, F. and Magdalena, L. (2001) *Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases*, World Scientific Publishing Company, Singapore.

Donald, W.A. (1986) *A Guide to Expert Systems*, Addison-Wesley, Boston, MA.

Fuzzy Logic Toolbox, *The MathWorks* (2004) http://www.mathworks.com/products/fuzzylogic/.

Giarratano, J. and Riley, G. (1989) *Expert Systems: Principles and Programming*, PWS-Kent Publishing Co, Boston, MA.

Ignizio, J.P. (1991) *Introduction to Expert Systems: The Development and Implementation of Rule-Based Expert Systems*, McGraw-Hill, Inc, USA.

Jang, J.S.R., Sun, C.T. and Mizutani, E. (1997) *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*, Prentice Hall Inc, USA.

Mamdani, E.H. and Assilian, S. (1975) An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller. *International Journal of Man-Machine Studies*, **7**(1), 1–13.

Michalewicz, Z. and Fogel, D.B. (1999) *How to Solve It: Modern Heuristics*, Springer Verlag, Germany.

Militallo, L.G., Hutton, R.J.B. (1998) Applied Cognitive Task Analysis (ACTA): A Practitioner's Toolkit for Understanding Cognitive. *Ergonomics*, **41**(11), 1618–1642.

Niwa, K., Sasaki, K. and Ihara, H. (1988) An Experimental Comparison of Knowledge Representation Schemes, in *Principles of Expert Systems*, (Eds A., Gupta and E.B., Prasad), IEEE Press, New York (pp. 133–140).

Schneider, M., Langholz, G., Kandel, A. and Chew, G. (1996) *Fuzzy Expert System Tools*, John Wiley & Sons, USA.

Takagi, T. and Sugeno, M. (1985) Fuzzy identification of systems and its applications of modeling and control, *IEEE Transactions of Systems. Man and Cybernetics*, USA (pp. 116–132).

Vicente, K.J. (1999) *Cognitive Work Analysis:∼ Towards Safe, Productive, and Healthy Computer-Based Work*, Lawrence Erlbaum Associates, Inc. Press, USA.

Zadeh, L.A. (1965) Fuzzy Sets. *Information and Control*, **8**, 338–353.

131: Evolutionary Computation

Ajith Abraham
Oklahoma State University, Stillwater, OK, USA

1 Introduction and Biological Motivation 920
2 Genetic Algorithms 921
3 Schema Theorem 922
4 Selection and Reproduction 922
5 GA Demonstrations 924
6 Evolution Strategies 925
7 Evolutionary Programming 928
8 Genetic Programming 928
9 Genetic Programming Basics 929
10 Summary 931
References 931
Further Reading 931

1 INTRODUCTION AND BIOLOGICAL MOTIVATION

A general introduction to artificial intelligence methods of measurement signal processing is given in **Article 128, Nature and Scope of AI Techniques, Volume 2**.

In nature, evolution is mostly determined by natural selection or different individuals competing for resources in the environment. Those individuals that are better are more likely to survive and propagate their genetic material. The encoding for genetic information (genome) is done in a way that admits asexual reproduction, which results in offspring that are genetically identical to the parent. Sexual reproduction allows some exchange and reordering of chromosomes, producing offspring that contain a combination of information from each parent. This is the recombination operation, which is often referred to as *crossover* because of the way strands of chromosomes cross over during the exchange. The diversity in the population is achieved by mutation operation.

Evolutionary algorithms are ubiquitous nowadays, having been successfully applied to numerous problems from different domains, including optimization, automatic programming, signal processing, bioinformatics, social systems, and so on. In many cases, the mathematical function, which describes the problem, is not known, and the values at certain parameters are obtained from simulations. In contrast to many other optimization techniques, an important advantage of evolutionary algorithms is they can cope with multimodal functions.

Usually found grouped under the term evolutionary computation or evolutionary algorithms (Bäck, 1996), are the domains of genetic algorithms (GA) (Holland, 1975), evolution strategies (Rechenberg, 1973; Schwefel, 1977), evolutionary programming (Fogel, Owens and Walsh, 1966), and genetic programming (Koza, 1992).

These all share a common conceptual base of simulating the evolution of *individual* structures via processes of *selection*, recombination, and mutation *reproduction*, thereby producing better solutions. The processes depend on the perceived performance of the individual structures as defined by the problem.

A population of candidate solutions (for the optimization task to be solved) is initialized. New solutions are created by applying reproduction operators (crossover and/or mutation). The fitness (how good the solutions are) of the resulting solutions is evaluated and suitable selection strategy is then applied to determine which solutions will be maintained into the next generation. The procedure is then iterated, as illustrated in Figure 1.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

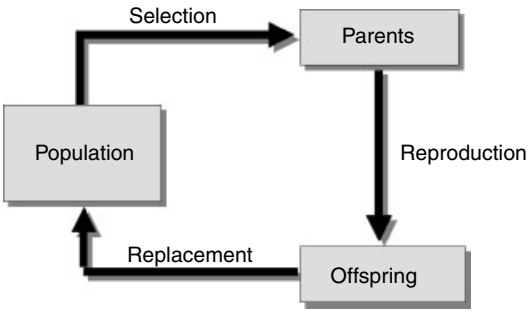


Figure 1. Flow chart of an evolutionary algorithm.

A primary advantage of evolutionary computation is that it is conceptually simple.

The procedure may be written as the difference equation:

$$x(t+1) = s\{v[x(t)]\} \tag{1}$$

where $x(t)$ is the population at time t under a representation x, v is a random variation (reproduction) operator, and s is the selection operator (Fogel, 1999).

1.1 Advantages of evolutionary algorithms

Following are some of the advantages of using evolutionary algorithms while compared to other global optimization techniques (Fogel, 1999).

1. Evolutionary algorithm performance is representation independent in contrast to other numerical techniques, which might be applicable for only continuous values or other constrained sets.
2. Evolutionary algorithms offer a framework such that it is comparably easy to incorporate prior knowledge about the problem. Incorporating such information focuses the evolutionary search, yielding a more efficient exploration of the state space of possible solutions.
3. Evolutionary algorithms can also be combined with more traditional optimization techniques. This may be as simple as the use of a gradient minimization after primary search with an evolutionary algorithm (e.g. fine tuning of weights of an evolutionary neural network) or it may involve simultaneous application of other algorithms (e.g. hybridizing with simulated annealing or Tabu search to improve the efficiency of basic evolutionary search).
4. The evaluation of each solution can be handled in parallel and only selection (which requires at least pair-wise competition) requires some serial processing. Implicit parallelism is not possible in many global optimization algorithms like simulated annealing and Tabu search.
5. Traditional methods of optimization are not robust to the dynamic changes in the problem of the environment and often require a complete restart in order to provide a solution (e.g. dynamic programming). In contrast, evolutionary algorithms can be used to adapt solutions to changing circumstance.
6. Perhaps, the greatest advantage of evolutionary algorithms comes from the ability to address problems for which there are no human experts. Although human expertise should be used when it is available, it often proves less than adequate for automating problem-solving routines.

2 GENETIC ALGORITHMS

A typical flowchart of a genetic algorithm is depicted in Figure 2. One iteration of the algorithm is referred to as a *generation*. The basic GA is very generic, and there are many aspects that can be implemented differently according to the problem (e.g. representation of solution (chromosomes), type of encoding, selection strategy, type of crossover and mutation operators, etc.). In practice, GAs are implemented by having arrays of bits or characters to represent the chromosomes. The individuals in the population

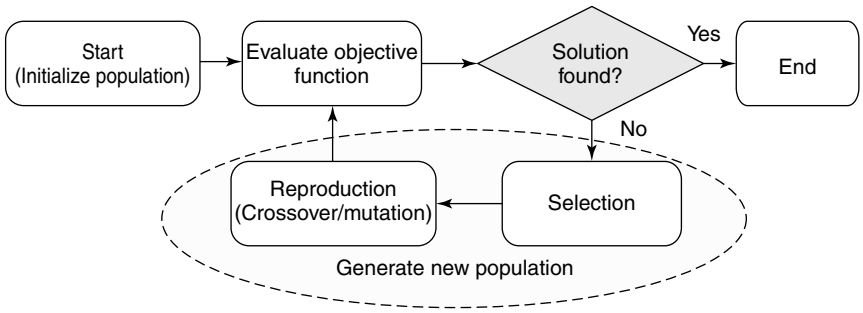


Figure 2. Flow chart of basic genetic algorithm iteration.

then go through a process of simulated evolution. Simple bit manipulation operations allow the implementation of crossover, mutation, and other operations. The number of bits for every gene (parameter) and the decimal range in which they decode are usually the same, but nothing precludes the utilization of a different number of bits or range for every gene.

When compared to other evolutionary algorithms, one of the most important GA feature is its focus on fixed-length character strings, although variable-length strings and other structures have been used.

2.1 Encoding and decoding

In a typical application of GAs, the given problem is transformed into a set of genetic characteristics (parameters to be optimized) that will survive in the best possible manner in the environment. For example, optimizing a function

$$\min f(x_1, x_2) = (x_1 - 5)^2 + (x_2 - 2)^2 \text{for} - 3 \leq x_1 3 \leq; \\ -8 \leq x_2 8 \leq \quad (2)$$

The parameters of the search are identified as x_1 and x_2, which are called the *phenotypes* in evolutionary algorithms. In genetic algorithms, the *phenotypes* (parameters) are usually converted to *genotypes* by using a coding procedure. Knowing the ranges of x_1 and x_2, each variable is to be represented using a suitable binary string. This representation using binary coding makes the parametric space independent of the type of variables used. The genotype (chromosome) should in some way contain information about solution, which is also known as *encoding*. GAs use a binary string encoding, as shown below.

Chromosome A: 110110111110100110110

Chromosome B: 11011110101010011110

Each bit in the chromosome strings can represent some characteristic of the solution. There are several types of encoding (e.g. direct integer or real numbers encoding), which directly depends on the problem.

Permutation encoding can be used in ordering problems, such as the traveling salesman problem (TSP) or task-ordering problem. In permutation encoding, every chromosome is a string of numbers, which represents numbers in a sequence. A chromosome using permutation encoding for a 9-city TSP problem will appear as follows:

Chromosome A: 4 5 3 2 6 1 7 8 9

Chromosome B: 8 5 6 7 2 3 1 4 9

The chromosome represents the order in which the salesman will visit the cities. Special care is taken to ensure that the strings represent real sequences after crossover and mutation. Floating-point representation is very useful for numeric optimization (e.g. for encoding the weights of a neural network).

It should be noted that in many recent applications, more sophisticated genotypes are appearing (e.g. chromosome can be a tree of symbols or a combination of a string and a tree, some parts of the chromosome are not allowed to evolve, etc.).

3 SCHEMA THEOREM

Theoretical foundations of evolutionary algorithms can be partially explained by the schema theorem (Holland, 1975), which relies on the concept of schemata. Schemata are templates that partially specify a solution (more strictly, a solution in the genotype space). If genotypes are strings built using symbols from an alphabet A, schemata are strings whose symbols belong to $A \cup (*)$. This extra-symbol $(*)$ must be interpreted as a wildcard, being loci occupied by it, called *undefined*. A chromosome is said to match a schema if they agree in the defined positions.

For example, the string 10011010 matches the schemata 1******* and **011*** among others but does not match *1*11*** because they differ in the second gene (the first defined gene in the schema).

A schema can be viewed as a hyperplane in a k-dimensional space, representing a set of solutions with common properties. Obviously, the numbers of solutions that match a schema H depend on the number of defined positions in it. Another related concept is the *defining-length* of a schema, defined as the distance between the first and the last defined positions in it.

The GA works by allocating strings to best schemata exponentially through successive generations, this being the selection mechanism mainly responsible for this behavior. On the other hand, the crossover operator is responsible for exploring new combinations of the present schemata in order to get the fittest individuals. Finally, the purpose of the mutation operator is to introduce fresh genotypic material in the population.

4 SELECTION AND REPRODUCTION

Individuals for producing offspring are chosen using a selection strategy after evaluating the fitness value of each individual in the selection pool. Each individual in the selection pool receives a reproduction probability

depending on its own fitness value and the fitness value of all other individuals in the selection pool. This fitness is used for the actual selection step afterwards. Some of the popular selection schemes are discussed below.

4.1 Roulette-wheel selection

The simplest selection scheme is the roulette-wheel selection, also called *stochastic sampling with replacement*. This technique is analogous to a roulette wheel with each slice proportional in size to the fitness. The individuals are mapped to contiguous segments of a line such that each individual's segment is equal in size to its fitness. A random number is generated and the individual whose segment spans the random number is selected. The process is repeated until the desired number of individuals is obtained. As illustrated in Figure 3, chromosome 1 has the highest probability for being selected since it has the highest fitness.

4.2 Rank-based fitness assignment

In rank-based fitness assignment, the population is sorted according to the objective values. The fitness assigned to each individual depends only on the position of the objective values in the individual's rank. Ranking introduces a uniform scaling across the population.

4.3 Tournament selection

In tournament selection, a number of individuals are chosen randomly from the population and the best individual from this group is selected as the parent. This process is repeated as often until there are sufficient individuals to choose. These selected parents produce uniformly random offspring. The tournament size will often depend on the problem, population size, and so on. The parameter for tournament selection is the tournament size. Tournament size takes values ranging from two to the total number of individuals in the population.

4.4 Elitism

When creating a new population by crossover and mutation, there is a big chance that we will lose the best chromosome. Elitism is the name of the method that first copies the best chromosome (or a few best chromosomes) to the new population. The rest is done in the classical way. Elitism can very rapidly increase performance of GA because it prevents losing the best-found solution.

4.5 Genetic operators

Crossover and mutation are two basic operators of GA. Performance of GA depends very much on the genetic operators. Type and implementation of operators depends on encoding and also on the problem. There are many ways of doing crossover and mutation. In this section, we will demonstrate some of the popular methods, with some examples and suggestions as to how to do it for different encoding schemes.

4.6 Crossover

Crossover selects genes from parent chromosomes and creates a new offspring. The simplest way to do this is to choose randomly some crossover point and everything before this point is copied from the first parent and then, everything after a crossover point is copied from the second parent. A single point crossover is illustrated as follows (| is the crossover point):

Chromosome A: 11011|00100110110

Chromosome B: 11011|11000011110

Offspring A: 11011|11000011110

Offspring B: 11011|00100110110

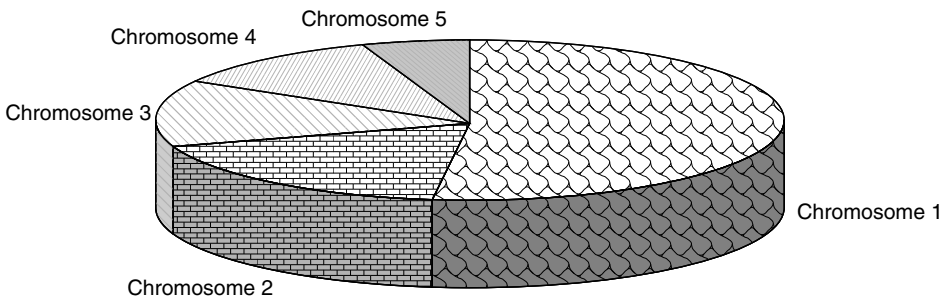


Figure 3. Roulette-wheel selection.

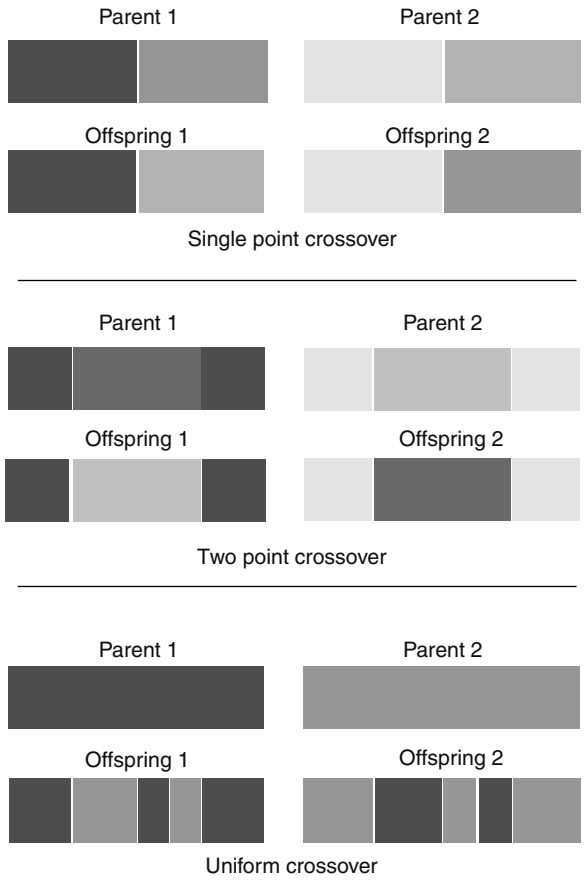


Figure 4. Types of crossover operators.

As illustrated in Figure 4, there are several crossover techniques. In a uniform crossover, bits are randomly copied from the first or the second parent. Specific crossover made for a specific problem can improve the GA performance.

4.7 Mutation

After crossover operation, mutation takes place. Mutation randomly changes the new offspring. For binary encoding, mutation is performed by changing a few randomly chosen bits from 1 to 0 or from 0 to 1. Mutation depends on the encoding as well as the crossover. For example, when we are encoding permutations, mutation could be exchanging two genes. A simple mutation operation is illustrated as follows:

Chromosome A: 1101111000011110

Chromosome B: 1101100100110110

Offspring A: 1100111000011110

Offspring B: 1101101100110110

For many optimization problems, there may be multiple, equal, or unequal optimal solutions. Sometimes, a simple GA cannot maintain stable populations at different optima of such functions. In the case of unequal optimal solutions, the population invariably converges to the global optimum. Niching helps to maintain subpopulations near global and local optima. A niche is viewed as an organism's environment and a species as a collection of organisms with similar features. Niching helps to maintain subpopulations near global and local optima by introducing a controlled competition among different solutions near every local optimal region. Niching is achieved by a sharing function, which creates subdivisions of the environment by degrading an organism's fitness proportional to the number of other members in its neighborhood. The amount of sharing contributed by individuals to their neighbor is determined by their proximity in the decoded parameter space (phenotypic sharing) based on a distance measure (Goldberg, 1989).

5 GA DEMONSTRATIONS

5.1 Rastrigin function

The Rastrigin function is a typical example of nonlinear multimodal function. It was first proposed by Rastrigin (Törn and Zilinskas, 1989) as a two-dimensional function and has later been generalized. This function is a fairly difficult problem due to its large search space and its large number of local minima. Rastrigin's function is defined as

$$F(x) = 10n + \sum_{i=1}^{n} x_i^2 - 10\cos(2\pi x_i), -5.12 \leq x_i \leq 5.12 \tag{3}$$

The function has just one global minimum, which occurs at the origin where the value of the function is 0. At any local minimum other than [0, 0], the value of Rastrigin's function is greater than 0. The farther the local minimum is from the origin, the larger the value of the function is at that point. Figure 5 illustrates the surface of the function for two input variables.

A real-value representation was used to encode the two input variables. The following parameters were used for the GA experiments.

- Mutation: 0.05, crossover: 0.90
- Population size: 20, number of iterations: 50, selection method: Roulette-wheel selection.

Figure 6 illustrates how the best fitness values were evolved during the 50 generations. As evident after 30 generations, the GA algorithm has succeeded in finding the best optimal solution.

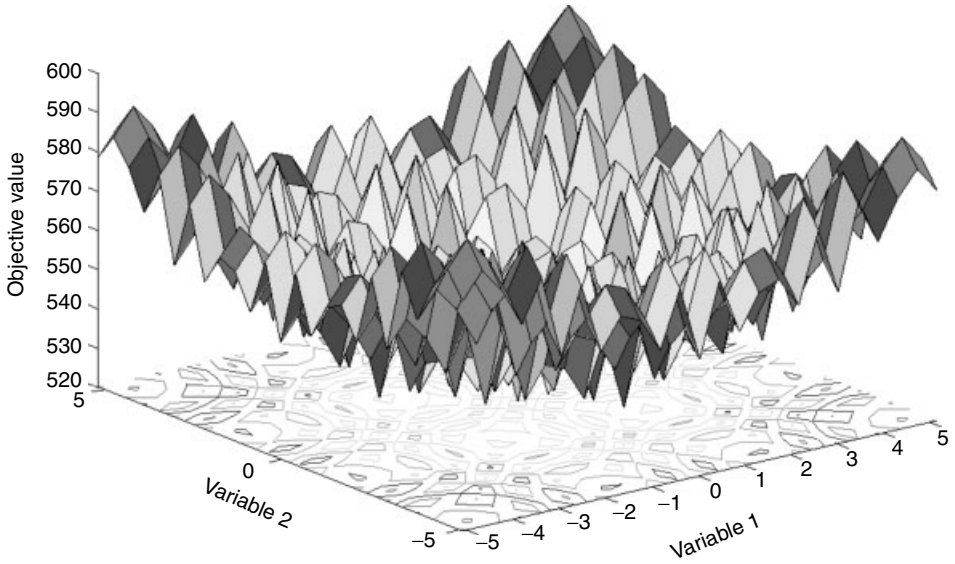


Figure 5. Rastrigin's function for two variables.

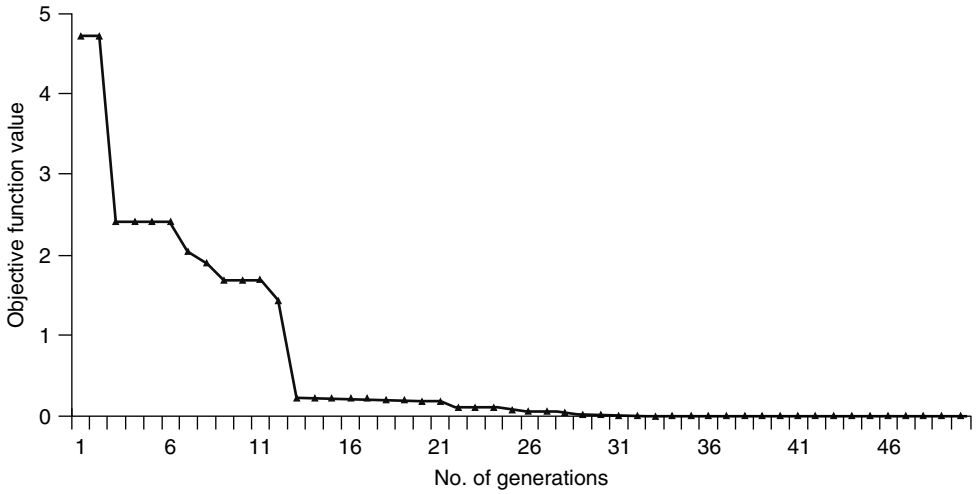


Figure 6. GA learning during the 50 generations.

5.2 Peaks function

Peaks function is a function of two variables, obtained by translating and scaling Gaussian distributions (Jang, Sun and Mizutani, 1997).

$$F(x, y) = 3(1-x)^2 \exp(-(x^2) - (y+1)^2) - 10\left(\frac{x}{5 - x^3 - y^5}\right) \exp(-x^2 - y^2) - \tfrac{1}{3}\exp(-(x+1)^2 - y^2) \quad \text{for } -3 \le x \le 3 \quad \text{and} \quad -3 \le y \le 3 \tag{4}$$

The Peak function surface is plotted in Figure 7, and the task is to find the optimum value (maximum) for the given range of x and y values. Using a population size of 30, the genetic algorithm was run for 25 iterations. Each input variable was represented using 8 bit. Crossover and mutation rates were set as 0.9 and 0.1 respectively.

Figure 8(a), (b), and (c), illustrate the convergence of the solutions on a contour plot of the surface. After 10 iterations, almost all the solutions were near the optimal point.

6 EVOLUTION STRATEGIES

Evolution strategy (ES) was developed by Rechenberg (1973) and Schwefel (1977). ES tends to be used for

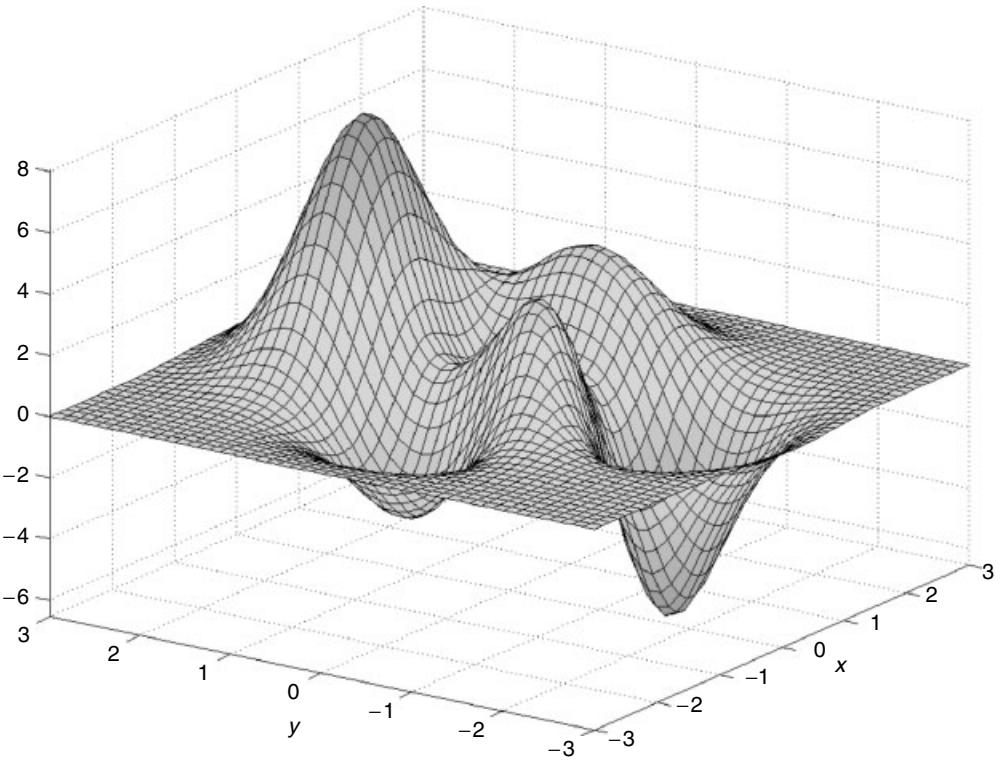


Figure 7. Surface of Peaks function.

empirical experiments that are difficult to model mathematically. The system to be optimized is actually constructed and ES is used to find the optimal parameter settings.

Evolution strategies merely concentrate on translating the fundamental mechanisms of biological evolution for technical optimization problems. The parameters to be optimized are often represented by a vector of real numbers (object parameters – o_p). Another vector of real numbers defines the strategy parameters (s_p), which controls the mutation of the objective parameters. Both object and strategic parameters form the data structure for a single individual.

A population P of n individuals could be described as follows:

$$P = (c_1, c_2, \ldots, c_{n-1}, c_n) \tag{5}$$

where the ith chromosome c_i is defined as: $c_i = (o_p, s_p)$

$$o_p = (o_1, o_2, \ldots, o_{n-1}, o_n) \quad \text{and}$$

$$s_p = (s_1, s_2, \ldots, s_{n-1}, s_n) \tag{6}$$

6.1 Mutation in evolution strategies

The mutation operator is defined as component-wise addition of normal distributed random numbers. Both the objective parameters and the strategy parameters of the chromosome are mutated. Objective parameter vector is calculated as follows:

$$o_{p(mut)} = o_p + N_0(s_p) \tag{7}$$

where $N_0(s_i)$ is the Gaussian distribution of mean value 0 and standard deviation s_i.

Usually, the strategy parameters mutation step size is done by adapting the standard deviation s_i. This may be done (for example) as follows:

$$s_{p(mut)} = ({s_1}^* A_1, {s_2}^* A_2, \ldots, {s_{n-1}}^* A_{n-1}, {s_n}^* A_n) \tag{8}$$

where A_i is randomly chosen from α or $1/\alpha$, depending on the value of equally distributed random variable E of [0,1]

$$A_i = \alpha \quad \text{if} \quad E < 0.5$$

$$A_i = \frac{1}{\alpha} \quad \text{if} \quad E \geq 0.5 \tag{9}$$

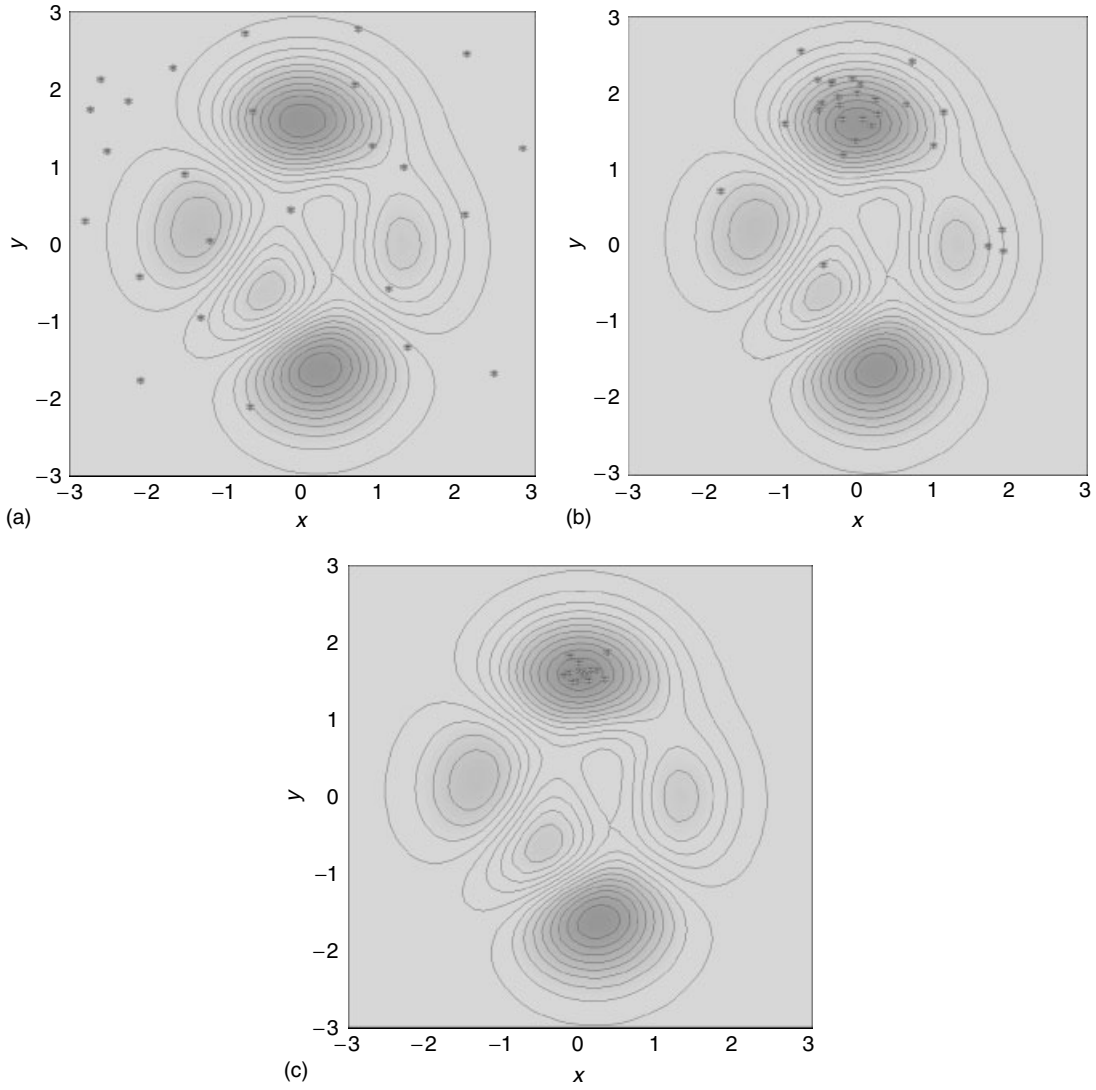


Figure 8. Convergence of solutions (a) generation 0; (b) after 5 generations; (c) after 20 generations (solution points are marked with *).

α is usually referred to as *strategy parameters adaptation value*.

6.2 Crossover (recombination) in evolution strategies

For two chromosomes $c_1 = (o_{p(c1)}, s_{p(c1)})$ and $c_2 = (o_{p(c2)}, s_{p(c2)})$, the crossover operator x is defined as follows:

$$\begin{aligned} R(c_1, c_2) &= c = (o_p, s_p) \\ \text{with } o_{p(i)} &= (o_{p(c1),i} | o_{p(c2),i}) \\ \text{and } s_{p(i)} &= (s_{p(c1),i} | s_{p(c2),i}) \end{aligned} \tag{10}$$

By defining $o_{p(i)}$ and $s_{p(i)} = (x|y)$, a value is randomly assigned for either x or y (50% selection probability for x and y).

6.3 Controlling the evolution

Let P be the number of parents in generation 1 and let C be the number of children in generation i. There are basically four different types of evolution strategies: P,C; $P + C$; $P/R, C$; and $P/R + C$, as discussed below. They mainly differ in how the parents for the next generation are selected and in the usage of crossover operators.

6.3.1 P,C *strategy*

The P parents produce C children, using mutation. Fitness values are calculated for each of the C children and the best P children become next-generation parents. The best individuals of C children are sorted by their fitness value and the first P individuals are selected to be next-generation parents ($C \geq P$).

6.3.2 P + C *strategy*

The P parents produce C children, using mutation. Fitness values are calculated for each of the C children and the best P individuals of both parents and children become next-generation parents. Children and parents are sorted by their fitness value and the first P individuals are selected to be next-generation parents.

6.3.3 P/R, C *strategy*

The P parents produce C children, using mutation and crossover. Fitness values are calculated for each of the C children and the best P children become next-generation parents. The best individuals of C children are sorted by their fitness value and the first P individuals are selected to be next-generation parents ($C \geq P$). Except for the usage of crossover operator, this is exactly the same as P,C strategy.

6.3.4 P/R + C *strategy*

The P parents produce C children, using mutation and recombination. Fitness values are calculated for each of the C children and the best P individuals of both parents and children become next-generation parents. Children and parents are sorted by their fitness value and the first P individuals are selected to be next-generation parents. Except for the usage of crossover operator, this is exactly the same as $P + C$ strategy.

7 EVOLUTIONARY PROGRAMMING

The book *Artificial Intelligence Through Simulated Evolution* by Fogel, Owens and Walsh (1966) is the landmark publication for evolutionary programming (EP). In this book, finite state automata are evolved to predict symbol strings generated from Markov processes and nonstationary time series (AI–FAQ-Genetic). The basic evolutionary programming method involves the following steps:

1. Choose an initial population (possible solutions at random). The number of solutions in a population is highly relevant to the speed of optimization, but no definite answers are available as to how many solutions are appropriate (other than > 1).
2. New offspring are created by mutation. Each offspring solution is assessed by computing its fitness. Typically, a stochastic tournament is held to determine the N solutions to be retained for the population of solutions. It should be noted that, typically, evolutionary programming method does not use any crossover as a genetic operator.

7.1 Evolutionary programming versus genetic algorithms

1. GA is implemented by having arrays of bits or characters to represent the chromosomes. In EP, there are no such restrictions for the representation. In most cases, the representation follows from the problem.
2. EP typically uses an adaptive mutation operator in which the severity of mutations is often reduced as the global optimum is approached, while GAs use a prefixed mutation operator. Among the schemes to adapt the mutation step size, the most widely studied is the 'meta-evolutionary' technique in which the variance of the mutation distribution is subject to mutation by a fixed variance mutation operator that evolves along with the solution.

7.2 Evolutionary programming versus evolution strategies

1. When implemented to solve real-valued function optimization problems, both typically operate on the real values themselves and use adaptive reproduction operators.
2. EP typically uses stochastic tournament selection, while ES typically uses deterministic selection.
3. EP does not use crossover operators, while ES (P/R, C and $P/R + C$ strategies) crossover. However, the effectiveness of the crossover operators depends on the problem at hand.

8 GENETIC PROGRAMMING

The genetic programming (GP) technique provides a framework for automatically creating a working computer program from a high-level statement of the problem (Koza, 1992).

Genetic programming achieves this goal of automatic programming by genetically breeding a population of computer programs, using the principles of Darwinian natural selection and biologically inspired operations. The operations include most of the techniques discussed in the previous sections. The main difference between GP and GA is the representation of the solution. GP creates computer programs in the LISP or scheme computer languages as the solution. LISP is an acronym for LISt Processor and was developed in the late 1950s (History of LISP, 2004). Unlike most languages, LISP is usually used as an interpreted language. This means that, unlike compiled languages, an interpreter can process and respond directly to programs written in LISP.

The main reason for choosing LISP to implement GP is because of the advantage that the programs and data have the same structure, which could provide easy means for manipulation and evaluation.

In GP, the individual population members are not fixed-length character strings that encode possible solutions to the problem at hand, they are programs that, when executed, are the candidate solutions to the problem. These programs are expressed in genetic programming as parse trees rather than as lines of code. For example, the simple program '$a + b * c$' would be represented as shown in Figure 9. The terminal and function sets are also important components of genetic programming. The terminal and function sets are the alphabets of the programs to be made. The terminal set consists of the variables and constants of the programs (e.g. A, B, and C in Figure 9).

The most common way of writing down a function with two arguments is the infix notation. That is, the two arguments are connected with the operator symbol between them as follows.

$$A + B$$

A different method is the prefix notation. Here, the operator symbol is written down first, followed by its required arguments.

$$+AB$$

While this may be a bit more difficult or just unusual for human eyes, it opens some advantages for computational uses. The computer language LISP uses symbolic expressions (or S-expressions) composed in prefix notation. Then, a simple S-expression could be

$$(operator, argument),$$

where *operator* is the name of a function and *argument* can be either a constant or a variable or another symbolic expression, as shown below.

$$(operator, argument\,(operator, argument)$$
$$(operator, argument))$$

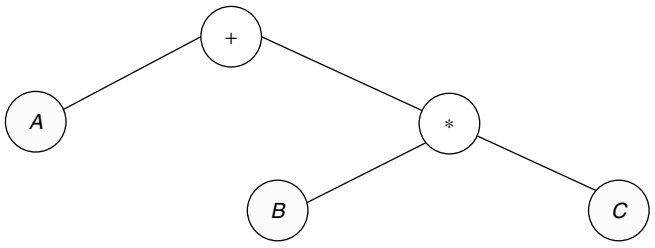


Figure 9. A simple tree structure of GP.

9 GENETIC PROGRAMMING BASICS

A parse tree is a structure that develops the interpretation of a computer program. Functions are written down as nodes and their arguments as leaves. A subtree is the part of a tree that is under an inner node of this tree, as illustrated in Figure 10. If this tree is cut out from its parent, the inner node becomes a root node and the subtree is a valid tree of its own.

There is a close relationship between these parse trees and S-expression; in fact, these trees are just another way of writing down expressions. While functions will be the nodes of the trees (or the operators in the S-expressions) and can have other functions as their arguments, the leaves will be formed by terminals, that is, symbols that may not be further expanded. Terminals can be variables, constants, or specific actions that are to be performed. The process of selecting the functions and terminals that are needed or are useful for finding a solution to a given problem is one of the key steps in GP. Evaluation of these structures is straightforward. Beginning at the root node, the values of all subexpressions (or subtrees) are computed, descending the tree down to the leaves. GP procedure could be summarized as follows:

- generate an initial population of random compositions of the functions and terminals of the problem;
- compute the fitness values of each individual in the population;
- using some selection strategy and suitable reproduction operators, produce offsprings;
- iterate the procedure until the required solution is found or the termination conditions have been reached (specified number of generations).

The creation of an offspring from the crossover operation is accomplished by deleting the crossover fragment of the first parent and then inserting the crossover fragment of the second parent. The second offspring is produced in a symmetric manner. A simple crossover operation is illustrated in Figure 11. In GP, the crossover operation is

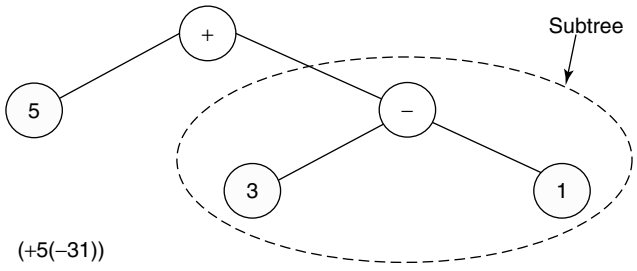


Figure 10. Illustration of a parse tree and a subtree.

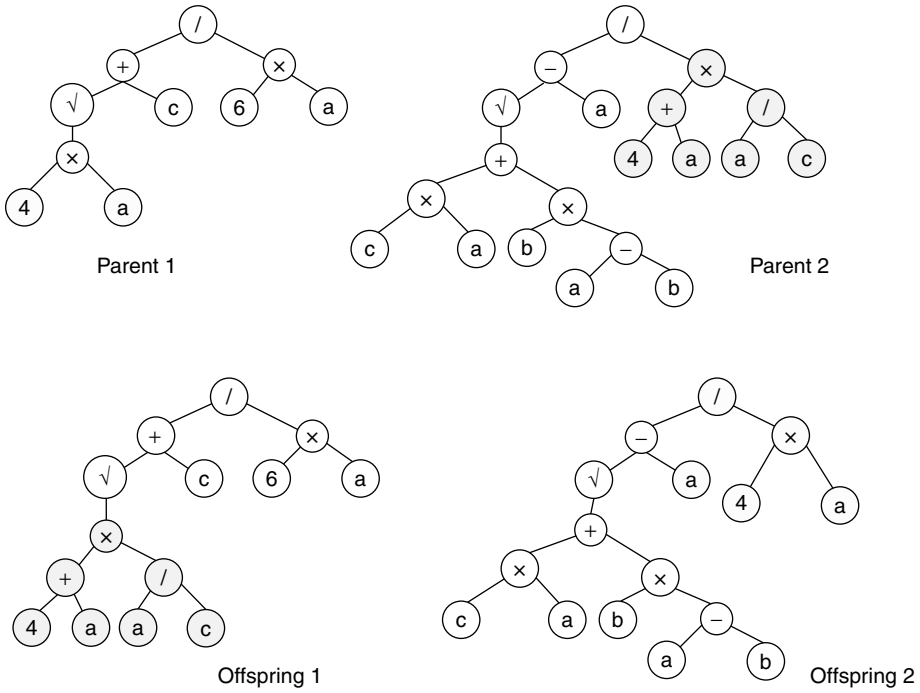


Figure 11. Illustration of crossover operator.

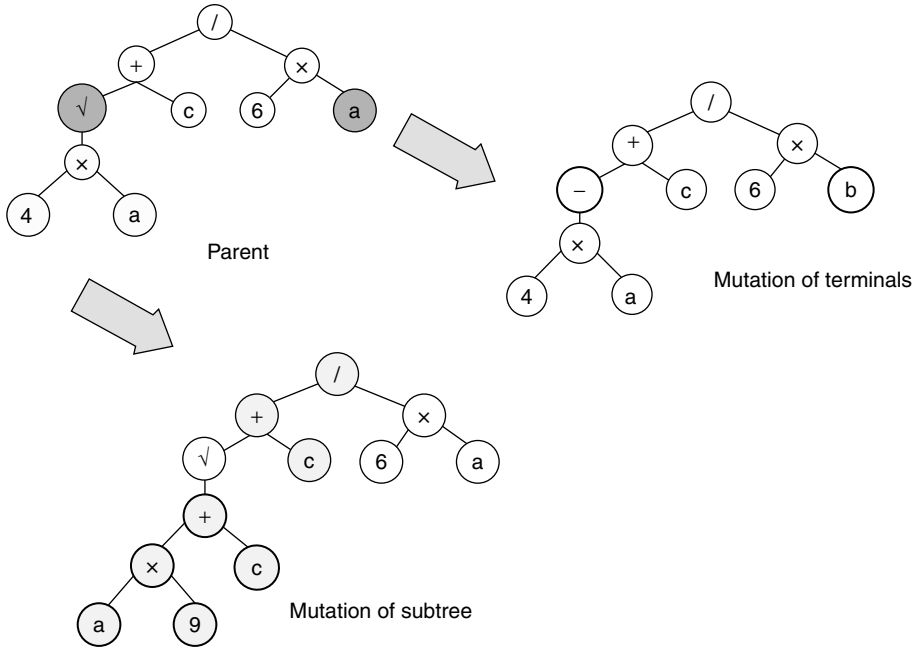


Figure 12. Illustration of mutation operator in GP.

implemented by taking randomly selected subtrees in the individuals and exchanging them.

Mutation is another important feature of genetic programming. Two types of mutations are commonly used. The simplest type is to replace a function or a terminal by another function or a terminal respectively. In the second kind, an entire subtree can replace another subtree. Figure 12 explains the concepts of mutation.

GP requires data structures that are easy to handle and evaluate and are robust to structural manipulations. These are among the reasons why the class of S-expressions was chosen to implement GP. The set of functions and terminals

that will be used in a specific problem has to be chosen carefully. If the set of functions is not powerful enough, a solution may be very complex or may not be found at all. Like in any evolutionary computation technique, the generation of the first population of individuals is important for successful implementation of GP. Some of the other factors that influence the performance of the algorithm are the size of the population, percentage of individuals that participate in the crossover/mutation, maximum depth for the initial individuals and the maximum allowed depth for the generated offspring, and so on. Some specific advantages of genetic programming are that no analytical knowledge is needed and still accurate results could be obtained. GP approach does scale with the problem size. GP does impose restrictions on how the structure of solutions should be formulated.

10 SUMMARY

This article presents the biological motivation and fundamental aspects of evolutionary algorithms and its constituents, namely, genetic algorithm, evolution strategies, evolutionary programming, and genetic programming. Performance of genetic algorithms is demonstrated using two function optimization problems. Important advantages of evolutionary computation as compared to classical optimization techniques are also discussed.

REFERENCES

AI – FAQ-Genetic. http://www.faqs.org/faqs/ai-faq/genetic/, accessed on September 10, 2004.

Bäck, T. (1996) *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic algorithms*, Oxford University Press, New York.

Fogel, D.B. (1999) *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*, 2nd edn, IEEE Press, Piscataway, NJ.

Fogel, L.J., Owens, A.J. and Walsh, M.J. (1966) *Artificial Intelligence Through Simulated Evolution*, John Wiley & Sons, New York.

Goldberg, D.E. (1989) *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley Publishing Corporation, Inc, Reading, MA.

History of LISP. (2004) http://www-formal.stanford.edu/jmc/history/lisp/lisp.html.

Holland, J. (1975) *Adaptation in Natural and Artificial Systems*, University of Michican Press, Ann Harbor, MI.

Jang, J.S.R., Sun, C.T. and Mizutani, E. (1997) *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*, Prentice Hall, USA.

Koza, J.R. (1992) *Genetic Programming*, MIT Press, Cambridge, MA.

Rechenberg, I. (1973) *Evolutionsstrategie: Optimierung Technischer Systeme nach Prinzipien der biologischen Evolution*, Fromman-Holzboog, Stuttgart.

Schwefel, H.P. (1977) *Numerische Optimierung von Computermodellen Mittels der Evolutionsstrategie*, Birkhaeuser, Basel.

Törn, A. and Zilinskas, A. (1989) *Global Optimization, Lecture Notes in Computer Science*, Vol. 350, Springer-Verlag, Berlin.

FURTHER READING

Michalewicz, Z. (1992) *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag, Berlin.

132: Data Acquisition Systems (DAS) in General

Gerd Wöstenkühler
Hochschule Harz, Wernigerode, Germany

1 Introduction 937
2 Basic Data Acquisition System 937
3 Basic Data Distribution System 939
Further Reading 939

1 INTRODUCTION

Data acquisition systems (DAS) interface between the real world of physical parameters, which are analog, and the artificial world of digital computation and control. With current emphasis on digital systems, the interfacing function has become an important one; digital systems are used widely because complex circuits are low cost, accurate, and relatively simple to implement. In addition, there is rapid growth in the use of microcomputers to perform difficult digital control and measurement functions.

Computerized feedback control systems are used in many different industries today in order to achieve greater productivity in our modern industrial societies. Industries that presently employ such automatic systems include steel making, food processing, paper production, oil refining, chemical manufacturing, textile production, cement manufacturing, and others.

The devices that perform the interfacing function between analog and digital worlds are analog-to-digital (A/D) and digital-to-analog (D/A) converters, which together are known as *data converters*. Some of the specific applications in which data converters are used include data telemetry systems, pulse code modulated communications, automatic test systems, computer display systems, video signal processing systems, data logging systems, and sampled data control systems. In addition, every laboratory digital multimeter or digital panel meter contains an A/D converter.

2 BASIC DATA ACQUISITION SYSTEM

Besides A/D and D/A converters, data acquisition and distribution systems may employ one or more of the following circuit functions:

- transducers,
- amplifiers,
- filters,
- nonlinear analog functions,
- analog multiplexers,
- sample-holds.

The interconnection of these components is shown in the diagram of the data acquisition portion of a computerized feedback control system in Figure 1.

The input to the system is a physical parameter such as temperature, pressure, flow, acceleration, and position, which are analog quantities. The parameter is first converted into an electrical signal by means of a transducer; once in electrical form, all further processing is done by electronic circuits.

Next, an amplifier boosts the amplitude of the transducer output signal to a useful level for further processing (*see also* **Article 133, Amplifiers and Filters for DAS, Volume 3**). Transducer outputs may be microvolt or millivolt level signals, which are then amplified to 1 to 10 V levels. Furthermore, the transducer output may be a high-impedance signal, a differential signal with common-mode noise, a current output, a signal superimposed on a high

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

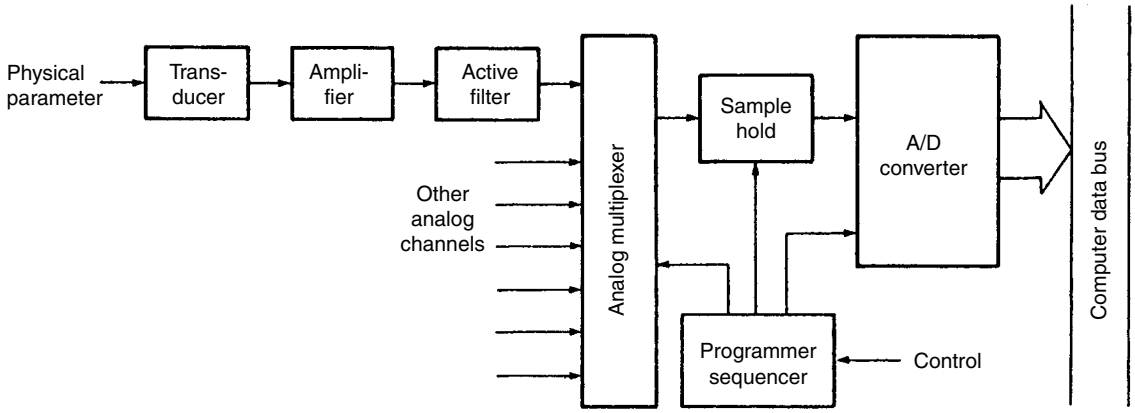


Figure 1. Data acquisition system. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

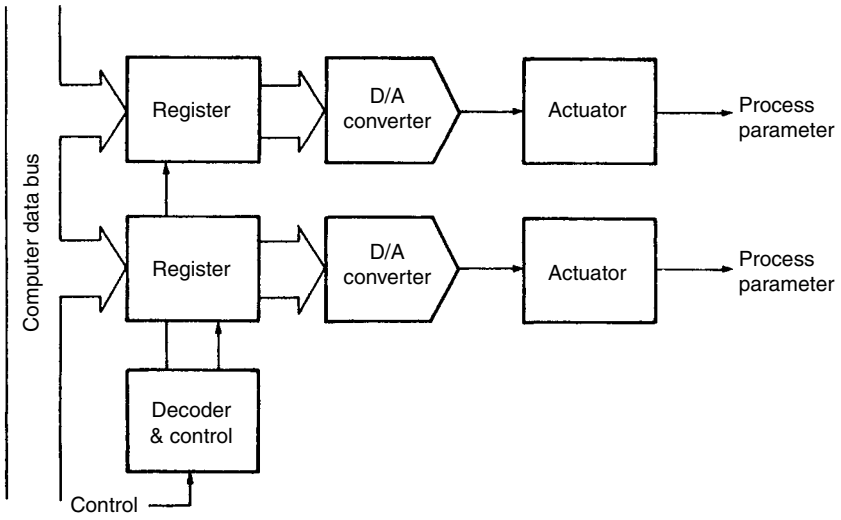


Figure 2. Data distribution system. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

voltage, or a combination of these. The amplifier, in order to convert such signals into a high-level voltage, may be one of several specialized types.

The amplifier is frequently followed by a low-pass active filter that reduces high-frequency signal components, unwanted electrical interference noise, or electronic noise from the signal (*see also* **Article 133, Amplifiers and Filters for DAS, Volume 3**). The amplifier is sometimes also followed by a special nonlinear analog function circuit that performs a nonlinear operation on the high-level signal. Such operations include squaring, multiplication, division, rms conversion, log conversion, or linearization.

The processed analog signal next goes to an analog multiplexer, which switches sequentially between a number of different analog input channels (*see also* **Article 134, Analog Multiplexers, Volume 3**). Each input is in turn connected to the output of the multiplexer for a specified period of time by the multiplexer switch. During this connection time, a sample-hold circuit acquires the signal voltage and then holds its value while an A/D converter converts the value into digital form. The resultant digital word goes to a computer data bus or to the input of a digital circuit.

Thus the analog multiplexer, together with the sample-hold, time shares the A/D converter with a number of analog input channels (*see also* **Article 135, Sample-hold Circuits, Volume 3**). The timing and control of the complete DAS is done by a digital circuit called a *programmer-sequencer*, which in turn is under the control of the computer. In some cases, the computer itself may control the entire DAS.

While this is perhaps the most commonly used DAS configuration, there are alternative ones. Instead of multiplexing high-level signals, low-level multiplexing

is sometimes used with the amplifier following the multiplexer. In such cases, just one amplifier is required, but its gain may have to be changed from one channel to the next during multiplexing. Another method is to amplify and convert the signal into digital form at the transducer location and send the digital information in serial form to the computer. Here, the digital data must be converted to parallel form and then multiplexed onto the computer data bus.

3 BASIC DATA DISTRIBUTION SYSTEM

The data distribution portion of a feedback control system, illustrated in Figure 2, is the reverse of the data acquisition system. The computer, based on the inputs of the data acquisition system, must close the loop on a process and control it by means of output control functions. These control outputs are in digital form and must, therefore, be converted into analog form in order to drive the process. The conversion is accomplished by a series of D/A converters as shown (also often called DAC's). Each D/A converter is coupled to the computer data bus by means of a register, which stores the digital word until the next update. The registers are activated sequentially by a decoder and control circuit, which is under computer control.

The D/A converter outputs then drive actuators that directly control the various process parameters such as temperature, pressure, and flow. Thus, the loop is closed on the process and the result is a complete automatic process control system under computer control.

FURTHER READING

The general principles of DAS are found explained in various titles.

Austerlitz, H. (2003) *Data Acquisition Techniques Using PCs*, Academic Press, San Diego.

Mihura, B. (2001) *LabVIEW for Data Acquisition*, Prentice Hall PTR, Upper Saddle River NJ.

James, K. (2000) *PC Interfacing and Data Acquisition: Techniques for Measurement, Instrumentation and Control*, Newnes, Oxford.

Park, J. and MacKay, S. (2003) *Practical Data Acquisition for Instrumentation and Control Systems*, Newnes, Oxford.

Taylor, H.R. and Taylor, J.R. (1997) *Data Acquisition for Sensor Systems*, Chapman & Hall, London.

This article is based in part on a chapter by E.L. Zuch originally published in Handbook of Measurement Science, P.H. Sydenham and R. Thorn, Wiley, 1982–92.

133: Amplifiers and Filters for DAS

Gerd Wöstenkühler
Hochschule Harz, Wernigerode, Germany

1 Operational and Instrumentation Amplifiers 940
2 Common-mode Rejection Principle and Application 940
3 Specialized Amplifiers Used in DAS 941
4 Filters for Use with DAS 941
Further Reading 942

1 OPERATIONAL AND INSTRUMENTATION AMPLIFIERS

The front end of a data acquisition system extracts the desired analog signal from a physical parameter by means of a transducer and then amplifies and filters it. An amplifier and filter are critical components in this initial signal processing. The amplifier must perform one or more of the following functions: boost the signal amplitude, buffer the signal, convert a signal current into a voltage, or extract a differential signal from common-mode noise.

To accomplish these functions, requires a variety of different amplifier types. The most popular type of electronic amplifier is an operational amplifier (opamp), which is a general purpose gain block having differential inputs. The opamp may be connected in many different closed-loop configurations to make it provide a range of functions; a few are shown in Figure 1. The gain and bandwidth of the circuits shown depend on the external resistors connected around the amplifier. An operational amplifier is a good choice in general where a single-ended signal is to be amplified, buffered, or converted from current to voltage.

In the case of differential signal processing, the instrumentation amplifier is a better choice since it maintains high impedance at both its differential inputs and the gain is set by a precision resistor located elsewhere in the amplifier circuit. One type of instrumentation amplifier circuit is shown in Figure 2. Notice that no gain-setting resistors are connected to either of the input terminals. Instrumentation amplifiers have the following important characteristics:

- high impedance differential inputs;
- low input offset voltage drift;
- low input bias currents;
- gain easily set by means of one or two external resistors;
- high common-mode rejection ratio.

2 COMMON-MODE REJECTION PRINCIPLE AND APPLICATION

The common-mode rejection ratio CMRR is an important parameter of differential amplifiers. An ideal differential input amplifier responds only to the voltage difference between its input terminals and does not respond at all to any voltage that is common to both input terminals (common-mode voltage). In nonideal amplifiers, however, the common-mode input signal causes some output response even though small compared to the response to a differential input signal.

The common-mode rejection ratio is defined as the ratio of differential and common-mode responses. The CMRR of an amplifier is the ratio of differential voltage gain to common-mode voltage gain and is generally expressed in dB:

$$\mathrm{CMRR} = 20 \log_{10} \left(\frac{A_{\mathrm{D}}}{A_{\mathrm{CM}}} \right) \tag{1}$$

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

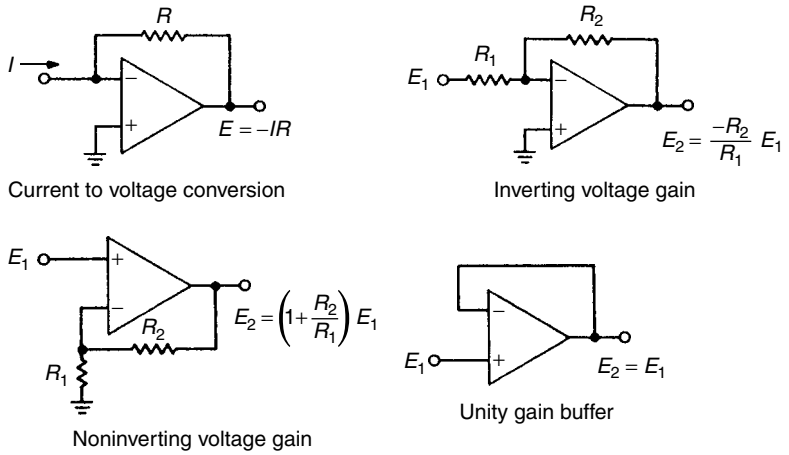


Figure 1. Operational amplifier configurations. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

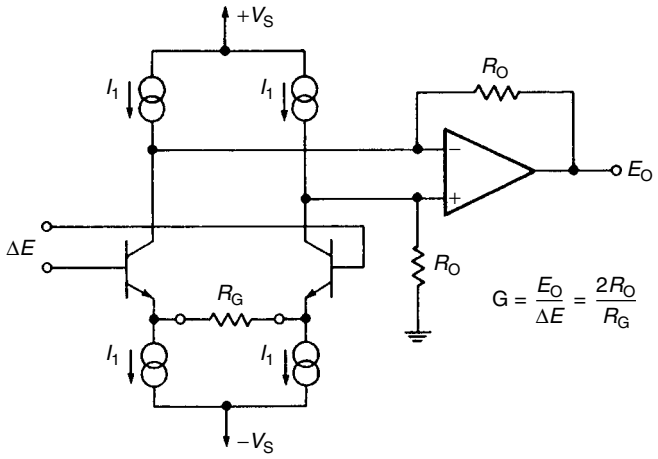


Figure 2. Simplified instrumentation amplifier circuit. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

where A_D is differential voltage gain and A_{CM} is common-mode voltage gain. CMRR is a function of frequency and, therefore, also a function of the impedance balance between the two amplifier's input terminals. At even moderate frequencies, CMRR can be significantly degraded by small unbalances in the source series resistance and shunt capacitance.

3 SPECIALIZED AMPLIFIERS USED IN DAS

There are several other special amplifiers that are useful in conditioning the input signal in a data acquisition system.

An isolation amplifier is used to amplify a differential signal, which is superimposed on a very high common-mode voltage, perhaps of several hundred or even of several thousand volts. The isolation amplifier has the characteristics of an instrumentation amplifier with a very high common-mode input voltage capability. Another special amplifier, the chopper stabilized amplifier, is used to accurately amplify microvolt level signals to the required amplitude. This amplifier employs a special switching stabilizer, which gives extremely low input offset voltage drift. Another useful device, the electrometer amplifier, has ultralow input bias currents, generally less than one picoampere and is used to convert extremely small signal currents into a high-level voltage.

4 FILTERS FOR USE WITH DAS

A low-pass filter frequently follows the signal processing amplifier to reduce signal noise. Low-pass filters are used for the following reasons: to reduce man-made electrical interference noise, to reduce electronic noise, and to limit the bandwidth of the analog signal to less than half the sampling frequency in order to eliminate frequency folding. When used for the last reason, the filter is called a *presampling filter* or an *antialiasing filter* (*see also* **Article 138, Sampling Theory Relevant to DAS, Volume 3**).

Man-made electrical noise is generally periodic, as, for example, in power line interference, and is sometimes reduced by means of a special filter such as a notch filter. Electronic noise, on the other hand, is random noise with noise power proportional to bandwidth and is present in transducer resistances, circuit resistances, and in amplifiers themselves. It is reduced by limiting the bandwidth of the

system to the minimum required to pass the desired signal components (*see also* **Article 181, Amplitude Modulated Signals: The Lock-in Amplifier, Volume 3**; **Article 180, Bandwidth Reduction of Baseband DC Signals, Volume 3**; **Article 179, Input Connections; Grounding and Shielding, Volume 3**; **Article 176, Signals and Signal-to-noise Ratio, Volume 3**; **Article 178, Noise Matching and Preamplifier Selection, Volume 3**; and **Article 121, Signals in the Presence of Noise, Volume 2**).

Grounding and shielding of amplifiers is an important topic in amplifier design – *see* **Article 177, Grounding and Shielding, Volume 3**.

No filter does a perfect job of eliminating noise or other undesirable frequency components, and therefore the choice of a filter is always a compromise. Ideal filters, frequently used as analysis examples, have flat passband response with infinite attenuation at the cutoff frequency, but are mathematical filters only and not physically realizable.

In practice, the systems engineer has a choice of cutoff frequency and attenuation rate. The attenuation rate and resultant phase response depend on the particular filter characteristic and the number of poles in the filter function. Some of the more popular filter characteristics used include the Butterworth, Chebyshev, Bessel, and elliptic alternatives. In making this choice, the effect of overshoot and nonuniform phase delay must be carefully considered. Figure 3 illustrates some practical low-pass filter response characteristics (*see also* **Article 125, Outline of Purpose of Analog Data Filters, Volume 2**).

Passive RLC filters are seldom used in signal processing applications today chiefly due to the undesirable characteristics of inductors. Active filters are generally used now since they permit the filter characteristics to be accurately set by precision and stable resistors and capacitors. Inductors, with their undesirable saturation and temperature drift characteristics, are thereby eliminated. Also, because active filters use operational amplifiers, the problems of insertion loss and output loading are also eliminated.

Filters are often implemented in digital form. These are based on the analog design concepts but the digitization, which introduced sampling issues, increases the complexity of operation considerably. Fortunately, the user is buffered from the mathematics and usually can work from data sheets to select hardware and software settings for the IC modules provided.

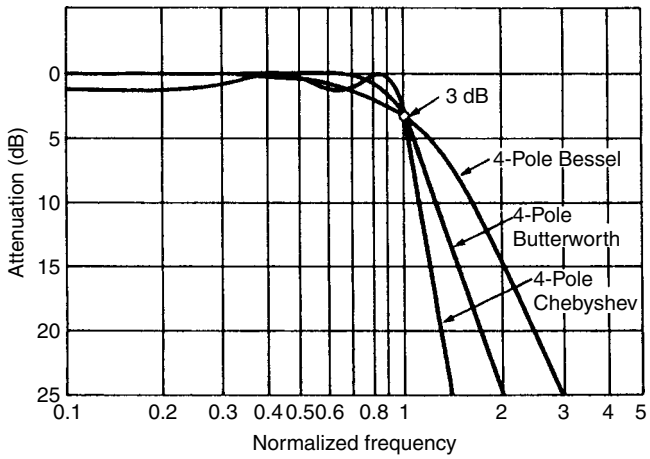


Figure 3. Some practical low-pass filter characteristics. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

FURTHER READING

Electronic operational amplifiers are covered as a single topic in many titles; the following may be useful.

Coughlin, R.F. and Driscoll, F.F. (2001) *Operational Amplifiers and Linear Integrated Circuits*, Prentice Hall College Division, Upper Saddle River, NJ.

Franco, S. (2001) *Design with Operational Amplifiers and Analog Integrated Circuits*, McGraw-Hill, Dubuque, Iowa.

Jerald, G., Graeme, J.G. and Graeme, J.E. (1997) *Optimizing Op Amp Performance*, McGraw-Hill, New York.

McGraw-Hill Science/Engineering/Math.

Mancini, R. (2003) *Op Amps for Everyone*, Newnes, Boston.

Morrison, R. (1970) *DC Amplifiers in Instrumentation*, Wiley-Interscience, New York.

Stanley, W.D. (2002) *Operational Amplifiers with Linear Integrated Circuits*, Prentice Hall, Upper Saddle River, NJ.

This article is based in part on a chapter by E.L. Zuch originally published in Handbook of Measurement Science Vol 1, P.H. Sydenham, Wiley, 1982.

134: Analog Multiplexers

Gerd Wöstenkühler

Hochschule Harz, Wernigerode, Germany

1 Analog Multiplexer Operation 943
2 Analog Multiplexer Characteristics 943
3 Analog Multiplexer Applications 944
Further Reading 945

1 ANALOG MULTIPLEXER OPERATION

Analog multiplexers are the circuits that time-share an analog-to-digital (A/D) converter among a number of different analog channels. Since the A/D converter in many cases is the most expensive component in a data acquisition system, multiplexing analog inputs to the A/D converter is an economical approach. Usually, the analog multiplexer operates in a sample-hold circuit (*see also* **Article 135, Sample-hold Circuits, Volume 3**), which holds the required analog voltage long enough for A/D conversion.

As shown in Figure 1, an analog multiplexer consists of an array of parallel electronic switches connected to a common output line. Only one switch is turned on at a time. Popular switch configurations include 4, 8, and 16 channels that are connected in single single-ended or dual differential configurations.

The multiplexer also contains a decoder-driver circuit that decodes a binary input word and turns on the appropriate switch. This circuit interfaces with standard TTL inputs and drives the multiplexer switches with the proper control voltages. For the 8-channel analog multiplexer shown, a one-of-eight decoder circuit is used.

Most analog multiplexers today employ the CMOS switch circuits based on the one shown in Figure 2. A CMOS driver controls the gates of the parallel-connected P-channel and N-channel MOSFETs. Both switches turn on together with the parallel connection giving relatively uniform on-resistance over the required analog input voltage range. The resulting on-resistance may vary from about 50 Ω to 2 kΩ depending on the multiplexer; this resistance increases with temperature.

2 ANALOG MULTIPLEXER CHARACTERISTICS

Because of the series resistance, it is common practice to operate an analog multiplexer into a very high load resistance such as the input of a unity gain buffer amplifier shown in Figure 1. The load impedance must be large compared with the switch on-resistance and any series source resistance in order to maintain high transfer accuracy. Transfer error is the input to output error of the multiplexer with the source and load connected; error is expressed as a percent of input voltage.

Transfer errors of 0.1 to 0.01% or less are required in most data acquisition systems. This is readily achieved by using operational amplifier buffers with typical input impedances from 10^8 to 10^{12} Ω. Many sample-hold circuits also have very high input impedances.

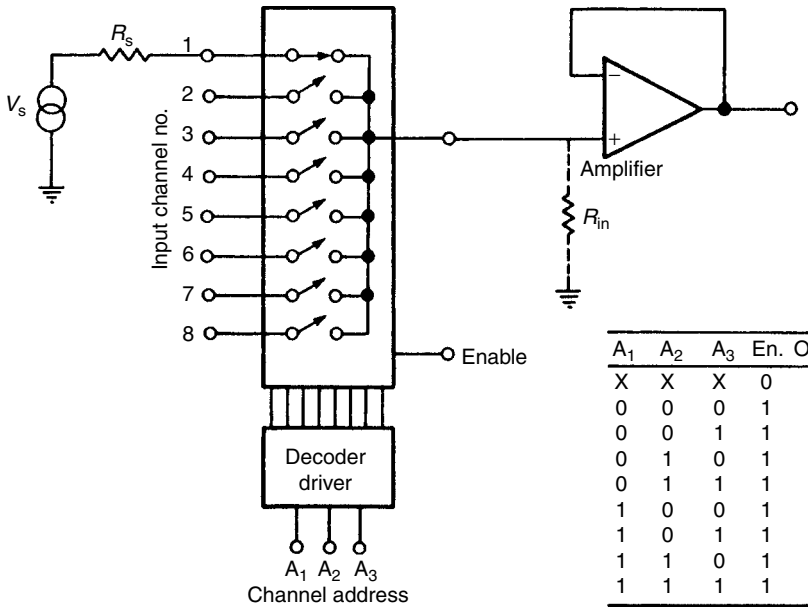
Another important characteristic of analog multiplexers is break-before-make switching. There is a small time delay between disconnection from the previous channel and connection to the next channel, which assures that two adjacent input channels are never instantaneously connected together.

Settling time is another important specification for analog multiplexers; it is the same definition given for amplifiers

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.

© 2005 John Wiley & Sons, Ltd.

ISBN: 0-470-02143-8.



A_1	A_2	A_3	En.	On-chan
X	X	X	0	None
0	0	0	1	1
0	0	1	1	2
0	1	0	1	3
0	1	1	1	4
1	0	0	1	5
1	0	1	1	6
1	1	0	1	7
1	1	1	1	8

Figure 1. Analog multiplexer circuit. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

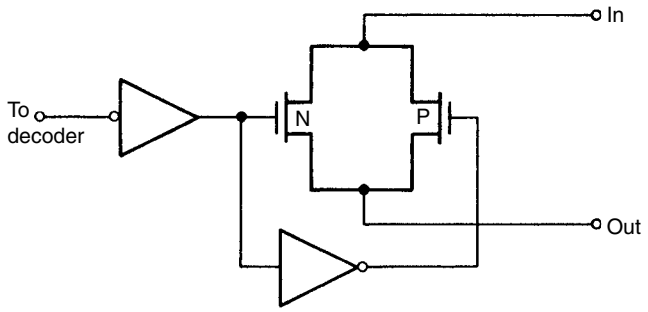


Figure 2. CMOS analog switch circuit. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

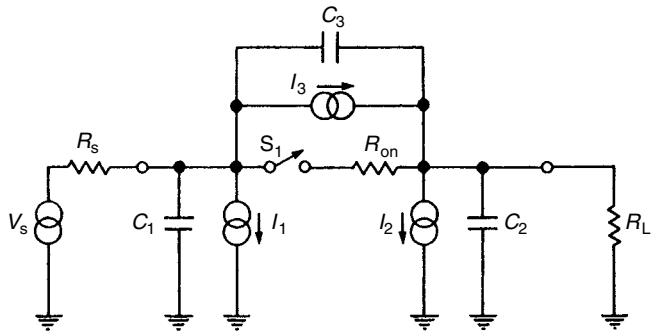


Figure 3. Equivalent circuit of analog multiplexer switch. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

except that it is measured from the time the channel is switched on. Throughput rate is the highest rate at which a multiplexer can switch from channel to channel with the output settling to its specified accuracy. Crosstalk is the ratio of output voltage to input voltage with all channels connected in parallel and off; it is generally expressed as an input to output attenuation ratio in decibels.

As shown in the representative equivalent circuit of Figure 3, analog multiplexer switches have a number of leakage currents and capacitances associated with their operation. These parameters are specified on data sheets and must be considered in the operation of the devices. Leakage currents, generally in picoamperes at room temperature, become troublesome only at high temperatures. Capacitances affect crosstalk and settling time of the multiplexer.

3 ANALOG MULTIPLEXER APPLICATIONS

Analog multiplexers are employed in two basic types of operation; low level and high level. In high-level multiplexing, the most popular type, the analog signal is amplified to the 1 to 10 V range ahead of the multiplexer. This has the advantage of reducing the effects of noise on the signal during the remaining analog processing. In low-level multiplexing, the signal is amplified after multiplexing; therefore, great care must be exercised in handling the low-level signal up to the multiplexer. Low-level multiplexers generally use two-wire differential switches in order to minimize noise pickup. Reed relays, because of essentially zero series

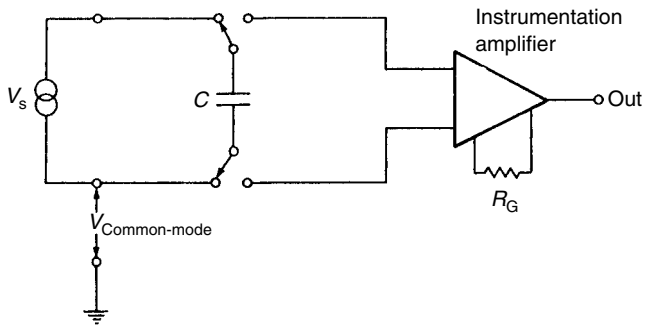


Figure 4. Flying-capacitor multiplexer switch. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

resistance and absence of switching spikes are frequently employed in low-level multiplexing systems. They are also useful for high common-mode voltages.

A useful specialized analog multiplexer is the flying-capacitor type. This circuit, shown as a single channel in Figure 4 has differential inputs and is particularly useful with high common-mode voltages. The capacitor connects first to the differential analog input, charging up to the input voltage, and is then switched to the differential output, which goes to a high input impedance instrumentation amplifier. The differential signal is, therefore, transferred to the amplifier input without the common mode voltage and is then further processed up to A/D conversion.

In order to realize large numbers of multiplexed channels, it is possible to connect analog multiplexers in parallel using the enable input to control each device. This is called *single-level multiplexing*. The output of several multiplexers can also be connected to the inputs of another to expand the number of channels; this method is *double-level multiplexing*.

FURTHER READING

The general principles of DAS are found explained in various titles. Multiplexing is usually covered in these books.

Austerlitz, H. (2003) *Data Acquisition Techniques Using PCs*, Academic Press, San Diego.

James, K. (2000) *PC Interfacing and Data Acquisition: Techniques for Measurement, Instrumentation and Control*, Newnes, Oxford.

Mihura, B. (2001) *LabVIEW for Data Acquisition*, Prentice Hall PTR, Upper Saddle River, NJ.

Park, J. and MacKay, S. (2003) *Practical Data Acquisition for Instrumentation and Control Systems*, Newnes, Oxford.

Taylor, H.R. and Taylor, J.R. (1997) *Data Acquisition for Sensor Systems*, Chapman & Hall, London.

This article is based in part on a chapter by E.L. Zuch originally published in Handbook of Measurement Science Vol 1, P.H. Sydenham, Wiley, 1982.

135: Sample-hold Circuits

Gerd Wöstenkühler
Hochschule Harz, Wernigerode, Germany

1 Role and Operation of Sample-hold Circuit in DAS 946
2 Sample-hold Unit As a Data Recovery Filter 946
3 Specialized Sample-hold Circuits 947
4 Sample-hold Characteristics 947
Further Reading 948

1 ROLE AND OPERATION OF SAMPLE-HOLD CIRCUIT IN DAS

Sample-hold circuits are the devices that store analog information and reduce the aperture time of an analog-to-digital (A/D) converter (*see also* **Article 138, Sampling Theory Relevant to DAS, Volume 3**; **Article 141, Digital-to-Analog (D/A) Converters, Volume 3**).

A sample hold is simply a voltage-memory device in which an input voltage is acquired and then stored on a high quality capacitor. A popular circuit is shown in Figure 1.

A_1 is an input buffer amplifier with a high input impedance so that the source, which may be an analog multiplexer (*see also* **Article 134, Analog Multiplexers, Volume 3**), is not loaded. The output of A_1 must be capable of driving the hold capacitor with stability and enough drive current to charge it rapidly. S is an electronic switch, generally field effect transistor, FET, which is rapidly switched on or off by a driver circuit that interfaces with transistor transistor logic (TTL) inputs.

C is a capacitor with low leakage and low dielectric absorption characteristics; it is a polystyrene, polycarbonate, polypropylene, or Teflon type. In the case of hybrid sample holds, the metal oxide semiconductor, MOS type capacitor is frequently used.

A_2 is the output amplifier that buffers the voltage on the hold capacitor. It must, therefore, have extremely low input bias current, and for this reason a FET input amplifier is required.

There are two modes of operation for a sample hold: sample mode or tracking mode, when the switch is closed; and hold mode, when the switch is open. Sample holds are usually operated in one of two basic ways. The device can continuously track the input signal and be switched into the hold mode only at certain specified times, spending most of the time in tracking mode. This is the case for a sample hold being employed as a deglitcher at the output of a digital-to-analog (D/A) converter, for example, *see* **Article 141, Digital-to-Analog (D/A) Converters, Volume 3**.

Alternatively, the device can stay in the hold mode most of the time and go to the sample mode just to acquire a new input signal level. This is the case for a sample hold used in a data acquisition system following the multiplexer (*see also* **Article 132, Data Acquisition Systems (DAS) in General, Volume 3**).

2 SAMPLE-HOLD UNIT AS A DATA RECOVERY FILTER

A common application for sample-hold circuits is in data recovery, or signal reconstruction, filters. The problem is to reconstruct a train of analog samples into the original signal; when used as a recovery filter, the sample hold is known as a *zero-order hold*. It is a useful filter because it fills in the space between samples, providing data smoothing.

As with other filter circuits, the gain and phase components of the transfer function are of interest. By an analysis

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

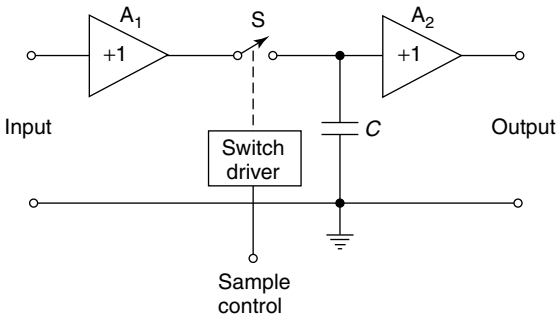


Figure 1. Popular sample-hold circuit. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

based on the impulse response of a sample hold and use of the Laplace transform, the transfer function is found to be

$$G_0(f) = \frac{1}{f_s} \frac{\sin\left(\dfrac{\pi \times f}{f_s}\right)}{\dfrac{\pi \times f}{f_s}} e^{\left(-j\frac{\pi \times f}{f_s}\right)} \qquad (1)$$

where f_s is the sampling frequency. This function contains the familiar $(\sin x)/x$ term plus a phase term, both of which are plotted in Figure 2.

The sample hold is, therefore, a low-pass filter with a cutoff frequency slightly less than $f_s/2$ and a linear phase that results in a constant delay time of $T/2$, where T is the time between samples. Notice that the gain function also has significant response lobes beyond f_s. For this reason, a sample-hold reconstruction filter is frequently followed by another conventional low-pass filter.

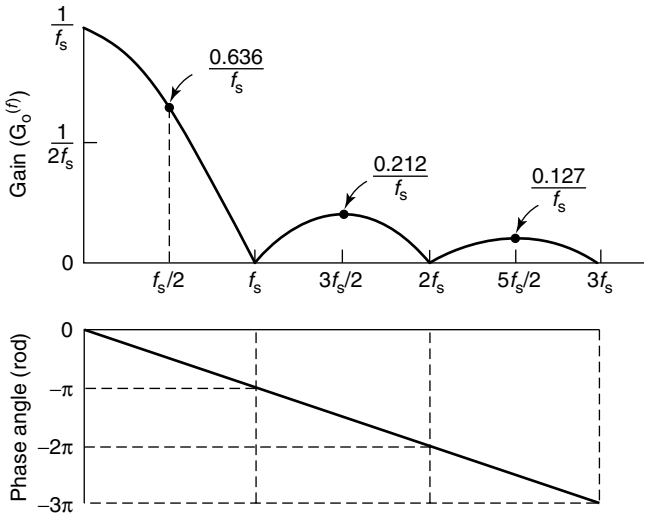


Figure 2. Gain and phase components of zero-order hold transfer function. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

3 SPECIALIZED SAMPLE-HOLD CIRCUITS

In addition to the basic circuit of Figure 1, several other sample-hold circuit configurations are frequently used. Figure 3 shows two such circuits that are closed-loop circuits as contrasted with the open-loop circuit of Figure 1. Figure 3(a) uses an operational integrator and another amplifier to make a fast, accurate inverting sample hold. A buffer amplifier is sometimes added in front of this circuit to give high input impedance. Figure 3(b) shows a high input impedance, noninverting sample-hold circuit.

The circuit in Figure 1, although generally not as accurate as those in Figure 3, can be used with a diode-bridge switch to realize ultrafast acquisition sample holds.

4 SAMPLE-HOLD CHARACTERISTICS

A number of parameters are important in characterizing sample-hold performance. Probably the most important of these is the acquisition time. The definition is similar to the definition of settling time for an amplifier (*see also* **Article 133, Amplifiers for DAS, Volume 3**). It is the time required, after the sample command is given, for the hold capacitor to charge to a full-scale

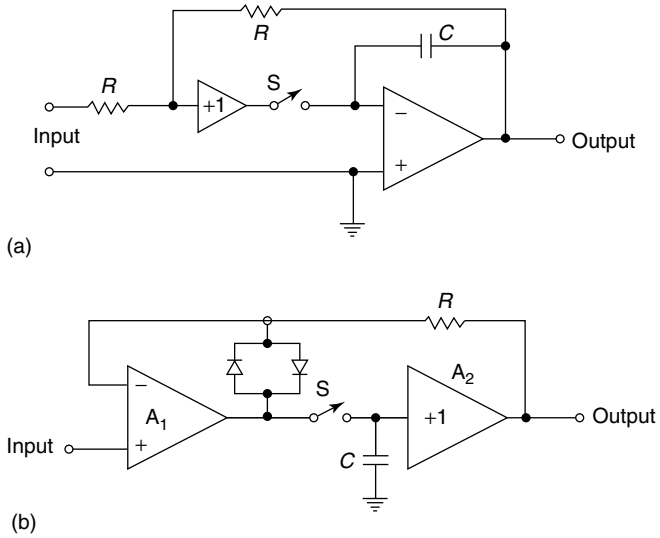


Figure 3. Accurate closed-loop sample-hold circuits: (a) inverting and (b) noninverting. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

voltage change and remain within a specified error band around final value.

Several hold-mode specifications are also important. Hold-mode droop is the output voltage change per unit time when the sample switch is open. This droop is caused by the leakage currents of the capacitor and switch, and the output amplifier bias current. Hold-mode feedthrough is the percentage of input signal transferred to the output when the sample switch is open. It is measured with a sinusoidal input signal and caused by capacitive coupling.

The most critical phase of sample-hold operation is the transition from the sample mode to the hold mode. Several important parameters characterize this transition. Sample-to-hold offset or step error is the change in output voltage from the sample mode to the hold mode, with a constant input voltage. It is caused by the switch still transferring some charge onto the hold capacitor as it turns off.

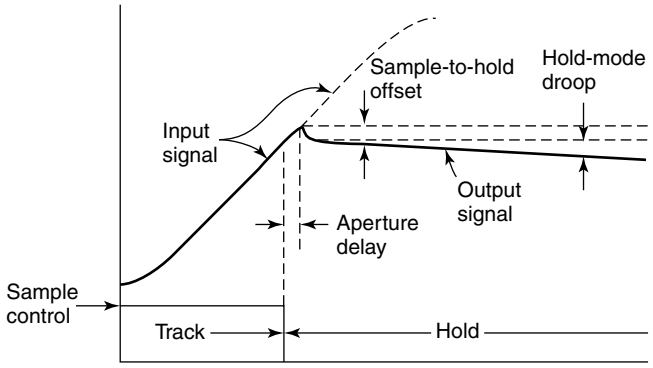


Figure 4. Some sample-hold characteristics. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

Aperture delay is the time elapsed from the hold command to when the switch actually opens; it is generally much less than a microsecond.

Aperture uncertainty or aperture jitter is the time variation, from sample-to-sample, of the aperture delay. It is the limit on how precise is the point in time of opening the switch. Aperture uncertainty is the time used to determine the aperture error due to rate of change of the input signal (*see also* **Article 138, Sampling Theory Relevant to DAS, Volume 3**).

Several of the above specifications are illustrated in the diagram of Figure 4.

Sample-hold circuits are simple in concept, but generally difficult to fully understand and apply. Their operation is full of subtleties, and they must, therefore, be carefully selected and then tested in a given application. Their simplicity is a strong reason they are not always supplanted by a digital storage device alternative.

FURTHER READING

The general principles of DAS are found explained in various titles. As the sample and hold device is usually needed in a DAS, the general titles usually cover them.

Austerlitz, H. (2003) *Data Acquisition Techniques Using PCs*, Academic Press, San Diego.

James, K. (2000) *PC Interfacing and Data Acquisition: Techniques for Measurement, Instrumentation and Control*, Newnes, Oxford.

Mihura, B. (2001) *LabVIEW for Data Acquisition*, Prentice Hall PTR, Upper Saddle River, NJ.

Park, J. and MacKay, S. (2003) *Practical Data Acquisition for Instrumentation and Control Systems*, Newnes, Oxford.

Taylor, H.R. and Taylor, J.R. (1997) *Data Acquisition for Sensor Systems*, Chapman & Hall, London.

This article is based in part on a chapter by E.L. Zuch originally published in Handbook of Measurement Science Vol 1, P.H. Sydenham, Wiley, 1982.

136: Quantizing Theory Relevant to DAS

Gerd Wöstenkühler
Hochschule Harz, Wernigerode, Germany

1 Quantizer Transfer Function 949
2 Quantizer Resolution and Error 949
Further Reading 950

1 QUANTIZER TRANSFER FUNCTION

The nonlinear transfer function shown in Figure 1 is that of an ideal quantizer with eight output states; with output code words assigned, it is also that of a 3-bit A/D converter. The eight output states are assigned the sequence of binary numbers from 000 through to 111. The analog input range for this quantizer is 0 to +10 V.

There are several important points concerning the transfer function of Figure 1.

First, the resolution of the quantizer is defined as the number of output states expressed in bits; in this case it is a simple 3-bit quantizer.

The number of output states for a binary coded quantizer is 2^n, where n is the number of bits. Thus, an 8-bit quantizer has 256 output states and a 12-bit quantizer has 4096 output states.

As shown in the diagram, there are $2^n - 1$ analog decision points (or threshold levels) in the transfer function. These points are at voltages of +0.625, +1.875, +3.125, +4.375, +5.625, +6.875, and +8.125 V. The decision points must be precisely set in a quantizer in order to divide the analog voltage range into the correct quantizer values.

The voltages 0.00, +1.25, +2.50, +3.75, +5.00, +6.25, +7.50, and +8.75 V are the center points of each output code word. The analog decision point voltages are precisely halfway between the code word center points. The quantizer staircase function is the best approximation that can be made to a straight line drawn through the origin and full-scale point; notice that the line passes through all of the code word center points.

2 QUANTIZER RESOLUTION AND ERROR

At any part of the input range of the quantizer, there is a small range of analog values within which the same output code word is produced. This small range is the voltage difference between any two adjacent decision points and is known as the *analog quantization size*, or *quantum*, Q. In Figure 1, the quantum is 1.25 V and is found in general by dividing the full-scale analog range by the number of output states. Thus,

$$Q = \frac{\text{FSR}}{2^n} \tag{1}$$

where FSR is the full-scale range, or 10 V in this case. Q is the smallest analog difference that can be resolved, or distinguished, by the quantizer. In the case of a 12-bit quantizer, the quantum is much smaller and is found to be

$$Q = \frac{\text{FSR}}{2^n} = \frac{10\,\text{V}}{4096} = 2.44\,\text{mV} \tag{2}$$

If the quantizer input is moved through its entire range of analog values and the difference between output and input is taken, a sawtooth error function results, as shown in Figure 1. This function is called the *quantizing error* and it is the irreducible error that results from the quantizing process. It can be reduced only by increasing the number of output states (or the resolution) of the quantizer, thereby

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

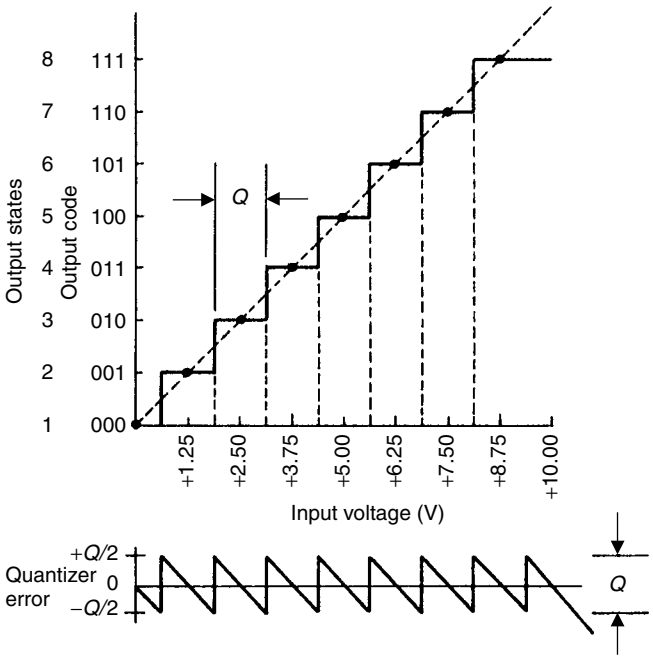


Figure 1. Transfer function and error of ideal 3-bit quantizer. (*Source*: reproduced by permission of Sydenham, P.H. (1983) *Handbook of Measurement Science*, Volume 1 Theoretical Fundamentals. John Wiley & Sons Ltd., Chichester.)

making the quantization finer and the system increasingly sophisticated.

For a given analog input value of the quantizer, the output error will vary anywhere from 0 to $\pm Q/2$; the error is zero only at analog values corresponding to the code center points. This error is also frequently called *quantization uncertainty* or *quantization noise*.

The quantizer output can be thought of as the analog input with quantization noise added to it. The noise has a peak-to-peak value of Q but, as with other types of noise, the average value is zero. Its rms value, however, is useful in analysis and can be computed from the triangular waveshape to be $Q/(2 \times \sqrt{3})$.

FURTHER READING

The general principles of DAS are explained in various titles. Quantization is usually covered at different depths of explanation.

Austerlitz, H. (2002) *Data Acquisition Techniques Using PCs*, Academic Press.

Mihura, B. (2001) *LabVIEW for Data Acquisition*, Prentice Hall PTR.

James, K. (2000) *PC Interfacing and Data Acquisition: Techniques for Measurement, Instrumentation and Control*, Newnes.

Park, J. and MacKay, S. (2003) *Practical Data Acquisition for Instrumentation and Control Systems*, Newnes.

Taylor, H.R. and Taylor, J.R. (1997) *Data Acquisition for Sensor Systems*, Kluwer Academic Publishers.

This article is based in part on a chapter by E.L. Zuch originally published in Handbook of Measurement Science Vol 1, P.H. Sydenham, Wiley, 1982.

137: Coding for Data Converters

Gerd Wöstenkühler
Hochschule Harz, Wernigerode, Germany

1 Natural Binary Code 951
2 Other Binary Codes 952
3 Popular Bipolar Codes Used in Data Converters 953
4 Binary Coded Decimal (BCD) Codes 953
Further Reading 954

1 NATURAL BINARY CODE

Analog-to-digital (A/D) and digital-to-analog (D/A) converters interface with digital systems by means of an appropriate digital code. While there are many possible codes to select, a few standard ones are almost exclusively used with data converters.

The most popular code is natural binary, or straight binary, which is used in its fractional form to represent a number

$$N = a_1 2^{-1} + a_2 2^{-2} + a_3 2^{-3} + \cdots + a_n 2^{-n} \qquad (1)$$

where each coefficient a assumes a value of zero or one. N has a value between zero and one (*see also* **Article 33, Coding Theory and its Application to Measurement, Volume 1**).

A binary fraction is normally written as 0.110101, but with data converter codes, the decimal point is omitted and the code word is written as 110101. This code word represents a fraction of the full-scale value of the converter and has no other numerical significance.

The binary code word 110101, therefore, represents the decimal fraction

$$\begin{aligned}(1 \times 0.5) + (1 \times 0.25) + (0 \times 0.125) + (1 \times 0.0625) \\ + (0 \times 0.03125) + (1 \times 0.015625) \\ = 0.828125 \text{ or } 82.8125\% \end{aligned} \qquad (2)$$

of full scale for the converter. If full scale is +10 V, then the code word represents +8.28125 V. The natural binary code belongs to a class of codes known as *positive-weighted codes* since each coefficient has a specific weight none of which is negative.

The leftmost bit has the most weight, 0.5 of full scale, and is called the *most significant bit* (MSB); the rightmost bit has the least weight, 2^{-n} of full scale, and is, therefore, called the *least significant bit* (LSB). The bits in a code word are numbered left to right from 1 to n.

The LSB has the same analog equivalent value as Q (*see* **Article 136, Quantizing Theory Relevant to DAS, Volume 3**), namely,

$$\text{LSB(analog value)} = \frac{\text{FSR}}{2^n} \qquad (3)$$

Table 1 is a useful summary of the resolution, number of states, LSB weights, and dynamic range (DR) for data converters from 1 to 20 bits resolution.

The dynamic range of a data converter in decibels (dB) is found as follows:

$$\begin{aligned}\text{DR} &= 20 \times \log 2^n = 20 \times n \times \log 2 \approx 20 \times n \times 0.301 \\ &= 6.02 \times n \end{aligned} \qquad (4)$$

where DR is dynamic range, n is the number of bits, and 2^n is the number of states of the converter. Since 6.02 dB corresponds to a factor of 2, it is simply necessary to multiply the resolution of a converter in bits by 6.02. A 12-bit converter, for example, has a dynamic range of 72.24 dB.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

Table 1. Resolution, number of states, LSB weight, and dynamic range for data converters.

Resolution bits n	Number of states 2^n	LSB weight 2^{-n}	Dynamic range (dB)
0	1	1	0.0
1	2	0.5	6.0
2	4	0.25	12.0
3	8	0.125	18.1
4	16	0.0625	24.1
5	32	0.03125	30.1
6	64	0.015625	36.1
7	128	0.0078125	42.1
8	256	0.00390625	48.2
9	512	0.001953125	54.2
10	1024	0.0009765625	60.2
11	2048	0.00048828125	66.2
12	4096	0.000244140625	72.2
13	8192	0.0001220703125	78.3
14	16 384	0.00006103515625	84.3
15	32 768	0.000030517578125	90.3
16	65 536	0.0000152587890625	96.3
17	131 072	0.00000762939453125	102.3
18	262 144	0.000003814697265625	108.4
19	524 288	0.0000019073486328125	114.4
20	1 048 576	0.00000095367431640625	120.4

Source: Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.

An important point to notice is that the maximum value of the digital code, namely all 1's, does not correspond with analog full scale, but rather with one LSB less than full scale, or FS $(1-2^{-n})$. Therefore, a 12-bit converter with a 0 to +10 V analog range has a maximum code of 1111 1111 1111 and a maximum analog value of $+10\,\text{V}\ (1-2^{-12}) \approx +9.99756\,\text{V}$. In other words, the maximum analog value of the converter, corresponding to all 1's in the code, never quite reaches the point defined as analog full scale.

2 OTHER BINARY CODES

Several other binary codes are used with A/D and D/A converters in addition to straight binary. These codes are offset binary, two's complement, binary coded decimal (BCD), and their complemented versions. Each code has a specific advantage in certain applications. BCD coding, for example, is used where digital displays must be interfaced such as in digital panel meters and digital multimeters. Two's complement coding is used for computer arithmetic logic operations, and offset binary coding is used with bipolar analog measures.

Not only the digital codes are standardized with data converters but also the analog voltage ranges. Most converters use unipolar voltage ranges of 0 to +5 V and 0 to +10 V although some devices use the negative ranges 0 to −5 V and 0 to −10 V. The standard bipolar voltage ranges are ±2.5 V, ±5 V, and ±10 V. Many converters offered are pin-programmable between these various ranges.

Table 2 shows straight binary and complementary binary codes for a unipolar 8-bit converter with a 0 to +10 V analog FS range. The maximum analog value of the converter is +9.961 V, or one LSB less than +10 V. Note that the LSB size is 0.039 V as shown near the bottom of the table. The complementary binary coding used in some converters is simply the logic complement of straight binary.

When A/D and D/A converters are used in bipolar operation, the analog range is offset by half scale, or by the MSB value. The result is an analog shift of the converter transfer function as shown in Figure 1. Notice for this 3-bit A/D converter transfer function that the code 000 corresponds with −5 V; 100 with 0 V; and 111 with +3.75 V. Since the

Table 2. Binary coding for 8-bit unipolar converters.

Fraction of FS	+10 V FS	Straight binary	Complementary binary
+FS − 1 LSB	+9.961	1111 1111	0000 0000
+3/4 FS	+7.500	1100 0000	0011 1111
+1/2 FS	+5.000	1000 0000	0111 1111
+1/4 FS	+2.500	0100 0000	1011 1111
+1/8 FS	+1.250	0010 0000	1101 1111
+1 LSB	+0.039	0000 0001	1111 1110
0	0.000	0000 0000	1111 1111

Source: Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.

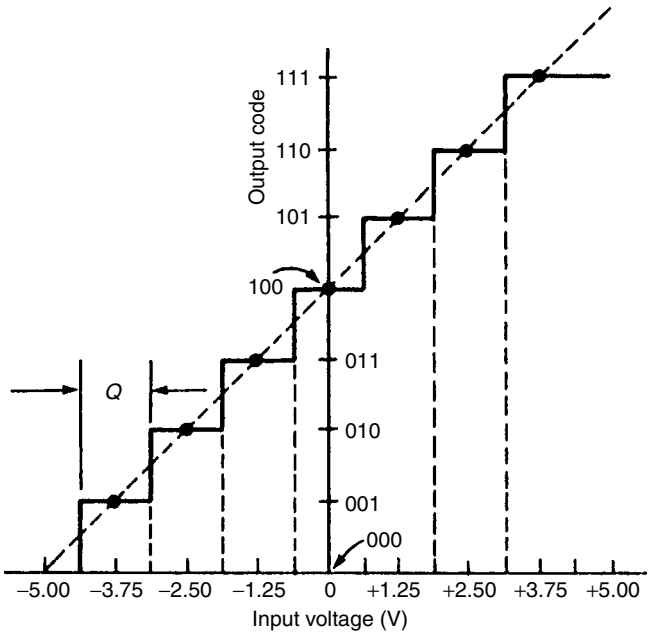


Figure 1. Transfer function for bipolar 3-bit A/D converter. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

Table 3. Popular bipolar codes used with data converters.

Fraction of FS	±5 V FS	Offset binary	Complementary offset binary	Two's complement	Sign-magnitude binary
+FS − 1 LSB	+4.961	1111 1111	0000 0000	0111 1111	1111 1111
+3/4 FS	+3.750	1110 0000	0001 1111	0110 0000	1110 0000
+1/2 FS	+2.500	1100 0000	0011 1111	0100 0000	1100 0000
+1/4 FS	+1.250	1010 0000	0101 1111	0010 0000	1010 0000
0	0.000	1000 0000	0111 1111	0000 0000	1000 0000[a]
−1/4 FS	−1.250	0110 0000	1001 1111	1110 0000	0010 0000
−1/2 FS	−2.500	0100 0000	1011 1111	1100 0000	0100 0000
−3/4 FS	−3.750	0010 0000	1101 1111	1010 0000	0110 0000
−FS + 1 LSB	−4.961	0000 0001	1111 1110	1000 0001	0111 1111
−FS	−5.000	0000 0000	1111 1111	1000 0000	–

Source: Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.
[a]Sign magnitude binary has two code words for zero as shown here:
0+ 1000 0000
0− 0000 0000.

output coding is the same as before the analog shift, it is now appropriately called *offset binary coding*.

3 POPULAR BIPOLAR CODES USED IN DATA CONVERTERS

Table 3 shows the offset binary code together with complementary offset binary, two's complement, and sign-magnitude binary codes. These are the most popular codes employed in bipolar data converters.

The two's complement code has the characteristic that the sum of the positive and negative codes for the same analog magnitude always produces all zeros and a carry. This characteristic makes the two's complement code useful in arithmetic computations. Notice that the only difference between two's complement and offset binary is the complementing of the MSB. In bipolar coding, the MSB becomes the sign bit.

The sign-magnitude binary code, infrequently used, has identical code words for equal magnitude analog values except that the sign bit is different. As shown in Table 3, this code has two possible code words for zero: 1000 0000 or 0000 0000. The two are usually distinguished as 0+ and 0−, respectively. Because of this characteristic, the code has maximum analog values of ±(FS − 1 LSB) and reaches neither analog +FS nor −FS.

4 BINARY CODED DECIMAL (BCD) CODES

Table 4 shows BCD coding and complementary BCD coding for a three-decimal digit data converter. These are the codes used with integrating type A/D converters employed in digital panel meters, digital multimeters, and other

Table 4. BCD and complementary BCD coding.

Fraction of FS	+10 V FS	Binary coded decimal	Complementary BCD
+FS − 1 LSB	+9.99	1001 1001 1001	0110 0110 0110
+3/4 FS	+7.50	0111 0101 0000	1000 1010 1111
+1/2 FS	+5.00	0101 0000 0000	1010 1111 1111
+1/4 FS	+2.50	0010 0101 0000	1101 1010 1111
+1/8 FS	+1.25	0001 0010 0101	1110 1101 1010
+1 LSB	+0.01	0000 0000 0001	1111 1111 1110
0	0.00	0000 0000 0000	1111 1111 1111

Source: Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.

decimal display applications. Here 4 bits are used to represent each decimal digit. BCD is a positive-weighted code but is relatively inefficient since in each group of 4 bits, only 10 out of a possible 16 states are utilized (*see also* **Article 33, Coding Theory and its Application to Measurement, Volume 1**).

The LSB analog value (or quantum, Q) for BCD is

$$\text{LSB(analog value)} = Q = \frac{\text{FSR}}{10^d} \tag{5}$$

where FSR is the full-scale range and d is the number of decimal digits. For example, if there are three digits and the FSR is 10 V, the LSB value is

$$\text{LSB(analog value)} = \frac{10\text{V}}{10^3} = 0.01\text{V} = 10\,\text{mV} \tag{6}$$

BCD coding is frequently used with an additional over-range bit, which has a weight equal to full scale and produces a 100% increase in range for the A/D converter. Thus, for a converter with a decimal full scale of 999, an over-range bit provides a new full scale of 1999, twice that of the previous one. In this case, the maximum output code is 1 1001 1001 1001. The additional range is commonly

referred to as $\frac{1}{2}$ digit, and the resolution of the A/D converter in this case is $3\frac{1}{2}$ digits.

Likewise, if this range is again expanded by 100%, a new full scale of 3999 results and is called $3\frac{3}{4}$ digits resolution. Here, two over-range bits have been added and the full-scale output code is 11 1001 1001 1001. When BCD coding is used for bipolar measurements, another bit, a sign bit, is added to the code and the result is sign-magnitude BCD coding.

FURTHER READING

The general principles of DAS are found explained in various titles. Coding methods are usually covered. Coding is covered in **Article 33, Coding Theory and its Application to Measurement, Volume 1**.

Austerlitz, H. (2003) *Data Acquisition Techniques Using PCs*, Academic Press, San Diego.

James, K. (2000) *PC Interfacing and Data Acquisition: Techniques for Measurement, Instrumentation and Control*, Newnes, Oxford.

Mihura, B. (2001) *LabVIEW for Data Acquisition*, Prentice Hall PTR, Upper Saddle River, NJ.

Park, J. and MacKay, S. (2003) *Practical Data Acquisition for Instrumentation and Control Systems*, Newnes, Oxford.

Taylor, H.R. and Taylor, J.R. (1997) *Data Acquisition for Sensor Systems*, Chapman & Hall, London.

This article is based in part on a chapter by E.L. Zuch originally published in Handbook of Measurement Science Vol 1, P.H. Sydenham, Wiley, 1982.

138: Sampling Theory Relevant to DAS

Gerd Wöstenkühler
Hochschule Harz, Wernigerode, Germany

1 Aperture Time 955
2 Sample-holds and Aperture Error 955
3 Sampled-data System and Sampling Theorem 956
4 Frequency Folding and Aliasing 957
Further Reading 957

1 APERTURE TIME

Conversion time is frequently referred to as *aperture time*. In general, aperture time refers to the time uncertainty (or time window) in making a measurement and results in an amplitude uncertainty (or error) in the measurement if the signal is changing during this time.

As shown in Figure 1, the input signal to the analog-to-digital (A/D) converter changes by ΔV during the aperture time t_a in which the conversion is performed. The error can be considered as an amplitude error or a time error; the two are related as follows:

$$\Delta V = t_a \frac{dV(t)}{dt} \tag{1}$$

where $dV(t)/dt$ is the rate of change with time of the input signal.

It should be noted that ΔV represents the maximum error due to signal change, since the actual error depends on how the conversion is done. At some point in time within t_a, the signal amplitude corresponds exactly with the output code word produced.

For the specific case of a sinusoidal input signal, the maximum rate of change occurs at the zero crossing of the waveform, and the amplitude error is

$$\Delta V = t_a \frac{d}{dt}(A \sin(\omega t))_{t=0} = t_a A \omega \tag{2}$$

The resultant error as a fraction of the peak-to-peak full-scale value is

$$\varepsilon = \frac{\Delta V}{2A} = \pi f t_a \tag{3}$$

As an example from this result, the aperture time required to digitize a 1 kHz signal to 10-bits resolution can be found. The resolution required is one part in 2^{10} or approximately 0.001.

$$t_a = \frac{\varepsilon}{\pi f} \approx \frac{0.001}{3.14 \times 10^3} \approx 320 \times 10^{-9} \tag{4}$$

The result is a required aperture time of just 320 ns! A 1 kHz signal is not a particularly fast one, yet it is difficult to find a 10-bit A/D converter to perform this kind of conversion at a reasonable price! Fortunately, there is a relatively simple and inexpensive way around this dilemma by using a sample-hold circuit.

2 SAMPLE-HOLDS AND APERTURE ERROR

A sample-hold circuit samples the signal voltage and then stores it on a capacitor for the time required to perform the A/D conversion. The aperture time of the A/D converter is, therefore, greatly reduced by the much shorter aperture time of the sample-hold circuit. In turn, the aperture time of the sample-hold is a function of its bandwidth and the switching time.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

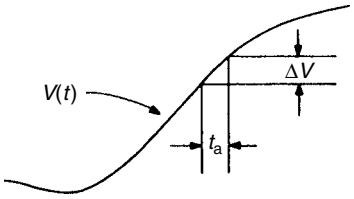


Figure 1. Aperture time and amplitude uncertainty. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

Figure 2 is a useful graph of (4). It gives the aperture time required for converting sinusoidal signals to a maximum error less than one part in 2^n where n is the resolution of the converter in bits. The peak-to-peak value of the sinusoid is assumed to be the full-scale range of the A/D converter. The graph is useful for selecting a sample-hold by aperture time or an A/D converter by conversion time.

3 SAMPLED-DATA SYSTEM AND SAMPLING THEOREM

In data acquisition and distribution systems, and other sampled-data systems, analog signals are sampled on a periodic basis as illustrated in Figure 3. The train of sampling pulses in Figure 3(b) represents a fast-acting switch, which connects to the analog signal for a very short time and then disconnects for the remainder of the sampling period.

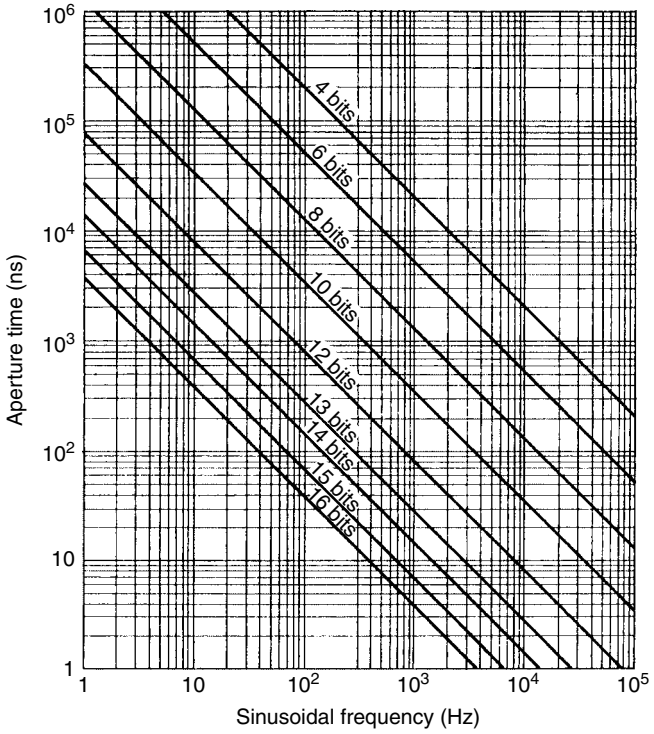


Figure 2. Graph of aperture error for sinusoidal signals. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

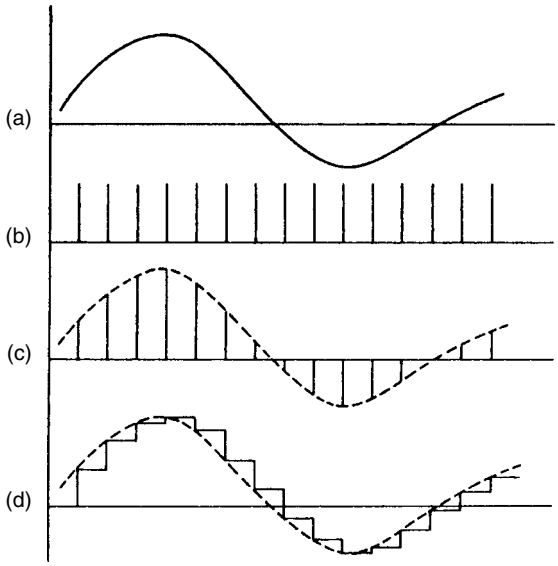


Figure 3. Signal sampling: (a) signal, (b) sampling pulses, (c) sampled signal, and (d) sampled and held signal. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

The result of the fast-acting sampler is identical with the result of multiplying the analog signal by a train of sampling pulses of unity amplitude, giving the modulated pulse train of Figure 3(c). The amplitude of the original signal is preserved in the modulation envelope of the pulses. If the switch-type sampler is replaced by a switch and capacitor (a sample-hold circuit), then the amplitude of each sample is stored between the samples and a reasonable reconstruction of the original analog signal results, as shown in Figure 3(d).

A common use of sampling is in the efficient use of data processing equipment and data transmission facilities. A single data transmission link, for example, can be used to transmit many different analog channels on a sampled, time-multiplexed basis, whereas it would be uneconomical to devote a complete transmission link to the continuous transmission of a single signal.

Likewise, a data acquisition and distribution system is used to measure and control the many parameters of a process control system by sampling the parameters and by updating the control inputs periodically. In data conversion systems, it is common to use a single, expensive A/D converter of high speed and precision, and then multiplex a number of analog inputs into it.

An important fundamental question to answer about sampled-data systems is this: 'How often must I sample an analog signal in order not to lose information from it'? It is obvious that all useful information can be extracted if a slowly varying signal is sampled at a rate such that little or no change takes place between samples. Equally obvious is the fact that the information is being lost if there is a significant change in signal amplitude between samples.

The answer to the question is contained in the well-known sampling theorem, which may be stated as follows: if a continuous bandwidth-limited signal contains no frequency components higher than f_c, then the original signal can be recovered without distortion if it is sampled at a rate of at least $2f_c$ samples per second.

4 FREQUENCY FOLDING AND ALIASING

The sampling theorem can be demonstrated by the frequency spectra illustrated in Figure 4. Figure 4(a) shows the frequency spectrum of a continuous bandwidth-limited analog signal with frequency components out to f_c. When this signal is sampled at a rate f_s, the modulation process shifts the original spectrum out to $\ldots, -3f_s, -2f_s, -f_s, f_s, 2f_s, 3f_s, \ldots$, in addition to the one at the origin. A portion of this resultant spectrum is shown in Figure 4(b).

If the sampling frequency f_s is not high enough, part of the spectrum centered about f_s will fold over into the original signal spectrum. This undesirable effect is called *frequency folding*. In the process of recovering the original signal, the folded part of the spectrum causes distortion in the recovered signal, which cannot be eliminated by filtering the recovered signal.

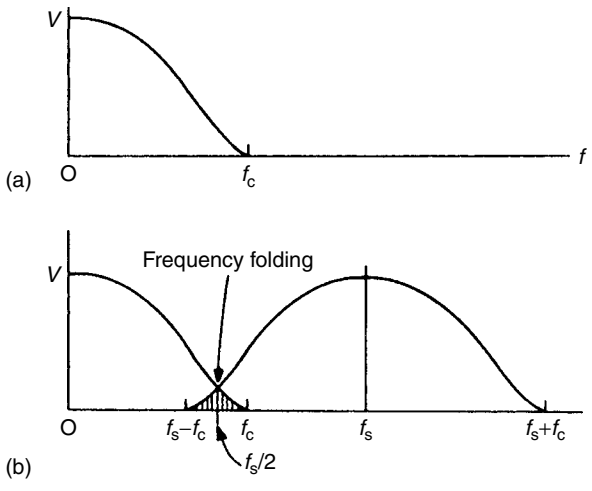


Figure 4. Frequency spectra demonstrating the sampling theorem: (a) continuous signal spectrum and (b) sampled signal spectrum. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

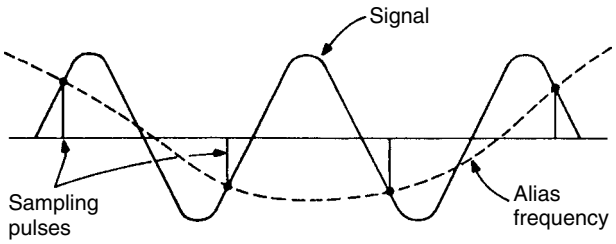


Figure 5. Alias frequency caused by inadequate sampling rate. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

From Figure 4, if the sampling rate is increased such that $f_s - f_c > f_c$, then the two spectra are separated and the original signal can be recovered without distortion. This demonstrates the results of the sampling theorem that $f_s > 2f_c$. Frequency folding can be eliminated in two ways: first, by using a high enough sampling rate, and second, by filtering the signal before sampling to limit its bandwidth to $f_s/2$.

It must be appreciated that in practice there is always some frequency folding present owing to high-frequency signal components, noise, and nonideal presample filtering. The effect must be reduced to negligible amounts for a particular application by using a sufficiently high sampling rate. The required rate, in fact, may be much higher than the minimum indicated by the sampling theorem.

The effect of an inadequate sampling rate on a sinusoid is illustrated in Figure 5; an alias frequency in the recovered signal results (*see also* **Article 31, Discrete Signal Theory, Volume 1**). In this case, sampling at a rate slightly less than twice per cycle gives the low-frequency sinusoid shown by the dotted line in the recovered signal. This alias frequency can be significantly different from the original frequency. From the figure, it is easy to see that if the sinusoid is sampled at least twice per cycle, as required by the sampling theorem, the original frequency is preserved.

FURTHER READING

The general principles of DAS are explained in various titles. They usually contain information on sampling effects. IC converter chip manufacturers offer technical data on their products.

Austerlitz, H. (2003) *Data Acquisition Techniques Using PCs*, Academic Press, San Diego.

Mihura, B. (2001) *LabVIEW for Data Acquisition*, Prentice Hall PTR, Upper Saddle River NJ.

James, K. (2000) *PC Interfacing and Data Acquisition: Techniques for Measurement, Instrumentation and Control*, Newnes, Oxford.

Park, J. and MacKay, S. (2003) *Practical Data Acquisition for Instrumentation and Control Systems*, Newnes, Oxford.

Taylor, H.R. and Taylor, J.R. (1997) *Data Acquisition for Sensor Systems*, Chapman & Hall, London.

This article is based in part on a chapter by E.L. Zuch originally published in Handbook of Measurement Science Vol 1, P.H. Sydenham, Wiley, 1982.

139: Analog-to-Digital (A/D) Converters

Gerd Wöstenkühler
Hochschule Harz, Wernigerode, Germany

1 Counter Type A/D Converter 959
2 Successive-approximation A/D Converters 959
3 Parallel (Flash) A/D Converters 960
Further Reading 961

1 COUNTER TYPE A/D CONVERTER

Analog-to-digital (A/D) converters, also called ADCs or encoders, employ a variety of different circuit techniques to implement the conversion function. As with digital-to-analog (D/A) converters, however, relatively few of the many originally devised circuits are widely used today. Of the various techniques available, the choice depends on the resolution and speed required.

One of the simplest A/D converters is the counter, or servo type. This circuit employs a digital counter to control the input of a D/A converter (*see also* **Article 141, Digital-to-Analog (D/A) Converters, Volume 3**). Clock pulses are applied to the counter and the output of the D/A is stepped up one LSB (least significant bit) at a time. A comparator compares the D/A output with the analog input and stops the clock pulses when they are equal. The counter output is then the converted digital word.

While this converter is simple, it is also relatively slow. An improvement on this technique is shown in Figure 1 and is known as a *tracking A/D converter*, a device commonly used in control systems. Here an up/down counter controls the DAC, and the clock pulses are directed to the pertinent counter input depending on whether the D/A output must increase or decrease to reach the analog input voltage.

The obvious advantage of the tracking A/D converter is that it can continuously follow the input signal and give updated digital output data if the signal does not change too rapidly. Also, for small input changes, the conversions can be quite fast. The converter can be operated in either the track or hold modes by a digital input control.

2 SUCCESSIVE-APPROXIMATION A/D CONVERTERS

By far, the most popular A/D conversion technique in general use for moderate to high-speed applications is the successive-approximation type A/D. This method falls into a class of techniques known as *feedback type A/D converters*, to which the counter type also belongs. In both cases, a D/A converter (*see also* **Article 141, Digital-to-Analog (D/A) Converters, Volume 3**) is in the feedback loop of a digital control circuit, which changes its output until it equals the analog input. In the case of the successive-approximation converter, the DAC is controlled in an optimum manner to complete a conversion in just n steps, where n is the resolution of the converter in bits.

The operation of this converter is analogous to weighing an unknown on a laboratory balance scale using standard weights in a binary sequence such as $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, . . . , $\frac{1}{2^n}$ kg. The correct procedure is to begin with the largest standard weight and proceed in order down to the smallest one.

The largest weight is placed on the balance pan first; if it does not tip, the weight is left on and the next largest weight is added. If the balance does tip, the weight is removed and the next one added. The same procedure is used for

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

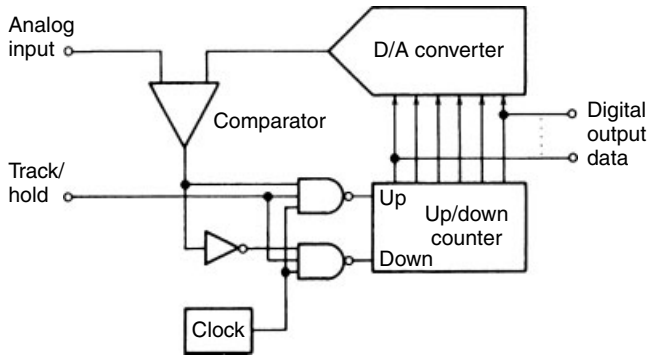


Figure 1. Tracking-type A/D converter. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

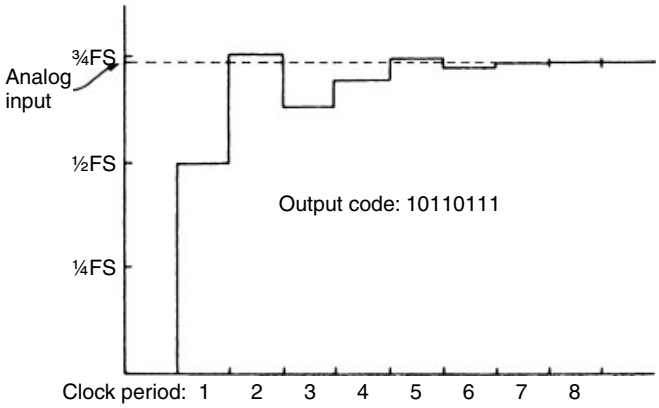


Figure 3. D/A output for 8-bit successive-approximation conversion. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

the next largest weight and so on down to the smallest. After the *n*th standard weight has been tried and a decision made, the weighing is finished. The total of the standard weights remaining on the balance is the closest possible approximation to the unknown.

In the successive-approximation A/D converter illustrated in Figure 2, a successive-approximation register (SAR) controls the D/A converter by implementing the weighing logic just described. The SAR first turns on the MSB (most significant bit) of the DAC and the comparator tests this output against the analog input. A decision is made by the comparator to leave the bit on or turn it off after which bit 2 is turned on and a second comparison is made. After *n* comparisons, the digital output of the SAR indicates all those bits that remain on and produce the desired digital code. The clock circuit controls the timing of the SAR. Figure 3 shows the D/A converter output during a typical conversion.

The conversion efficiency of this technique means that high-resolution conversions can be made in a very short time. For example, it is possible to perform a 10-bit conversion in 1 μs or less. Of course, the speed of the internal circuitry, in particular the D/A and comparator, are critical for high-speed performance.

3 PARALLEL (FLASH) A/D CONVERTERS

For ultrafast conversions required in video signal processing and radar applications where up to 8-bits resolution is required, a different technique is employed; it is known as the *parallel* (also *flash*, or *simultaneous*) method and is illustrated in Figure 4. This circuitry employs $2^n - 1$ analog comparators to directly implement the quantizer transfer function of an A/D converter.

The comparator trip-points are spaced 1 LSB apart by the series resistor chain and voltage reference. For a given analog input voltage, all comparators that are biased below the voltage turn on and all those biased above it remain off. Since all comparators change state simultaneously, the quantization process is a one-step operation.

A second step is required, however, since the logic output of the comparators is not in binary form. Therefore,

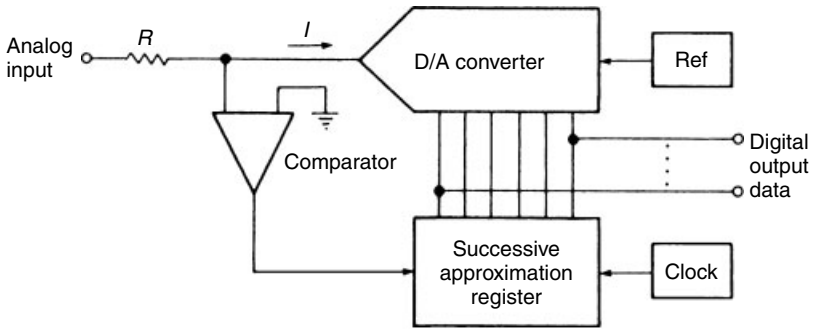


Figure 2. Successive-approximation A/D converter. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

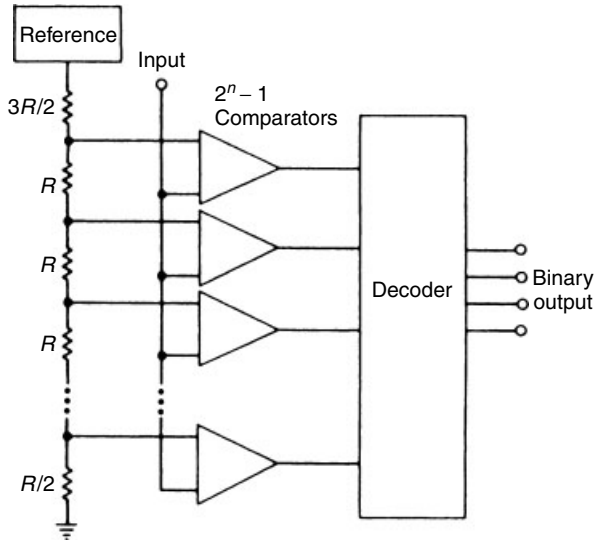


Figure 4. 4-bit parallel A/D converter. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

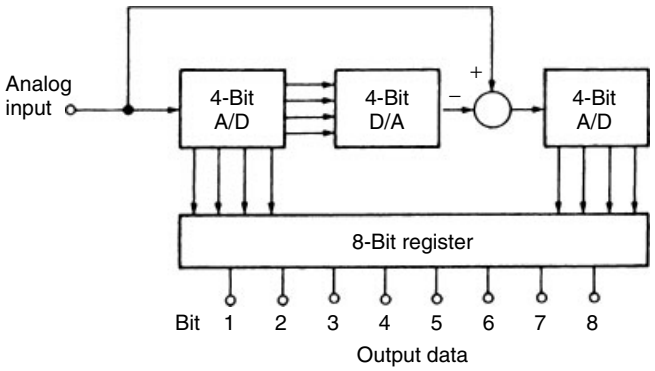


Figure 5. Two-stage parallel 8-bit A/D converter. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

an ultrafast decoder circuit is employed to make the logic conversion to binary. The parallel technique reaches the ultimate at high speed because only two sequential operations are required to make the conversion.

The limitation of the method, however, is in the large number of comparators required for even moderate resolutions. A 4-bit converter, for example, requires only 15 comparators, but an 8-bit converter needs 255. For this reason, it is common practice to implement an 8-bit A/D with two 4-bit stages as shown in Figure 5.

The result of the first 4-bit conversion is converted back to analog by means of an ultrafast 4-bit D/A and then subtracted from the analog input. The resulting residue is then converted by the second 4-bit A/D and the two sets of data are accumulated in the 8-bit output register. Converters of this type achieve 8-bit conversions at rates of 20 MHz and higher, while single-stage 4-bit conversions can reach 50 to 100 MHz rates.

FURTHER READING

A/D converters are usually covered in the general titles in DAS. Manufacturers provide data sheets and application notes for their own products.

Austerlitz, H. (2003) *Data Acquisition Techniques Using PCs*, Academic Press, San Diego.

James, K. (2000) *PC Interfacing and Data Acquisition: Techniques for Measurement, Instrumentation and Control*, Newnes, Oxford.

Mihura, B. (2001) *LabVIEW for Data Acquisition*, Prentice Hall PTR, Upper Saddle River, NJ.

Park, J. and MacKay, S. (2003) *Practical Data Acquisition for Instrumentation and Control Systems*, Newnes, Oxford.

Taylor, H.R. and Taylor, J.R. (1997) *Data Acquisition for Sensor Systems*, Chapman & Hall, London.

Converters are covered in

Hoeschele, D.F. (1994) *Analog-to-Digital and Digital-to-Analog Conversion Techniques*, Wiley-Interscience, New York.

Jespers, P.G.A. (2001) *Integrated Converters: D to A and A to D Architectures, Analysis and Simulation*, Oxford University Press, Oxford.

Moscovici, A. (2001) *High Speed A/D Converters – Understanding Data Converters Through SPICE*, Kluwer Academic Publishers, Boston.

Razavi, B. (1995) *Principles of Data Conversion System Design*, IEEE Press, New York.

This article is based in part on a chapter by E.L. Zuch originally published in Handbook of Measurement Science Vol 1, P.H. Sydenham, Wiley, 1982.

140: Integrating Type A/D Converters

Gerd Wöstenkühler
Hochschule Harz, Wernigerode, Germany

1 Indirect A/D Conversion 962
2 Dual-slope A/D Conversion 962
3 Charge-balancing A/D Conversion 962
Further Reading 964

1 INDIRECT A/D CONVERSION

A class of analog-to-digital (A/D) converters, known as *integrating type*, operates by an indirect conversion method. The unknown input voltage is converted into a time period, which is then measured by a clock and counter. A number of variations exist on the basic principle such as single-slope, dual-slope, and triple-slope methods. In addition, there is another technique – completely different – that is known as the *charge-balancing*, *quantized-feedback method* or sigma-delta modulator.

The most popular of these methods are dual-slope and charge balancing; although both are slow, they have excellent linearity characteristics with the capability of rejecting input noise. Because of these characteristics, integrating type A/D converters are almost exclusively used in digital panel meters, digital multimeters, and other comparatively slow measurement applications.

2 DUAL-SLOPE A/D CONVERSION

The dual-slope technique, shown in Figure 1, is perhaps the best known. Conversion begins when the unknown input voltage is switched to the integrator input; at the same time, the counter begins to count clock pulses and counts up to overflow. At this point, the control circuit switches the integrator to the negative reference voltage, which is integrated until the output is back to zero. Clock pulses are counted during this time until the comparator detects the zero crossing and turns them off.

The counter output is then the converted digital word. Figure 2 shows the integrator output waveform where T_1 is a fixed time and T_2 is the time proportional to the input voltage. The times are related as follows:

$$T_2 = T_1 \frac{V_{\text{in}}}{V_{\text{ref}}} \tag{1}$$

The digital output word, therefore, represents the ratio of the input voltage to the reference voltage.

Dual-slope conversion has several important features. First, conversion accuracy is independent of the stability of the clock and integrating capacitor so long as they are constant during the conversion period. Accuracy depends only on the reference accuracy and the integrator circuit linearity. Second, the periodic noise rejection of the converter can be infinite if T_1 is set to equal the period of the noise. To reject 60 Hz power noise, therefore, requires that T_1 be 16.667 ms or its multiples.

3 CHARGE-BALANCING A/D CONVERSION

The charge-balancing, or quantized-feedback, method of conversion is based on the principle of generating a pulse train with a frequency proportional to the input voltage and then counting the pulses for a fixed period of time. This circuit is shown in Figure 3. Except for the counter and timer,

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

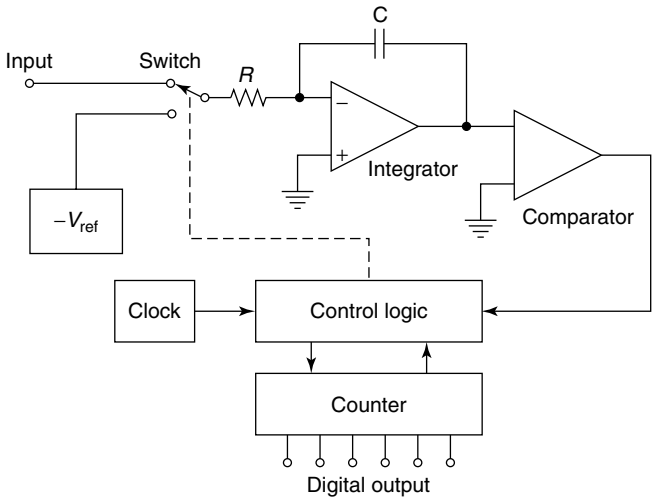


Figure 1. Dual-slope A/D converter. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

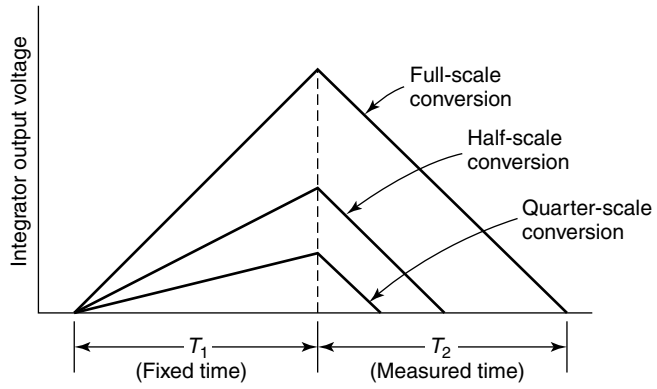


Figure 2. Integrator output waveform for dual-slope A/D converter. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

the circuit is a voltage-to-frequency (V/F) converter that generates an output pulse rate proportional to input voltage.

The circuit operates as follows. A positive input voltage causes a current to flow into the operational integrator through R_1. This current is integrated, producing a negative going ramp at the output. Each time the ramp crosses zero, the comparator output triggers a precision pulse generator, which puts out a constant width pulse.

The pulse output controls switch S_1, which connects R_2 to the negative reference for the duration of the pulse. During this time, a pulse of current flows out of the integrator-summing junction, producing a fast, positive ramp at the integrator output. This process is repeated, generating a train of current pulses, which exactly balances the input current – hence the name charge balancing. This balance

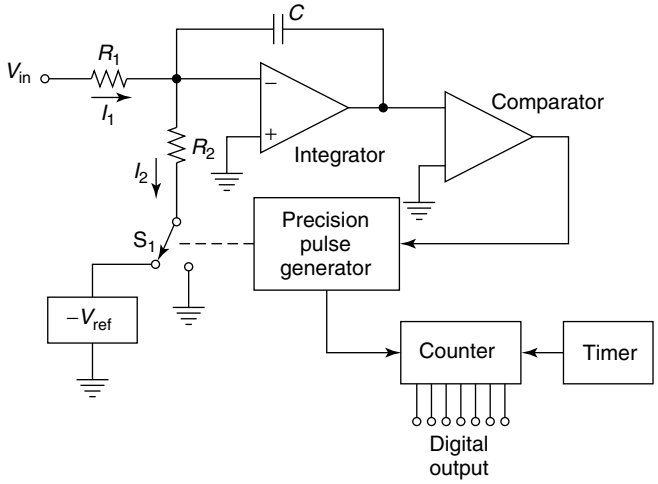


Figure 3. Charge-balancing A/D converter. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

has the following relationship:

$$f = \frac{1}{\tau}\frac{V_{\text{in}}}{V_{\text{ref}}}\frac{R_2}{R_1} \tag{2}$$

where τ is the pulse width and f the frequency.

A higher input voltage, therefore, causes the integrator to ramp up and down faster, producing higher frequency output pulses. The timer circuit sets a fixed time period for counting and the number of pulses within this period represents the digital output.

If the output of the comparator is synchronized with a clock signal and directly used to control the switch S_1 as the pulse generator and more than one clock period is needed to balance the current of maximum input voltage over two clock periods, the circuit becomes a sigma-delta modulator. The synchronized comparator output signal generates a sequence of 1 and 0, and the output pulse density balances the input voltage. The counter for the digital output counts the periods n where the switch is connected to the reference voltage and its relation to the used number N of clock periods in the time window represents the digital output signal.

$$n = N\frac{V_{\text{in}}}{V_{\text{ref}}}\frac{R_2}{R_1} \tag{3}$$

Like the dual-slope converter, the circuit also integrates input noise, and if the timer is synchronized with the noise frequency, infinite rejection results. Figure 4 shows the noise rejection characteristic of all integrating type A/D converters with rejection plotted against the ratio of integration period to noise period.

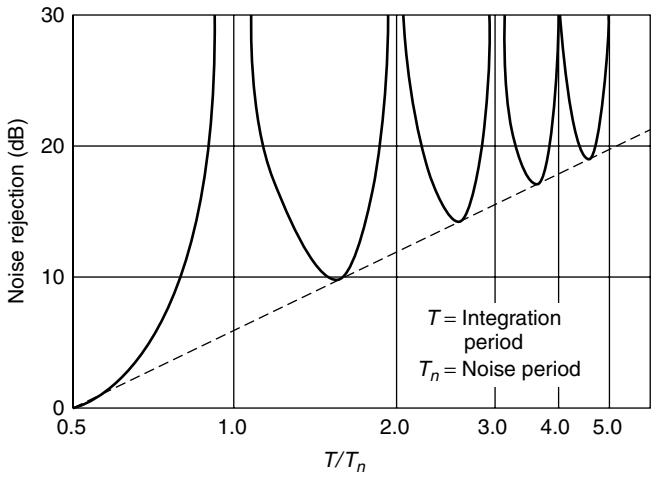


Figure 4. Noise rejection for integrating type A/D converters. (Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

FURTHER READING

The general principles of DAS are found explained in various titles. These often cover converters.

Austerlitz, H. (2003) *Data Acquisition Techniques Using PCs*, Academic Press, San Diego.

Mihura, B. (2001) *LabVIEW for Data Acquisition*, Prentice Hall PTR, Upper Saddle River, NJ.

James, K. (2000) *PC Interfacing and Data Acquisition: Techniques for Measurement, Instrumentation and Control*, Newnes, Oxford.

Park, J. and MacKay, S. (2003) *Practical Data Acquisition for Instrumentation and Control Systems*, Newnes, Oxford.

Taylor, H.R. and Taylor, J.R. (1997) *Data Acquisition for Sensor Systems*, Chapman & Hall, London.

Converters are covered in the following:

Hoeschele, D.F. (1994) *Analog-to-Digital and Digital-to-Analog Conversion Techniques*, Wiley-Interscience, New York.

Jespers, P.G.A. (2001) *Integrated Converters: D to A and A to D Architectures, Analysis and Simulation*, Oxford University Press, Oxford.

Moscovici, A. (2001) *High Speed A/D Converters – Understanding Data Converters Through SPICE*, Kluwer Academic Publishers, Boston, MA.

This article is based in part on a chapter by E.L. Zuch originally published in Handbook of Measurement Science Vol 1, P.H. Sydenham, Wiley, 1982.

141: Digital-to-Analog (D/A) Converters

Gerd Wöstenkühler
Hochschule Harz, Wernigerode, Germany

1 Introduction to D/A Converters 965
2 Weighted Current Source D/A Converter 965
3 $R-2R$ D/A Converter 966
4 Multiplying and Deglitching D/A Converters 968
Further Reading 970

1 INTRODUCTION TO D/A CONVERTERS

Digital-to-analog (D/A) converters are the devices by which computers communicate with the outside analog world. They are employed in a variety of applications from CRT display systems and voice synthesizers to automatic test systems, digitally controlled attenuators, and process control actuators. In addition, they are key components inside most analog-to-digital (A/D) converters. D/A converters are also referred to as *DACs* and are termed *decoders* by communications engineers.

The transfer characteristic of an ideal 3-bit D/A converter is shown in Figure 1. Each input code word produces a single, discrete analog output value, generally, but not always, a voltage. Over the output range of the converter 2^n different values are produced including zero; and the output has a one-to-one correspondence with input, which is not true for A/D converters.

There are many different circuit techniques used to implement D/A converters, but a few popular ones are widely used today. Virtually all D/A converters in use are of the *parallel type* where all bits change simultaneously upon application of an input code word; *serial type* D/A converters, on the other hand, produce an analog output only after receiving all digital input data in sequential form.

2 WEIGHTED CURRENT SOURCE D/A CONVERTER

The most popular D/A converter design in use today is the weighted current source circuit illustrated in Figure 2. An array of switched transistor current sources is used with binary weighted currents. The binary weighting is achieved by using emitter resistors with binary related values of $R, 2R, 4R, 8R, \ldots 2^{n-1}R$. The resulting collector currents are then added together at the current summing line.

The current sources are switched on or off from standard TTL semiconductor device inputs by means of the control diodes connected to each emitter. When the TTL input is high, the current source is on; when the input is low, it is off, with the current flowing through the control diode. Fast switching speed is achieved because there is direct control of the transistor current, and the current sources never go into saturation.

To interface with standard TTL levels, the current sources are biased to a base voltage of $+1.2$ V. The emitter currents are regulated to constant values by means of the control amplifier and a precision voltage reference circuit together with a binary transistor.

The summed output currents from all current sources that are on go to an operational amplifier summing junction; the amplifier converts this output current into an output voltage. In some D/A converters, the output current is used to directly drive a resistor load for maximum speed, but the positive output voltage in this case is limited to about $+1$ V.

The weighted current source design has the advantage of simplicity and high speed. Both PNP and NPN transistor current sources can be used with this technique although the TTL interfacing is more difficult with NPN sources. This

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

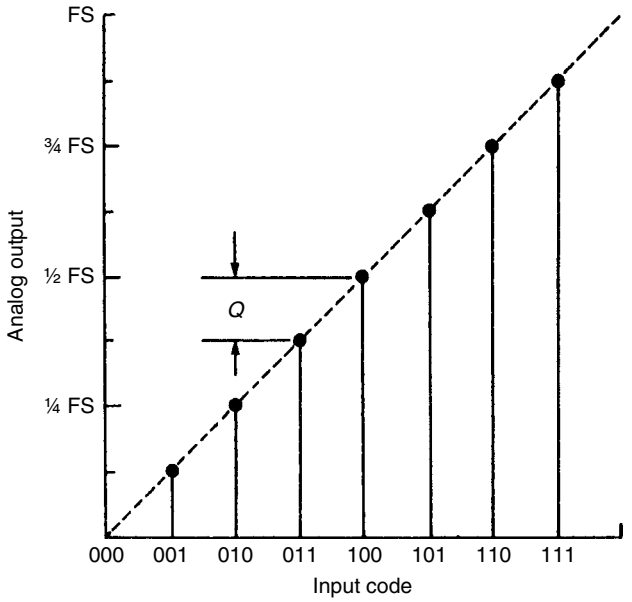


Figure 1. Transfer characteristic of ideal 3-bit D/A converter. (*Source*: Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

technique is used in most monolithic, hybrid, and modular D/A converters.

A difficulty in implementing the higher resolution D/A converter designs to this concept is that a wide range of emitter resistors is required and very high value resistors cause problems with both temperature stability and switching speed. To overcome these problems, weighted current sources are used in identical groups, with the output of each group divided down by a resistor divider as shown in Figure 3.

The resistor network, R_1 through R_4, divides the output of group 3 down by a factor of 256 and the output of group 2 down by a factor of 16 with respect to the output of group 1. Each group is identical, with four current sources of the type shown in Figure 2, having binary current weights of 1, 2, 4, 8. Figure 3 also illustrates the method of achieving a bipolar output by deriving an offset current from the reference circuit, which is then subtracted from the output current line through resistor R_O. This current is set to exactly one half of the full scale output current.

3 *R*–2*R* D/A CONVERTER

A second popular technique for D/A conversion is the $R-2R$ ladder method. As shown in Figure 4, the network

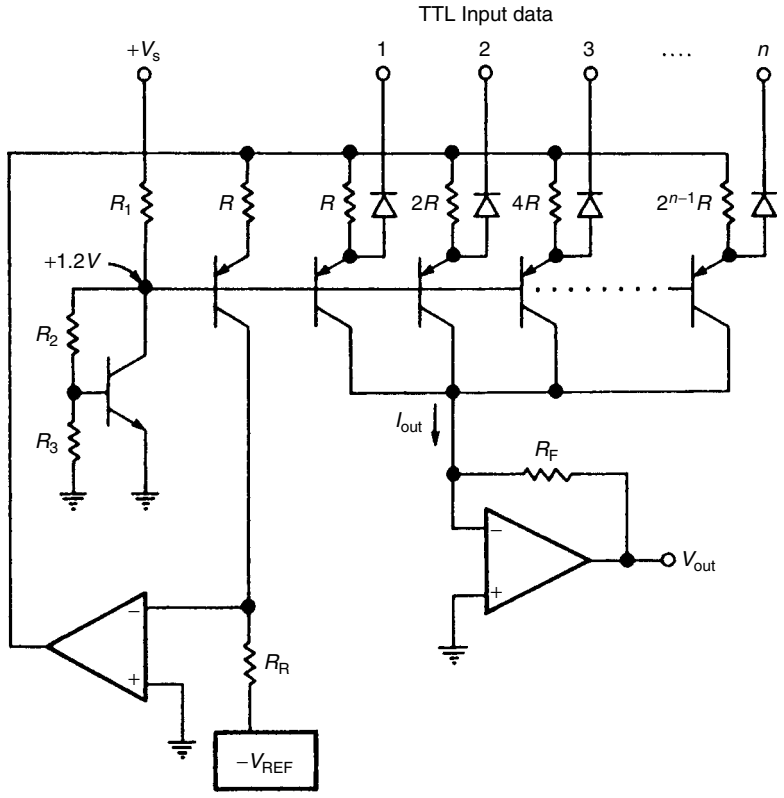


Figure 2. Weighted current source D/A converter. (*Source*: Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

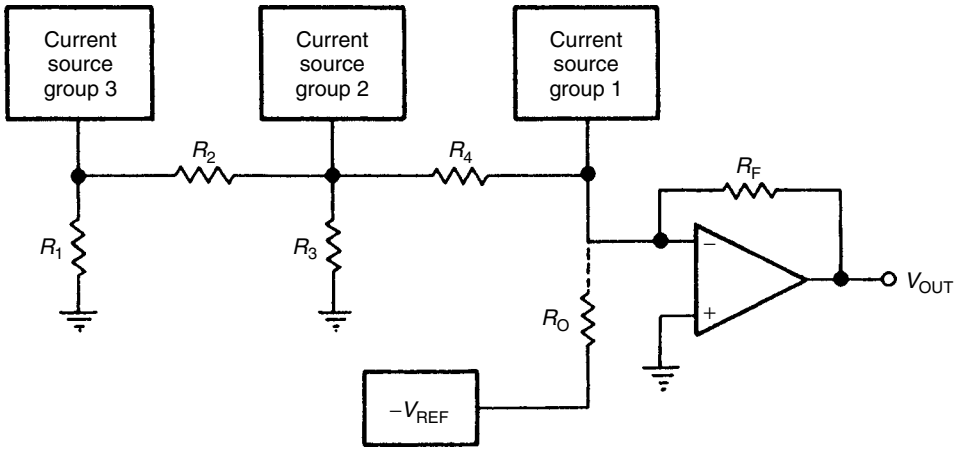


Figure 3. Current dividing the outputs of weighted current source groups. (*Source*: Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

consists of series resistors of value R and shunt resistors of values $2R$. The bottom of each shunt resistor has a single-pole double-throw electronic switch, which connects the resistor to either ground or to the output current summing line. As in the previous circuit, the output current summing line goes to an operational amplifier that converts current to voltage.

The operation of the $R-2R$ ladder network is based on the binary division of current as it flows down the ladder. Examination of the ladder configuration reveals that at point A, looking to the right, one measures a resistance of $2R$; therefore, the reference input to the ladder has a resistance of R. At the reference input, the current splits into two equal parts since it sees equal resistances in either direction. Likewise, the current flowing down the ladder to the right continues to divide into two equal parts at each resistor junction.

The result is binary weighted currents flowing down each shunt resistor in the ladder. The digitally controlled switches direct the currents to either the summing line or to ground. Assuming all bits are on, as shown in the diagram, the output current is

$$I_{OUT} = \frac{V_{REF}}{R}\left(\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots + \frac{1}{2^n}\right) \qquad (1)$$

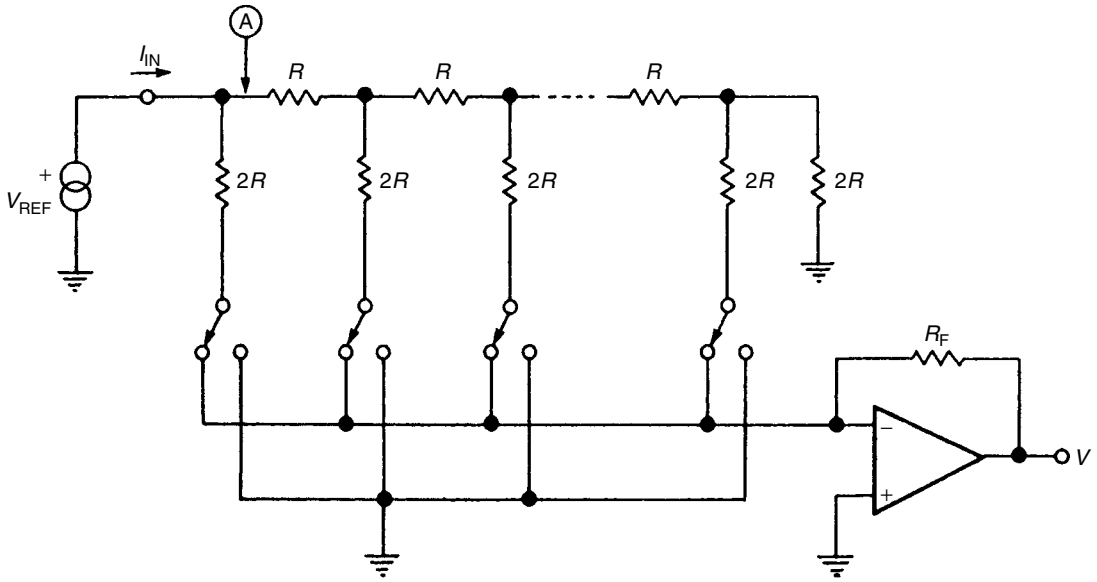


Figure 4. $R-2R$ ladder D/A converter. (*Source*: Reproduced from *Handbook of Measurement Science, Volume One*, Sydenham, 1983 © John Wiley & Sons Ltd.)

which is a binary series. The sum of all currents is then

$$I_{\text{OUT}} = \frac{V_{\text{REF}}}{R}(1 - 2^{-n}) \tag{2}$$

where the 2^{-n} term physically represents the portion of the input current flowing through the $2R$ terminating resistor to ground at the far right.

The advantage of the R–$2R$ ladder technique is that only two values of resistors are required, with the resultant ease of matching or trimming and excellent temperature tracking. In addition, for high speed applications, relatively low resistor values can be used. Excellent results can be obtained for high resolution D/A converters by using laser-trimmed thin film resistor networks.

4 MULTIPLYING AND DEGLITCHING D/A CONVERTERS

The R–$2R$ ladder method is specifically used for multiplying D/A converters. With these converters, the reference voltage can be varied over the full range of $\pm V_{\text{max}}$ with the output being the product of the reference voltage and the digital input word. Multiplication can be performed in 1, 2, or 4 algebraic quadrants.

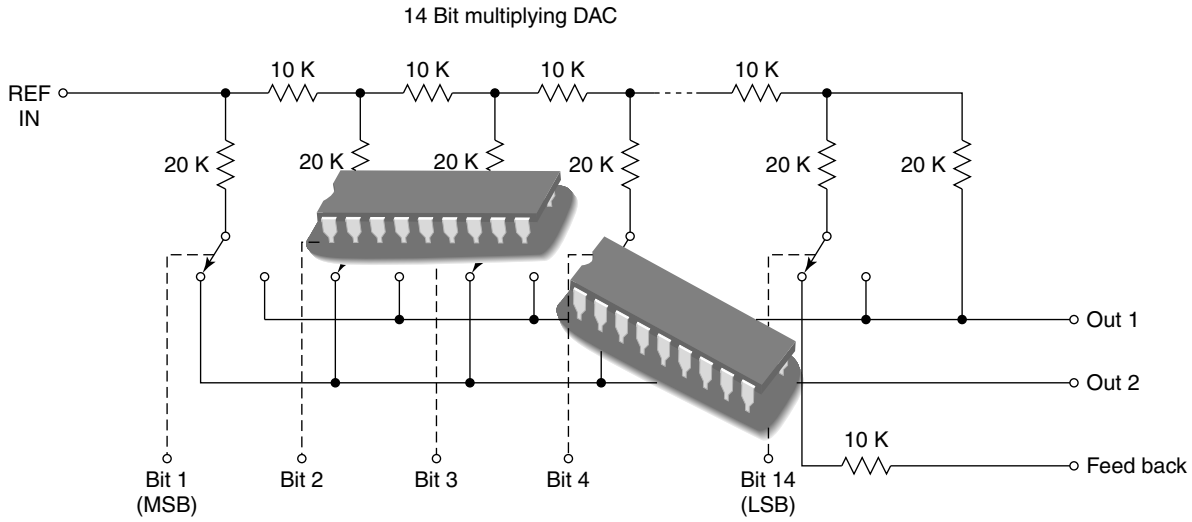


Figure 5. CMOS 14-bit multiplying D/A converters. (*Source*: Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

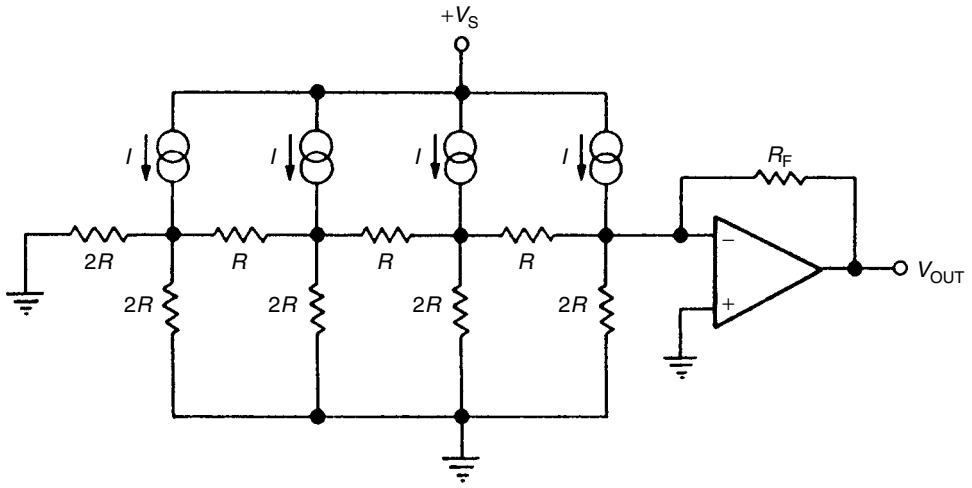


Figure 6. D/A converter employing R–$2R$ ladder with equal value switched current sources. (*Source*: Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

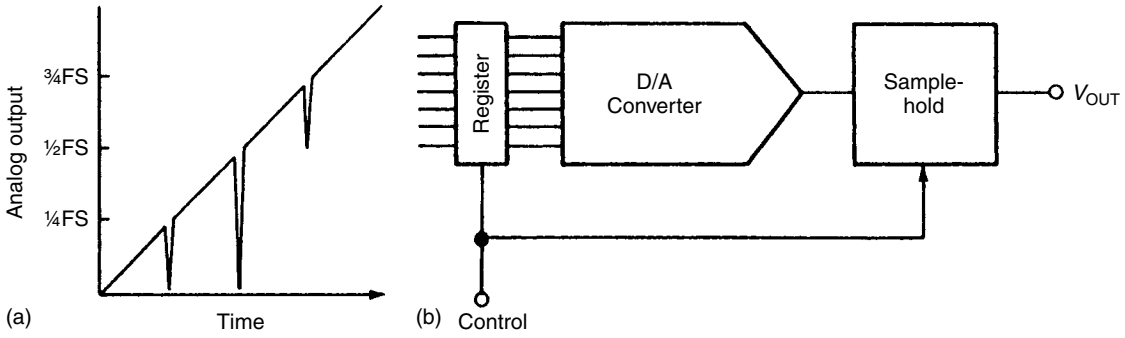


Figure 7. (a) Output glitches and (b) deglitched D/A converter. (*Source*: Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

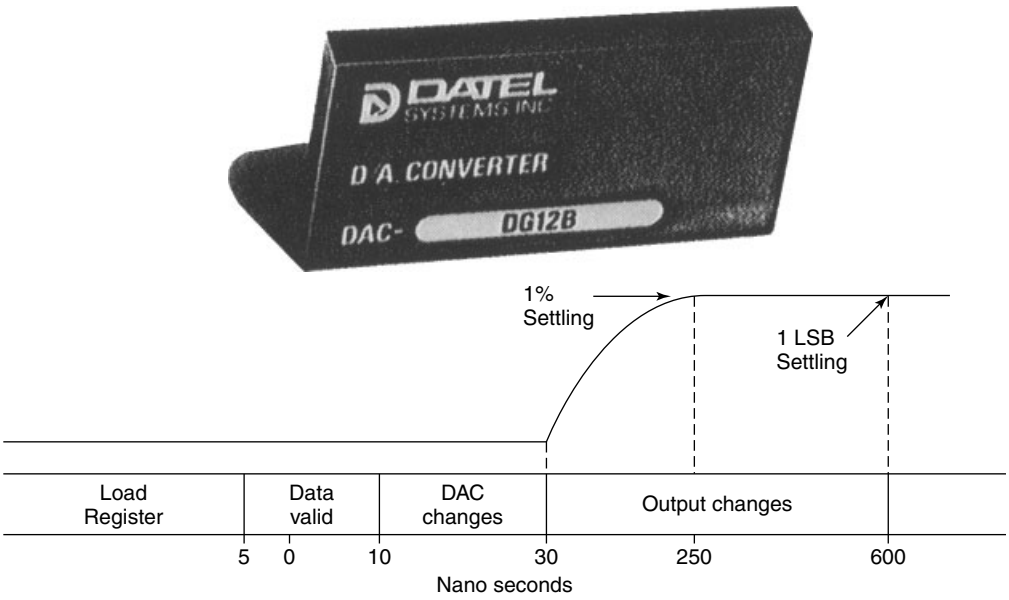


Figure 8. Modular deglitched D/A converter. (*Source*: Reproduced from *Handbook of Measurement Science*, Volume One, Sydenham, 1983 © John Wiley & Sons Ltd.)

If the reference voltage is unipolar, the circuit is a one-quadrant multiplying DAC; if it is bipolar, the circuit is a two-quadrant multiplying DAC. For a four-quadrant operation the two current summing lines shown in Figure 4 must be subtracted from each other by operational amplifiers.

In multiplying D/A converters, the electronic switches are usually implemented with CMOS devices. Multiplying DACs are commonly used in automatic gain controls, CRT character generation, complex function generators, digital attenuators, and divider circuits. Figure 5 shows two 14-bit multiplying CMOS D/A converters.

Another important D/A converter design takes advantage of the best features of both the weighted current source technique and the $R-2R$ ladder technique. This circuit, shown in Figure 6, uses equal value switched current sources to drive the junctions of the $R-2R$ ladder network. The advantage of the equal value current sources is obvious since all emitter resistors are identical and switching speeds are also identical. This technique is used in many ultrahigh speed D/A converters.

One other specialized type D/A converter used primarily in CRT display systems is the deglitched D/A converter. All D/A converters produce output spikes, or glitches, which are most serious at the major output transitions of $\frac{1}{4}$FS, $\frac{1}{2}$FS, and $\frac{3}{4}$FS as illustrated in Figure 7(a).

Glitches are caused by small time differences between some current sources turning off and others turning on. Take, for example, the major code transition at half scale from $0111\cdots1$ to $1000\cdots0$. Here, the MSB current source turns on while all other current sources turn off. The small difference in switching times results in a narrow half-scale glitch. Such a glitch produces distorted characters on CRT displays.

Glitches can be virtually eliminated by the circuit shown in Figure 7(b). The digital input to a D/A converter is controlled by an input register while the converter output goes to a specially designed sample-hold circuit. When the digital input is updated by the register, the sample-hold is switched into the hold mode. After the D/A converter has changed to its new output value and all glitches have settled out, the sample-hold is then switched back into the tracking mode. When this happens, the output changes smoothly from its previous value to the new value with no glitches present. Figure 8 shows a modular deglitched D/A converter, which contains the circuitry just described.

FURTHER READING

Converters used in DAS setups are explained in most general titles on data acquisition. Manufacturers supply application notes and data sheets for their products.

Austerlitz, H. (2003) *Data Acquisition Techniques Using PCs*, Academic Press, San Diego.

Mihura, B. (2001) *LabVIEW for Data Acquisition*, Prentice Hall PTR, Upper Saddle River, NJ.

James, K. (2000) *PC Interfacing and Data Acquisition: Techniques for Measurement, Instrumentation and Control*, Newnes, Oxford.

Park, J. and MacKay, S. (2003) *Practical Data Acquisition for Instrumentation and Control Systems*, Newnes, Oxford.

Taylor, H.R. and Taylor, J.R. (1997) *Data Acquisition for Sensor Systems*, Chapman & Hall, London.

Converters are covered in

Hoeschele, D.F. (1994) *Analog-to-Digital and Digital-to-Analog Conversion Techniques*, Wiley-Interscience, New York.

Jespers, P.G.A. (2001) *Integrated Converters: D to A and A to D Architectures, Analysis and Simulation*, Oxford University Press, Oxford.

Moscovici, A. (2001) *High Speed A/D Converters – Understanding Data Converters Through SPICE*, Kluwer Academic Publishers, Boston, MA.

Razavi, B. (1995) *Principles of Data Conversion System Design*, IEEE Press, New York.

This article is based in part on a chapter by E.L. Zuch originally published in Handbook of Measurement Science Vol 1, P.H. Sydenham, Wiley, 1982.

142: Z-transforms

Amar Bousbaine
University of Derby, Derby, UK

1 Purpose and Role of the Z-transform in Measurement Systems 973
2 Introduction to Z-transform Theory 973
3 Region of Convergence 974
4 Properties of the Region of Convergence 975
5 Properties of Z-transform 975
6 Some Common Z-transform Pairs 976
7 The Inverse Z-transform 976
8 Application of the Z-transform in Signal Processing 978
9 Causality and Stability 978
References 980

1 PURPOSE AND ROLE OF THE Z-TRANSFORM IN MEASUREMENT SYSTEMS

The z-transform is an invaluable tool for representing, analyzing, and designing discrete time signals and systems. Since most measurement systems now contain significant digital signal processing (DSP), the z-transform is an important element in measurement system design. Information about signals can be obtained by examining their z-transforms. In particular, the stability or instability of a linear time invariant discrete system is characterized by the location of the poles in the z-plane. It also indicates how the system responds to particular oscillatory components of the input.

Moreover, the use of z-transform techniques permits simple algebraic manipulations, that is, it facilitates the solution to difference equations because convolution operators are transformed into simple algebraic ones. In many cases, the sum of finite or infinite series for the z-transform can be expressed in a closed form expression, which offers a compact alternative representation of signals.

Prototype continuous time filters can be converted into high-quality digital filters by simple mapping from the s-domain to the z-domain. Consequently, the z-transform has become an important tool for the analysis and design of digital filters (**Article 148, General Performance of the Digital Filter, Volume 3**; **Article 149, Low-, High-, and Band-pass Digital Filters, Volume 3**; **Article 150, Finite Infinite Impulse Response (IIR) Digital Filters, Volume 3**; **Article 151, Finite Impulse Response (FIR) Digital Filters, Volume 3**).

2 INTRODUCTION TO Z-TRANSFORM THEORY

The z-transform of a sequence $x[n]$ is defined as

$$Z\{x[n]\} \equiv X(z) = \sum_{n=-\infty}^{\infty} x[n]z^{-n} \tag{1}$$

where z is a complex variable.

If z is defined as $z = re^{j\omega}$, where $r = |z|$ is the module of z and ω its argument, then (1) becomes

$$X(re^{j\omega}) = \sum_{n=-\infty}^{\infty} x[n]r^{-n}e^{-j\omega n} \tag{2}$$

Equation (2) can be interpreted as the discrete time Fourier transform (DTFT) of the modified sequence, $x[n]r^{-n}$.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

For $r = 1$, (2) reduces to the Fourier transform of $x[n]$, provided the latter exists. The contour $r = |z| = 1$ is a circle in the z-plane of unity radius and is called the *unit circle*.

The definition, (1), can be applied to the impulse response sequence, $h[n]$, of a digital filter. The z-transform of $h[n]$ is called the *transfer function* of the filter and is defined as

$$H(z) = \sum_{n=-\infty}^{\infty} h[n]z^{-n} \tag{3}$$

3 REGION OF CONVERGENCE

Since the z-transform is an infinite power series, it exists only for those values of z for which this series converges. For a given sequence, the set of values for which its z-transform converges is called the *region of convergence* (ROC).

The z-transform exists when the finite sum of (1) converges. A necessary condition for convergence is absolute summability of $x[n]z^{-n}$. Since $\left|x[n]z^{-n}\right| = \left|x[n]r^{-n}\right|$, then

$$\sum_{n=-\infty}^{\infty} \left|x[n]r^{-n}\right| < \infty \tag{4}$$

The region of convergence of a z-transform is important in that without knowledge of the ROC, there is no unique relationship between a sequence and its z-transform. Hence, a z-transform must always be specified with its region of convergence. Moreover, if the ROC of a z-transform of a sequence includes the unit circle, the Fourier transform of the sequence is obtained simply by evaluating the z-transform on the unit circle (Figure 1).

One-sided exponential signals are encountered frequently in engineering problems because one is interested in the behavior of a signal either before or after a given time instant. With such signals, the magnitude of one or more poles determines the ROC boundaries. Some of these properties of the ROC are illustrated in the examples that follow.

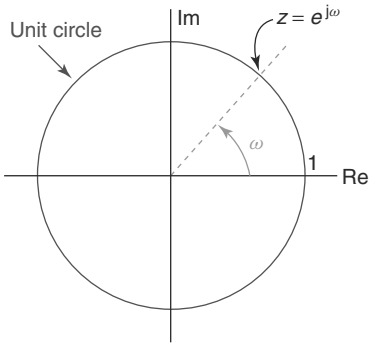


Figure 1. The unit circle.

Example 1. Right-sided sequence

Determine the z-transform of the signal,

$$x[n] = a^n u[n] = \begin{cases} a^n, & n \geq 0 \\ 0, & n < 0 \end{cases} = \sum_{n=0}^{\infty} a^n z^{-n}$$

Solution: Because $x[n]$ is nonzero only for $n \geq 0$, the sequence is known as a *right-sided* sequence.

From the definition, (1)

$$X(z) = \sum_{n=-\infty}^{\infty} x[n]z^{-n} = \sum_{n=0}^{\infty} a^n z^{-n} = \sum_{n=0}^{\infty} (az^{-1})^n$$

For convergence, $\sum_{n=0}^{\infty} |(az^{-1})^n| < \infty$

Thus the region of convergence is the range of values of z for which $|az^{-1}| < 1$ or equivalently, $|z| > |a|$, see Figure 2. Inside the ROC, the infinite series converges to

$$X(z) = \sum_{n=0}^{\infty} (az^{-1})^n = \frac{1}{1 - az^{-1}} = \frac{z}{z - a} \quad \text{for } |z| > |a|$$

Example 2. Left-sided sequence

Determine the z-transform of the signal,

$$x[n] = -a^n u[-n-1] = \begin{cases} -a^n, & n \leq -1 \\ 0, & n \geq 0 \end{cases}$$

$$= \sum_{n=-\infty}^{-1} (-a)^n z^{-n}$$

Solution: Because $x[n]$ is nonzero only for $n \leq -1$, the sequence is known as a *left-sided* sequence.

From the definition, (1)

$$X(z) = \sum_{n=-\infty}^{\infty} x[n]z^{-n} = -\sum_{n=-\infty}^{-1} a^n z^{-n}$$

$$= \sum_{n=1}^{\infty} (a^{-1}z)^n = 1 - \sum_{n=0}^{\infty} (a^{-1}z)^n$$

ROC is given by, $|a^{-1}z| < 1$ or equivalently, $|z| < |a|$ (see Figure 3). Inside the ROC, the infinite series converges to

$$X(z) = 1 - \frac{1}{1 - a^{-1}z} = \frac{z}{z - a} \quad \text{for} \quad |z| < |a|$$

Example 3. Two-sided sequence

Determine the z-transform of the signal,

$$x[n] = \left(-\frac{1}{3}\right)^n u[n] - \left(\frac{1}{2}\right)^n u[-n-1]$$

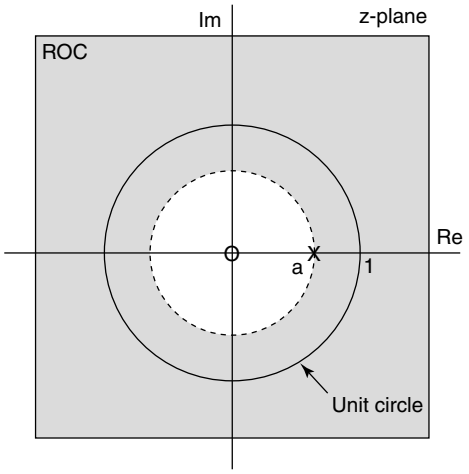


Figure 2. Pole-zero plot and ROC for a right-sided sequence.

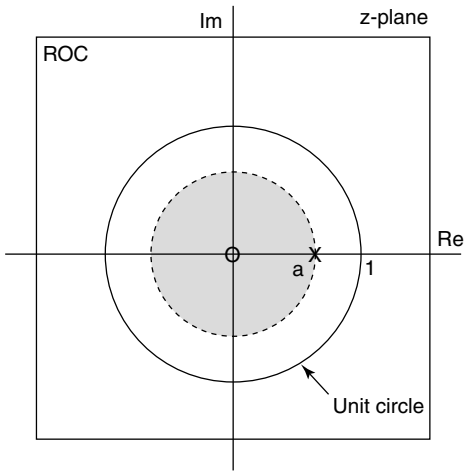


Figure 3. Pole-zero plot and ROC for a left-sided sequence.

Solution: From Examples 1, 2 and Table 2 the z-transform is

$$\left(-\frac{1}{3}\right)^n u[n] \overset{z}{\longleftrightarrow} \frac{1}{1+\frac{1}{3}z^{-1}}, \text{ ROC is } |z| > \frac{1}{3}$$

$$\left(\frac{1}{2}\right)^n u[-n-1] \overset{z}{\longleftrightarrow} \frac{1}{1-\frac{1}{2}z^{-1}}, \text{ ROC is } |z| < \frac{1}{2}$$

$$X(z) = \frac{1}{1+\frac{1}{3}z^{-1}} + \frac{1}{1-\frac{1}{2}z^{-1}}, \text{ ROC is } \frac{1}{3} < |z| < \frac{1}{2}$$

(see Figure 4).

4 PROPERTIES OF THE REGION OF CONVERGENCE

The basic properties of the ROC depend on the nature of the signal (for more details see Oppenheim and Schafer (1999) and Proakis and Manolakis (1996)).

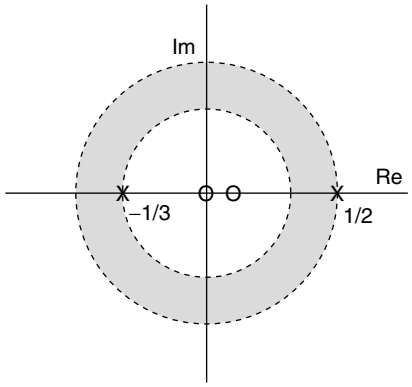


Figure 4. Pole-zero plot and ROC for a two-sided sequence.

Property 1: The ROC is a ring or a disk in the z-plane centered at the origin (Figures 2, 3, 4, and 6).

Property 2: The ROC cannot contain any poles, (Figures 2, 3, and 4).

Property 3: The Fourier transform of $x[n]$ converges absolutely if and only if the ROC of the z-transform of $x[n]$ includes the unit circle.

Property 4: If $x[n]$ is a finite duration sequence, that is, a sequence that is zero except in finite interval $-\infty < M_1 \leq n \leq M_2 < \infty$, then the ROC is the entire plane, except possibly $z = 0$ or $z = \infty$.

Property 5: If $x[n]$ is a right-sided sequence, that is, a sequence for $n < M_1 < \infty$, the region of convergence extends outward from the outermost finite pole in $X(z)$ to, and possibly include, $z = \infty$ (Figure 2).

Property 6: If $x[n]$ is a left-sided sequence, that is, a sequence for $-\infty < M_2 < n$, the region of convergence extends inward from the innermost nonzero pole in $X(z)$ to, and possibly include, $z = 0$ (Figure 3).

Property 7: A two-sided sequence is an infinite-duration sequence that is neither right sided nor left sided. If $x[n]$ is a two-sided sequence, the ROC will consist of a ring in the z-plane, bounded on the interior and exterior by a pole and, consistent with property 2, not containing any poles (Figure 4).

Property 8: The ROC must be a connected region (Figure 2, 3, 4, 5, and 6).

5 PROPERTIES OF Z-TRANSFORM

The z-transform possesses a number of properties that make it an extremely valuable tool in the study of discrete time signals and systems. Some of the useful properties that

Table 1. Properties of the z-transform.

Property	Sequence	z-transform	ROC
	$x[n]$	$X(z)$	ROC: $R_2 < \|z\| < R_1$
	$x_1[n]$	$X_1(z)$	R_{1c}
	$x_2[n]$	$X_2(z)$	R_{2c}
Linearity	$ax_1[n] + bx_2[n]$	$aX_1(z) + bX_2(z)$	Includes $R_{1c} \cap R_{2c}$
Time shift	$x_[n - n_o]$	$z^{-n_o}X(z)$	That of $X(z)$, except possibly $z = 0$ if $n_0 > 0$ and $z = \infty$ if $n_0 < 0$
Frequency scaling	$a^n x[n]$	$X(a^{-1}z)$	$\|a\|R_2 < \|z\| < \|a\|R_1$
Time reversal	$x[-n]$	$X(z^{-1})$	$\frac{1}{R_2} < \|z\| < \frac{1}{R_1}$
Multiplication by n	$nx[n]$	$-z\dfrac{d}{dz}X(z)$	ROC
Convolution	$x_1[n]^*x_2[n]$	$X_1(z)X_2(z)$	At least the intersection of R_{1c} and R_{2c}
Conjugation	$x^*[n]$	$X^*(z^*)$	ROC
Real part	$Re\{x[n]\}$	$\frac{1}{2}\{X(z) + X^*(z^*)\}$	Includes ROC
Imaginary part	$Im\{x[n]\}$	$\frac{1}{j2}\{X(z) - X^*(z^*)\}$	Includes ROC
Multiplication	$x_1[n]x_2[n]$	$\frac{1}{j2\pi}\oint X_1(v)X_2(\frac{z}{v})v^{-1}dv$	At least, $R_{1l}R_{2l} < \|z\| < R_{1u}R_{2u}$

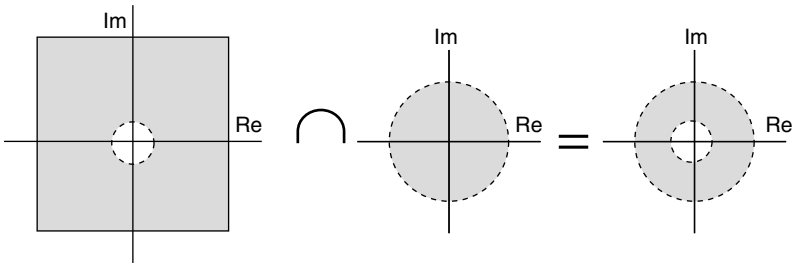


Figure 5. Intersection of the ROCs of the right-sided and left-sided sequences.

have found practical use in digital signal processing are summarized in Table 1.

6 SOME COMMON Z-TRANSFORM PAIRS

The recovery of a time domain signal from its z-transform, the inverse z-transform, can often be easily evaluated by expressing $X(z)$ as a linear combination of simpler basic z-transform pairs and the z-transform properties to invert a large class of z-transform. The z-transform pairs of common sequences are given in Table 2.

7 THE INVERSE Z-TRANSFORM

There are three methods that are often used for finding a sequence $x[n]$, given its z-transform $X(z)$.

7.1 Direct method (using the inversion integral)

$$x[n] = \frac{1}{j2\pi}\oint_c X(z)z^{n-1}dz \tag{5}$$

where $\oint_c$ represents integration along the closed contour, c, in the counter-clockwise direction in the z-plane. The contour must be chosen to lie in the region of convergence of $X(z)$.

7.2 Power series expansion method

The power series can be obtained by arranging the numerator and the denominator of $X(z)$ in descending powers of z and then dividing the numerator by the dominator using long division.

Given a z-transform $X(z)$ with its ROC, $X(z)$ can be expanded into a power series of the form $X(z) = \sum_{n=-\infty}^{\infty} c_n z^{-n}$, which converges in the ROC. Then by the uniqueness of the z-transform, $x[n] = c_n$ for all n.

Table 2. Some common z-transform pairs.

Signal, $x[n]$	z-transform, $X(z)$	Region of Convergence
$x[n] = \frac{1}{j2\pi} \oint X(z) z^{n-1} dz$	$X(z) = \sum_{-\infty}^{\infty} x[n] z^{-n}$	ROC
$\delta[n]$	1	All z
$\delta[n-m]$	z^{-m}	All z
$u[n]$	$\frac{1}{1-z^{-1}}$	$\|z\| > 1$
$a^n u[n]$	$\frac{1}{1-az^{-1}}$	$\|z\| > a$
$na^n u[n]$	$\frac{az^{-1}}{(1-az^{-1})^2}$	$\|z\| > a$
$-a^n u[-n-1]$	$\frac{1}{1-az^{-1}}$	$\|z\| < a$
$-na^n u[-n-1]$	$\frac{az^{-1}}{(1-az^{-1})^2}$	$\|z\| < a$
$\cos(\omega_o n) u[n]$	$\frac{1 - z^{-1}\cos(\omega_o)}{1 - 2z^{-1}\cos(\omega_o) + z^{-2}}$	$\|z\| > 1$
$\sin(\omega_o n) u[n]$	$\frac{z^{-1}\cos(\omega_o)}{1 - 2z^{-1}\cos(\omega_o) + z^{-2}}$	$\|z\| > 1$
$a^n \cos(\omega_o n) u[n]$	$\frac{1 - az^{-1}\cos(\omega_o)}{1 - 2az^{-1}\cos(\omega_o) + a^2 z^{-2}}$	$\|z\| > a$
$a^n \sin(\omega_o n) u[n]$	$\frac{az^{-1}\cos(\omega_o)}{1 - 2az^{-1}\cos(\omega_o) + a^2 z^{-2}}$	$\|z\| > a$

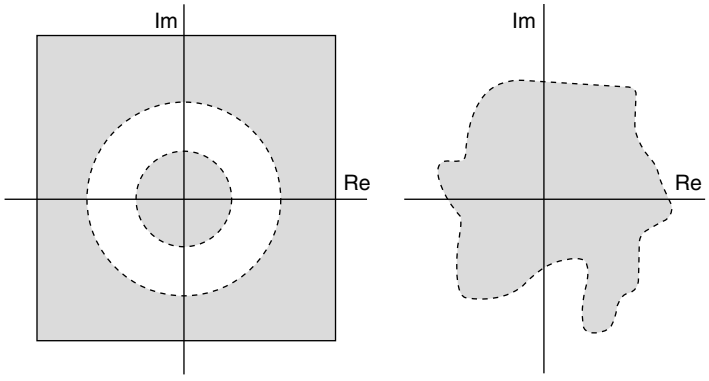


Figure 6. Unacceptable ROC.

7.3 Partial fraction expansion method

In the partial expansion method, the z-transform is first expanded into a sum of simple partial fractions. The inverse z-transform of each partial fraction is then obtained from tables, such as Table 1, and then summed to give the overall inverse z-transform.

In many practical cases of linear time invariant (LTI) systems, the z-transforms are rational functions of z^{-1} and given by

$$X(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_M z^{-M}}{a_0 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_N z^{-N}} = \frac{\sum_{k=0}^{M} b_k z^{-k}}{\sum_{k=0}^{N} a_k z^{-k}} \tag{6}$$

If the poles of $X(z)$ are of the first order and, for simplicity, $N = M$, then $X(z)$ can be rewritten as

$$X(z) = B_0 + \frac{A_1}{1 - p_1 z^{-1}} + \frac{A_2}{1 - p_2 z^{-1}} + \cdots + \frac{A_M}{1 - p_M z^{-1}}$$

$$X(z) = B_0 + \frac{A_1 z}{z - p_1} + \frac{A_2 z}{z - p_2} + \cdots + \frac{A_M z}{z - p_M} = B_0 + \sum_{k=1}^{M} \frac{A_k z}{z - p_k} \quad (7)$$

where p_k are the poles of $X(z)$, assumed distinct, A_k are the partial fraction coefficients, and $B_0 = b_M/a_M$.

The coefficients A_k may be obtained by multiplying both sides of (7) by $z - p_k/z$ and setting $z = p_k$, that is,

$$A_k = \left. \frac{X(z)}{z}(z - p_k) \right|_{z=p_k} \quad (8)$$

8 APPLICATION OF THE Z-TRANSFORM IN SIGNAL PROCESSING

One of the most important applications of the z-transform in DSP is in the design and analysis of the errors in digital filters, (Oppenheim and Schafer, 1999) and (Proakis and Manolakis, 1996). It is used extensively to determine the coefficients of digital filters, especially Infinite Impulse Response (IIR) filters (**Article 150, Finite Infinite Impulse Response (IIR) Digital Filters, Volume 3**). Another important application of the z-transform in discrete filter design is in the representation of digital filter structures. These are often represented in the form of block diagrams, which represent the difference equations or equivalently the transfer functions.

8.1 Pole and zero description in the discrete time systems

In most practical discrete time systems, the transfer function, $H(z)$, can be expressed in terms of its poles and zeros, as shown by (9).

$$X(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_M z^{-M}}{a_0 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_M z^{-M}} = \frac{\sum_{k=0}^{M} b_k z^{-k}}{\sum_{k=0}^{M} a_k z^{-k}} \quad (9)$$

If $H(z)$ has poles at $z = p_1, p_2, \ldots, p_M$ and zero at $z = z_1, z_2, \ldots, z_M$, then $H(z)$ can be factored and represented as

$$X(z) = K \frac{(z - z_1)(z - z_2)\ldots(z - z_M)}{(z - p_1)(z - p_2)\ldots(z - p_M)} \quad (10)$$

where K is the gain factor.

A plot of the poles and zeros of the transfer function is known as the *pole-zero diagram* and provides a useful way of representing and analyzing the filter. For example, for the filter to be stable, all its poles must lie within the unit circle. There is no restriction on the zero location.

8.2 Frequency response estimation

In the design of digital filters, it is often necessary to examine the spectrum of the filter to establish that the desired specifications are satisfied. The frequency response of a system can be readily obtained from the z-transform and given by

$$H(e^{j\omega t}) = \sum_{n=-\infty}^{\infty} h[n] z^{-n}|_{z=e^{j\omega t}} = \sum_{n=-\infty}^{\infty} h[n] e^{-jn\omega t} \quad (11)$$

where $H(e^{j\omega t})$ is the frequency response of the system. In general, $H(e^{j\omega t})$ is complex where the modulus represents its magnitude response and the phase response of the system.

9 CAUSALITY AND STABILITY

A causal linear time invariant system is one whose unit impulse response, $h[n]$, satisfies the condition

$$h[n] = 0, \quad n < 0 \quad (12)$$

Therefore, the impulse response of a causal LTI system is determined from the transfer function by using the right-sided inverse transforms. Consequently, an LTI system is causal if and only if the ROC of the system function is outside the maximum pole circle.

Anticausal signals, on the other hand, have ROC that is inside the minimum pole circle. If an LTI system is stable, then the impulse response is absolutely summable and the DTFT of the impulse response exists. Hence, the ROC must include the unit circle. Therefore, an LTI system is Bounded-Input Bounded-Output (BIBO) stable if and only if the ROC of the system function includes the unit circle.

Example 4. A discrete time system is characterized by the following transfer function

$$H(z) = \frac{3 - 4z^{-1}}{1 - 3.5z^{-1} + 1.5z^{-2}}$$

1. Specify the ROC of $H(z)$ and determine $h[n]$ for the following conditions
   (a) the system is stable
   (b) the system is causal
   (c) the system is anticausal.
2. Plot the pole-zero diagram of the system.
3. Plot the frequency response of the system. Assume a sampling frequency of 500 Hz and a resolution of less than 1 Hz.

Solution:

1. *ROC and $h[n]$*

$$H(z) = \frac{3 - 4z^{-1}}{1 - 3.5z^{-1} + 1.5z^{-2}} = \frac{z(3z - 4)}{z^2 - 3.5z + 1.5}$$

The poles of $H(z)$ are $p_1 = 0.5$ and $p_1 = 3$. Therefore, the expansion of the form, (6), is

$$\frac{H(z)}{z} = \frac{3z - 4}{(z - 0.5)(z - 3)} = \frac{A_1}{z - 0.5} + \frac{A_2}{z - 3}$$

$$A_1 = \frac{3z - 4}{(z - 0.5)(z - 3)} \cdot (z - 0.5)\Big|_{z=0.5}$$

$$= \frac{3 \times 0.5 - 4}{0.5 - 3} = 1$$

$$A_2 = \frac{3z - 4}{(z - 0.5)(z - 3)} \cdot (z - 3)\Big|_{z=3}$$

$$= \frac{3 \times 3 - 4}{3 - 0.5} = 2$$

$$\therefore \frac{H(z)}{z} = \frac{1}{z - 0.5} + \frac{2}{z - 3}$$

$$H(z) = \frac{1}{1 - 0.5z^{-1}} + \frac{2}{1 - 3z^{-1}}$$

(a) *Stable system*
For the system to be stable, the ROC must include the unit circle, hence $0.5 < |z| < 3$. Consequently, $h[n]$ can be obtained using Table 2 and is given by

$$h[n] = 0.5^n u[n] - 2(3)^n u[-n - 1].$$

The system is noncausal.

(b) *Causal system*
For the system to be causal, the ROC of the system function is outside the maximum pole circle, that is, $|z| > 3$. In this case, $h[n]$ is,

$$h[n] = 0.5^n u[n] + 2(3)^n u[n].$$

The system is unstable.

(c) *Anticausal system*
For the system to be anticausal, the ROC of the system function is inside the minimum pole circle, that is, $|z| < 0.5$. In this case, the impulse response

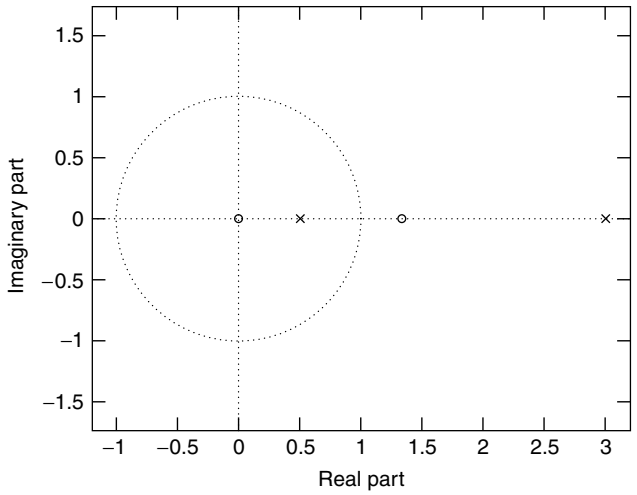


Figure 7. Pole-zero diagram.

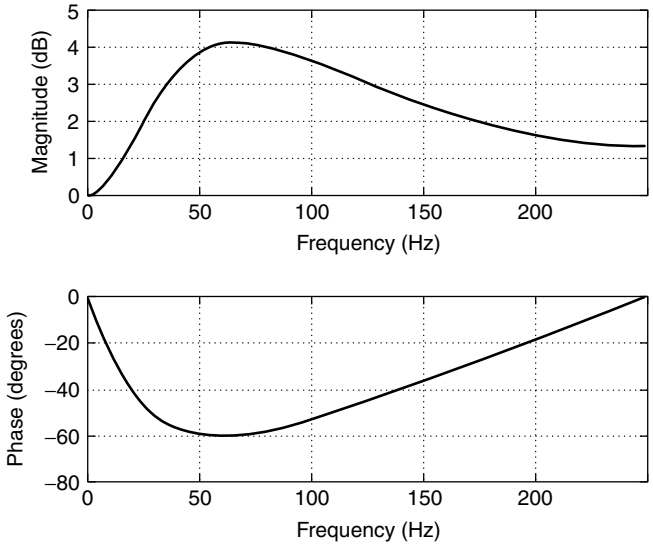


Figure 8. Frequency response.

is given by

$$h[n] = -0.5^n u[-n-1] - 2(3)^n u[-n-1].$$

The system is unstable.

2. *Plot of pole-zero diagram*

Figure 7 shows the locations of the poles and zeros of the discrete time system in the pole-zero diagram. MATLAB software is used for the plot shown in Figure 7 (for more details see Ifeachor and Jervis (2002)).

3. *Plot of frequency response*

Figure 8 shows the frequency response of the discrete time system. MATLAB software is used for the plot shown in Figure 8 (for more details see Ifeachor and Jervis (2002)).

REFERENCES

Ifeachor, E. and Jervis, B. (2002) *Digital Signal Processing, A Practical Approach*, Prentice Hall, Harlow, UK.

Oppenheim, A. and Schafer, R. (1999) *Discrete-Time Signal Processing*, 2nd edn, Prentice-Hall, Upper Saddle River, New Jersy.

Proakis, J.G. and Manolakis, D.G. (1996) *Digital Signal Processing, Principles, Algorithms, and Applications*, Prentice Hall, Upper Saddle River, New Jersy.

143: DFT and FFTs

Gerd Wöstenkühler

Hochschule Harz, Wernigerode, Germany

1 Fourier Spectrum of Periodic Signals	981
2 Aliasing Distortion	982
3 Discrete Fourier Transform	983
4 Fast Fourier Transform	984
5 Windowing	985
6 Averaging in Time and Frequency Domain	986
Further Reading	988

1 FOURIER SPECTRUM OF PERIODIC SIGNALS

Each periodic signal with $x(t) = x(t + T_0)$ can be described as a sum of harmonic functions shown in (1) (*see also* **Article 27, Signals in the Frequency Domain, Volume 1**). The time shift T_0 represents the fundamental frequency $\omega_0 = 2\pi f_0 = 2\pi / T_0$. The frequencies of the harmonic functions are the fundamental frequency and its multiples like $k\omega_0$ with $k \in \mathbb{N}$. The even components of the periodic signal are specified with the cosine functions and the odd components with the sine functions. The amplitudes are the values of the variables a_k and b_k for the kth frequency while the component $a_0/2$ represents the DC-value.

$$x(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} [a_k \cos(k\omega_0 t) + b_k \sin(k\omega_0 t)] \quad (1)$$

A typical example with the components of a square wave function is shown in Figure 1. It shows three harmonic functions and the sum of these functions and the desired square wave function, which can be achieved when all harmonic components would be used.

If the periodic signal is known, the coefficients a_k and b_k can be calculated with an integral over one period of the fundamental frequency according to (2) and (3). Quite often t_0 is set to zero or $-T_0/2$.

$$a_k = \frac{2}{T_0} \int_{t_0}^{t_0+T_0} x(t) \cos(k\omega_0 t)\, \mathrm{d}t \quad k \in \mathbb{N}_0 \quad (2)$$

$$b_k = \frac{2}{T_0} \int_{t_0}^{t_0+T_0} x(t) \sin(k\omega_0 t)\, \mathrm{d}t \quad k \in \mathbb{N} \quad (3)$$

The harmonic functions in (1) can be substituted according to the Euler's theorem shown in (4) and (5).

$$\cos(\varphi) = \frac{1}{2}(\mathrm{e}^{\mathrm{j}\varphi} + \mathrm{e}^{-\mathrm{j}\varphi}) \quad (4)$$

$$\sin(\varphi) = \frac{1}{\mathrm{j}2}(\mathrm{e}^{\mathrm{j}\varphi} - \mathrm{e}^{-\mathrm{j}\varphi}) \quad (5)$$

Using this description, the time function $x(t)$ can be described with (6).

$$x(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \left[\frac{1}{2}(a_k - \mathrm{j}b_k)\mathrm{e}^{\mathrm{j}k\omega_0 t} + \frac{1}{2}(a_k + \mathrm{j}b_k)\mathrm{e}^{-\mathrm{j}k\omega_0 t} \right] \quad (6)$$

The term $0.5(a_k - \mathrm{j}b_k)$ can be substituted by a complex value $\underline{c}_k$ while k has a positive value. The DC-value $a_0/2$ can be substituted by c_0 and the term $0.5(a_k + \mathrm{j}b_k)$ with a positive value of k can be substituted by a complex value $\underline{c}_{-k}$. The last part can be rewritten as $0.5(a_k - \mathrm{j}b_k)$ while k has a negative value and so it becomes $\underline{c}_k$ for $k < 0$. These substitutions are shown in (7) to (10).

$$\underline{c}_k = \frac{1}{2}(a_k - \mathrm{j}b_k) \quad k \in \mathbb{N} \quad \{1, 2, 3, \ldots, +\infty\} \quad (7)$$

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.

© 2005 John Wiley & Sons, Ltd.

ISBN: 0-470-02143-8.

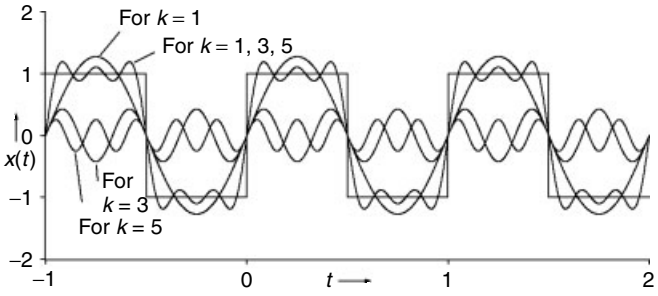


Figure 1. Square wave and 3 harmonics and sum of harmonics.

$$c_0 = \frac{1}{2}a_0 \tag{8}$$

$$\underline{c}_{-k} = \frac{1}{2}(a_k + \mathrm{j}b_k) \quad k \in \mathbb{N} \quad \{1, 2, 3, \ldots, +\infty\} \tag{9}$$

$$\underline{c}_k = \frac{1}{2}(a_k - \mathrm{j}b_k) \quad -k \in \mathbb{N} \quad \{-1, -2, -3, \ldots, -\infty\} \tag{10}$$

With this substitution, (1) becomes very compact and can be rewritten as (11).

$$x(t) = \sum_{k=-\infty}^{\infty} \underline{c}_k \mathrm{e}^{\mathrm{j}k\omega_0 t} \qquad k \in \mathbb{Z} \tag{11}$$

The variables $\underline{c}_k$ can be calculated using (7) or (10) with (2) and (3) for the variables a_k and b_k and the Euler's theorem to replace the cosine and sine functions. This leads to (12). If $x(t)$ is a real time function, the coefficients $\underline{c}_k = \underline{c}^*_{-k}$ (conjugate complex), based on the even cosine function of the real parts and the odd sine function of the imaginary parts.

$$\underline{c}_k = \frac{1}{T_0}\int_{t_0}^{t_0+T_0} x(t)\mathrm{e}^{-\mathrm{j}k\omega_0 t}\,\mathrm{d}t \qquad k \in \mathbb{Z} \tag{12}$$

The values of the Fourier coefficients a_k of an even pulse train with the amplitude 1 follows the si-function with the maximum amplitude of $2T_1/T_0$ for $k \in \mathbb{N}$ according to (13). The real values of the complex coefficients $\underline{c}_k$ follows the si-function with the maximum value of T_1/T_0 for $k \in \mathbb{Z}$ according to (14). Figure 2 shows a part of an even pulse train with $T_1 = T_0/4$, the coefficients of a_k (b_k are all zero) and the real part of $\underline{c}_k$ (imaginary parts of $\underline{c}_k$ are all zero).

$$a_k = \frac{2T_1}{T_0}\frac{\sin\left(\dfrac{k\pi T_1}{T_0}\right)}{\dfrac{k\pi T_1}{T_0}} = \frac{2T_1}{T_0}\,\mathrm{si}\left(\frac{k\pi T_1}{T_0}\right); \quad b_k = 0 \tag{13}$$

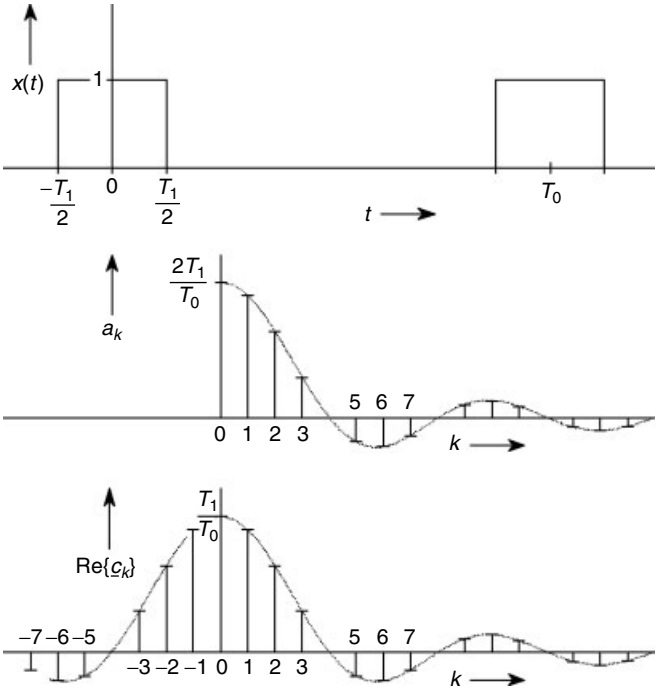


Figure 2. Even pulse train, Fourier coefficients a_k and complex Fourier coefficients $\underline{c}_k$.

$$\underline{c}_k = \frac{T_1}{T_0}\frac{\sin\left(\dfrac{k\pi T_1}{T_0}\right)}{\dfrac{k\pi T_1}{T_0}} + \mathrm{j}0 = \frac{T_1}{T_0}\,\mathrm{si}\left(\frac{k\pi T_1}{T_0}\right) + \mathrm{j}0 \tag{14}$$

2 ALIASING DISTORTION

For digital signal processing, it is necessary to convert an analog function into a digital representation. Therefore, the analog function is sampled with a constant sampling rate and each value is converted into a digital representation (*see also* **Article 138, Sampling Theory Relevant to DAS, Volume 3**). The function of sampling can be regarded as a multiplication of the original analog function with an impulse train of Dirac impulses as described in (15). Each sampled value x_n is the value of the original function $x(t)$ at the location nT_s of the time shifted Dirac impulses $\delta(t - nT_\mathrm{s})$.

$$x_n = x(nT_\mathrm{s}) = x(t) \times s(t) = x(t)\sum_{n=-\infty}^{\infty}\delta(t - nT_\mathrm{s}) \tag{15}$$

Such a multiplication in the time domain is equivalent to a convolution in the frequency domain. The spectrum of the original signal can be described with the frequencies up to a maximum frequency or cutoff frequency f_c. On the

basis of the description of a real-time function with complex Fourier coefficients, described in the previous section for a periodic signal, the frequency components cover an area between $-f_c$ to $+f_c$. The frequency of a Dirac impulse train has only frequency components at $f_s = 1/T_s$ and its multiples like $nf_s = n/T_s$. On the basis of the convolution, we can find the complex frequency values of the original analog signal around all frequencies of the Dirac impulse train. We get a periodic spectrum as shown in Figure 3.

The complete information of the analog signal can be found around each frequency component nf_s of the Dirac impulse train. Using an ideal filter, we can select a frequency area of, for example, $-f_c$ to $+f_c$. In this case, we get the original analog function back as shown with the dotted line in Figure 3. This works as long as the frequency areas of the reproduced original spectrum do not overlap. When the sampling frequency is not high enough ($f_s < 2f_c$), the spectrums will overlap. This effect is called *frequency folding* (*see also* **Article 138, Sampling Theory Relevant to DAS, Volume 3** and **Article 31, Discrete Signal Theory, Volume 1**). In this case, the ideal filter would select not only the original frequency components but also the overlapping part of the reproduced frequency components as shown in Figure 4.

The selection of an ideal filter for the frequency components between $-f_c$ and $+f_c$ causes a deviation within the reproduced analog signal. The frequencies f of the original spectrum appear at the frequency $f_s - f$. If these frequencies are in the overlapping part, they will be interpreted as frequency components with values of $f_s - f$. So you get frequency components with different frequencies out of the original signal. This effect is called *aliasing effect* of the sampling process or aliasing distortion. To avoid this, the frequency of the sampling rate has to fulfil (16). To meet this condition, you have to use a sampling rate that is high enough or you have to reduce the frequency area of the analog signal using analog low pass filters as antialias filters.

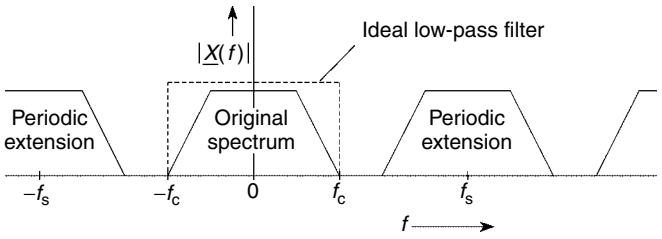


Figure 3. Spectrum of a sampled signal with $f_s > 2f_c$.

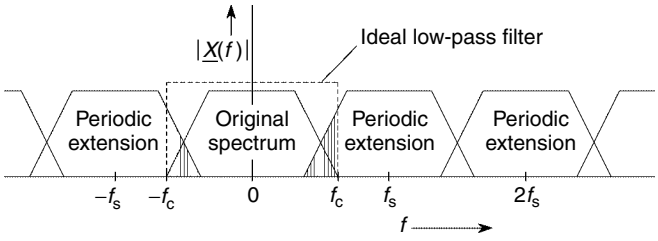


Figure 4. Spectrum of a sampled signal with $f_s < 2f_c$ with frequency folding.

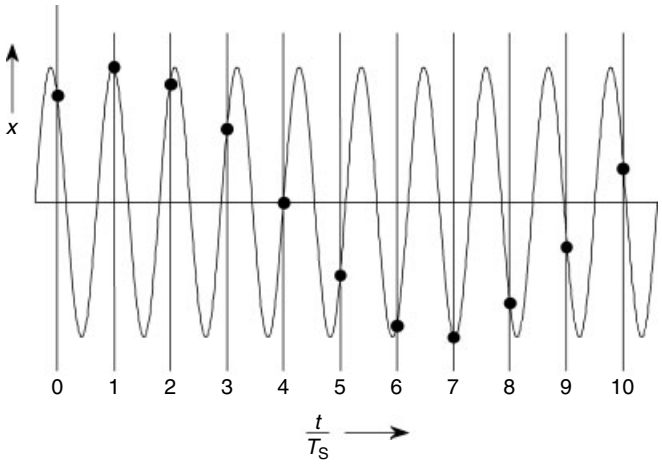


Figure 5. Sampled sine function with alias distortion.

$$f_s \geq 2f_c \tag{16}$$

A simple example of the aliasing effect is shown in Figure 5, where a sine function is sampled with an unsuitable sampling rate. It can be seen that the indicated sample points suggest a frequency component that is not part of the original function.

3 DISCRETE FOURIER TRANSFORM

If we use a digital system to convert the sampled time function into the frequency domain, (12) of the previous section becomes (17). The complex spectral component $\underline{X}_k$ is the sum of all sampled time values x_n multiplied with an arrow within the complex plane adjusted with its number of periods k to the corresponding time position n within the used time window of N samples.

$$\underline{X}_k = \sum_{n=0}^{N-1} x_n e^{-j2\pi nk/N} \quad N \in \mathbb{N} \tag{17}$$

Figure 6 illustrates the function of this equation. On the left-hand side, you see a sine function, which can be described with N samples from 0 to $N-1$. The dimension of the real part is the horizontal line and the time is the vertical line. You can imagine the imaginary part, which would be orthogonal to these two dimensions. The function

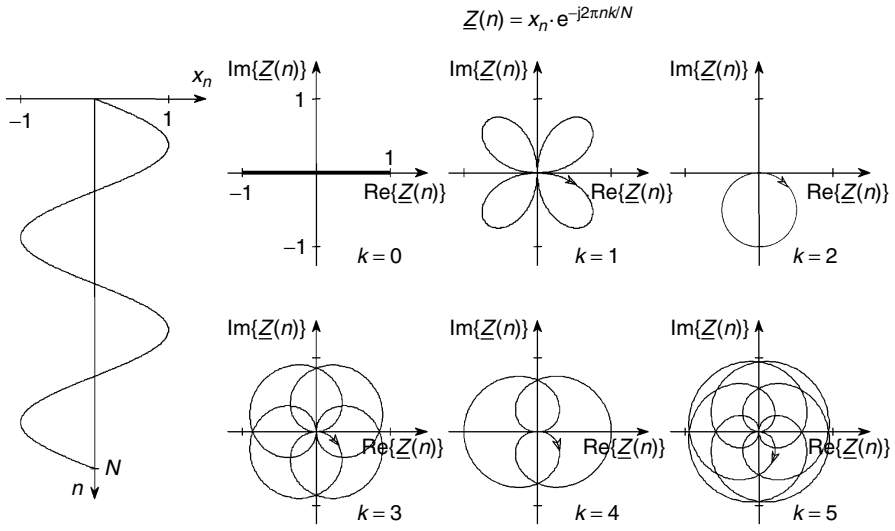


Figure 6. Graphical interpretation of the DFT.

$e^{-j2\pi nk/N}$ twists the time function with k revolutions at the end where the Nth sample would be. The other six diagrams show the projection of this twisted sine function into the complex plane where the real part is the horizontal line and the imaginary part is the vertical line. If $k = 0$, the function is not twisted and the projection describes a horizontal line as shown in the top left diagram. The sum $\underline{X}_0$ of all components is zero, which means that the sine function has no DC component. If $k = 1$, the resulting projection is shown in the top middle diagram. It describes a symmetric function and the sum $\underline{X}_1$ of all components is zero, which means that the analyzed sine function has no component with 1 period in the time window. If $k = 2$, the resulting projection is shown in the top right diagram. It describes a circle around $-j0.5$ with two revolutions and the sum $\underline{X}_2$ of all components is $-jN/2$, which means that the analyzed sine function has a component with 2 periods in the time window. Nevertheless, all other values of k up to $N - 1$ will produce a sum with the value of zero except $k = N - 2$ where the sum has the value $+jN/2$. Using (17), the sine function with two periods in the time window is transformed into a description with $\underline{X}_2 = -jN/2$ and $\underline{X}_{N-2} = +jN/2$.

There exist slightly different equations for this transformation. Some have used average values, which is the sum divided by the number of elements and so the values of the spectrum become independent of the number of samples. In this case, all following equations and scaling might be different based on this factor. The equations in this article have used the more often described equations of the discrete Fourier transform where the factor $1/N$ is used in the inverse Fourier transform. This is the sum of all complex frequency components $\underline{X}_k$ adjusted with the number of periods k and the used time position n within the used time window of N samples according to (18).

$$x_n = \frac{1}{N} \sum_{k=0}^{N-1} \underline{X}_k \, e^{j2\pi nk/N} \quad N \in \mathbb{N} \tag{18}$$

Figure 7 shows the inverse Fourier transform with the values determined in Figure 6. In the complex plane, the two components with the amplitude $N/2$ are shown. For the time position $n = 0$, the sum of both components is zero and this is the value of x_0. The value of $x_{N/16}$ where $n = N/16$ ($n \in \mathbb{N}$), the component of X_2 will rotate counterclockwise with 45° or $\pi/4$ and the component of X_{N-2} will rotate clockwise with 45° or $\pi/4$. The sum of both components yields $x_{N/16} = \sin(\pi/4) \approx 0.707$. On the basis of the symmetric frequency components of a real time function, all values x_n have an imaginary part of zero.

4 FAST FOURIER TRANSFORM

A central operation of the Fourier transform and the inverse Fourier transform is a complex multiplication with sine and cosine terms. For the calculation of all values, there are around N^2 complex multiplications and additions necessary. For a large number N of sample points, the calculation time of the frequency components becomes quite long and so several attempts have been made to reduce the amount of necessary computational time for solving this equation.

The used factor of $e^{2\pi nk/N}$ is a revolving arrow around the unit circle in the complex plane. The basic step size is $e^{2\pi/N}$. If N has the value of 2 to the power of ν

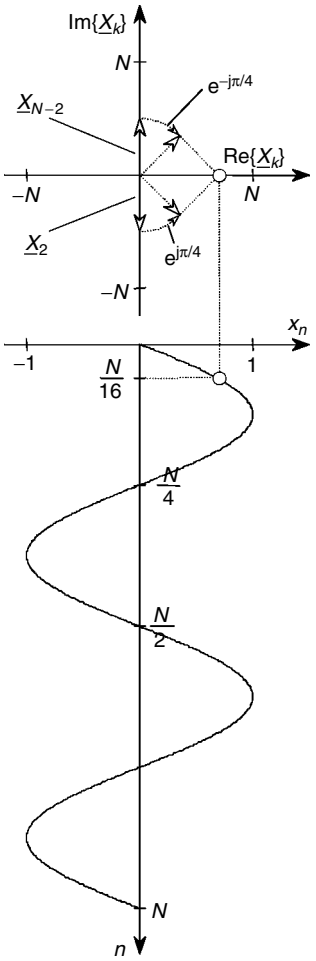


Figure 7. Graphical interpretation of the inverse DFT.

where $\nu \in \mathbb{N}$, you get the same positions of this arrow at different combinations of nk/N. Figure 8 shows this effect for $N = 16$ samples for 4 different values of k.

A very efficient algorithm using this effect was published by Cooley and Tukey in 1965. Their algorithm needs about $N \log_2(N)$ multiplications and additions. This is the reason why this algorithm is $N/\log_2(N)$ faster. Table 1 shows this factor for different numbers of samples. For $N = 16\,384$ samples, the relation of the time difference is nearly the same as 1 h for the discrete Fourier transform (DFT) down to 3 s for the fast Fourier transform (FFT). On the basis of this, it is highly recommended to use $N = 2^\nu$ with $\nu \in \mathbb{N}$ samples for data acquisition.

5 WINDOWING

One problem of data acquisition is to find the right size of the time window with N samples to truncate the sampled function. It is not a problem if it covers one period or its multiples of a periodic signal, but that is usually not the case. The used segment of the sampled

Table 1. Acceleration of the FFT algorithm compared with the DFT algorithm.

$N = 2^\nu$	ν	Acceleration ($t_{DFT}/t_{FFT} \approx N/\nu$)
16	4	4
256	8	32
1024	10	102.4
16 384	14	1170.3

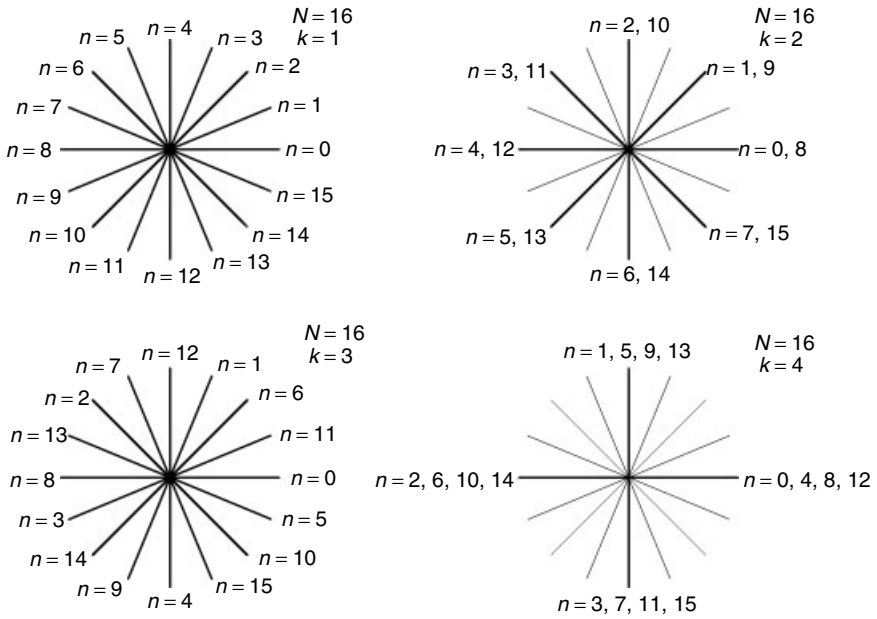


Figure 8. Arrow positions around the unit circle used for the FFT algorithm.

time signal can be regarded as a multiplication of the original analog signal with a single pulse selecting the time window. The spectrum of the samples is the convolution of the spectrum of the time function with the spectrum of the single pulse (time window or truncation window). In this case, the spectrum is continuous and the DFT (or FFT) algorithm determines values of the continuous spectrum at the frequencies $f_0 = 1/T_0$ and its multiples up to $(N-1)f_0$. The behavior of the continuous function between these points is unknown. I would prefer a different view. If we use the discrete spectrum of the DFT (mean values are zero), we get a description of a periodic signal in the time domain, where the N samples cover one period. In fact, the continuous signal of the analog function between these samples is unknown, but better to imagine. If this periodic signal describes the signal you want to analyze, the complex spectral values can be used. The problem with the periodic extension is to get a continuous periodic signal at the edges of the used period. Figure 9(a) shows two sampled sine waves of 20 periods and Figure 9(b) shows 20.5 periods in the time window. If we extend this period, we get the original signal in the case in Figure 9(a) and based on this we get a very good spectral description indicating 20 periods in the time window. If we extend the period in Figure 9(b), we get a nonharmonic connection between these periods. This produces a distribution or leakage distortion of the spectral line as shown in the amplitude spectrum in Figure 9(b). The reason is the nonharmonic connection of the periodic extension.

There have been several attempts at creating a weighting function or truncation window function of the used time window. Each can be regarded as a multiplication of this window with the time signal. The result is the convolution

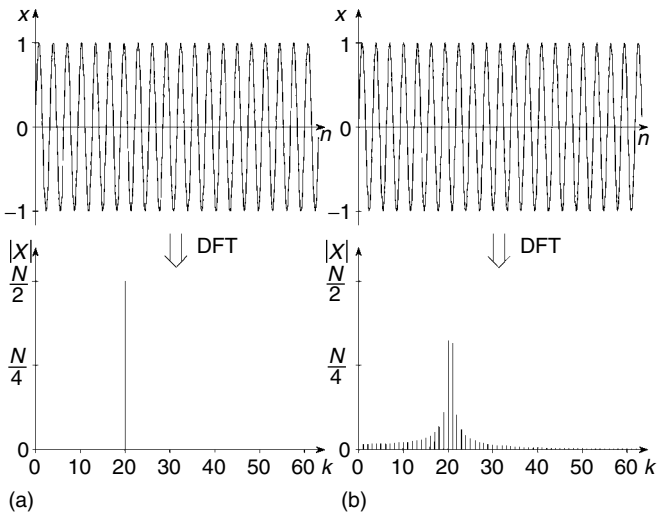


Figure 9. Amplitude spectrums of sampled sine functions.

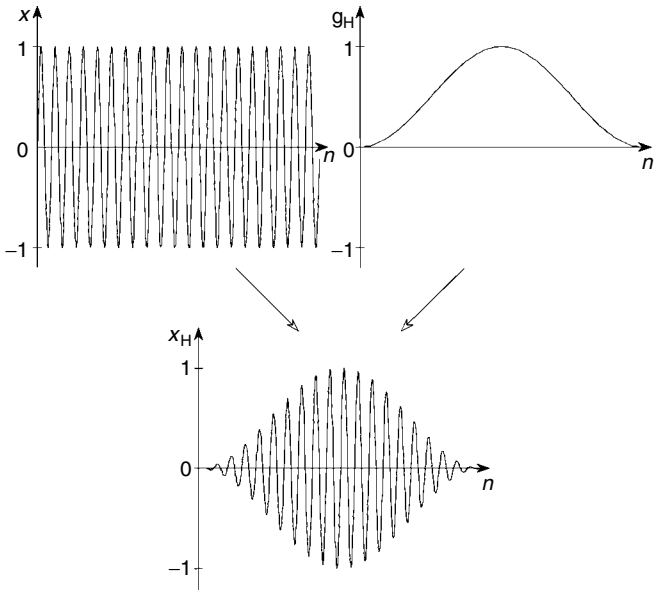


Figure 10. Function of the Hanning weighting window.

of both spectrums (*see also* **Article 31, Discrete Signal Theory, Volume 1**). An often used weighting function is the Hanning window, where each sampled value is weighted with a value according to (19).

$$g_{\mathrm{H}n} = \frac{1}{2}\left(1 - \cos\left(\frac{n \times 2\pi}{N}\right)\right) \qquad n \in \{0, 1, 2, 3, \ldots, N-1\} \quad (19)$$

Figure 10 shows the behavior in the time domain.

In this case, the periodic extended signal of the sampled time window has got an amplitude modulation with the frequency f_0. This creates side bands if we have only multiple periods in our time segment of N samples as shown in Figure 11(a). If we have fractions of such a signal as shown on the right-hand side in Figure 10, we get only a small leakage distortion of the original spectrum at 20.5 f_0 of the spectrum. It is much easier to analyze such a spectral description.

6 AVERAGING IN TIME AND FREQUENCY DOMAIN

Quite often the signal we want to analyze is covered in noise. This means that we have a signal, for example, a square wave function, we want to analyze and disturbing noise. One data acquisition process with N samples provides a spectrum, which does not give us the necessary accuracy. If the process is continuous, and if we can measure several time windows with the same conditions, we

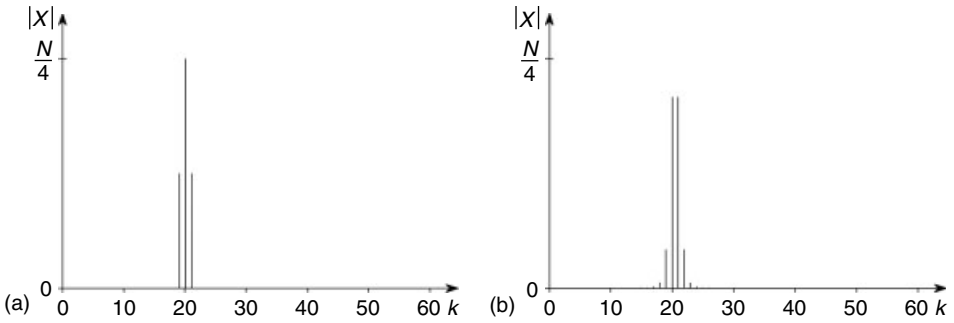


Figure 11. Amplitude spectrums of sampled sine functions with Hanning weighting window.

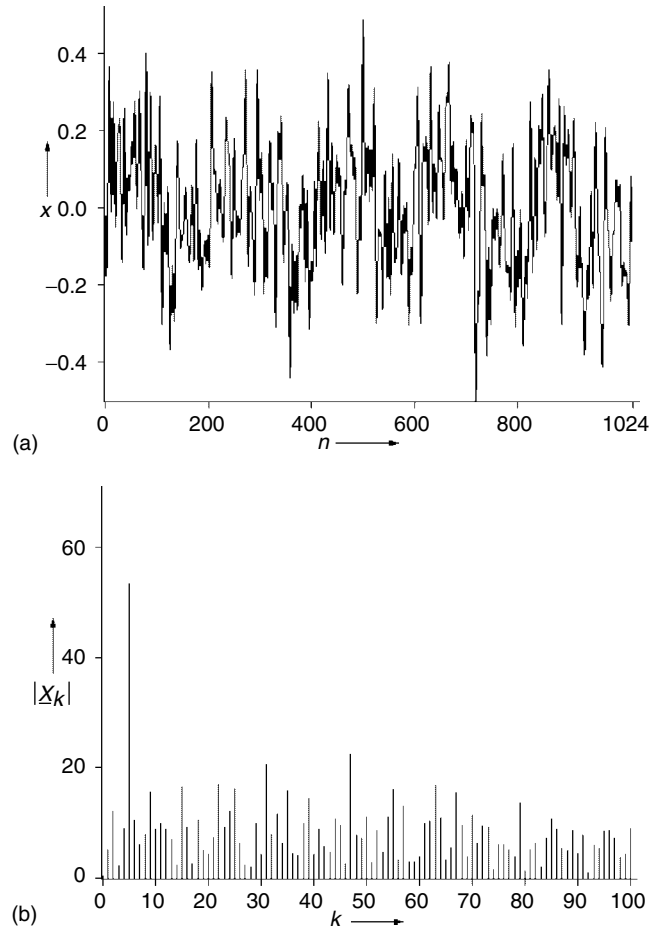


Figure 12. Square wave signal and noise in the time and frequency domain.

can use an average algorithm. Let us assume that the signal has a stable amplitude and frequency, and that the noise has a spectrum where the values in each time segment varies around a mean value. The behavior of such spectral components can be described with a mean value and a standard deviation. Figure 12 shows a measured square wave signal of 5 periods with noise sampled with 1024 samples and the spectrum up to $k = 100$.

If you cannot use a synchronization with the square wave signal, the phase of all components of our square wave function varies and we can only average the amplitude spectrum of all measured time windows as described in (20). If we use only the amplitude spectrum $|\underline{X}_k|$, we cannot apply the inverse DFT to regenerate the time signal based on the missing phase of the frequency components. Using this average function, we get the spectrum of the signal added with the average value of the noise. The standard deviation of the noise components around its mean value will be reduced. Such a spectrum is shown in Figure 13 with an average over 100 time windows ($M = 100$). The frequency components of the square wave signal can be seen easily.

$$\overline{|\underline{X}_k|} = \frac{1}{M}\sum_{m=1}^{M}|\underline{X}_{km}| \qquad k \in \{0, 1, 2, 3, \ldots, N-1\} \quad (20)$$

If we can get a synchronized time window for data acquisition, for example, with a separate trigger signal, the phase of the spectral components of our signal will not change in the spectrum and we can apply the average

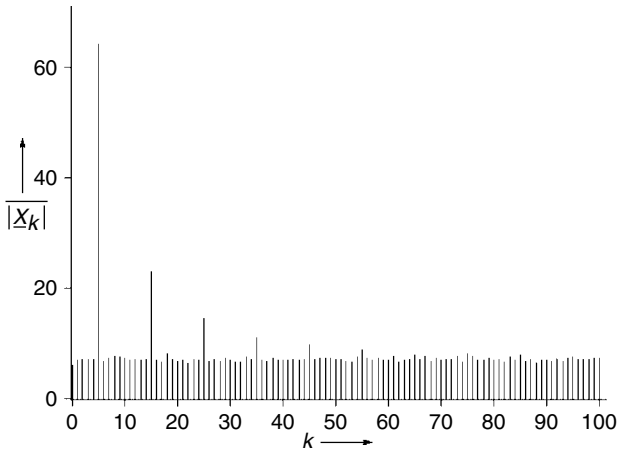


Figure 13. Averaged amplitude spectrum of a square wave signal and noise.

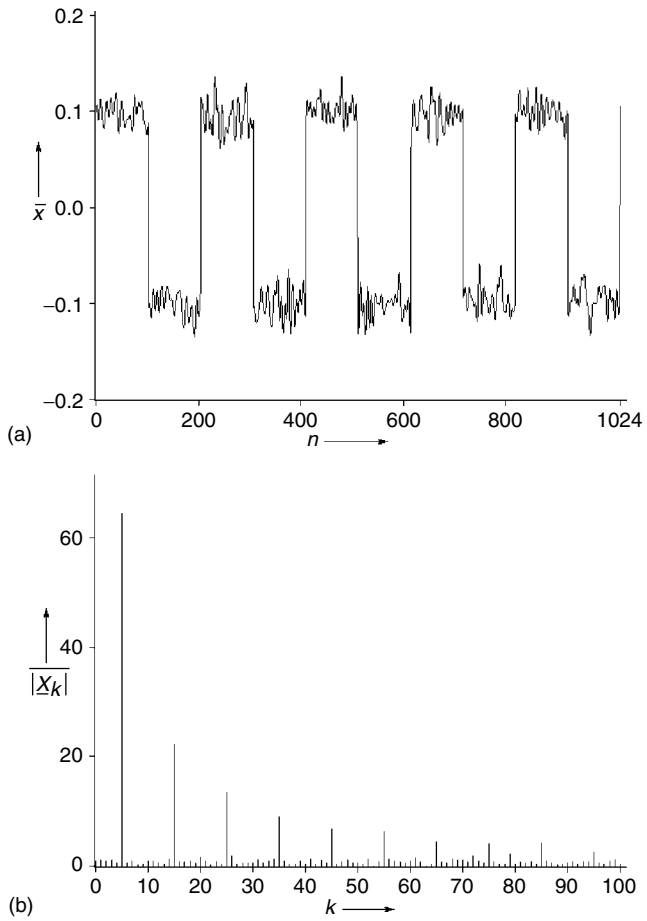


Figure 14. Averaged square wave signal and noise in the time and frequency domain.

algorithm to the time signal. The phase of the noise signal is not synchronized and on the basis of this effect the value of those spectral components will be reduced. The average algorithm can be applied to the time signal (21) or to the complex spectral values (22). Figure 14 shows a time signal and its spectrum of 100 time windows ($M = 100$) as presented in Figure 12. The remaining noise becomes much smaller and the frequency components of the square wave signal can be analyzed to higher frequency components compared with the spectrum shown in Figure 13.

$$\overline{x_n} = \frac{1}{M}\sum_{m=1}^{M} x_{nm} \qquad n \in \{0, 1, 2, 3, \ldots, N-1\} \qquad (21)$$

$$\overline{\underline{X}_k} = \frac{1}{M}\sum_{m=1}^{M} \underline{X}_{km} \qquad k \in \{0, 1, 2, 3, \ldots, N-1\} \qquad (22)$$

FURTHER READING

Romberg, T.M., Black, J.L. and Ledwidge, T.J. (1996) *Signal Processing for Industrial Diagnostics*, John Wiley & Sons, Chichester.

Sydenham, P.H. (1983) *Handbook of Measurement Science, Volume 1 Theoretical Fundamentals*, John Wiley & Sons, Chichester.

144: DSP Chip Sets

Iain Paterson-Stephens

University of Derby, Derby, UK

1 Elements of Digital Signal Processing (DSP) 989
2 Basic DSP Functions 990
3 Implementation of DSP Functions in Microelectronic Form 991
4 Programmable Chip Sets 992
5 Application Specific Integrated Circuits (ASICS) 993
6 Introduction to Gate Arrays 993
7 Processors in General 994
8 Measurement System Chip Sets 995
References 996
Further Reading 996

1 ELEMENTS OF DIGITAL SIGNAL PROCESSING (DSP)

Most Digital Signal Processing (DSP) devices incorporate similar functions and features, the reason for this is clear when we consider the fundamental algorithms for which they are designed to process. The majority of DSP algorithms currently in widespread use fall into the following categories: Filtering of signals (convolution), Transformation of signals (e.g. time to frequency domain) or Correlation (comparison of signals), (Ifeachor and Jervis, 2001).

Considering the first of these terms; 'Digital filtering', generally refers to a numerical operation governed by the following relationship

$$y(i) = \sum_{k=0}^{K} b(k)x(i-k) - \sum_{m=1}^{M} a(m)y(i-m) \qquad (1)$$

Without really going into the mathematics of this relationship, it can be seen that the process of digital filtering a signal requires a number of addition $\mathbf{\Sigma}$, subtraction, and multiplication $a(m)y(i-m)$ operations. In the relationship, the values $b(k)$ and $a(m)$ represent two tables of filter coefficients stored in memory. These will be multiplied by tables of past input $x(i-k)$ and output $y(i-m)$ samples respectively. Each time a new input sample arrives it will be placed on the top of the input sample table and the oldest sample removed. Similarly, new output samples will be placed onto the top of the output sample table and the oldest removed; hence, both data tables always have a fixed length for any given filter design. Figure 1 depicts the basic arrangement for a single table of coefficients and past output samples.

The process outlined above indicates the features that will be required on any general purpose DSP device. It should be able to efficiently access and manage its memory areas and simultaneously handle the flow of data into and out of its data ports. For real-time operation, the DSP device will also require a fast arithmetic unit, special memory addressing modes, and good support for interfacing to external I/O.

Another example of a common DSP algorithm is the Fast Fourier Transform (FFT), which is essentially an efficient method for calculating the Discrete Fourier Transform (DFT). The FFT was first described in 1963 by Cooley and Tukey, and has become widely used for obtaining the frequency domain representation of time domain signals. The mathematical relationship used to describe the DFT is given in (2).

$$X(K) = \sum_{n=0}^{N-1} x(n)W_N^{nK}, \quad \text{where } W_N^{nK} = \mathrm{e}^{-\frac{j2\pi nm}{N}} \qquad (2)$$

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.

© 2005 John Wiley & Sons, Ltd.

ISBN: 0-470-02143-8.

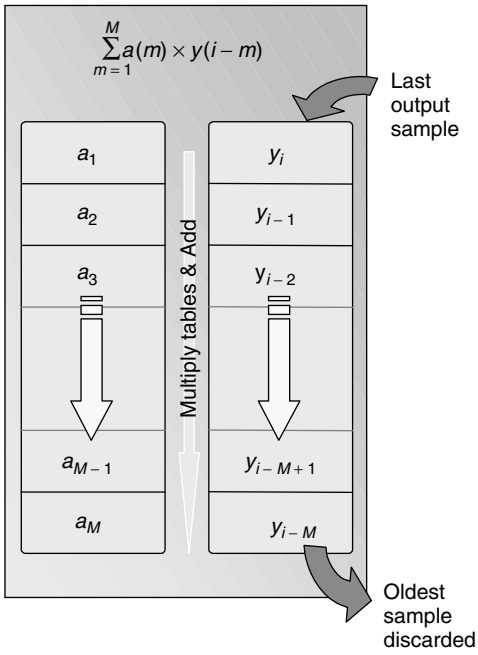


Figure 1. Filter coefficients and samples.

It can be seen that the relationship for the DFT again calls for storage of a list of coefficients W_N^{nK} (often referred to as *Twiddle factors*) in memory along with a list of sample data. The frequency resolution of the DFT is determined by the number of sample points used in the calculation and hence this directly determines the number of sample points/twiddle factor values that need to be stored in memory. The requirement for efficient data access and memory management is therefore a requirement for this DSP application as are a fast arithmetic unit and specialist memory addressing modes.

Correlation, Convolution, and Power Spectral Density are all examples of functions related to the Discrete Fourier Transform. These common signal processing functions can all be executed in the digital domain using schemes similar to that shown in the equation for the DFT, (2).

In a typical system using a DSP device, the sampled data will arrive with a regular period at the input to the DSP. On the arrival of each data sample, the DSP will carry out the calculations described by the algorithm and the resulting data will be presented to the DSP output, again at a regular period. Clearly, the DSP must be quick enough to complete algorithm processing before the next new data sample arrives. This characteristic is essential for any device that will be used successfully to implement DSP algorithms in real time.

It is apparent, from the expressions for the DFT and digital filtering, that the DSP device must carry out numerous arithmetic operations. In addition, extra steps must usually be taken by the DSP programmer to ensure that calculation errors will not occur under normal operation. Most DSP devices incorporate a number of arithmetic features that ease this process such as saturation arithmetic and overflow detection.

2 BASIC DSP FUNCTIONS

Following the previous discussion about the nature of typical DSP algorithms, it will be apparent that a number of characteristics are desirable for any DSP device, which is used to implement real-time DSP algorithms. Some of these fundamental DSP device functions are summarized in the following section.

2.1 The 'multiply accumulate' operation (MAC)

Each of the fundamental signal processing algorithms discussed in the previous section have operated upon data samples by calculating a running sum (accumulation) of a series of products. A DSP device must therefore be able to accumulate a series of multiplication results, this is known as the *Multiply Accumulate*, or MAC operation. The MAC is one of the most fundamental and distinct features that should be found on any dedicated DSP device.

2.2 Efficient data access–specialized memory addressing

The DSP must be able to access its memory areas as efficiently as possible so that the processor does not waste time waiting for data to be fetched. In the case of the MAC operation, which typically runs repeatedly very quickly for short bursts of time, it is important that an efficient MAC operation is not held up while waiting for data to be fetched from memory. Typically, DSP devices will provide a number of different methods for fetching data and determining its address within memory. The FFT, for example, makes use of so-called bit-reversed addressing. The output of an FFT algorithm is usually 'bit reversed', that is, the data points get out of sequence in a predictable way as a result of the FFT operation. This is rectified by accessing the data points using a bit-reversed addressing scheme that most DSP devices provide.

Another DSP specific addressing mode is the circular buffer, depicted in Figure 2, in which an area of memory can be made to appear as a continuous loop. Typically, data pointers are used to indicate the current data input location. Each time a data transfer takes place, the pointer is shifted one place further round the buffer.

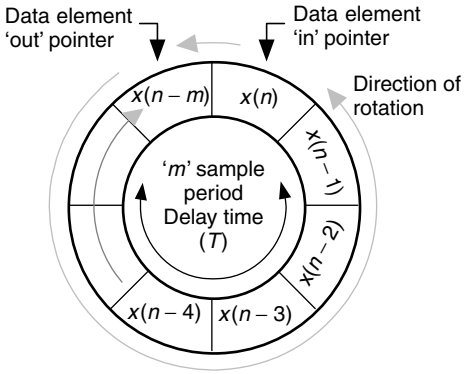


Figure 2. DSP circular buffer.

2.3 Deterministic performance

When a DSP algorithm is to be performed in real time using a DSP device, it is important that the programmer is able to predict the amount of processor time the function will take to complete, that is, DSP processing must be deterministic. If the processor operation is not deterministic, then problems can arise when large amounts of data are being processed in real time. Real-time performance is a term used to describe many DSP systems; it can be defined as

> Completing the required processing task quickly enough so that system performance can continue without being hindered by the processing overhead.

2.4 Extensive interfacing capabilities

Real-time DSP systems must be able to process signals that come from real-world sources such as data from an analog to a digital converter. The DSP device must therefore incorporate extensive interfacing capabilities so that transducers, data acquisition devices, memory or a system host can be connected efficiently. Most DSP devices provide on-chip interface components that require minimal external circuitry and only limited intervention by the main core of the DSP processor. The benefit of this is that processor time is not wasted interacting with slow peripherals

2.5 Superscalar operation

Superscalar mode of operation is a term often used in relation to DSP devices. Superscalar operation means that the device can execute multiple instructions in parallel. This explains why a DSP can be rated at 1600 MIPS even though the processor clock is only 200 MHz. In this case, the DSP executes eight parallel operations per instruction cycle. Superscalar operation is utilized heavily in most DSP devices, an example is the common MAC operation in which two instructions are executed in parallel, that is, the Multiply and Accumulate. Usually, the MAC is run in parallel with other operations such as data fetches and address pointer updates all of which operate simultaneously in the same instruction cycle. The result is that hundreds of millions of operations can (in principle) be performed per second.

2.6 Hardware repeat loop

Programmable DSP devices have instruction sets that contain unique features for DSP algorithm work. One particular feature, which is not generally available on a microprocessor, is hardware support for instruction looping. This recognizes that DSP algorithms consist of compact code segments that need to run very efficiently at full speed.

For example, in a filtering application, the MAC instruction will be run repeatedly in short bursts each time operating on different data sets. Without hardware looping, this process would be very inefficient as a loop counter variable would need to be maintained. Using a DSP device equipped with a zero overhead loop facility, the loop counter is maintained in hardware, easing the burden on the processor core. Also, the instruction is fetched from memory once and run many times thus giving a further potential speed increase.

3 IMPLEMENTATION OF DSP FUNCTIONS IN MICROELECTRONIC FORM

It is possible to perform DSP functions in real time using a standard microprocessor as opposed to a dedicated DSP device. In fact, many of the techniques pioneered on early DSP devices have now been adopted for use on general purpose microprocessors. In addition, many microprocessor and microcontroller manufacturers quote the ability of their device to perform DSP functions, after all, there is nothing sacred about a DSP function – it is merely a mathematical relationship. The issue is whether the device is capable of performing the function quickly enough for the intended application so as to give real-time performance. Typically, a standard microprocessor will not be able to efficiently perform DSP functions such as a MAC operation, simultaneous access of two data tables in memory or highly efficient bit-reversal addressing. In such a case, it would be necessary

for the programmer to devise some sort of software scheme to produce these functions, although this will never be as quick as the dedicated DSP hardware equivalent.

Take the humble multiply operation as an example; all general purpose DSP devices are provided with dedicated hardware to perform a multiply that typically provides calculation results within one processor cycle. In comparison, a typical microprocessor carries out its multiply operation by a binary long multiplication process. When it encounters a multiply instruction, an internal sequence of operations, called a *microcode*, is invoked. This microcode performs the multiplication as a sequence of shifts and adds on successive clock cycles until the result is complete. Since the microcode has many steps, the operation requires many clock cycles to perform. The resulting overhead for multiplication on a microprocessor could be approximately 80 processor clock cycles to perform a 16-bit multiplication. Both devices can perform a multiply; its just that the DSP device is generally quicker (Bateman and Paterson-Stephens, 2003). Of course, money is also a factor; if you are prepared to pay for an expensive microprocessor, then you are likely to get more hardware features such as a dedicated multiply unit.

4 PROGRAMMABLE CHIP SETS

The programmable, or general purpose DSP device, typically encompasses all, and more, of the features that have been presented in the preceding section. The programmable DSP may be reconfigured to suit an extremely wide range of different applications. Manufacturers of these devices usually provide software development tools and evaluation/development platforms that help engineers to design and develop new applications. Often manufacturers will provide extensive customer support, design services, access to third party vendors, and ready coded software libraries downloadable from the Internet. It has been the case that general purpose programmable DSP devices have been used in small volume applications where only tens of thousands, or fewer units are expected to be sold. DSP devices are produced by a number of manufacturers including, in no particular order, those identified in the following list by:

- Lucent Technologies
- Texas Instruments
- Philips
- Analog Devices
- Motorola
- ST Microelectronics
- Harris Semiconductors
- Microchip

Certain manufacturers have gained a reputation for their devices within particular industries, usually because they have incorporated some additional targeted features. For example, Analog Devices and Motorola DSP devices have gained wide acceptance in the professional audio product market, whereas devices from Texas Instruments are used extensively within telecoms applications. In the case of the pro-audio application, this has possibly been due to the development of low-cost 24- and 32-bit processors within the product range of Analog Devices and Motorola.

All programmable DSP chip sets have a range of software development options and tools available from the manufacturer. Often, DSP code is written using the C or C++ programming language, which helps to ensure portability between different DSP platforms, see Figure 3. C/C++ is often the first choice for developing signal processing functions; however, device specific assembly language is commonly used for low level operations that deal directly with I/O devices and other hardware elements. Assembly language is also commonly used for time-critical operations in which optimum performance cannot be achieved using C code, for example, processor interrupts associated with signal I/O are typically written using assembly language (Bateman and Paterson-Stephens, 2003).

Unfortunately, no standard DSP programming language has yet emerged so it is necessary for the designers to familiarize themselves in great detail with the specifics of each new device that is used. Although many DSP manufacturers produce a family of different devices covering fixed/floating point, different performance levels (MIPS) and I/O capabilities – often each device within a family will have its own individual quirks, which prevent easy movement of DSP code from one device to another.

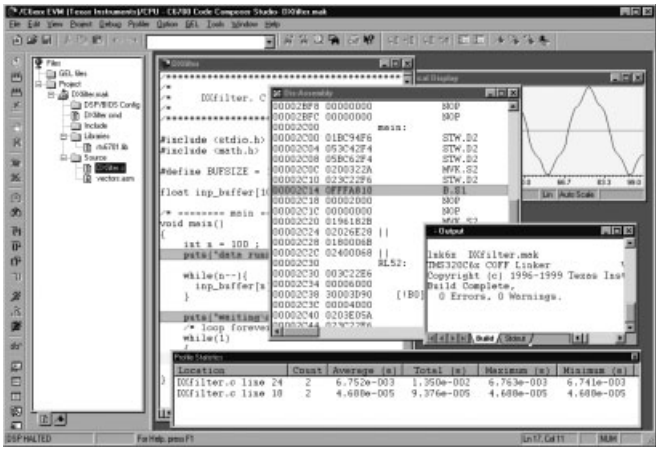


Figure 3. C/C++ program development for the TI C6x DSP processor.

5 APPLICATION SPECIFIC INTEGRATED CIRCUITS (ASICS)

This type of device is expensive to develop and production is only viable for high volume applications where sales targets are in the hundreds of thousands of units. Application Specific Integrated Circuit (ASIC) based DSP designs are expensive to develop because the level of simulation and testing required before going to fabrication is usually more rigorous – obviously, with a programmable device it is possible to make (software) changes to the design after the system has been manufactured, whereas an ASIC design, once fabricated, is fixed – mistakes are consequently very expensive. Development tools sets and test jigs for ASIC are also somewhat specialized and are consequently expensive. The ASIC option does however provide a cost effective route if the sales volumes are high since an ASIC can be developed to provide the specific functions and features required for an application and as such excess cost for unused features can be minimized (Bateman and Paterson-Stephens, 2003). The trend for general purpose DSP devices on the other hand is toward the integration of large amounts of memory, extensive, and varied I/O facilities and numerous other features. While this is useful for the general case, it is likely that many of these features will not be required in a large number of applications.

It is interesting that manufacturers of some general purpose DSP devices are now providing the option for user specified device features so that a device can be made up from a specific processing core and selected on-chip peripherals. ASIC devices also offer good reliability because they usually only perform the specific tasks for which they have been designed. Further more expensive silicon area can be minimized and because many different system functions can be brought onto the ASIC itself, the final system design can be optimized in terms of the circuit board (PCB) space required. One further issue relating to the minimization of circuit functions in silicon is that ASIC-DSP designs used for low power applications, such as mobile telephony, for example, can be optimally low power for the specific application, that is, no excess or unused circuitry means that power consumption can be kept to the minimum.

One example of an ASIC device is a surround sound decoder chip used for consumer audio applications developed and produced by consumer product manufacturers such as Yamaha or Sony. Clearly, in this type of application worldwide distribution of large numbers of units is predicted.

6 INTRODUCTION TO GATE ARRAYS

Gate Arrays and specifically Field Programmable Gate Arrays (FPGA) are finding many applications in new DSP designs. FPGAs in particular offer a number of potential advantages over the traditional programmable DSP device (Goslin, 1997). The FPGA offers the opportunity to optimize silicon use to a specific application in much the same way as an ASIC. Additional system functions that would conventionally exist outside the DSP device can potentially be brought onto the FPGA and hence minimization of circuit board space can be achieved. A Gate Array is essentially a grid of unconnected logic cells that can be linked together as required, by the use of an appropriate tool, into a functioning circuit. The process of configuring the circuitry usually fuses a mesh of links between the logic cells. FPGA devices are available in a range of sizes up to complex devices containing tens of thousands of logic elements. One of the advantages of the FPGA design route is that the configuration of circuitry can be very application specific, for example, if the designer knows that the application will only require 13-bit arithmetic, then that is all that needs to be designed. Also, greater use of parallelism can be achieved with an FPGA than a programmable DSP. For example, if a very fast convolution (filtering) processor is required that contains say 16 filter taps, the filter would require 16 multiplies and 16 adds (MACS) per new input sample. In a programmable DSP application, a fast single MAC unit will be used within a hardware loop to perform the filter function in just over 16 instruction cycles. For the FPGA approach, it would be possible to incorporate 16 independent MAC units within hardware design, which can all operate simultaneously within one system cycle. The greater potential use of parallelism in the hardware design can therefore offer greater performance than that is achievable on a general purpose programmable DSP device (Goslin, 1997).

Development and production costs are significantly lower than the ASIC route and for many engineers the development process is more familiar compared to the use of a programmable DSP device. Because the design of a DSP function within an FPGA comes down to a description of hardware, for example, multipliers, delays and so on, many engineers are already familiar with the development process and tools. Typically, a graphical CAD-based approach to design can be used or the use of an appropriate Hardware Descriptor Language (HDL). All potential DSP functions can be defined in terms of a description of hardware. Many manufacturers provide downloadable HDL libraries that define typical DSP functions such as Convolution, Correlation, and Frequency Transforms. Also, a number of programmable DSP cores from manufacturers such as Texas Instruments and Analog Devices can be purchased in HDL language form for integration into an FPGA design (Bateman and Paterson-Stephens, 2003).

7 PROCESSORS IN GENERAL

There is an important device architecture difference between a classic microprocessor compared to the DSP device, that is, the overall structure of the memory itself within the device. Most microprocessors are designed around the Von-Neuman architecture, shown in Figure 4, in which the program instructions and data items share the same memory space and hence are accessed via the same address and data buses. This strategy can be slow because the CPU needs to fetch each instruction (op-code) to be executed followed by fetching any associated data (operands) before it can start processing. Each of these tasks takes up precious time, that is, processor clock cycles.

Most DSP devices use a style of processor architecture known as the Harvard Architecture (shown in Figure 5). Using Harvard Architecture the program and data items are stored in separate memory spaces, and are accessed by independent processor buses. An example of this is DSP devices from Motorola, which make use of a Dual-Harvard Architecture with two independent data spaces, X and Y, as well as a program memory space. Using Harvard Architecture enables program and data memory to be accessed simultaneously, giving a speed advantage over a conventional microprocessor architecture (Bateman and Paterson-Stephens, 2003). DSP devices usually have sophisticated Address Generation Units (AGUs), which generate data addresses using specific addressing modes. Modulo addressing is unique to DSP devices, and is useful for automatically implementing circular buffer memory constructs and bit-reversed addressing accesses with no software overhead. The ability to auto increment and decrement values in the address registers is also useful. The term software overhead refers to an additional code, which would need to be in place to perform a function if the specific feature was not available.

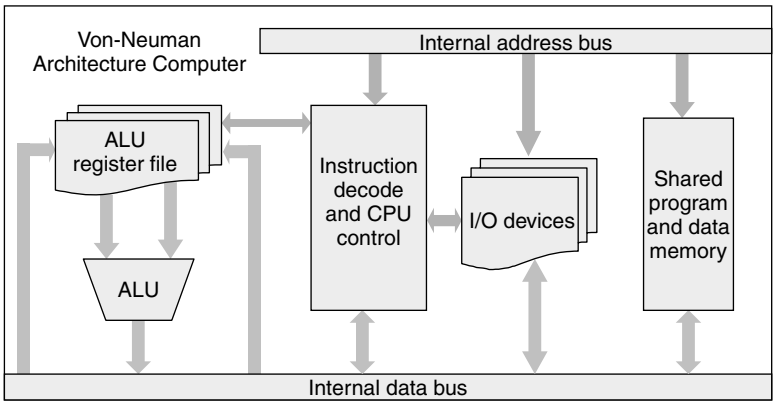


Figure 4. Von-Neuman Architecture Computer.

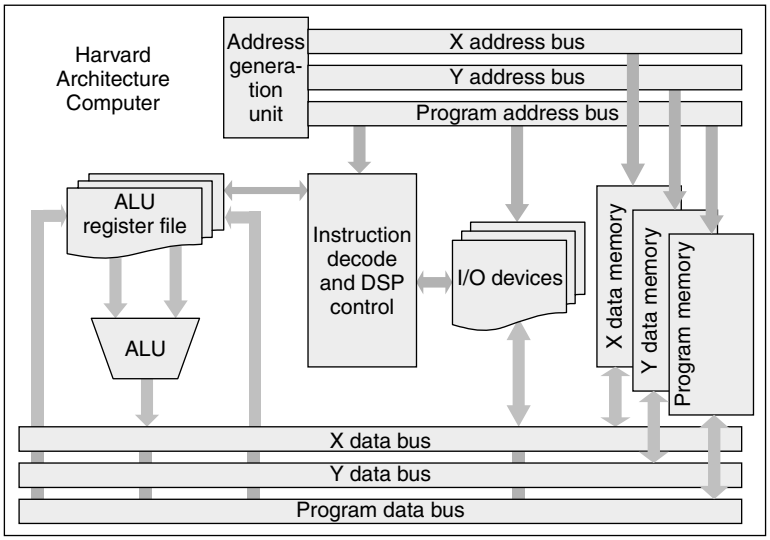


Figure 5. Harvard Architecture Computer.

Clock cycle	1	2	3	4	5	6	7	8	9
Fetch	F1			F2			F3		
Decode		D1			D2			D3	
Execute			E1			E2			E3
	Instruction cycle one			Instruction cycle two			Instruction cycle three		

Figure 6. Nonpipelined operation.

Clock cycle	1	2	3	4	5	6	7	8	9
Fetch	F1	F2	F3	F4	F5	Etc.			
Decode		D1	D2	D3	D4	D5	Etc.		
Execute			E1	E2	E3	E4	E5	Etc.	
	Instruction cycle one			Instruction cycle four					
		Instruction cycle two			Instruction cycle five				
			Instruction cycle three			Etc.			

Figure 7. Pipelined operation of DSP instructions.

DSP and Reduced Instruction Set Computer (RISC) architectures aim to deliver results from the processing core on every instruction cycle. However, in practice the processing of an instruction goes through a number of phases. The instruction must first be fetched from memory along with any associated operands, then the processing core will need to decode or interpret the instruction before finally dispatching it for execution. This scheme is shown diagrammatically in Figure 6.

The objective of pipelining is to overlap the different phases so that instructions can be fetched prior to the completion of other instructions in the pipeline. Figure 7 shows a pipelined arrangement in which the different phases of the three instructions previously shown in Figure 6 are now allowed to overlap. It can be seen that results start coming out of the core on every clock cycle. This is after a delay of two clock cycles while the first instruction ripples through the pipeline. The ability to pipeline instructions and data and the effectiveness of the pipeline comes largely from the fact that the DSP device has a very efficient memory architecture (Harvard Architecture). Many standard microprocessors attempt to pipeline instructions and data but it is doubtful that the pipeline will be as effective on a Von-Neuman compared to the Harvard Architecture (Bateman and Paterson-Stephens, 2003).

8 MEASUREMENT SYSTEM CHIP SETS

Many manufacturers produce system chip sets incorporating DSP hardware suitable for measurement applications. In fact, most general purpose DSP devices are well suited to this type of application. Standard interface methods have been adopted that allow DSP devices to be hooked up with I/O hardware for simple no fuss system development (TI-SPRA477, 1998). A standard high-speed serial interface known as I^2S has been adopted by most manufacturers, which allows a simple 3- or 4-wire interface between Analog-to-Digital (AD) and Digital-to-Analog (DA) converters and the DSP device (Bateman and Paterson-Stephens, 2003). Usually, AD and DA devices can be obtained for most applications that incorporate all the necessary timing logic and anti-aliasing filters. Furthermore, it is common practice to make use of a CODEC device that combines all the necessary analog input and output hardware, sampling logic, serial interface logic, filters, and so on. These devices offer a very simple and cost-effective way of integrating a complete measurement system into just a handful of chips. The word CODEC is used to describe a device that can be used to COde and DECode data between different formats. In the context we are considering here, the coding is from analog to digital form and the decoding is from digital back to analog. The term CODEC can also be used to describe any coder/decoder device or algorithm, for example, an algorithm that converts between a linear and compressed data format and back again could be described as a CODEC. The block diagram of a standard interface CODEC device is shown in Figure 8.

The CODEC shown in Figure 8 incorporates a parallel port that is generally used as a low speed connection through which the DSP device can send initialization information and monitor various aspects of the CODECs operation. The sample data is usually, though not necessarily, sent via the CODECs serial port, which is connected directly to one of the DSPs on-chip serial interfaces. The serial connection uses a four-wire bus comprising a data clock, frame sync, data in, and data out lines (Jordan, 1996).

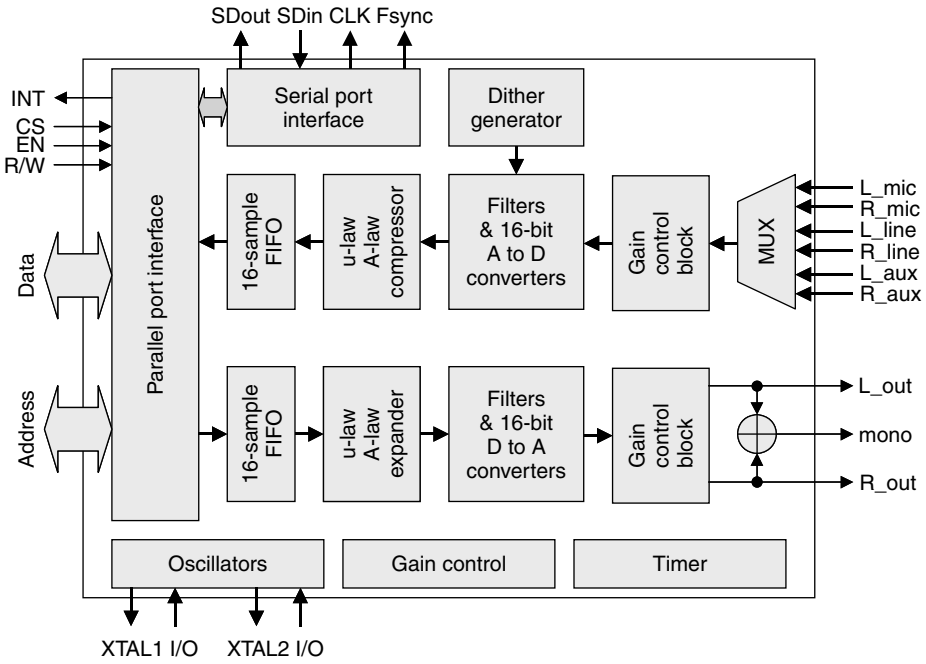


Figure 8. Standard interface CODEC for DSP-based measurement applications.

The CODEC requires only a minimal amount of external support components such as basic I/O buffering and filters, power smoothing, and a crystal clock. The I/O buffering can be made using standard op-amp designs and the filters only need to provide AC coupling that is, an RC filter will suffice. The crystal clocks are used to provide the sample rate for the device and this can be set to operate at any one of a number of standard rates (Crystal Semiconductor Corporation,1994). Because the CODEC incorporates its own sample rate generator, it is not necessary for the DSP to provide this from its internal clock source. Instead, it is common practice in this type of application for the CODEC serial port to be configured such that the CODEC generates the data clock and frame sync signals and the DSP simply synchronizes itself to the CODEC.

REFERENCES

Bateman, A. and Paterson-Stephens, I. (2003) *The DSP Handbook – Applications, Design and Implementation*, ISBN: 0-201-39851-6, Prentice Hall, UK.

TI-SPRA477 (1998) *TMS320C6000 McBSP Interface to the CS4231A Multimedia CODEC*, Texas Instruments.

Jordan, M. (1996) *Hardware and Software Interface Issues for DSPs and Serial Audio CODECs – Application Notes*, Crystal Semiconductor Corporation.

Ifeachor, E.C. and Jervis, B.W. (2001) *Digital Signal Processing – A Practical Approach*, ISBN 0 201 54413X, Addison-Wesley, UK.

Crystal Semiconductor Corporation (1994) *CS4231A Multimedia CODEC Applications Guide*, USA.

Goslin, R.G. (1997) *A Guide to Using FPGAs for Application Specific Digital Signal Processor Performance*, Xilinx Inc, USA.

FURTHER READING

Xilinx Logicore (2002) *MAC FIR V3.0 Data Sheet*, Xilinx Inc, USA.

145: DSP Tools

Iain Paterson-Stephens

University of Derby, Derby, UK

1 Tool Sets in Support of DSP 997
2 Overview of Selected Small- to-large Mainline DSP Tools 997
3 Choice of DSP Tools for an Application 998
4 Introducing DSP Tools in an Organization 999
5 Limitations of DSP Tools 1001
References 1002

1 TOOL SETS IN SUPPORT OF DSP

Design for DSP applications, as with any design, must go through a number of stages and iterations before the final product can be released. For DSP designs, a number of key stages will need to be considered; at a basic level, these can be summarized as algorithm development and design, hardware implementation and software coding and these must be carried out in the context of the end application (Bateman and Paterson-Stephens, 2003). There are, of course, many factors that will influence the path taken for each of these design stages, development cost and time to market being the main driving factors. Many tools are available, which will help the engineer to realize a particular design as quickly as possible and at minimum cost. In particular, tools are available, which will enable algorithms to be simulated at block diagram level in order to verify the design concept (examples include Matlab and Simulink, or Hyperception), see Figure 1. The block diagram can then be automatically translated to a high-level language (HLL), such as C or Ada. The block diagram algorithm to HLL translation can be target specific, that is to say the code generated can be optimized for a particular DSP device (Real-time workshop running under Simulink) (MathWorks–A, 1998). Once the design has been translated, it is compiled and tested on the target device.

The testing and verification process has also advanced somewhat in recent years so it is now possible to monitor, in real time, the operations taking place inside the DSP device. This allows software problems to be identified and resolved as quickly as possible so that the overall design is not delayed (Bateman and Paterson-Stephens, 2003). Standard libraries for software development are also emerging, which means that the design can be built using standard signal processing plug-in functions; for example, in the design of a real-time filter, the engineer can select from a library of standard filter algorithms, add the necessary coefficients to define the filter response, and then finally glue the application together with a small amount of extra software that will manage data flow and other aspects of DSP 'housekeeping.' All this sounds very attractive and in principle the tools are all here and available now to carry out the task as described. However, in reality, the initial cost of setting up this sort of design approach can be high not only in terms of equipment and software costs but also in terms of a high initial learning curve. The final choice of approach will often be determined by other factors not least of all the context of the application, that is, whether the design is likely to be mass produced and sold at a low cost or it is a one-off for research purposes.

2 OVERVIEW OF SELECTED SMALL-TO-LARGE MAINLINE DSP TOOLS

2.1 Visualization tools–algorithm design/verification

Matlab

- Command line and interpreted-code simulation interface

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.

© 2005 John Wiley & Sons, Ltd.

ISBN: 0-470-02143-8.

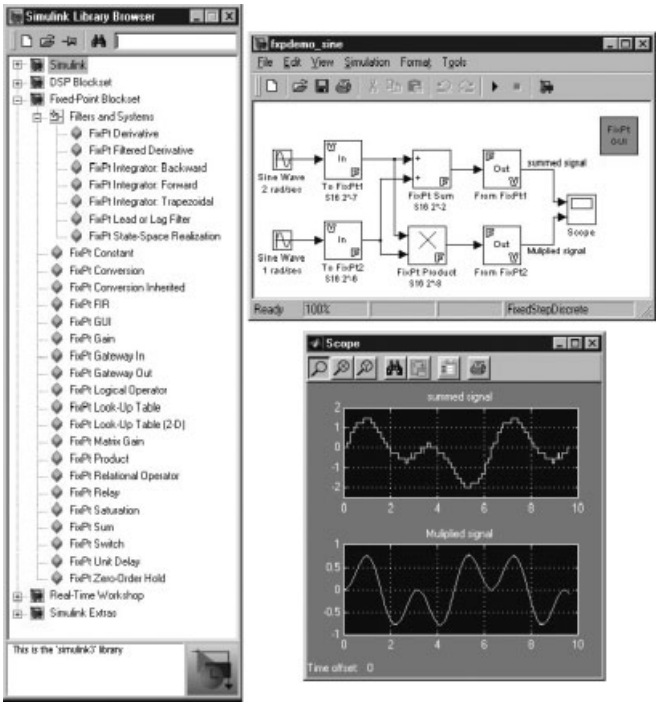


Figure 1. Simulink application modeling screen shot.

- Matrix-based programming language
- Extensive Toolbox support for a range of signal-processing tasks
- Complex graphing/plotting capabilities
- Massive library of support documentation and third party books
- Fast simulations.

Simulink

- 'Graphical front end' for Matlab – intuitive user interface
- All the power of Matlab with extra support for fast simulations
- Extensive range of toolboxes and support libraries
- Real-time mode simulation possible
- Compiled code blocks for very fast simulations
- Interfacing capabilities – able to connect algorithms to hardware
- Compilation of algorithms for target hardware 'Real-time Workshop'.

2.2 DSP system development tools

Visual DSP++ (Analog Devices)

- Can be used with entire range of Analog Devices DSPs
- Integrated design suite used for software and hardware debugging
- Mixed C/C++ and Assembly level programming and debugging
- Easy integration with Simulink target coder
- In-circuit emulation support – real-time debugging
- Support for customized host-control interface designs
- Memory data graphing support – useful for signal processing apps.

Code Composer Studio (Texas Instruments)

- Can be used with entire range of DSPs from Texas Instruments
- Integrated design suite used for software and hardware debugging
- Mixed C/C++ and Assembly level programming and debugging
- Real-time mode debugging – in-circuit emulation
- DSP/BIOS Plug-ins for fast application development/code reuse
- Real-time data exchange to host platform (PC).

3 CHOICE OF DSP TOOLS FOR AN APPLICATION

At the basic level, there are only four things that are required to start developing DSP applications: (i) a low specification host computer – usually windows PC; (ii) a text editor for writing software applications using assembly code; (iii) a low-cost DSP development board that connects to the PC via a parallel or serial cable, see TI-SPRU191 (1996), and (iv) the basic software tools usually supplied with the development board (Bateman and Paterson-Stephens, 2003).

Assuming the user already owns the PC and a simple text editor, the rest can usually be purchased for around $100, and this may or may not include a suitable power supply. Let us start by looking at the last two items from this list, that is, the basic hardware and software tools.

3.1 Basic hardware–a low-cost development board

Most manufacturers of DSP devices offer a low-cost development board, which is intended to be used for a basic evaluation of the technology, see TI-SPRU191 (1996) as an example of this. These boards are usually freestanding and connect to the PC via a slow data link, for example, a serial or parallel connection, see Figure 2.

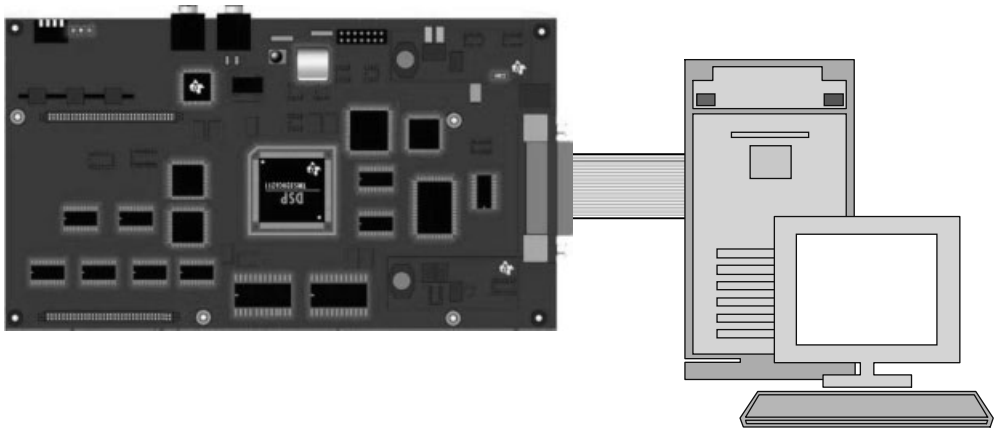


Figure 2. Simple DSP development system.

With the board comes a comprehensive set of software tools for assembling, linking, debugging, and downloading code. Typically, the development board will include data converters so that an audio bandwidth signal can be passed into and out of the DSP device, and generally these are suitable for voice or music applications thus enabling the user to evaluate algorithms using real signals in real time.

3.2 Basic software–a low-cost development environment

In order to get started with DSP software development, a number of basic tools will be needed. These will allow software routines to be written, converted to a form understandable to the DSP, and then downloaded to the target device. Usually, these tools are supplied with a development system as part of the complete package. The following list describes the basic tools required:

- *Basic text editor*: This can be a very simple application such as Windows notepad and is used for writing DSP programs in assembly language.
- *Assembler*: Available from the manufacturer of the DSP device, it is used to convert the users' text-based assembly program into a machine-readable format.
- *Linker*: Organizes the machine-readable code generated by the assembler so that it will match the memory configuration of the target DSP.
- *Debug environment*: Enables software to be tested for the particular DSP device; the debug environment may be in the form of a simulator or an emulator.
- *Downloader*: Often a downloader is provided with DSP development systems so that assembled and linked programs can be transferred to the DSP development board.
- *Hex conversion utility*: Converts assembled and linked DSP code into a form suitable for an EPROM programmer.

An overview of the steps taken during DSP software development is shown in Figure 3. The diagram shows more steps than might typically be required in a basic project; however, these are shown here for completeness (TI-SPRU102C, 1998). At intermediate stages within the diagram, the expected input and output file types are shown.

3.3 More advanced tools

Although the basic setup previously described is fine for the beginner who needs to gain a low-level understanding of the technology and processes involved, the more advanced designer working on a serious commercial project will require development tools that offer many more features (TI-SPRA520, 1999).

All DSP manufacturers offer an extensive range of DSP development tools, for example, Code Composer Studio offered by Texas Instruments, see Figure 4, is a fully integrated software package that can be used for the whole development process in a very similar way to Microsoft Visual C++. Visual DSP++ integrates editor, assembler, and compiler into one package along with extensive features to help manage projects, allow code reuse (via libraries) and target debugging (TI-SPRU303, 1999). In effect, the advanced tool sets integrate all of the stages identified in Figure 1 along with many more features.

4 INTRODUCING DSP TOOLS IN AN ORGANIZATION

Many of the tools and techniques used for designing DSP hardware and/or software will be familiar to engineers

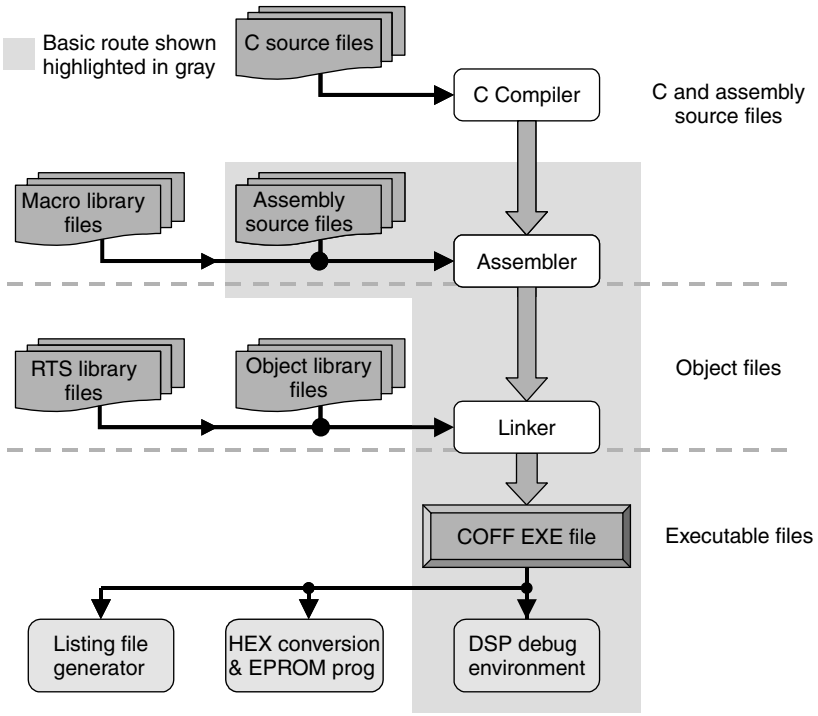


Figure 3. DSP program development flow.

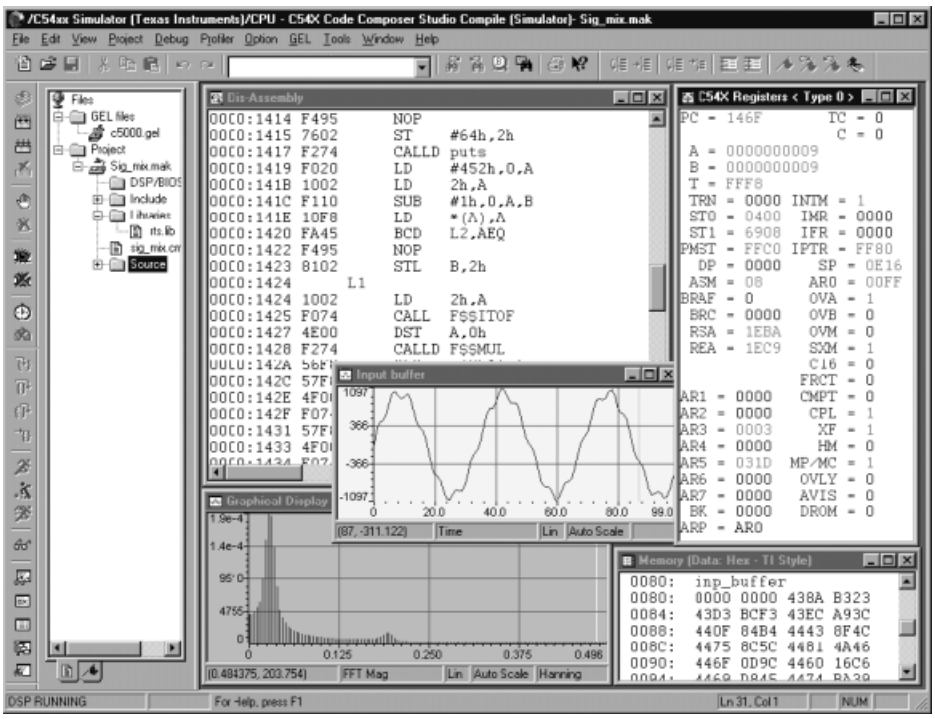


Figure 4. Screen shot of TI Code Composer Studio DSP design interface.

already involved in the design of conventional microprocessor-based electronic products. Similarly, engineers involved in design of analog electronic products, for example, control systems or filter designs, will be familiar with the algorithmic simulation and visualization tools, and complex mathematics. The problem often faced as far as a DSP application design is concerned is the very broad range of skills and knowledge that will be required of the engineer.

In addition to knowledge about the application, the DSP engineer should have a reasonably good grasp of assembly programming, C/C++ programming, hardware design and interfacing, microprocessor architectures, DSP algorithms, and discrete mathematics. That said, the DSP engineer will usually work as part of a design team where each individual will have his or her own spatiality.

When introducing DSP tools into an organization, the choice as to which tools will be the most appropriate is usually predetermined by the choice of DSP device and manufacturer. All DSP manufacturers and a number of third party organizations provide short courses on the use of their specific tools sets and hardware design with their own DSP chips. It is often very worthwhile arranging for key staff to attend manufacturers' own short courses, if for no other reason than to ensure that 'undocumented features' of the development tools can be understood at the outset rather than allowing them to plague a design for a period of time during the development process.

5 LIMITATIONS OF DSP TOOLS

On the face of it, a DSP processor is very similar to a conventional microprocessor; both have address and data buses, a central processing core and memory, and so on. In fact, if a DSP processor is programmed inefficiently without much regard for the subtle differences a DSP hardware architecture has to offer, the DSP will perform pretty much the same as a standard microprocessor. In order to exploit fully the potential speed increases for certain applications, the programmer must be truly aware of what the device has to offer. DSP features such as pipelined operation, multiple data paths and storage, Multiply and Accumulate (MAC) and other arithmetic features/modes of operation, and so on are only accessed when requested specifically by the programmer. When programming a DSP device, the safest path in many ways is to use the native assembly language offered by the manufacturer. Assembly language has the advantage that it can be very efficient and well optimized, and because it is so closely related to the specific DSP architecture, the programmer is likely to exploit the subtleties of the DSP architecture itself. The downside to assembly language is that it is quite difficult to learn, it is device specific and therefore the programmers will need to update their knowledge for each new device, and it is not easy to read and hence not easy to debug.

An alternative to coding using assembly language is to use an HLL such as C or C++; this has the advantage of not being target specific and so the skills of the programmer are more generic and usable across a range of devices. C/C++ code is clearly an industry standard programming language and so many programmers will find this approach to be a very accessible route into DSP programming. HLLs present

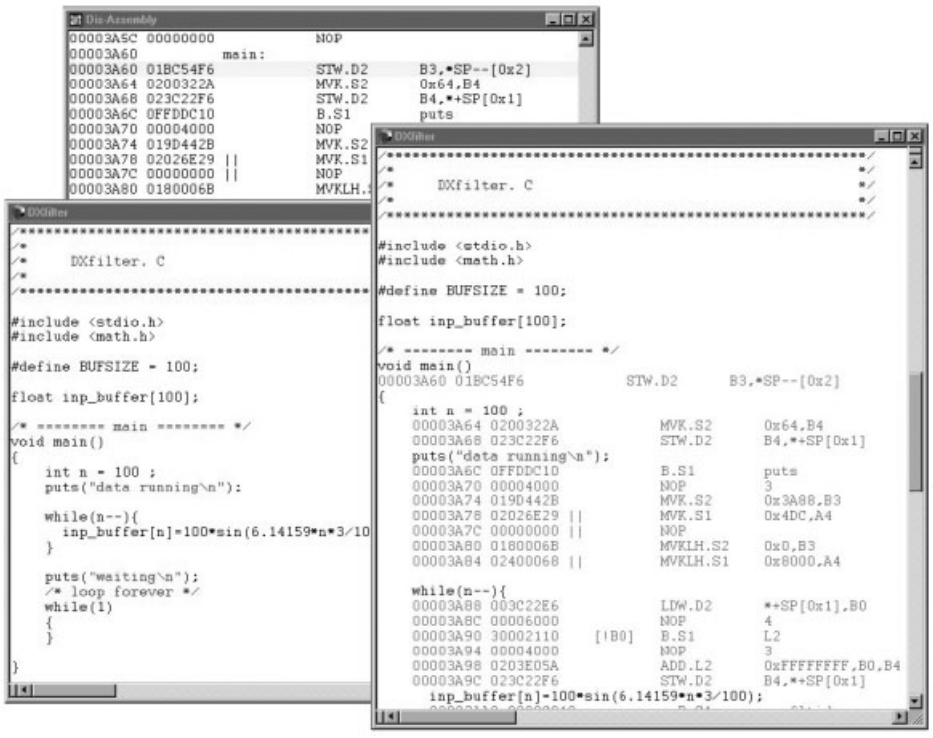


Figure 5. C source debugging using Code Composer Studio from Texas instruments.

their own problems, in particular that of code optimization (Bateman and Paterson-Stephens, 2003). Because HLLs are somewhat abstracted from the target hardware, the compiled code is usually not well optimized for the device. Usually, tool manufacturers provide a range of software switches that can be turned on/off to force the compiler to compile HLL code in a certain way so as to improve optimization. At times, however, the programmer will be required to hand code sections of HLL generated assembly language, see Figure 5. In particular, when designing routines that perform interfacing functions, it is necessary to resort to hand-coded assembly language in order to meet the timing requirements of the interface.

Simulation tools such as Matlab, Simulink, and Hyperception provide a useful platform for developing algorithms and the user to interact with elements of a design before it has been committed to a hardware or embedded software implementation. These are intuitive tools that allow the designer to 'play' with algorithms and to gain a greater understanding of how the final design will operate. These simulation tools work very well and give some good insight into a problem; however, there are potential pitfalls

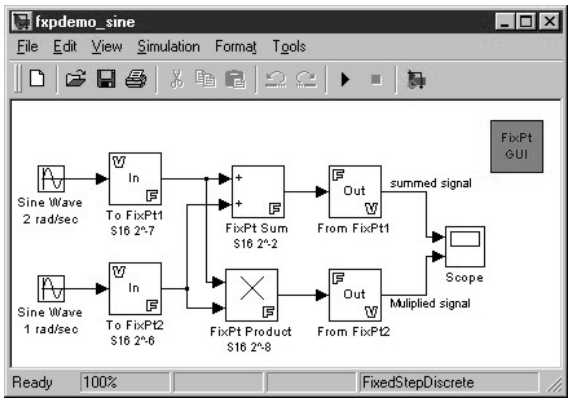


Figure 6. Fixed point DSP simulation tools within Simulink.

(MathWorks-B, 1999). A pertinent example for DSP application design is the effect of fixed point processing and how these are to be considered in the final design. When simulating a design using a visualization tool, such as Simulink, the simulation is performed perhaps using a 64-bit floating point desktop PC. The design may work well in simulation; however, when it is translated over to the target hardware, perhaps a 16-bit fixed point processor, the performance may be somewhat different (MathWorks-C, 1999). Simulation tools usually incorporate features to help incorporate fixed point processing into a simulation, see Figure 6, and this can help resolve any differences.

As far as limitations of design, simulation, and implementation tools are concerned, there are potential problems and pitfalls but if the user takes the time to familiarize themselves with tools, many problems can be avoided.

REFERENCES

Bateman, A. and Paterson-Stephens, I. (2003) *The DSP Handbook, Algorithms, Applications and Design Techniques*, ISBN: 0-201-39851-6, Pearson Education/Prentice Hall, UK.

TI-SPRU191 (1996) *TMS320C54x DSK Users Guide*, Texas Instruments, http://www.ti.com/.

TI-SPRU102C (1998) *TMS320C54x Assembly Language Tools Users Guide*, Texas Instruments, http://www.ti.com/.

TI-SPRA520 (1999) *Code Composer Studio White Paper*, Texas Instruments, http://www.ti.com/.

TI-SPRU303 (1999) *TMS320C6xxx DSP/BIOSII Users Guide*, Texas Instruments, http://www.ti.com/.

MathWorks–A (1998) *Guide to Rapid Prototyping for DSP Design, The MathWorks – DSP Design Technical Examples*.

MathWorks-B (1999) *Using Simulink – Version III Users Manual*, The MathWorks Inc.

MathWorks-C (1999) *Real-time Workshop – Users Guide*, The MathWorks Inc.

146: Principles of DSP Hardware Design

Iain Paterson-Stephens
University of Derby, Derby, UK

1 General Principles of DSP Hardware Design 1003
2 Integration of DSP Hardware with Software Element 1005
3 Suggested Process for DSP Design 1005
4 Tools in Support of DSP Development 1006
5 Installation and Testing of DSP Developments 1007
References 1007

1 GENERAL PRINCIPLES OF DSP HARDWARE DESIGN

Before proceeding to design any application specific digital signal processing (DSP) hardware, it is important to have a detailed understanding of the intended application. This goes beyond just the sheer number crunching requirements of the selected DSP; it is also important to fully appreciate in detail the I/O requirements and how these may impact on the algorithmic performance of the DSP. While the DSP device must be able to perform specific numerical functions very efficiently, it must also have the capability to interface with other devices and systems so that minimum disruption occurs to the processing core (Bateman and Paterson-Stephens, 2003). Most DSP devices incorporate functional units that can operate with some independence from the core and are able to carry out tasks such as interrupt handling, direct memory accesses (DMA), serial interfacing, and timing operations, (TI-SPRU131d, 1997). In this section, typical system architectures will be presented in order to put many of the DSP facilities into context within different applications from a hardware point of view. Single and multiprocessor architectures will be considered.

1.1 A generalized DSP system

The block diagram shown in Figure 1 represents a standard DSP system architecture in which provision has been made for connecting various peripheral devices.

Looking at Figure 1, it is apparent that in addition to processing signals coming from an external source such as an ADC or CODEC, the DSP may need to handle events originating from a host computer or a user control surface. In normal operation, the DSP runs synchronously with the ADC or CODEC as this provides the regular sampling interval. On arrival of each new sample, the DSP carries out the necessary processing and makes the new output sample available to the DAC or output side of the CODEC. For example, in a professional audio system, the new samples would arrive at a rate of 48 000 per second and the DSP must keep up with this flow of data. If the DSP loses synchronization, then, in the case of this audio example, an audible click would be heard, that is, data would be lost. It is, therefore, imperative that the DSP is able to synchronize with external data flowing into and out of its ports. Taking this audio example a little further, at the same time as samples arriving at 48 000 per second, the DSP may be interrupted by other external devices; the host interface may have new control information updating a particular algorithm; the control surface may have updates from the user requiring a filter response to be recalculated or signal levels to be changed. The host and user events are not likely to be synchronous with the arrival of sampled data and yet the DSP must be

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

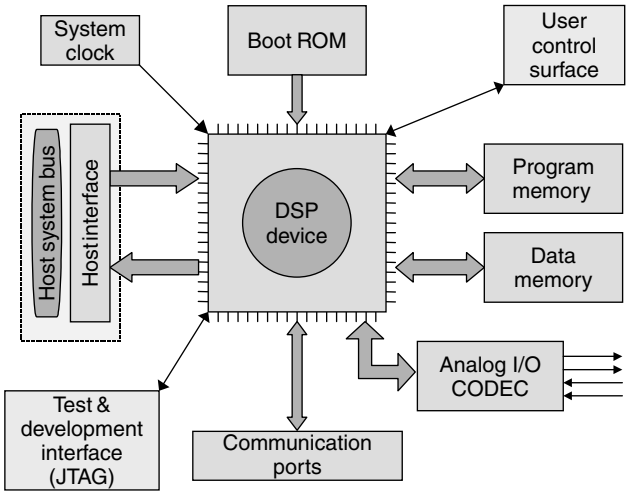


Figure 1. Generalized DSP system architecture.

able to attend to these external interruptions and still maintain full synchronization. There are a number of hardware strategies that can be used by a DSP device to handle transactions with its peripherals. These are summarized briefly as follows.

1.1.1 POLLED I/O

The DSP can POLL external devices to see if they have new data available. In a POLLED system, the DSP simply checks an external pin or memory location to see if a predetermined condition is met, for example, is bit one set high? If the condition is TRUE, then an action can be taken, for example, collect data items from memory. Although the POLLING system is very simple, it does have the disadvantage that the DSP is continually checking for the condition being met rather than getting on with more useful work.

1.1.2 Interrupt driven

All DSP devices incorporate a strategy for interrupting the normal processor flow. External interrupts are provided in the form of a number of physical pins on the outside of the DSP chip each of which will be allocated an individual interrupt number and an associated location in program memory at which a service routine will be stored. The interrupt service routine is usually a short program provided by the user, which carries out a task associated with a particular interrupt.

For example (see Figure 2), if a press button is provided for the user of the DSP system, the button could be connected to an interrupt input pin. When the button is pressed, the DSP is halted from its normal flow and switches to run an alternative routine stored at a different location in memory, called the *interrupt service routine*. On completion of the interrupt service routine, the DSP goes back to where it was prior to the interrupt.

1.2 Multiprocessor systems

Real-time performance for some applications is beyond the capabilities of a single DSP device. In this situation, the only possibility is to use a multiprocessing configuration in which a number of DSP devices are linked together (see Figure 3), and the algorithm is divided into subtasks and distributed among them.

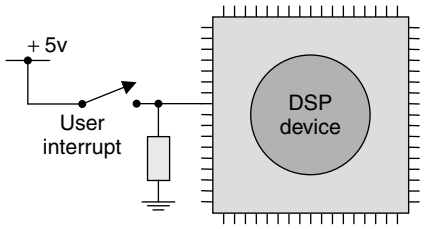


Figure 2. User interrupt.

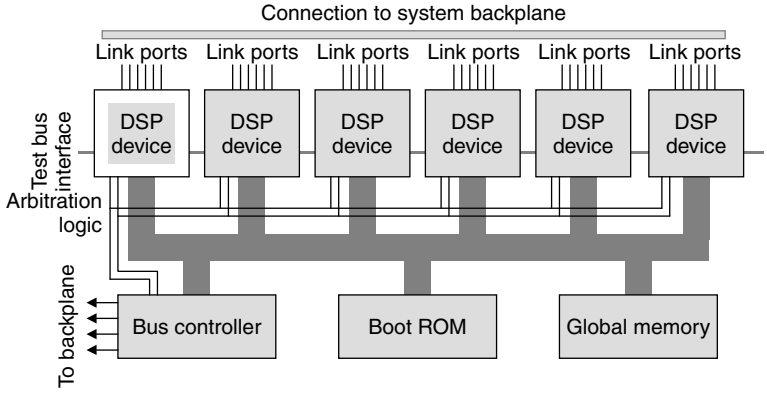


Figure 3. Multiprocessor DSP system design.

Some DSP processors designed with inbuilt multiprocessing features are available. These include devices from Texas Instruments such as the TMS320C4x and C6xxx family. The Analog Devices SHARC family also provides extensive multiprocessing capability through its six serial Link Ports each with a data bandwidth of 40 Mbps.

2 INTEGRATION OF DSP HARDWARE WITH SOFTWARE ELEMENT

The process of integrating the software and hardware elements of a DSP project starts with the generation of source code or a source file. The term *source file* refers to a text-based file written by the programmer, which contains a sequence of instructions; this is the starting point for many programs written for DSP devices. The source file, if written using assembly language instructions, is then converted, using an assembler, into an object file. Object files contain the actual machine readable instructions understood by the particular DSP device targeted. Usually, an object file is said to be relocatable and as such cannot be used directly by the DSP device; instead, it must first be passed through a linker program that produces the final executable DSP program. Although a relocatable program contains all the correct machine instructions, they have not been allocated to specific memory locations within the DSP. One of the functions of the linker is to resolve this memory allocation issue so that all instructions within the final executable program have a specific memory address. The benefit offered by this approach is that relocatable code is not tied to a specific memory structure or even system architecture; therefore, the design, either hardware or software element, can be updated during the life of the product without much limitation (Bateman and Paterson-Stephens, 2003).

The name often used to reference a final executable program is *object module* – note the difference with an object file – the module is the final packaged program that will run on the DSP device. When the object module has been produced, it can be loaded in to the DSP device; this may be via a debug environment hosted on a PC or through the use of an EPROM and a HEX conversion utility. The standard object file and object module format used on the Texas Instruments DSP devices is the COFF format, this stands for Common Object File Format. The HEX conversion utility is used to convert the object module into correctly ordered binary data, which is structured in such a way as to suit the EPROM, EEPROM, or other nonvolatile memory devices (TI-SPRU190b, 1998).

A common approach used in current DSP designs is to use a serial EEPROM to contain the system boot code and any boot-up data that will be required for the application. When the DSP is powered up, the DSP device can be configured to immediately download all of the system code from the serial EEPROM into the DSP's own 'on-chip' memory. The DSP then performs a soft reset and immediately runs the recently downloaded code that is sitting in its memory. One of the benefits of using a serial EPROM like this is that they are often very compact, 8-pin, devices that consume very little PCB real estate and have a very simple 3-wire connection interface (see Figure 4), which keeps the circuit complexity to a minimum.

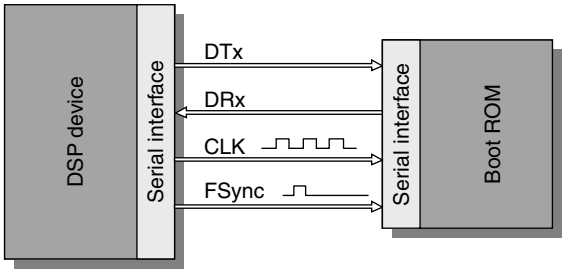


Figure 4. Serial boot ROM using standard 3-wire interface.

3 SUGGESTED PROCESS FOR DSP DESIGN

Every design is different and each engineer will have his or her own preferred methods, but a good starting point for any DSP application is to first consider the basic algorithms that will be required and to design a functional simulation of the system using Matlab and Simulink or similar visualization packages, see Figure 5. This step helps the designer to

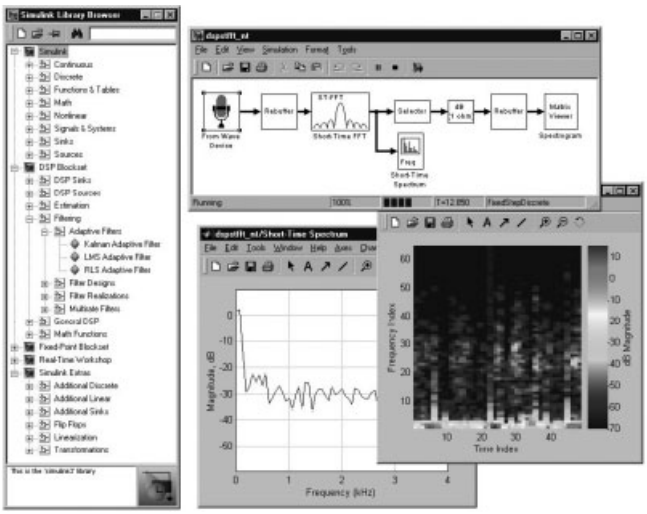


Figure 5. A screen shot showing Simulink while running a basic simulation.

fully understand all of the algorithmic elements and likely number crunching demands of the processor that will need to be used.

The benefit of this initial stage is that the application can be considered in detail without the complications of sometimes unpredictable I/O hardware or the quirks of a specific DSP device. Although this stage is a purely simulated one, it is still possible to make some initial predictions about the various aspects of the final design, for example: the memory storage requirements for filter coefficient tables, data storage, and even code storage requirements.

In order to select an appropriate processor and final system architecture, it is helpful to understand the algorithmic demands as detailed above, and the interfacing requirements for the application must also be considered in detail. For example, does the application require multiple channels of fast analog I/O, a user interface, a host port, or other interfacing mechanisms?

After considering a range of issues about the application, its interfacing requirements and algorithmic demands, it is possible to draw up a broad system specification. This can then be used to help the designer make choices about an appropriate DSP device. Once a device or at least a family of devices has been selected, it is advisable to rework the system specification and to consider in detail how the performance of the algorithms will be affected by the specific architecture of the DSP device selected (Bateman and Paterson-Stephens, 2003). It should be possible, at this point, to make pretty accurate predictions about the number of machine cycles that will be required for each signal processing task taking into account the limitations and overheads inflicted by the I/O, memory, and other hardware structures on the device.

When an appropriate decision has been made about the device to be used, the detailed hardware design can follow. Usually, manufacturers provide very detailed application notes giving recommended hardware configuration and in many cases complete system designs that can be adapted to suit a particular application. This is always a good starting point because much of the initial design has been done, leaving only the application specifics requiring attention. Most manufacturers even provide complete hardware solutions in CAD form so that the initial design can be quickly imported and adapted.

4 TOOLS IN SUPPORT OF DSP DEVELOPMENT

One of the biggest factors in the development of a DSP application is the suitability of the design tools available. In the past, DSP development tools were a rather clumsy collection of different applications that needed to be used in the correct order so that the final design could be realized. Nowadays, DSP development tools have been merged such that initial coding and syntax debugging tools have been integrated into a complex suite usually containing a compiler, assembler, code debugging tool, and a project management tool (TI-SPRU269, 1998). Most of these new integrated DSP development packages allow real hardware to be connected such that code can be tested in real time with 'online' monitoring of performance, this has greatly eased the process of debugging complex software with multiple events occurring in real time.

Code profiling allows the user to investigate in detail the timing of different sections of code within an application. This is particularly useful for complex real-time applications in which multiple threads of execution may be taking place. The designer is able to run the application in real time and investigate where most of the processor time is being spent. The designer can then make choices about optimization of code within specific algorithms in order to bring the final design within the specification of the device and/or application.

Although DSP development tools such as those already described allow very detailed analysis of the code running on the DSP device itself, there is still a useful place for the DSP device simulator in the armory of tools available to the designer. A simulator is simply a software application running on a host computer that simulates all of the function of the target DSP. Simulators do not run in real time, but they are used to run assembly and C code in the same way as the target device. One of the benefits of using the simulator is that it can be easier to test certain elements of an algorithm 'offline' without the distraction of interfacing components and associated software getting in the way. Also, the simulator usually has more detailed tools for profiling code and assessing where efficiency gains can be made.

The mechanism by which DSP hardware can be monitored, probed, and debugged in real time is the JTAG standard and its associated interface. JTAG actually stands for the Joint Test Action Group, which is a working group composed mainly of leading electronic manufacturers who set out to establish a common standard for in-circuit testing and emulation. The result was the IEEE 1149.1 JTAG standard. Of particular interest to the DSP developer is the ability to use the JTAG interface when testing software. In this mode of operation, the DSP device can be single stepped, register values and memory areas monitored, and forced and peripheral devices configured. Integrated packages such as Texas Instruments Code Composer Studio and Analog Devices Visual DSP++ are examples of hosted

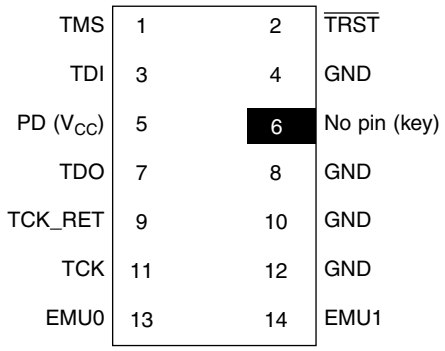


Figure 6. Standard 14-pin JTAG interface connector.

software packages that communicate with target hardware entirely through a fast JTAG emulation port. All aspects of software debugging, loading software, testing, and uploading data to the host are performed via the JTAG link. The JTAG support provided on the target DSP will run in parallel with the operation of the DSP device so that, for example, the value within a target register could be monitored without the need to halt the DSP itself.

The physical connection between the host computer and the DSP's JTAG port is made via a standard 14-pin header, see Figure 6. The header carries serial data between the target device and the host via the TDO (test data out) and TDI (test data in) connectors. The test clock, which is used to synchronize all data transfers, is carried on the TCK connection and a return clock signal is carried on the TCK_RET connection. Pin six acts as key so that the test cable cannot be installed incorrectly and other connection pins such as TMS, EMU0, and EMU1 are used to control various operating modes of the interface.

5 INSTALLATION AND TESTING OF DSP DEVELOPMENTS

There is no substitute for 'in-the-field' testing of DSP designs as is the case for any other design. An application may work perfectly well in principle on the bench back at the design lab, but when it is taken out into the field and tested in its intended environment, many other factors can affect performance. Noise induced onto power supplies, electromagnetic interference and inappropriate/unexpected use by users can all be factors.

The JTAG interface mentioned in the previous section is also useful for in-the-field testing just as it was for the development and debugging stage of the project. Usually, a 14-pin JTAG interface connector will be included on the final DSP board so that in-the-field testing and problem solving can be carried out. The IEEE 1149.1 JTAG standard, in addition to the 'software level' debugging mode, describes a set of rules by which a common test bus system can operate throughout a system level design. The idea of the test bus system is that complex components within a system can be linked to a common bus through which they can be probed and tested while still in circuit. This provides a simple mechanism by which modern systems containing a collection of highly integrated densely packed chips can be tested with relative ease. The JTAG standard allows the user to force conditions within a device as required to perform a particular test, so, for example, using the JTAG interface, a user could force the logic level of certain connection pins on a device to specific value and subsequently observe the resulting operation. This technique is often referred to as *boundary scan*, because the boundary conditions, that is, those of the external connection pins can be monitored and controlled. A subset of the JTAG standard also allows other aspects of a device's performance and operation to be controlled (TI-SPRU173, 1996). Using the JTAG interface, the final DSP system installation can be fully tested and evaluated at software, hardware, and system levels.

REFERENCES

Bateman, A. and Paterson-Stephens, I. (2003) *The DSP Handbook, Algorithms, Applications and Design Techniques*, ISBN: 0-201-39851-6, Pearson Education/Prentice Hall, UK.

TI-SPRU131d (1997) *TMS320C54x CPU and Peripherals*, Reference set, Vol. 1. Texas Instruments, USA, http://www.ti.com/.

TI-SPRU190b (1998) *TMS320C62xx/C67xx Peripherals Reference Guide*, Texas Instruments, USA, http://www.ti.com/.

TI-SPRU173 (1996) *TMS320C54x Applications Guide – Reference Set*, Vol. 4, Texas Instruments, USA, http://www.ti.com/.

TI-SPRU269 (1998) *TMS320CC6xxx EVM Users Manual and Technical Reference*, Texas Instruments, USA, http://www.ti.com/.

147: Ideal Digital Filter Approximation

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Digital Filter Categories 1008
2 Response in the time and Frequency Domains 1008
3 Approximation to the Ideal Digital Filter 1009
Related Articles 1011
Further Reading 1011

1 DIGITAL FILTER CATEGORIES

This is the first article on digital filters. It provides an introduction to their nature and how they are characterized in the time and frequency domains. A general bibliography on the subject is given at the end of this article. Articles **Article 148, General Performance of the Digital Filter, Volume 3**; **Article 149, Low-, High-, and Band-pass Digital Filters, Volume 3**; **Article 150, Finite Infinite Impulse Response (IIR) Digital Filters, Volume 3** and **Article 151, Finite Impulse Response (FIR) Digital Filters, Volume 3** provide detail of types.

A filter is a system that selectively changes the waveshape, amplitude, and phase spectra of a signal in a desired manner. The purposes of the changes are to improve the quality of the signal, for example, to reduce noise, to extract information from the signal, or to separate signals previously combined to make an efficient use of a communication channel.

Filters can be classified in two groups according to signal form:

- *Analog filters* – these filters process continuous signals and are therefore an example of continuous linear systems. They are constructed from linear electrical components.
- *Digital filters* – these are filters formed as sampled data signals. They may be realized by hardware or by suitable programming of a digital computer.

Classification of filters according to the frequency ranges, which are transmitted or rejected, is as follows:

- A *low-pass filter* – has a passband in the low-frequency region;
- A *high-pass filter* – transmits only high frequency input signals;
- A *band-pass filter* – has a passband in a particular frequency band;
- A *band-stop filter* – rejects only a particular frequency band.

Idealized response magnitude characteristics of various types of analog filters are sketched in Figure 1.

A sampled signal has frequency components up to $\omega = \pi/T$ rad s^{-1} and, therefore, the digital filter is classified according to its effect on frequency components in the range $-\pi/T < \omega < \pi/T$. Idealized response magnitude characteristics of various types of the digital filters are sketched in Figure 2.

2 RESPONSE IN THE TIME AND FREQUENCY DOMAINS

A filter may be described either by its frequency response, transfer function, $H(j\omega)$, or by its impulse response, $I(t)$. For example, the impulse responses of Figure 3 have a

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

form of decaying oscillations at a frequency of about ω_0 rad s^{-1}, with zero average value. The corresponding frequency responses will therefore display strong transmission in the region of $\omega = \omega_0$, and rejection of zero frequency inputs.

The pole-zero approach may be used to demonstrate the relationship between the magnitude and phase response of the linear filter. The pole-zero configuration and frequency response characteristic of a typical reactive filter are shown in Figure 4.

Analog filters are composed of passive electrical elements: resistors, inductors, and capacitors. The effect of practical components and terminations is to move the poles and zeros slightly away from the imaginary axis. This means that infinite peaks and zeros will not be observed in practice.

Response characteristics for the elementary low-pass, high-pass, and band-pass filters are shown in Figure 5.

The use of three types of linear electric circuit element, resistors, capacitors, and inductors, enables zeros to be placed anywhere in the s-plane and poles to be placed to the left of the imaginary axis. The use of an active element in a filter obviates the need for inductors. Low- and high-pass digital filters are covered in **Article 149, Low-, High-, and Band-pass Digital Filters, Volume 3**.

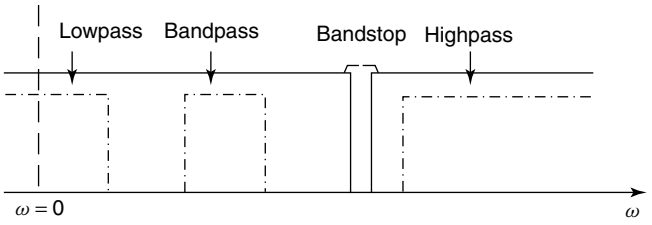


Figure 1. Idealized response magnitude characteristics of various types of analog filters.

3 APPROXIMATION TO THE IDEAL DIGITAL FILTER

The transfer function of a filter is given by

$$H(s) = \frac{1}{a_n s^n + a_{n-1}s^{n-1} + a_{n-2}s^{n-2} + \ldots a_1 s + a_0} \quad (1)$$

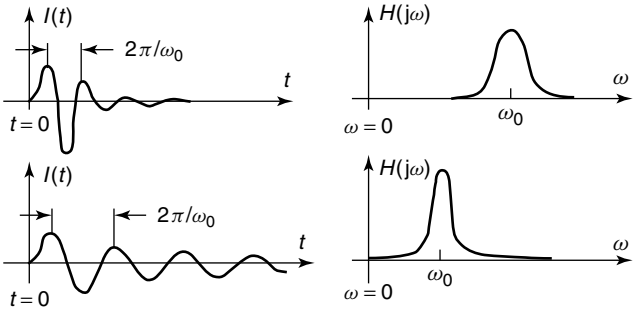


Figure 3. Impulse and frequency responses of two analog band-pass filters.

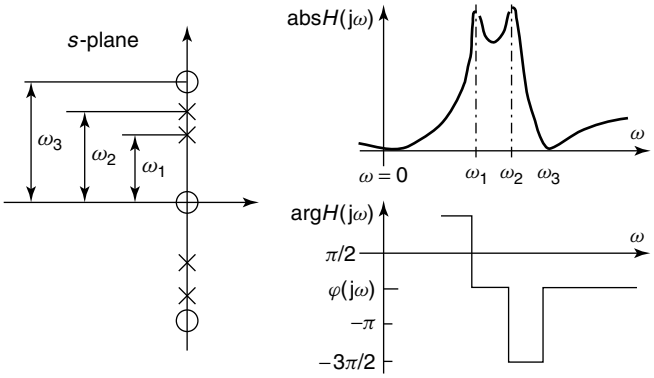


Figure 4. Characteristic of a reactive band-pass filter.

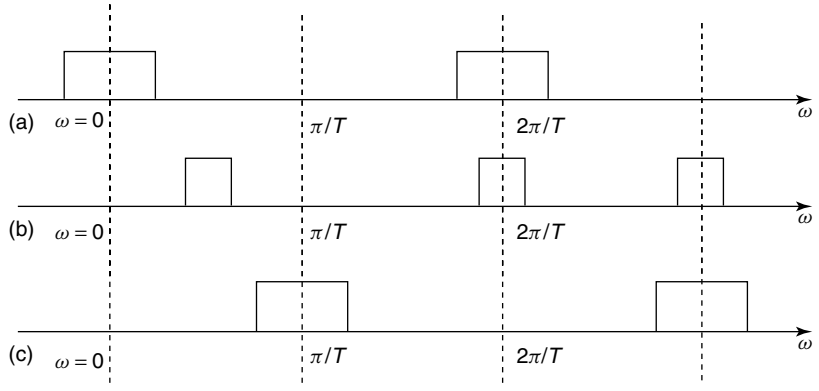


Figure 2. Idealized response magnitude characteristics of digital filters; (a) low-pass, (b) band-pass, and (c) high-pass.

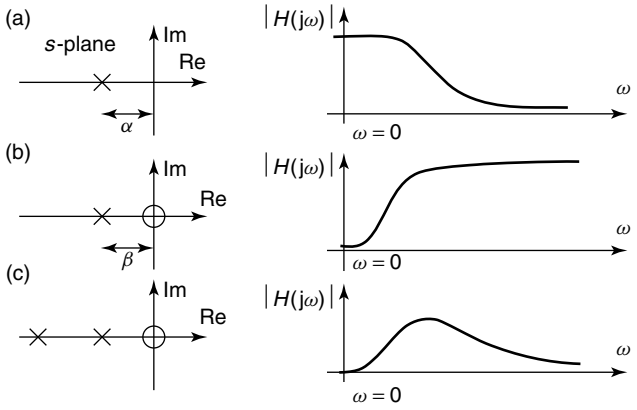


Figure 5. Typical resistors–capacitor (R–C) filter characteristics; (a) low-pass, (b) high-pass, and (c) band-pass.

where the factors of the denominator polynomial represent the poles of $H(s)$. Its frequency response is

$$H(s)|_{s=j\omega} H(j\omega) = \frac{1}{an(j\omega)^n + a_{n-1}(j\omega)^{n-1} + a_{n-2}(j\omega)^{n-2} + \ldots a_1(j\omega) + a_0} \quad (2)$$

and the square of its response magnitude is given by

$$|H(s)|^2 = H(j\omega)H*(j\omega) = H(j\omega)H(-j\omega) = \left\{\frac{1}{a_n(j\omega)^n + a_{n-1}(j\omega)^{n-1} + a_{n-2}(j\omega)^{n-2} + \ldots a_1(j\omega) + a_0}\right\} \times \left\{\frac{1}{a_n(-j\omega)^n + a_{n-1}(-j\omega)^{n-1} + a_{n-2}(-j\omega)^{n-2} + \ldots a_1(-j\omega) + a_0}\right\} \quad (3)$$

This approximates the ideal low-pass filter (Figure 6).

Figure 7 illustrates two types of polynomials having these characteristics. The denominator polynomial should have a magnitude close to 1.0 over the passband and as large as possible over the stopband.

The left-hand figure shows several functions of the form

$$f_1(\omega) = \left[1 + \left(\frac{\omega}{\omega_0}\right)^{2n}\right] \quad (4)$$

where n is a positive integer.

The right-hand plot of Figure 7 shows functions of the form

$$f_2(\omega) = \left[1 + \varepsilon^2 C_n^2\left(\frac{\omega}{\omega_0}\right)\right] \quad (5)$$

where C_n denotes the so-called Chebyshev polynomial of order n in the range $-\omega_0 < \omega < \omega_0$. The value of expression $\varepsilon^2 C_n^2(\omega/\omega_0)$ oscillates between 0 and ε^2 for any value of n.

The use of the polynomials, (4) and (5), gives rise to the well-known Butterworth and Chebyshev filters, (see Figure 8), which are defined by the following squared

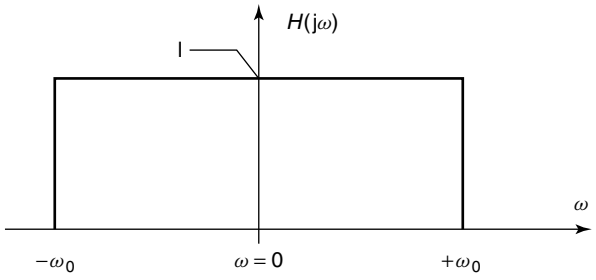


Figure 6. Approximation of the ideal low-pass filter.

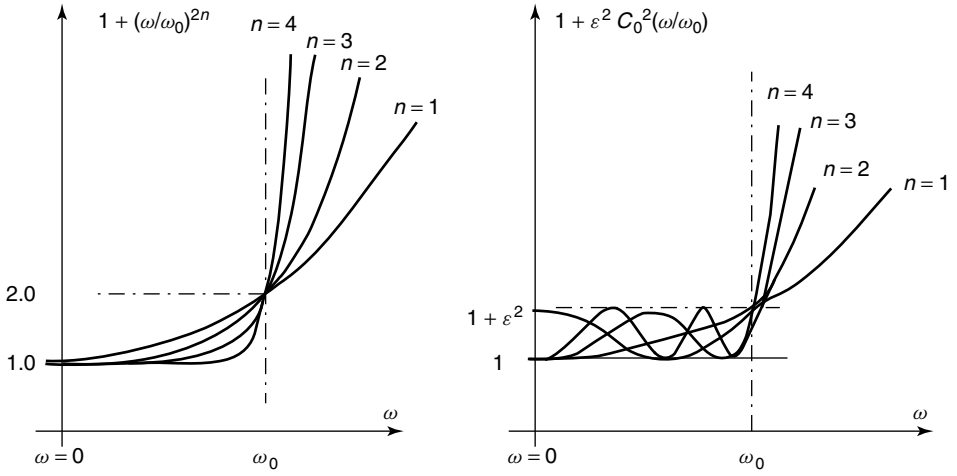


Figure 7. Two types of polynomial characteristics.

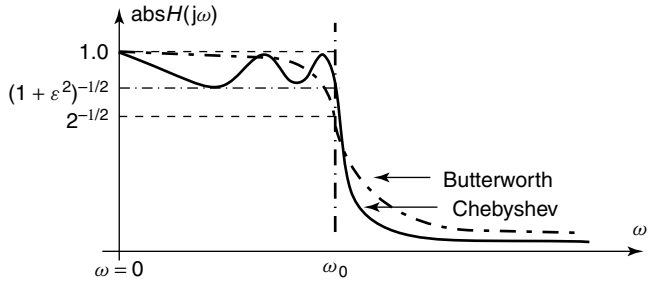


Figure 8. Frequency response, magnitude characteristic of the low-pass Butterworth and Chebyshev filters of fifth-order.

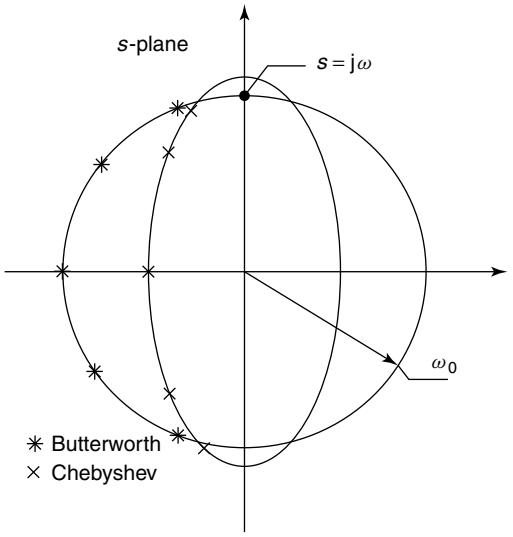


Figure 9. Pole locations of the low-pass Butterworth and Chebyshev filters.

magnitude functions: the Butterworth function

$$|H(\mathrm{j}\omega)|^2 = \frac{1}{[1 + (\omega/\omega_0)^{2n}]} \tag{6}$$

and the Chebyshev function

$$|H(\mathrm{j}\omega)|^2 = \frac{1}{[1 + \varepsilon^2 C_n^2(\omega/\omega_0)]} \tag{7}$$

In the case of the Butterworth filter, we have

$$\begin{aligned} |H(\mathrm{j}\omega)|^2 = H(\mathrm{j}\omega)H(-\mathrm{j}\omega) &\xrightarrow{\mathrm{j}\omega = s} H(s)H(-s) \\ &= \frac{1}{1 + (s/\mathrm{j}\omega_0)^{2n}} \\ &= \frac{1}{1 + (-\mathrm{j})^n(s/\mathrm{j}\omega_0)^n} \frac{1}{1 - (\mathrm{j})^n(s/\mathrm{j}\omega_0)^n} \end{aligned} \tag{8}$$

$$|H(s)|^2 = \begin{cases} \dfrac{1}{1 - (s/\omega_0)^n} \dfrac{1}{1 + (s/\omega_0)^n} \\ \quad = \dfrac{1}{1 - (s/\omega_0)^{2n}} & \text{for even } n \\ \dfrac{1}{1 - \mathrm{j}(s/\omega_0)^n 1 + \mathrm{j}(s/\omega_0)^n} \\ \quad = \dfrac{1}{1 + (s/\omega_0)^{2n}} & \text{for odd } n \end{cases} \tag{9}$$

This function has $2n$ poles equally spaced around a circle of radius ω_0 in the s-plane.

For even n	For odd n
$(s/\omega_0)^{2n} = 1$	$(s/\omega_0)^{2n} = -1$
$(s/\omega_0)^{2n} = \mathrm{e}^{\mathrm{j}2\pi}$	$(s/\omega_0)^{2n} = \mathrm{e}^{\mathrm{j}\pi}$
$s^n = \omega_0^n\ \mathrm{e}^{\mathrm{j}\pi}$	$s^n = \omega_0^n\ \mathrm{e}^{\mathrm{j}\frac{\pi}{2}}$
$s_i = \omega_0\ \mathrm{e}^{\mathrm{j}\frac{\pi}{n}i}$	$s_i = \omega_0\ \mathrm{e}^{\mathrm{j}\frac{\pi}{2n}(i+1)}$
where $i = 0, 1, \ldots, 2n-1$	where $i = 0, 1, \ldots, 2n-1$

(10)

The n poles to the left of the imaginary axis define the filter. It may be shown that the poles of the Chebyshev filter are arranged on an ellipse whose major axis lies along the imaginary axis in the s-plane. The pole locations of the low-pass Butterworth and Chebyshev filters are shown in Figure 9.

RELATED ARTICLES

Article 125, Outline of Purpose of Analog Data Filters, Volume 2; **Article 133, Amplifiers and Filters for DAS, Volume 3**; **Article 148, General Performance of the Digital Filter, Volume 3**; **Article 149, Low-, High-, and Band-pass Digital Filters, Volume 3**.

FURTHER READING

Antoniou, A. (1993) *Digital Filters: Analysis, Design and Applications*, McGraw-Hill Companies, New York.

Baher, H. (1993) *Selective Linear-Phase Switched-Capacitor and Digital Filters*, Kluwer Academic Publishers, New York.

Chen, Chi-Tsong (2000) *Digital Signal Processing: Spectral Computation and Filter Design*, Oxford University Press, Oxford.

Datta, J., Karley, B., Lane, J. and Norwood, J. (2000) *DSP Filter Cookbook*, Delmar Learning, Clifton Park, New York.

Hamming, R.W. (1998) *Digital Filters*, Dover Publications, London.

Jackson, L.B. (1996) *Digital Filters and Signal Processing: With Matlab Exercises*, Kluwer Academic, Dordrecht.

Mano, M.M. (2001) *Digital Design*, Prentice Hall, New Jersey.

Parks, T.W. and Burrus, C.S. (1987) *Digital Filter Design*, Topics in Digital Signal Processing, Wiley-Interscience, New York.

Phillips, L.C. and Nagle, T. (1994) *Digital Control System Analysis and Design*, Pearson Education, Upper Saddle River, NJ.

Rorabaugh, C.B. (1993) *Digital Filter Designer's Handbook: Featuring C Routines*, McGraw-Hill, New Jersey.

Schlichtharle, D. (2000) *Digital Filters: Basics and Design*, Springer-Verlag, Telos, Berlin.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

This article is based in part on a chapter by J. McGhee *et al* originally published in Measurement Data Handling: Hardware Techniques Vol. 2, J. McGhee *et al*, Technical University of Lodz, 2001.

148: General Performance of the Digital Filter

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Forms of Realization of the Digital Filter 1013
2 Poles and Zeros in Digital Filter Design 1013
3 Impulse Response of the Digital Filter 1014
Related Articles 1014
Further Reading 1014

1 FORMS OF REALIZATION OF THE DIGITAL FILTER

Digital filters can be realized in two different ways:

- *Hardware* – realized by a digital logic circuit containing storage, delay, and addition–subtraction and multiplication by constants. The advantage of the digital filter is its speed, especially if some of the necessary operations are performed in parallel. Its very costly realization is its disadvantage.
- *Software* – as a program. The filter operates rather slowly in comparison with hardware forms, but on the other hand, this solution is more flexible and inexpensive.

There are two methods of digital filter implementation. The first that one should consider is filtering in the frequency domain. This is an important and widely used approach, which allows great flexibility in the choice of the filter characteristics. The signal spectrum is simply multiplied by the choice of filter characteristic.

2 POLES AND ZEROS IN DIGITAL FILTER DESIGN

The alternative implementation considers filtering in the time domain. The is done by convolution of the input signal with the impulse response of the appropriate filter.

A transfer function, $H(z)$, represents a useful frequency response characteristic. Knowing $H(z)$, it is a simple matter to derive the time domain formula, which describes the operation of the filter. When considering the digital filter, we are concerned with impulse response rather than its frequency spectra; we use a set of z-plane poles and zeros of transfer function, $H(z)$.

If a digital filter and its transfer function have the poles and zeros, as shown in Figure 1, then its response for some frequency ω_1 may be found by drawing vectors from the poles and zeros to the point $z = \exp(j\omega_1 T_s)$. The response magnitude at frequency ω_1 is equal to $a_1 a_2 / b_1$, where a_1, a_2, and b_1 are the lengths of vectors from the poles and zeros to point z.

Point A in the figure represents sinusoidal frequencies $\omega = 0, 2\pi/T_s, 4\pi/T_s, \ldots$ and point B represents frequencies $\omega = \pi/T_s, 3\pi/T_s, 5\pi/T_s, \ldots$ The frequency response characteristic of any digital filter repeats indefinitely at intervals in ω of $2\pi/T$ rad s^{-1}.

Figure 2(a) shows a simple low-pass filter specified by single zero at $z = -1$. At frequency $\omega = 0$, the zero vector has a maximum length of 2.0, and hence, the response amplitude (Figure 2b) must also be a maximum, and it becomes shorter when we move the zero vector anticlockwise around the unit circle. The vector vanishes at point $z = -1$, which corresponds to the sinusoidal frequency $\omega = 2\pi/T$. By considering the phase angle that the zero

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

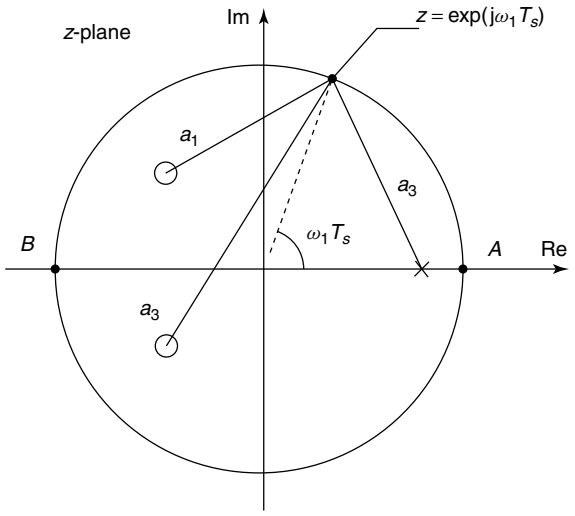


Figure 1. The z-plane poles and zeros of digital filter.

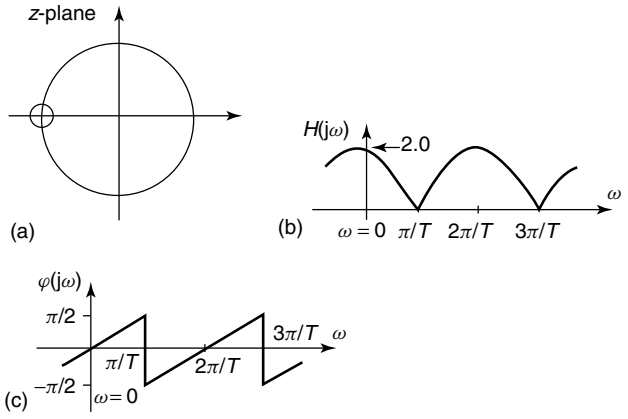


Figure 2. An elementary low-pass digital filter (a) pole-zero configuration; (b) magnitude response characteristic; and (c) phase response characteristic.

vector makes with the positive real axis, we may also infer the filter's phase response (Figure 2c).

The transfer function of this elementary digital filter is $H(z) = (z + 1)$ and equals the transform of the output signal divided by the input. If we write it as $X(z)$ and $Y(z)$ respectively, then $Y(z) = (z + 1)X(z) = zX(z) + X(z)$.

3 IMPULSE RESPONSE OF THE DIGITAL FILTER

The impulse response of a such filter is shown in Figure 3(a). We convolve this impulse response with a typical input signal having successive samples of value. . . $x(n-1), x(n), x(n+1), x(n+2), \ldots$ To calculate the output sample $y(n)$, which coincides with the input samples, $x(n)$, we lay a reversed version of the impulse

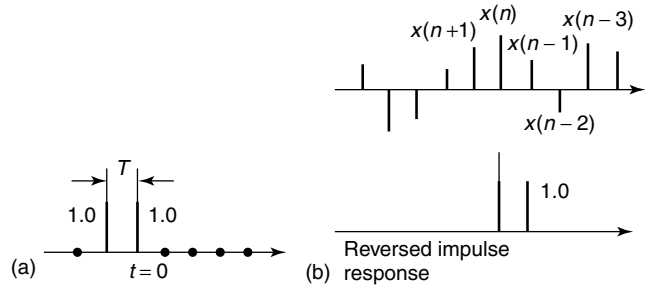


Figure 3. (a) Impulse response of the low-pass filter; (b) derivation of the filter's output by convolution.

response. Cross multiplying point by point and the sum terms gives

$$y(n) = 1 \times x(n) + 1 \times x(n+1) \tag{1}$$

An expression such as this is called a *recurrence formula*.

The filter from the example is 'unrealizable', in the sense that a given output value $y(n)$ depends upon the coincident input $x(n)$ and the next input $x(n+1)$. This difficulty may be solved by dividing $H(z)$ by a term z^m, where m equals the excess number of zeros. This is equivalent to adding m poles at the origin of the z-plane; such poles do not affect the frequency response magnitude characteristic but merely have the effect of delaying the output by m sampling intervals. The new transfer function has the form

$$H'(z) = \frac{(z+1)}{z} = 1 + z^{-1} \tag{2}$$

hence the recurrence formula becomes

$$y(n) = x(n) + x(n+1) \tag{3}$$

which means that any output sample is found by summing the present and previous inputs. For details, refer to **Article 147, Ideal Digital Filter Approximation, Volume 3**.

RELATED ARTICLES

Article 25, Introduction to Signals in Physical Systems, Volume 1; **Article 26, Signal Classification, Volume 1**; **Article 27, Signals in the Frequency Domain, Volume 1**; **Article 147, Ideal Digital Filter Approximation, Volume 3**; **Article 149, Low-, High-, and Band-pass Digital Filters, Volume 3**.

FURTHER READING

Antoniou, A. (1993) *Digital Filters: Analysis, Design and Applications*, McGraw-Hill Companies, New York.

Baher, H. (1993) *Selective Linear-Phase Switched-Capacitor and Digital Filters*, Kluwer Academic Publishers, New York.

Chen, Chi-Tsong (2000) *Digital Signal Processing: Spectral Computation and Filter Design*, Oxford University Press, Oxford.

Datta, J., Karley, B., Lane, J. and Norwood, J. (2000) *DSP Filter Cookbook*, Delmar Learning, Clifton Park, New York.

Hamming, R.W. (1998) *Digital Filters*, Dover Publications, London.

Jackson, L.B. (1996) *Digital Filters and Signal Processing: With Matlab Exercises*, Kluwer Academic, Dordrecht.

Mano, M.M. (2001) *Digital Design*, Prentice Hall, New Jersey.

Parks, T.W. and Burrus, C.S. (1987) *Digital Filter Design*, Topics in Digital Signal Processing, Wiley-Interscience, New York.

Phillips, L.C. and Nagle, T. (1994) *Digital Control System Analysis and Design*, Pearson Education, Upper Saddle River, NJ.

Rorabaugh, C.B. (1993) *Digital Filter Designer's Handbook: Featuring C Routines*, McGraw-Hill, New Jersey.

Schlichtharle, D. (2000) *Digital Filters: Basics and Design*, Springer-Verlag, Telos, Berlin.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

This article is based in part on a chapter by J. McGhee *et al* originally published in Measurement Data Handling: Hardware Techniques Vol. 2, J. McGhee *et al*, Technical University of Lodz, 2001.

149: Low-, High-, and Band-pass Digital Filters

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Low-pass Digital Filter 1016
2 High-pass Digital Filter 1016
3 Analyzing a Digital Filter from a Pole-zero Plot 1017
4 Building a Digital Filter from a Pole-zero Plot 1017
Related Articles 1018
Further Reading 1018

1 LOW-PASS DIGITAL FILTER

The low-pass filter with a single pole placed at $z = \alpha$ is presented in Figure 1.

If $0 < \alpha \leqslant 1$, then the pole vector will be very small when $\omega = 0$, giving a large response, and relatively large when $\omega = \pi/T$, giving a small response.

The transfer function of the considered filter is

$$H(z) = \frac{1}{(z-\alpha)} = \frac{Y(z)}{X(z)} \tag{1}$$

Hence,

$$(z-\alpha)Y(z) = X(z) \quad \text{or} \quad zY(z) - \alpha Y(z) = X(z) \tag{2}$$

The recurrence formula is therefore

$$y(n+1) - \alpha = x(n) \tag{3}$$

which is equivalent to

$$y(n) - \alpha(n-1) = x(n-1) \tag{4}$$

or

$$y(n) = \alpha y(n-1) + x(n-1) \tag{5}$$

The output sample $y(n)$ is found from the previous output and previous input. A filter, which calculates a new output using one or more previous outputs is called *recursive* and arises whenever a transfer function has a pole placed other than at the origin of the z-plane. It can be also written as a nonrecursive recurrence formula

$$y(n) = x(n-1) + \alpha x(n-2) + \alpha^2 x(n-3) + \alpha^3 x(n-4) + \cdots\cdots \tag{6}$$

2 HIGH-PASS DIGITAL FILTER

A high-pass filter has the transfer function

$$H(z) = \frac{(z-1)}{(z+\alpha)} = \frac{Y(z)}{X(z)} \tag{7}$$

Figure 2 gives its (a) pole-zero configuration, (b) magnitude response characteristic, and (c) impulse response.

From the pole-zero plot,

$$(z+\alpha)Y(z) = (z-1)X(z) \tag{8}$$

The recurrence formula is therefore

$$y(n+1) + \alpha y(n) = x(n+1) - x(n) \tag{9}$$

or

$$y(n) + \alpha y(n-1) = x(n) - x(n-1) \tag{10}$$

giving

$$y(n) = -\alpha y(n-1) + x(n) - x(n-1) \tag{11}$$

The first nine terms of the impulse response are shown in Figure 2(c).

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

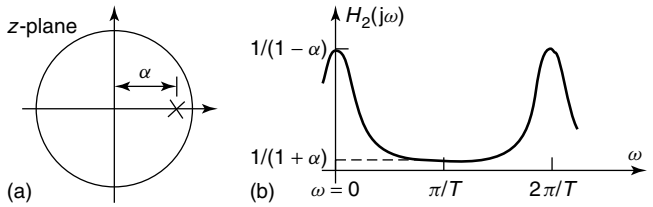


Figure 1. Elementary low-pass digital filter. (a) Pole-zero configuration and (b) magnitude response characteristic.

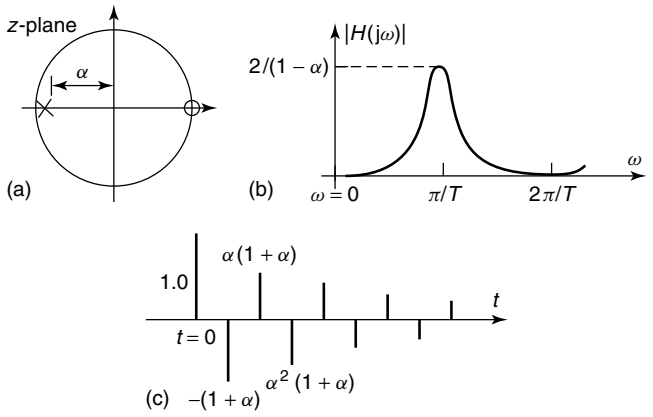


Figure 2. A simple high-pass digital filter. (a) Pole-zero configuration, (b) magnitude response characteristic, and (c) impulse response.

3 ANALYZING A DIGITAL FILTER FROM A POLE-ZERO PLOT

Figure 3(a) shows an example of how to analyze the digital filter with a pole-zero configuration.

The complex conjugate pole pair placed on a circle of radius r, where r is close to unity, gives rise to a band-pass characteristic centered on ω_0 rad s^{-1} (see Figure 3b). Zero-frequency and high-frequency rejection are provided by the zeros at $z = 1$ and $z = -1$. As it is quite often convenient,

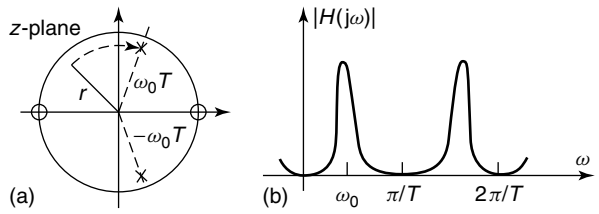


Figure 3. A band-pass digital filter. (a) Pole-zero configuration and (b) magnitude response characteristic.

we express the pole position in polar coordinates; the transfer function is then

$$H(z) = \frac{(z-1)(z+1)}{[z - r\exp(j\omega_0 T)][z - r\exp(-j\omega_0 T)]}$$
$$= \frac{z^2 - 1}{z^2 - rz[\exp(j\omega_0 T) + \exp(-j\omega_0 T)] + r^2} \quad (12)$$

and

$$H(z) = \frac{z^2 - 1}{z^2 - 2rz\cos\omega_0 T + r^2} \quad (13)$$

From this expression, we may derive the recurrence formula

$$y(n) = 2r\cos(\omega_0 T)y(n-1) - r^2 y(n-2) + x(n) - x(n-2) \quad (14)$$

4 BUILDING A DIGITAL FILTER FROM A POLE-ZERO PLOT

For hardware realization (see Figure 4), we choose the following parameter values:

- radius $r = 0.95$
- sampling frequency $f_s = 1\,\text{kHz}$, that is, $T_s = 1\,\text{ms}$

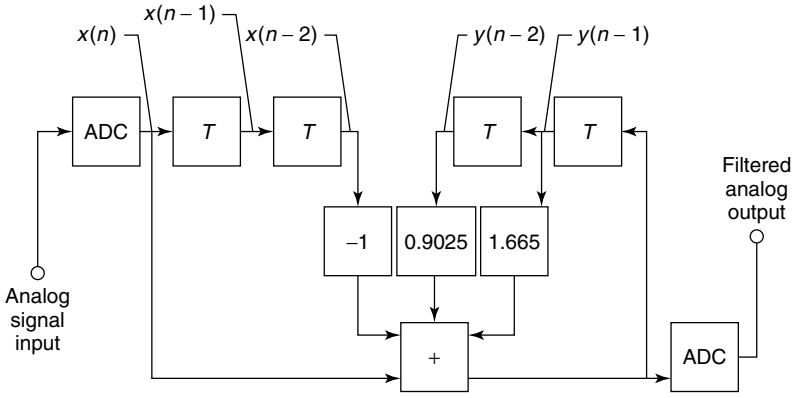


Figure 4. A hardware implementation of the band-pass filter.

- center frequency $\omega_0 = 80\,\text{Hz}$, giving $\omega_0 T = 0.503\,\text{rad}$ or 28.8°.

The recurrence formula becomes

$$y(n) = 1.6650y(n-1) - 0.9025y(n-2) + x(n) - x(n-2) \quad (15)$$

For details, refer **Article 147, Ideal Digital Filter Approximation, Volume 3**.

RELATED ARTICLES

Article 25, Introduction to Signals in Physical Systems, Volume 1; **Article 146, Principles of DSP Hardware Design, Volume 3**; **Article 147, Ideal Digital Filter Approximation, Volume 3**; **Article 148, General Performance of the Digital Filter, Volume 3**; **Article 150, Finite Infinite Impulse Response (IIR) Digital Filters, Volume 3**; **Article 151, Finite Impulse Response (FIR) Digital Filters, Volume 3**.

FURTHER READING

Antoniou, A. (1993) *Digital Filters: Analysis, Design and Applications*, McGraw-Hill Companies, New York.

Baher, H. (1993) *Selective Linear-Phase Switched-Capacitor and Digital Filters*, Kluwer Academic Publishers, New York.

Chen, Chi-Tsong (2000) *Digital Signal Processing: Spectral Computation and Filter Design*, Oxford University Press, Oxford.

Datta, J., Karley, B., Lane, J. and Norwood, J. (2000) *DSP Filter Cookbook*, Delmar Learning, Clifton Park, New York.

Hamming, R.W. (1998) *Digital Filters*, Dover Publications, London.

Jackson, L.B. (1996) *Digital Filters and Signal Processing: With Matlab Exercises*, Kluwer Academic, Dordrecht.

Mano, M.M. (2001) *Digital Design*, Prentice Hall, New Jersey.

Parks, T.W. and Burrus, C.S. (1987) *Digital Filter Design*, Topics in Digital Signal Processing, Wiley-Interscience, New York.

Phillips, L.C. and Nagle, T. (1994) *Digital Control System Analysis and Design*, Pearson Education, Upper Saddle River, NJ.

Rorabaugh, C.B. (1993) *Digital Filter Designer's Handbook: Featuring C Routines*, McGraw-Hill, New Jersey.

Schlichtharle, D. (2000) *Digital Filters: Basics and Design*, Springer-Verlag, Telos, Berlin.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

This article is based in part on a chapter by J. McGhee *et al* originally published in Measurement Data Handling: Hardware Techniques Vol. 2, J. McGhee *et al*, Technical University of Lodz, 2001.

150: Finite Infinite Impulse Response (IIR) Digital Filters

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 The Infinite Impulse Response (IIR) 1019
2 Design of IIR Digital Filters from Analog Filters 1020
3 Impulse Invariance Method 1020
4 Determination of a Digital Filter from an Analog Filter 1021
5 Determination of a Digital Filter Transform from an Analog Filter 1022
6 Bilinear Transformation Model 1022
7 Converting an Analog Filter into a Digital Equivalent Filter 1022
8 Properties of Design of Digital Butterworth Filters 1023
9 Designing a Digital Butterworth Filter 1023
10 Design of a Digital Filter Using the Bilinear Transformation Method 1024
Related Articles 1025
Further Reading 1026

1 THE INFINITE IMPULSE RESPONSE (IIR)

Any digital filter specified in terms of one or more z-plane poles has the infinite impulse response, IIR. It is a recursive filter, because no infinite impulse response could be implemented in a purely nonrecursive operation.

The recurrence formulae of IIR filters normally involve relatively few terms. It is clear that the IIR of any realizable filter cannot be symmetrical in form and that the IIR filter cannot display pure linear-phase characteristics.

The advantage of its design is that the positioning of just one or a few poles inside but close to the unit circle enables very selective filter passband characteristics to be achieved.

Many practical IIR filters are based upon analog equivalents because analogous filters based upon lumped circuit elements have infinite impulse response. The design problem is to find a suitable transformation for mapping the s-plane poles and zeros of the analog filter representation onto the z-plane.

There are several techniques for deriving a digital filter from an analog one. Two of the widely used ones are as follows:

- Impulse invariance
- Bilinear transformation.

It is possible to choose a suitable z-plane pole-zero configuration for a filter without direct reference to analog designs. The least confusing point of view toward digital filter design is to consider the filter as being specified in terms of angle around the unit circle rather than in terms of analog frequencies.

A problem in determining an appropriate set of specifications for a low-pass digital filter is presented in Figure 1.

Specifications for a required filter often take the form of such a given tolerance scheme. The dashed curve represents the frequency response of a system that meets the prescribed specifications. In this case, there is a passband wherein the magnitude of response must approximate to 1, with an error of $\pm\delta_1$, which can be written as

$$1 - \delta_1 \leq \left|H\left(e^{j\omega}\right)\right| \leq 1 + \delta_1 \quad \text{where} \quad |\omega| \leq \omega_p \qquad (1)$$

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

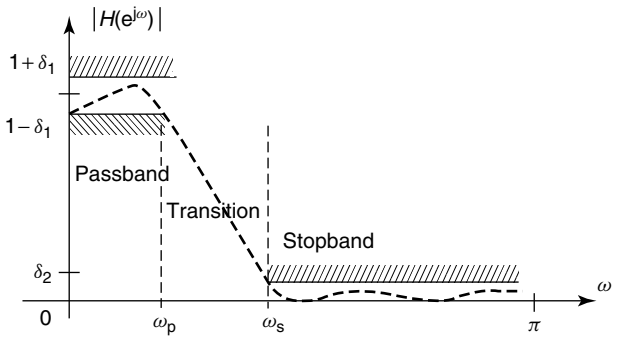


Figure 1. Tolerance limits for approximation of ideal low-pass filter.

There is also a stopband in which the magnitude response must approximate to zero, with an error less than δ_2, for which

$$\left|H\left(e^{j\omega}\right)\right| \leq \delta_2 \quad \text{where} \quad \omega_s \leq |\omega| \leq \pi \mid \tag{2}$$

The passband cut-off frequency, ω_p, and the stopband cut-off frequency, ω_s, are given in terms of z-plane angles.

The next step is to find the discrete-time linear system whose frequency response falls within the prescribed tolerance.

2 DESIGN OF IIR DIGITAL FILTERS FROM ANALOG FILTERS

This method is often used because analog filter design is highly developed. Many useful analog design methods have relatively simple closed-form design formulae. In many applications, it is of interest to use a digital filter to simulate the performance of an analog linear time-invariant filter.

Consider an analog system function given by

$$H_a(s) = \frac{\sum_{k=0}^{M} e_k s^k}{\sum_{k=0}^{N} c_k s^k} = \frac{Y_a(s)}{X_a(s)} \tag{3}$$

where $x_a(t)$ is the input and $y_a(t)$ is the output of the filter, and $X_a(s)$ and $Y_a(s)$ are their respective Laplace transforms. It is assumed that $H_a(s)$ has been obtained through one of the established approximation methods used in analog filter design. The input and output of such a system are related by the convolution integral,

$$y_a(t) = \int_{-\infty}^{\infty} x(\tau) h_a(t-\tau)\, \mathrm{d}\tau \tag{4}$$

where the impulse response, $h_a(t)$, is the inverse Laplace transform of $H_a(s)$.

Alternatively, an analog system having a system function $H_a(s)$ can be described by the differential equation

$$\sum_{k=0}^{N} c_k \frac{\mathrm{d}^k y_a(t)}{\mathrm{d}t^k} = \sum_{k=0}^{M} e_k \frac{\mathrm{d}^k x_a(t)}{\mathrm{d}t^k} \tag{5}$$

The corresponding rational system function for digital filters has the form

$$H(z) = \frac{\sum_{k=0}^{M} b_k z^{-k}}{\sum_{k=0}^{N} c_k z^{-k}} = \frac{Y(z)}{X(z)} \tag{6}$$

The input and output are related by the convolution sum

$$y(n) = \sum_{-\infty}^{+\infty} x(k) n(n-k) \tag{7}$$

or equivalently by the difference equation

$$\sum_{k=0}^{N} a_k y(n-k) = \sum_{k=0}^{M} b_k x(n-k) \tag{8}$$

In transforming an analog system to a digital system, we must therefore obtain either $H(z)$ or $h(n)$ from the analog filter design.

3 IMPULSE INVARIANCE METHOD

This procedure for transforming an analog filter to a digital filter design corresponds to choosing the unit-sample response of the digital filter as equally spaced samples of the impulse response of the analog filter. That is,

$$h(n) = h_a(nT) \tag{9}$$

where T is the sampling period.

The resulting relationship between the response of analog and digital version of a typical low-pass filter is illustrated in Figure 2.

To investigate the interpretation of impulse invariant design in terms of a relationship between the s-plane and the z-plane, consider the system functions of the analog filter expressed in terms of the fraction expansion, so that

$$H_a(s) = \sum_{k=1}^{N} \frac{A_k}{s - s_k} \tag{10}$$

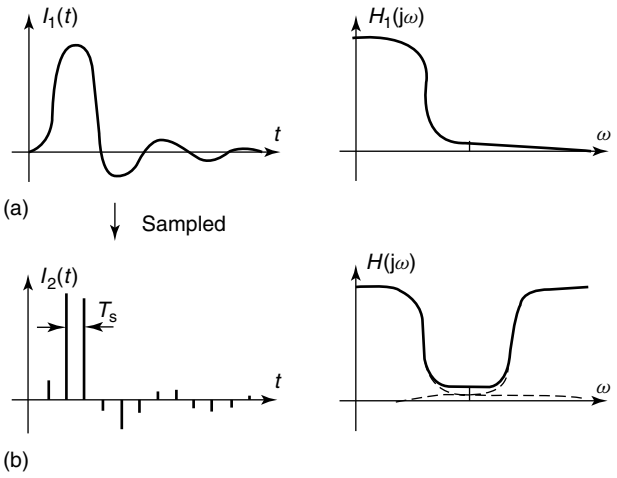


Figure 2. The impulse invariance method. (a) The impulse and frequency response of the analog filter; (b) its digital equivalent. $I_2(t)$ is a sampled version of $I_1(t)$.

The corresponding impulse response is

$$h_a(t) = \sum_{k=1}^{N} A_k e^{s_k T} u(t) \tag{11}$$

where $u(t)$ is a continuous-time unit step function.

The response of the digital filter is then

$$h(n) = h_a(nT) = \sum_{k=1}^{N} A_k e^{s_k nT} u(n) = \sum_{k=1}^{N} A_k (e^{s_k T})^n u(n) \tag{12}$$

The system function of the digital filter $H(z)$ is consequently given by

$$H(z) = \sum_{k=1}^{N} \frac{A_k}{1 - e^{s_k T} z^{-1}} \tag{13}$$

If the analog filter is stable, corresponding to the real part of s_k less than zero, then the magnitude of $e^{s_k T}$ will be less than unity. The corresponding pole in the digital filter will be inside the unit circle, and consequently the digital filter will be also stable. It can be shown that strips of width $2\pi/T$ in the s-plane map into the entire z-plane, as depicted in Figure 3.

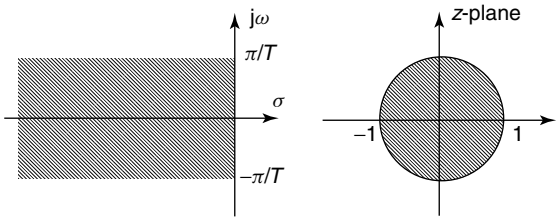


Figure 3. Representation of periodic sampling.

4 DETERMINATION OF A DIGITAL FILTER FROM AN ANALOG FILTER

As an example of the determination of a digital filter from an analog filter by means of the impulse invariance, consider the analog system $H_a(s)$ given by

$$H_a(s) = \frac{s + a}{(s + a)^2 + b^2} = \frac{0.5}{s + a + jb} + \frac{0.5}{s + a - jb} \tag{14}$$

The corresponding transfer function of the impulse invariant digital filter is then

$$H(z) = \frac{0.5}{1 - e^{-aT} e^{-jbT} z^{-1}} + \frac{0.5}{1 - e^{-aT} e^{jbT} z^{-1}}$$
$$= \frac{1 - (e^{-aT} \cos bT) z^{-1}}{(1 - e^{-aT} e^{-jbT} z^{-1})(1 - e^{-aT} e^{-jbT} z^{-1})} \tag{15}$$

from which

$$H(z) = \frac{1 - (e^{-aT} \cos bT) z^{-1}}{1 - 2(e^{-aT} \cos bT) z^{-1} + (e^{-aT})^2 z^{-2}} \tag{16}$$

The digital filter has one zero at the origin and a zero at

$$z = e^{-aT} \cos bT \tag{17}$$

Figure 4 shows the s-plane pole-zero plot for $H_a(s)$ and the z-plane pole-zero plot of $H(z)$, along with the corresponding analog and digital frequency response functions.

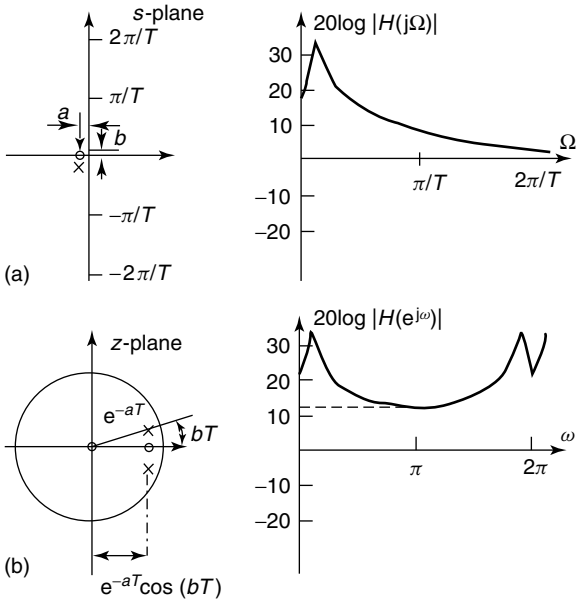


Figure 4. (a) Pole-zero plot and frequency response of a second-order analog system; (b) Pole-zero plot and frequency response of the discrete-time system obtained by sampling the impulse response of the above system.

In this case, the frequency response of the analog system falls off rather slowly relative to the sampling frequency and thus the effects of aliasing are apparent in the digital frequency response.

5 DETERMINATION OF A DIGITAL FILTER TRANSFORM FROM AN ANALOG FILTER

Here is determined the digital filter transform function from the analog filter of the previous case, this time by means of the impulse invariance.

The analog system function $H_a(s)$ given by

$$H_a(s) = \frac{-b}{(s+a)^2 + b^2} = \frac{\text{j}0.5}{s+a-\text{j}b} - \frac{\text{j}0.5}{s+a+\text{j}b} \quad (18)$$

The corresponding transfer function of the impulse invariant digital filter is then

$$\begin{aligned} H(z) &= \frac{\text{j}0.5}{1 - \text{e}^{-aT}\text{e}^{\text{j}bT}z^{-1}} - \frac{\text{j}0.5}{1 - \text{e}^{-aT}\text{e}^{-\text{j}bT}z^{-1}} \\ &= \frac{\text{e}^{-aT}\sin bT}{(1 - \text{e}^{-aT}\text{e}^{-\text{j}bT}z^{-1})(1 - \text{e}^{-aT}\text{e}^{-\text{j}bT}z^{-1})} \\ &= \frac{(\text{e}^{-aT}\sin bT)z^2}{(z - \text{e}^{-aT}\text{e}^{-\text{j}bT})(z - \text{e}^{-aT}\text{e}^{-\text{j}bT})} \\ &= \frac{\text{e}^{-aT}\sin bT}{1 - 2(\text{e}^{-aT}\cos bT)z^{-1} + (\text{e}^{-aT})^2 z^{-2}} \end{aligned} \quad (19)$$

The digital filter has a double zero at the origin.

6 BILINEAR TRANSFORMATION MODEL

The bilinear transformation is an example of the so-called frequency transformation method. A basis of the approach can be clarified by an example. Consider first the function

$$s = \frac{(z-1)}{(z+1)} \quad \text{where} \quad z = \text{e}^{sT} \quad (20)$$

which is bilinear in the sense that both its numerator and denominator polynomials are linear in the variable z.

In order to explain the value of this function in converting an analog filter into a digital equivalent, we need to evaluate its spectrum. This is found by putting $s = \text{j}\omega$, or $z = \text{e}^{\text{j}\omega T}$, which gives

$$\begin{aligned} s &= \frac{\text{e}^{\text{j}\omega T} - 1}{\text{e}^{\text{j}\omega T} + 1} = \frac{\text{e}^{\text{j}\omega T/2}(\text{e}^{\text{j}\omega T/2} - \text{e}^{-\text{j}\omega T/2})}{\text{e}^{\text{j}\omega T/2}(\text{e}^{\text{j}\omega T/2} + \text{e}^{-\text{j}\omega T/2})} \\ &= \text{j}\tan\frac{\omega T}{2} = \sigma + \text{j}\Omega \end{aligned} \quad (21)$$

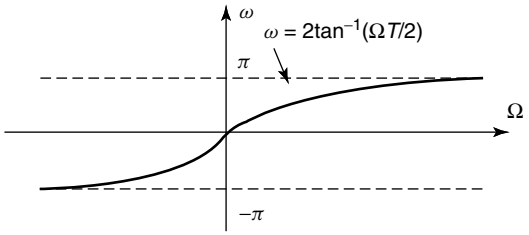


Figure 5. Mapping of the s-plane into z-plane using the bilinear transformation.

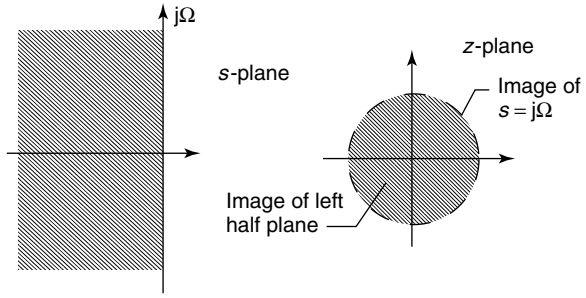


Figure 6. Mapping of the analog frequency axis onto the unit circle using the bilinear transformation.

thus for z on the unit circle, $\sigma = 0$, for which Ω and ω are related by

$$\Omega = \tan\left(\frac{\omega}{2}\right) \quad (22)$$

This relationship is plotted in Figure 5.

In addition to the fact that the imaginary axis in the s-plane maps to the unit circle in the z-plane, the left half of the s-plane maps to the inside of the unit circle and the right half of the s-plane maps to the outside of the unit circle, as depicted in Figure 6.

7 CONVERTING AN ANALOG FILTER INTO A DIGITAL EQUIVALENT FILTER

Suppose the need is to convert an analog filter with transfer function

$$H_a(s) = \frac{1}{s+\alpha} \quad (23)$$

into an equivalent digital filter.

Wherever s appears in $H_a(s)$, replace it by s from (20), giving us the digital filter transfer $H(z)$ to obtain

$$\begin{aligned} H(z) &= \frac{1}{\dfrac{z-1}{z+1} + \alpha} = \frac{z+1}{z-1+\alpha(z+1)} \\ &= \frac{1}{\alpha+1}\frac{z+1}{z+\dfrac{\alpha-1}{\alpha+1}} \end{aligned} \quad (24)$$

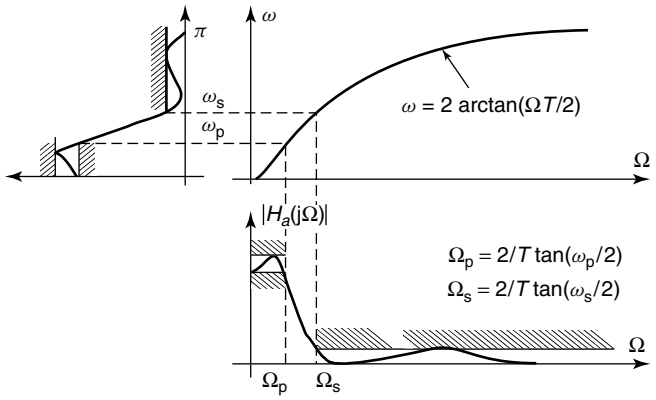


Figure 7. Frequency response and tolerance schemes for analog and corresponding digital filter.

from which the frequency response is given by

$$H(\mathrm{j}\omega) = \frac{1}{\mathrm{j}\tan\omega T/2 + \alpha} \tag{25}$$

The complete frequency response characteristics of the analog filter are compressed into the frequency range $0 < \omega < \pi/T$ in the corresponding digital filter. Figure 7 shows the mapping of an analog frequency response and its corresponding digital frequency response and tolerance scheme.

8 PROPERTIES OF DESIGN OF DIGITAL BUTTERWORTH FILTERS

The squared magnitude function for an analog Butterworth filter is given by

$$\left|H_a(\mathrm{j}\Omega)\right|^2 = \frac{1}{1 + \left(\dfrac{\mathrm{j}\Omega}{\mathrm{j}\Omega_\mathrm{c}}\right)^{2N}} \tag{26}$$

as sketched in Figure 8.

The roots of the denominator polynomial, the poles of the squared magnitude function, are then at

$$s_\mathrm{p} = (-1)^{\frac{1}{2N}}\ \mathrm{j}\Omega_\mathrm{c} \tag{27}$$

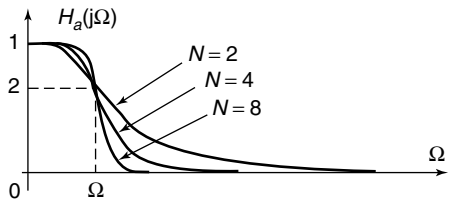


Figure 8. Dependence of the Butterworth magnitude characteristic on the order N.

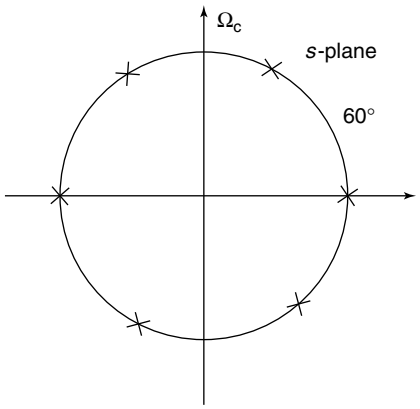


Figure 9. s-plane pole location for a third-order Butterworth filter.

which can be written in the form of poles

$$s_\mathrm{pk} = \left\{ \begin{array}{ll} \Omega_\mathrm{c} \mathrm{e}^{\mathrm{j}\frac{\pi}{N}k} & \text{for } N \text{ odd} \\ \Omega_\mathrm{c} \mathrm{e}^{\mathrm{j}\frac{\pi}{2N}(2k+1)} & \text{for } N \text{ even} \end{array} \right\} \tag{28}$$

where $k = 0, \ldots, N-1$

The properties of a Butterworth filter in terms of poles and zeros are then as follows:

- There are $2N$ poles equally spaced in angle on the circle of radius Ω_c in the s-plane.
- The poles are symmetrically located with respect to the imaginary axis.
- A pole never falls on the imaginary axis.
- A pole occurs on the real axis for odd N but not for even N.
- The angular spacing between the poles on the circle is π/N radians (for $N = 3$, it is indicated in Figure 9).
- The filter is stable when poles occur on the left-half-plane part of the Butterworth circle.

9 DESIGNING A DIGITAL BUTTERWORTH FILTER

Consider the need to design a filter such that the passband magnitude is constant within 1 dB for a frequency below 0.2π and the stopband attenuation is greater than 15 dB for frequencies between 0.3π and π.

If the passband magnitude is normalized to unity at $\omega = 0$, then we require that

$$\begin{aligned} &20\log_{10}|H(\mathrm{j}0.2\pi)| \geq -1 \quad \text{and} \\ &20\log_{10}|H(\mathrm{j}0.3\pi)| \leq -15 \end{aligned} \tag{29}$$

The filter design consists essentially of determining the parameters N and Ω_c to meet the desired specification. For

the Butterworth filter,

$$20\log_{10}|H(\mathrm{j}\Omega)| = 10\log_{10}|H(\mathrm{j}\Omega)|^2 \tag{30}$$

then

$$10\log_{10}\frac{1}{1+(\Omega/\Omega_\mathrm{c})^{2N}} = a[\mathrm{dB}] \tag{31}$$

and

$$1+\left(\frac{\Omega}{\Omega_\mathrm{c}}\right)^{2N} = 10^{0.1a} \tag{32}$$

In our example we have

$$1+\left(\frac{0.2\pi}{\Omega_\mathrm{c}}\right)^{2N} = 10^{0.1} \tag{33}$$

and

$$1+\left(\frac{0.3\pi}{\Omega_\mathrm{c}}\right)^{2N} = 10^{1.5} \tag{34}$$

The solution of these two equations leads to the value $N = 5.88$ and $\Omega_\mathrm{c} = 0.70474\,\mathrm{rad\,s^{-1}}$. The parameter N, however, must be an integer and, consequently, in order for the specifications to be met or exceeded, we round N up to the nearest integer so that $N = 6$.

Then, however, both of the passband and stopband specifications cannot be met exactly. If we substitute $N = 6$ in (34), we obtain $\Omega_\mathrm{c} = 0.7032\,\mathrm{rad\,s^{-1}}$. With this value, the passband specification will be met exactly and the stopband specification will be exceeded for the analog filter. This allows some margin for the aliasing in the digital filter. With the value of Ω_c and with $N = 6$, there are three pole pairs in the left half of the s-plane with coordinates at

- pole pair 1: $-0.1820 \pm \mathrm{j}0.6792$
- pole pair 2: $-0.4972 \pm \mathrm{j}0.4972$
- pole pair 3: $-0.6792 \pm \mathrm{j}0.1820$.

so that

$$\begin{aligned} H_a(s) = {} & \frac{0.120931}{(s^2+0.3640s+0.4945)} \\ & \times (s^2+0.9945s+0.4945) \\ & \times (s^2+1.3585s+0.4945) \end{aligned} \tag{35}$$

Now consider the design of the desired filter above this time using the impulse invariance method.

If we express $H_a(s)$ as a partial fraction and perform a transformation of the above equation, the resulting system function of the digital filter is

$$\begin{aligned} H(z) = {} & \frac{0.2871-0.4466z^{-1}}{1-1.2971z^{-1}+0.6949z^{-2}} \\ & + \frac{-2.1429+1.1454z^{-1}}{1-1.0691z^{-1}+0.3699z^{-2}} \\ & + \frac{1.8558-0.6304z^{-1}}{1-0.9972z^{-1}+0.2570z^{-2}} \end{aligned} \tag{36}$$

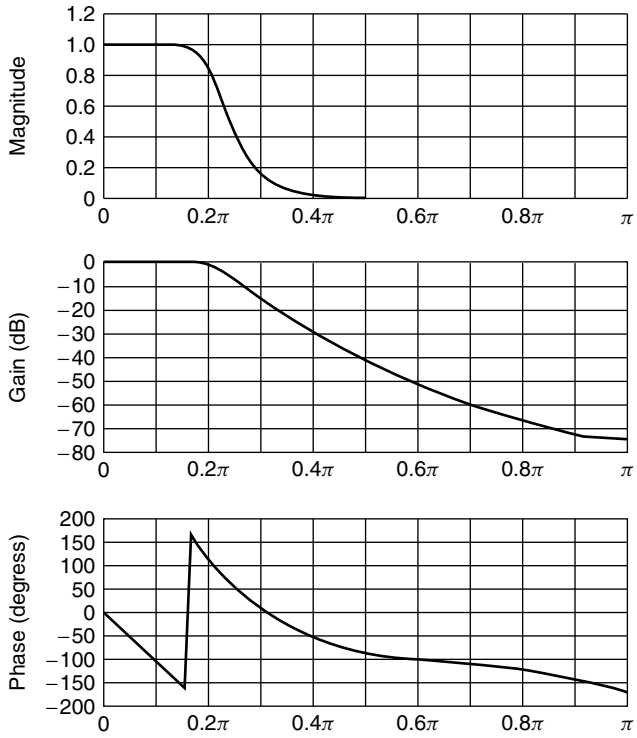


Figure 10. Frequency response of sixth-order Butterworth filter transform by impulse invariance.

It is evident from (36) that the output signal can be derived from the input signal by a recurrence formula. The frequency response of the above system is shown in Figure 10.

Recall now that the filter was designed to exactly meet the specification at the passband edge and to exceed the specification at the stopband edge, and in fact this is the case. This is an indication that the analog filter is sufficiently band limited such that aliasing presents no problems.

10 DESIGN OF A DIGITAL FILTER USING THE BILINEAR TRANSFORMATION METHOD

Given below is the design of a digital filter using the bilinear transformation method.

The digital frequency specification must be prewarped to the corresponding analog frequencies so that with the frequency distortion inherent in the bilinear transformation the critical analog frequencies will map to the correct critical digital frequencies. For the specific filter, we require that

$$20\log_{10}\left|H_a\left(\mathrm{j}2\tan\left(\frac{0.2\pi}{2}\right)\right)\right| \geq -1 \tag{37}$$

and

$$20\log_{10}\left|H_a\left(\mathrm{j}2\tan\left(\frac{0.3\pi}{2}\right)\right)\right| \leq -15 \qquad (38)$$

where we have consequently assumed that $T = 1$. Solving the equations with equality

$$1 + \left(\frac{2\tan(0.1\pi)}{\Omega_c}\right)^{2N} = 10^{0.1} \qquad (39)$$

and

$$1 + \left(\frac{2\tan(0.15\pi)}{\Omega_c}\right)^{2N} = 10^{1.5} \qquad (40)$$

so that

$$N = \frac{1}{2}\,\frac{\log\left[(10^{1.5}-1)\,/\,(10^{0.1}-1)\right]}{\log\left[\tan(0.15\pi)\,/\tan(0.1\pi)\right]} = 5.305 \qquad (41)$$

N must be chosen as 6. If we determine Ω_c by substituting $N = 6$ into (39), we obtain $\Omega_c = 0.76622$. For this value of Ω_c, the passband specifications are exceeded and the stopband specifications are met exactly. In the s-plane, 12 poles of the squared magnitude function are uniformly distributed in angle on a circle of radius 0.76622, as shown in Figure 11.

The transfer function in the s-plane corresponding to the left-half-plane poles is

$$H_a(s) = \frac{0.20238}{(s^2 + 0.396s + 0.5871)\times(s^2 + 1.083s + 0.5871)\times(s^2 + 1.4802s + 0.5871)} \qquad (42)$$

The transfer function $H(z)$ for the digital filter is then obtained by applying the bilinear transformation to $H_a(s)$

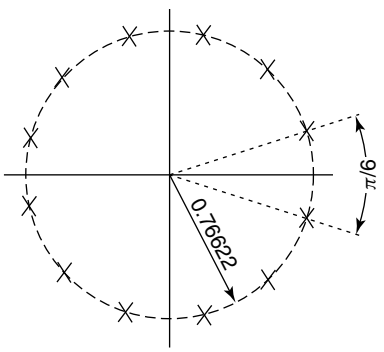


Figure 11. s-plane locations of sixth-order Butterworth filter.

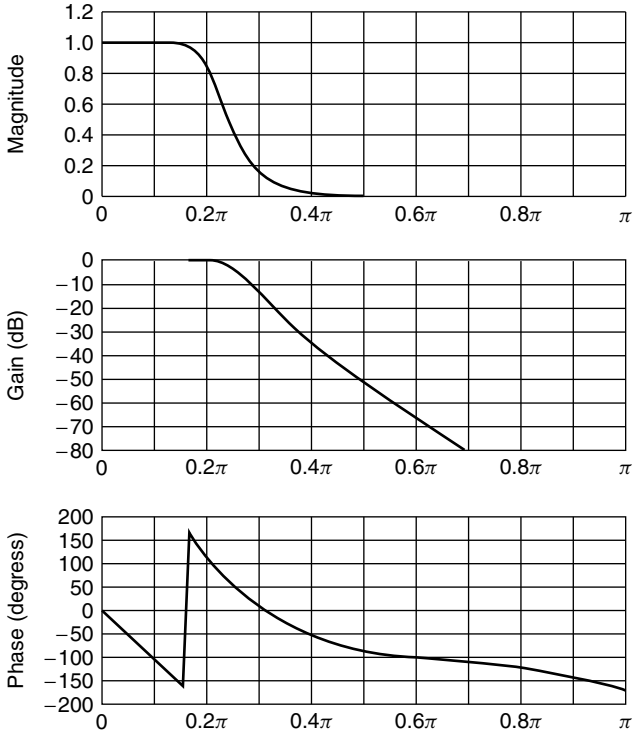


Figure 12. Frequency transfer of sixth-order Butterworth filter transformed by bilinear transformation.

with T chosen as unity, with the result that

$$H(z) = \frac{0.0007378(1 - z^{-1})^6}{(1 - 1.2686z^{-1} + 0.7051z^{-2})\times(1 - 1.0106z^{-1} + 0.3583z^{-2})\times\frac{1}{(1 - 0.9044z^{-1} + 0.2155z^{-2})}} \qquad (43)$$

The magnitude and phase of the digital frequency response is shown in Figure 12.

At $\omega = 0.2\pi$, the magnitude is down 0.5632 dB and at $\omega = 0.3\pi$ the magnitude is down exactly 15 dB.

It should be noted that the magnitude function in Figure 12 falls off much more rapidly than the one in Figure 10. This is because the bilinear transformation maps the entire jΩ axis of the s-plane onto the unit circle. Since the analog Butterworth filter has sixth-order zero at $s = \infty$, the resulting digital filter has a sixth-order zero at $z = -1$.

For further information on digital filter design, consult the bibliography given in **Article 147, Ideal Digital Filter Approximation, Volume 3**.

RELATED ARTICLES

Article 125, Outline of Purpose of Analog Data Filters, Volume 2; **Article 133, Amplifiers and Filters for DAS,**

Volume 3; **Article 147, Ideal Digital Filter Approximation, Volume 3**; **Article 148, General Performance of the Digital Filter, Volume 3**; **Article 149, Low-, High-, and Band-pass Digital Filters, Volume 3**.

FURTHER READING

Antoniou, A. (1993) *Digital Filters: Analysis, Design and Applications*, McGraw-Hill Companies, New York.

Baher, H. (1993) *Selective Linear-Phase Switched-Capacitor and Digital Filters*, Kluwer Academic Publishers, New York.

Chen, Chi-Tsong (2000) *Digital Signal Processing: Spectral Computation and Filter Design*, Oxford University Press, Oxford.

Datta, J., Karley, B., Lane, J. and Norwood, J. (2000) *DSP Filter Cookbook*, Delmar Learning, Clifton Park, New York.

Hamming, R.W. (1998) *Digital Filters*, Dover Publications, London.

Jackson, L.B. (1996) *Digital Filters and Signal Processing: With Matlab Exercises*, Kluwer Academic, Dordrecht.

Mano, M.M. (2001) *Digital Design*, Prentice Hall, New Jersey.

Parks, T.W. and Burrus, C.S. (1987) *Digital Filter Design*, Topics in Digital Signal Processing, Wiley-Interscience, New York.

Phillips, L.C. and Nagle, T. (1994) *Digital Control System Analysis and Design*, Pearson Education, Upper Saddle River, NJ.

Rorabaugh, C.B. (1993) *Digital Filter Designer's Handbook: Featuring C Routines*, McGraw-Hill, New Jersey.

Schlichtharle, D. (2000) *Digital Filters: Basics and Design*, Springer-Verlag, Telos, Berlin.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

This article is based in part on a chapter by J. McGhee *et al* originally published in Measurement Data Handling: Hardware Techniques Vol. 2, J. McGhee *et al*, Technical University of Lodz, 2001.

151: Finite Impulse Response (FIR) Digital Filters

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Outline of Digital Filter with Finite Impulse Response (FIR) 1027
2 The Moving-average Filter 1027
3 FIR Design Based on Window Function 1027
4 Transversal Filter Design by The Window Method 1028
5 Digital Filter with Linear-phase Spectrum 1029
6 Summary of Digital Filter Methods 1029
Related Articles 1030
Further Reading 1030

1 OUTLINE OF DIGITAL FILTER WITH FINITE IMPULSE RESPONSE (FIR)

A digital transversal filter is a filter, which has m delay stages, $(m+1)$ positive or negative multipliers, and an adder or summing junction. It is nonrecursive, with a time domain recurrence formula given by

$$y(n) = a_0x(n) + a_1x(n-1) + a_2x(n-2) + \cdots\cdots + a_mx(n-m) \quad (1)$$

and

$$y(n) = \sum_{i=0}^{m} a_i x(n-i) \quad (2)$$

The filter impulse response is just made up to the multiplier coefficient sequence a_0 to a_m, and is finite in duration.

The disadvantage of a nonrecursive finite impulse response (FIR) filter is that it requires many more delay and multiplier elements than a recursive filter. However, the filter has significant advantages. For example, an impulse response, which is finite in duration, can also be symmetrical in form. This produces a pure linear-phase characteristic; there is no phase distortion. It does not have to be linear phase, but most practical designs incorporate this advantage, which is not available in analog filters. The filter is also inherently stable, since it does not involve feedback from output to input.

2 THE MOVING-AVERAGE FILTER

A simple form of digital filter in which all the multipliers are equal is often called a *moving-average filter*.

As an example, consider the transversal filter in which all multipliers are equal. We use 19 delay and 20 multipliers equal to 0.05.

The transfer function, as a sum of finite power series, is given by

$$H(z) = \frac{(1-z^{-20})}{(1-z^{-1})} = \frac{(z^{20}-1)}{z^{19}(z-1)} \quad (3)$$

The filter has 19 poles in origin and $z = 1$. The zeros of the function are uniformly distributed in angle on a unit circle in the z-plane and the angular spacing between the zeros on the circle is $\pi/10$T rad. The first transmission zero is at $\omega = \pi/10$T. The characteristics of this particular design are shown in Figure 1.

3 FIR DESIGN BASED ON WINDOW FUNCTION

Digital filter design starts from an ideal desired frequency response specification such as

$$H_d(e^{j\omega}) = \sum_{-\infty}^{\infty} h_d(n)e^{-j\omega n} \quad (4)$$

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

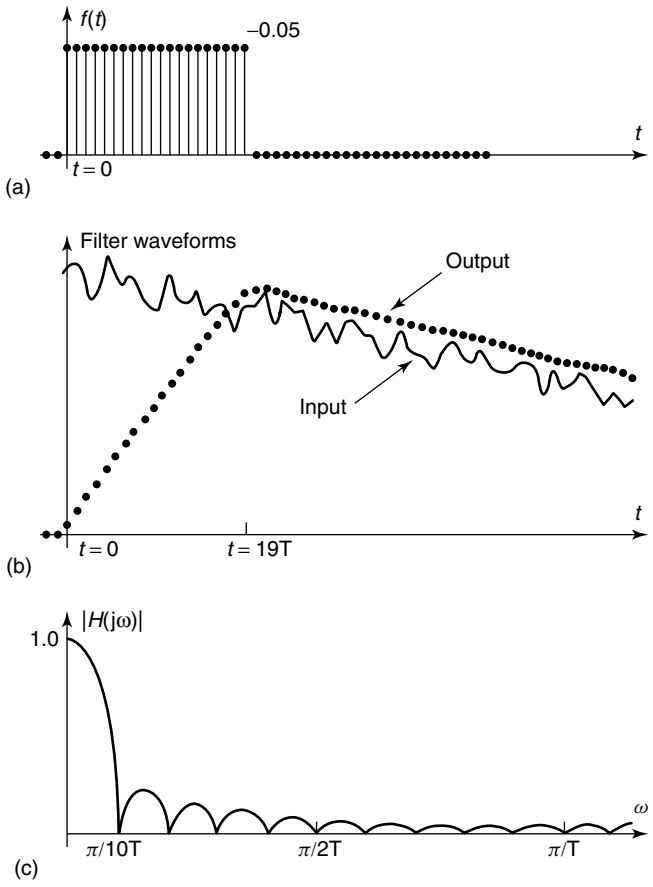


Figure 1. A moving-average filter (a) impulse response, (b) typical input and output waveform, (c) frequency response magnitude characteristic.

The equivalent impulse response may then be found by the inverse discrete Fourier transform, and it is used to define the multipliers of the digital transversal filter.

$$h_d(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} H_d(\mathrm{e}^{\mathrm{j}\omega n})\mathrm{e}^{\mathrm{j}\omega n}\,\mathrm{d}\omega \tag{5}$$

The impulse response will often contain an unacceptable number of terms. It is possible to simplify the impulse response by ignoring the smaller terms.

$$h(n) = \begin{cases} h_d(n), & 0 < n < N-1 \\ 0, & \text{elsewhere} \end{cases} \tag{6}$$

Such truncation is equivalent to multiplying the function by a rectangular observation window; it causes spreading of the frequency domain characteristic. An alternative solution to the truncation problem is to select a more effective window function, $w(n)$, such as a Hamming, Hanning, or Keiser window.

$$H(n) = h(n)w(n) \tag{7}$$

Now follows some commonly used window functions. These are specified by the following equations:

Rectangular window function

$$w(n) = \begin{cases} 1, & 0 \le n \le N-1 \\ 0, & \text{elsewhere} \end{cases} \tag{8}$$

Bartlett window function

$$w(n) = \begin{cases} \dfrac{2n}{N-1}, & 0 \le n \le \dfrac{N-1}{2} \\ 2 - \dfrac{2n}{N-1}, & \dfrac{2n}{N-1} \le n \le N-1 \end{cases} \tag{9}$$

Hanning window function

$$w(n) = \begin{cases} \dfrac{1}{2}\left[1 - \cos\left(\dfrac{2\pi n}{N-1}\right)\right], & 0 \le n \le N-1 \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

Hamming window function

$$w(n) = \begin{cases} 0.54 - 0.46\cos\left(\dfrac{2\pi n}{N-1}\right), & 0 \le n \le N-1 \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

Blackman window function

$$w(n) = \begin{cases} 0.42 - 0.5\cos\left(\dfrac{2\pi n}{N-1}\right) \\ \quad +0.08\cos\left(\dfrac{4\pi n}{N-1}\right), & 0 \le n \le N-1 \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

Windowing can be applied at different stages of spectrum estimation:

- to the time record, called *linear windowing* in the time domain,
- to the results of the Discrete Fourier Transform called *linear windowing* in the frequency domain.

4 TRANSVERSAL FILTER DESIGN BY THE WINDOW METHOD

To illustrate the techniques, now consider the design of a low-pass filter, which takes as its starting point an ideal low-pass filter characteristic $G_1(\omega)$ with a cutoff frequency of $\omega_0 = \pi/4\mathrm{T}$ (Figure 2(a)).

The impulse response $I(t)$ is symmetrical about $t = 0$ and of sampled $(\sin x)/x$ form (Figure 2(b)). We next decide how many impulse response terms we can accommodate in the final design, and multiply by a window function with this number of terms.

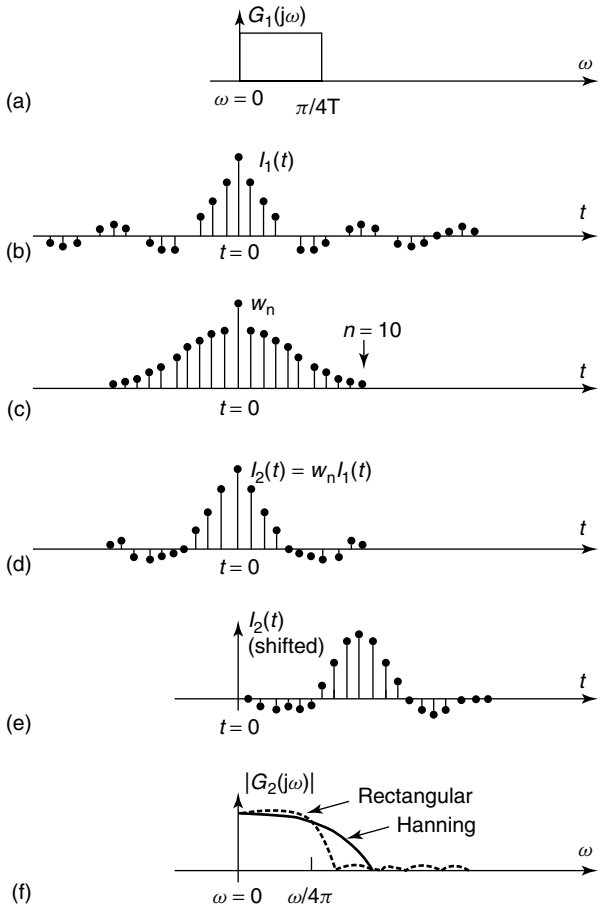


Figure 2. Transversal filter design by window method.

A Hamming window can be used that is defined as:

$$w(n) = \begin{cases} 0.54 + 0.46\cos n\pi/N, & -N < n < n \\ 0, & \text{elsewhere} \end{cases} \tag{13}$$

This has $(2N - 1)$ terms. Use 21 transversal multipliers so that $N = 11$. The impulse response can be shifted forward (Figure 2(e)) so that it begins at $t = 0$. This makes the filter physically releasable and converts the original zero-phase filter into a pure linear-phase one. The resulting frequency response magnitude characteristic is shown in Figure 2(f).

5 DIGITAL FILTER WITH LINEAR-PHASE SPECTRUM

This time, the design filter has to achieve a linear-phase spectrum.

The desired frequency response is defined as

$$H_d(e^{-j\omega}) = \begin{cases} e^{-j\omega\alpha}, & |\omega| < \omega_c \\ 0, & \text{elsewhere} \end{cases} \tag{14}$$

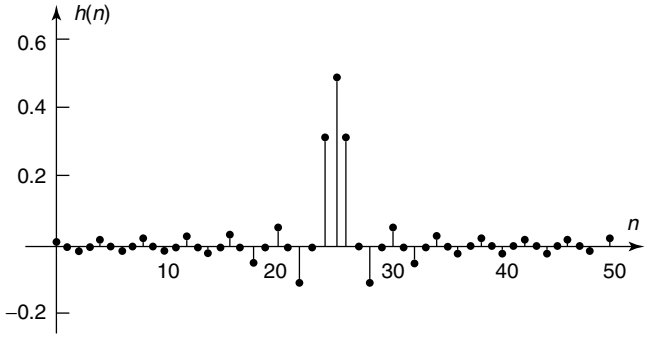


Figure 3. Truncated impulse response of an ideal low-pass filter. (Delay is 25 samples, total length is 51 samples and cutoff frequency is $\omega_0 = \pi/2$.)

The corresponding impulse response is

$$h_d(n) = \begin{cases} \dfrac{1}{2\pi}\displaystyle\int_{-\omega}^{\omega} e^{j\omega(n-\alpha)}\,d\omega \\ \dfrac{\sin[\omega_c(n-\alpha)]}{\pi(n-\alpha)}, n \neq \alpha \end{cases} \tag{15}$$

Clearly, the impulse response has infinite duration. To create a finite duration linear phase causal filter of length N, it is necessary to define $h(n) = h_d(n)w(n)$ and $\alpha = (N-1)/2$. Figure 3 shows a plot of $h(n)$ for a rectangular window where $N = 51$, and $\omega_c = \pi/2$.

6 SUMMARY OF DIGITAL FILTER METHODS

It is useful to summarize the methods and techniques presented in this part on digital data filtering. It started with a comment on filter classifications. Several types of filters have been distinguished according to the signal form and frequency range.

It continued with the description of the filter in the frequency domain by its frequency response transfer function, and in the time domain by its impulse response. The Butterworth and Chebyshev filters have been presented as examples of an ideal filter approximation.

The general performance of digital filters has been outlined using two different methods of digital filter realization and two different ways of the filter implementation. The method of filter amplitude and phase spectra estimation from its zero-pole configuration has been presented.

Low-pass and high-pass filters are described as typical examples of the digital filters in the z-plane. Their magnitude response characteristics have been considered. The problem of determining an appropriate set of specifications with a low-pass filter as the example has been given.

Filters with infinite impulse response, IIR, and filters with finite impulse response, have been described. Since many

practical IIR filters are based upon analog equivalents, the design problem is to find a suitable transformation for mapping the s-plane poles and zeros of the analog filter into the z-plane. The impulse invariance method or bilinear transformation has been introduced. The design principles and procedures for both methods have been presented. Several examples of low-pass Butterworth filter design have been shown.

The coverage also includes the concept of the moving-average filter as an example of an FIR filter. FIR filter design based on a window function is the most useful form. Design examples have been given.

For further information on digital filter design, consult the bibliography given in **Article 147, Ideal Digital Filter Approximation, Volume 3**.

RELATED ARTICLES

Article 125, Outline of Purpose of Analog Data Filters, Volume 2; **Article 133, Amplifiers and Filters for DAS, Volume 3**; **Article 147, Ideal Digital Filter Approximation, Volume 3**; **Article 149, Low-, High-, and Band-pass Digital Filters, Volume 3**; **Article 150, Finite Infinite Impulse Response (IIR) Digital Filters, Volume 3**.

FURTHER READING

Antoniou, A. (1993) *Digital Filters: Analysis, Design and Applications*, McGraw-Hill Companies, New York.

Baher, H. (1993) *Selective Linear-Phase Switched-Capacitor and Digital Filters*, Kluwer Academic Publishers, New York.

Chen, Chi-Tsong (2000) *Digital Signal Processing: Spectral Computation and Filter Design*, Oxford University Press, Oxford.

Datta, J., Karley, B., Lane, J. and Norwood, J. (2000) *DSP Filter Cookbook*, Delmar Learning, Clifton Park, New York.

Hamming, R.W. (1998) *Digital Filters*, Dover Publications, London.

Jackson, L.B. (1996) *Digital Filters and Signal Processing: With Matlab Exercises*, Kluwer Academic, Dordrecht.

Mano, M.M. (2001) *Digital Design*, Prentice Hall, New Jersey.

Parks, T.W. and Burrus, C.S. (1987) *Digital Filter Design*, Topics in Digital Signal Processing, Wiley-Interscience, New York.

Phillips, L.C. and Nagle, T. (1994) *Digital Control System Analysis and Design*, Pearson Education, Upper Saddle River, NJ.

Rorabaugh, C.B. (1993) *Digital Filter Designer's Handbook: Featuring C Routines*, McGraw-Hill, New Jersey.

Schlichtharle, D. (2000) *Digital Filters: Basics and Design*, Springer-Verlag, Telos, Berlin.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

This article is based in part on a chapter by J. McGhee *et al* originally published in Measurement Data Handling: Hardware Techniques Vol. 2, J. McGhee *et al*, Technical University of Lodz, 2001.

152: Fundamentals of the Stored Program Digital Computer

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Data and Information 1033
2 Basic Organization of the Stored Program Computer 1033
3 Purpose and Format of Macroinstructions 1035
Related Articles 1035
References 1035

1 DATA AND INFORMATION

Data are the basic elements used to represent the formation or measurement of facts, which are described by numbers, characters, assertions and so on, in a digital form. Data by themselves have little if any meaning. Only processed data gives meaningful information, as represented in Figure 1. The raw material to gain knowledge is information, which is extracted from data.

Consider the example of the manufacture of a light bulb to illustrate this transformation from data to information. The information required is whether the lamp is either too bright or too dull. Two pieces of information are required to perform the assessment. In the first place, a standard brightness is required and in the second, the actual brightness must be measured. Obtaining information about the lamp's brightness means that knowledge about the lamp has been acquired and a specification of uncertainty becomes possible. In a wider context, this knowledge can assist in production control so helping in running the business, in improving health and safety, or in making strategic decisions about long-term planning among many other possibilities. Such a view means that information is necessary to control almost everything. Hence, control is central to the question of data and information. More on the differences between data, information, and knowledge is given in **Article 13, Signals, Information and Knowledge, and Meaning, Volume 1**.

2 BASIC ORGANIZATION OF THE STORED PROGRAM COMPUTER

The stored program computer (Conger, 1999; Gaonkar, 1999; Karalis, 1997; Wilkinson and Makki, 1992), which is an indispensable component in measurement systems (Lang, 1987, 1991; Morrison, 1997) is also important in measurement systems design (Sydenham, Hancock and Thorn, 1989). A common basic architecture for a stored program computer, which is shown in Figure 2, consists of three main subsystems. A *central processing unit*, or CPU, executes instructions. The set of instructions make up the instruction set. Instructions, which are to be executed in a prescribed manner, form part of a program stored in main memory.

Main memory may be divided into two parts. A primary memory is used to store the program to be executed as well as the data to be operated on. This primary memory, which is normally of a semiconductor form, is referred to as *random access memory*, or RAM. It consists of a volatile, *read/write memory*, or R/W memory, and a nonvolatile *read only memory*, or ROM.

Secondary memory, which is often called a backing store, may have different recording forms. Of the three kinds of magnetic recording forms, the first kind is called the *hard*

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

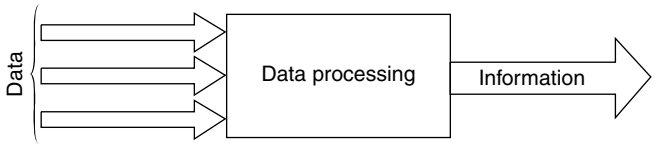


Figure 1. A processor converts data to information.

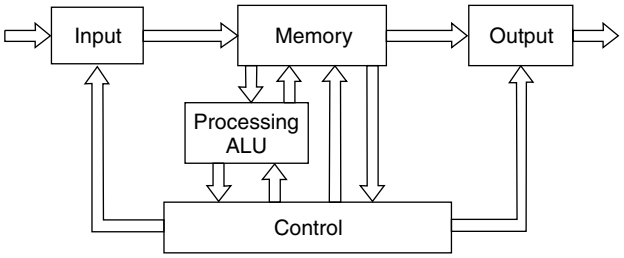


Figure 2. Architecture of a stored program computer.

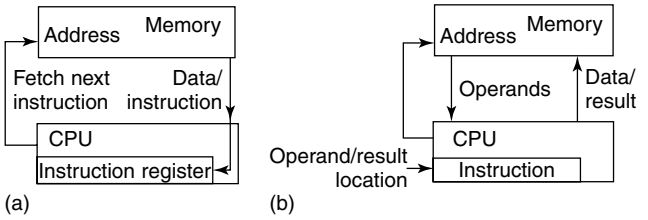


Figure 3. The fetch principle in (a) and the execute principle in (b) of stored program operation.

disk drive, or HDD, which is the main operational form of secondary memory. The other two magnetic recording forms are either removable floppy disk drives or removable tape drives. Removable floppy diskettes may be used for transferring programs and data from one computer to another. Originally, they took the form of 5.25-inch platters. Later, the 3.5-inch high density disk, first developed by

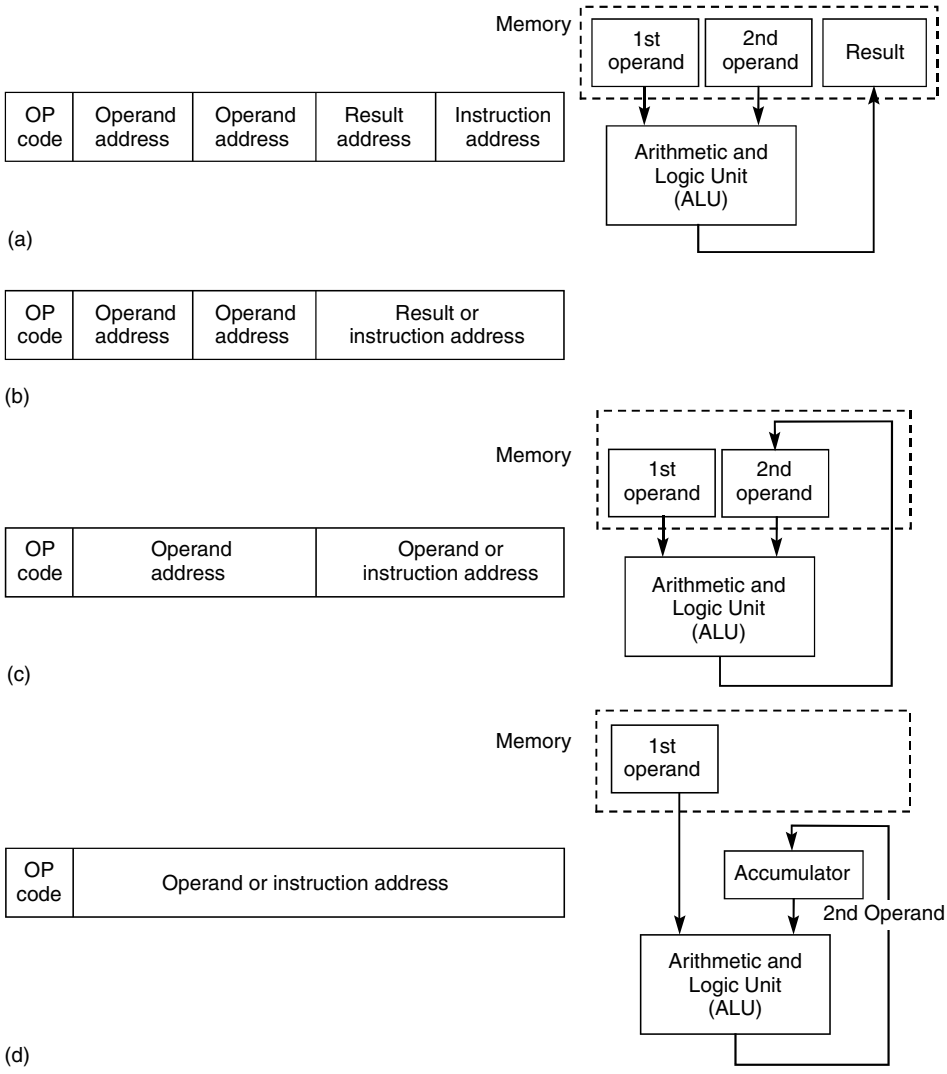


Figure 4. The four macroinstruction formats.

the Sony Corporation, became the most common form of removable media for backing store.

Unfortunately, these media are not all interchangeable or even compatible with the 3.5-inch, 1.44 Mb type.

Tape systems are almost all exclusively used for the purposes of backing up either the complete contents or selected parts of HDD contents. This is mainly used for data security purposes. However, it can also be used as a means of fast recovery of original Operating System, or OS, installations in the event of irrecoverable system crashes. INPUT DEVICES and OUTPUT DEVICES are interfaces to video display units, or VDUs, floppy disk drives, and other peripherals.

The basic mode of operation, consisting of two separate control cycles as shown in Figure 3, are

- FETCH CYCLE when instructions are fetched,
- EXECUTE CYCLE when the instruction is processed.

More details are given in **Article 154, Internal Operation of the Microprocessor, Volume 3**.

Instructions are executed sequentially. The memory is used to store both the program and data. It should be noted that alternative computer architectures are possible. These depend on data availability and not the order of the instructions. Thus, they can operate concurrently and are known as *data flow computers*.

3 PURPOSE AND FORMAT OF MACROINSTRUCTIONS

A number of instruction formats are available depending on the machine design and the compromises made. The most general format is shown in Figure 4(a), where an operational code, or op-code, is specified along with the addresses of the Arithmetic and Logic Unit, or ALU, source operands, and the destination of the result. The next instruction address can also be specified. Thus, data transfer and control can occur together. This format is used for microprogramming, which is a sequence of microinstructions used to execute one top level machine instruction.

If the program control and data processing instructions are separated, it is possible to reduce the instruction fields by one. Figure 4(b) shows this format of separate control and processing instructions.

A further reduction is possible if the destination address is assumed. Figure 4(c) can be interpreted as '*take operands 1 and 2 from the addresses given and return the result to operand 2 address*'. This is somewhat restrictive since memory data is overwritten. An alternative approach is to include a separate register to store the result. This is called the *accumulator*.

The presence of the accumulator, as shown in Figure 4(d) can be used to produce a one-address instruction format, where one of the source operands comes from instruction and the other is the accumulator. The result is returned to the accumulator. In practice, a register file is used for providing a number of detectable registers determined by the op-code.

RELATED ARTICLES

Article 78, Embedded Systems, Volume 2; **Article 153, Single Address Instruction Microcomputer, Volume 3**; **Article 158, Microcontroller Systems, Volume 3**; **Article 159, Designing and Building Software for Measuring Systems, Volume 3**.

REFERENCES

Conger, D. (1999) *Fundamentals of Microcomputers for Technology Students*, Prentice Hall, NJ.

Gaonkar, R. (1999) *Microprocessor Architecture, Programming, and Applications with the 8085*, Prentice Hall, NJ.

Karalis, E. (1997) *Digital Design Principles and Computer Architecture*, Prentice Hall, NJ.

Lang, T.T. (1987) *Electronics of Measuring Systems*, John Wiley & Sons, UK.

Lang, T.T. (1991) *Computerized Instrumentation*, John Wiley & Sons, UK.

Morrison, T. (1997) *The Art of Computerised Measurement*, Oxford University Press, Oxford.

Sydenham, P.H., Hancock, N.H. and Thorn, R. (1989) *Introduction to Measurement Science and Technology*, John Wiley & Sons, UK.

Wilkinson, W. and Makki, R. (1992) *Digital System Design*, Prentice Hall, Europe.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

153: Single Address Instruction Microcomputer

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Data Width and Address Range in Computer Buses 1036
2 Instruction Set 1036
3 External Buses – Synchronous and Asynchronous Transfers 1036
4 Overcoming Problems in the Synchronous Protocol 1038
Related Articles 1038
References 1038

1 DATA WIDTH AND ADDRESS RANGE IN COMPUTER BUSES

Figure 1 shows a typical microprocessor architecture consisting of an arithmetic logic unit (ALU), accumulator and register file, which contains the *stack pointer*, or SP, and the *program counter*, or PC. Microcomputers usually handle 8, 16, 24, 32 or 64 bit wide data words. It is obviously uneconomical to choose a 32-bit bus for use in a system intended for 8 bits (Holdsworth, 1987; Lang, 1991; Rangan, Sarama and Mani, 1983; Tocci and Laskowski, 1987; Wilkinson and Makki, 1992). Component level buses, which are merely extensions of the local processor bus to a back-plane, normally have 16 to 20 address lines giving 64 kB to 1 MB of addressable RAM. Although an 8-bit bus system may be adequate for small process or machine tool systems, it is totally inadequate for modern business applications, which usually include versatile graphic facilities. Hence, allow at least 24 address lines up to say 32 or even 64.

The external data bus and external address bus are buffered by a data buffer, or DBF register and an address buffer, or ABF register, while instructions are loaded into the instruction register, or IR, for use by the control unit. Data transfer between registers occurs via an internal or local processor bus. Control lines from the control unit are available for interfacing to external devices.

2 INSTRUCTION SET

The instruction set for the processor can be divided into the following:

- TRANSFER: movement of data from one location to another such as register or memory locations,
- ARITHMETIC AND LOGICAL: ADD, SUB, OR, AND, NOT, and so on,
- SHIFT AND ROTATE: moving each bit in a binary number,
- JUMP/BRANCH: alters sequence of instructions including program control determined by conditions of the last ALU result,
- INPUT/OUTPUT: transfer from processor registers to external devices,
- CONTROL: interrupts, bit test, external conditions affecting the instruction flow.

3 EXTERNAL BUSES – SYNCHRONOUS AND ASYNCHRONOUS TRANSFERS

Information must be distributed or transferred within the system. The various bus signals, which are used for this purpose, are divided into three main groups. One group, called the *data bus*, is used to carry data from and to various locations within the system. This data needs to be

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

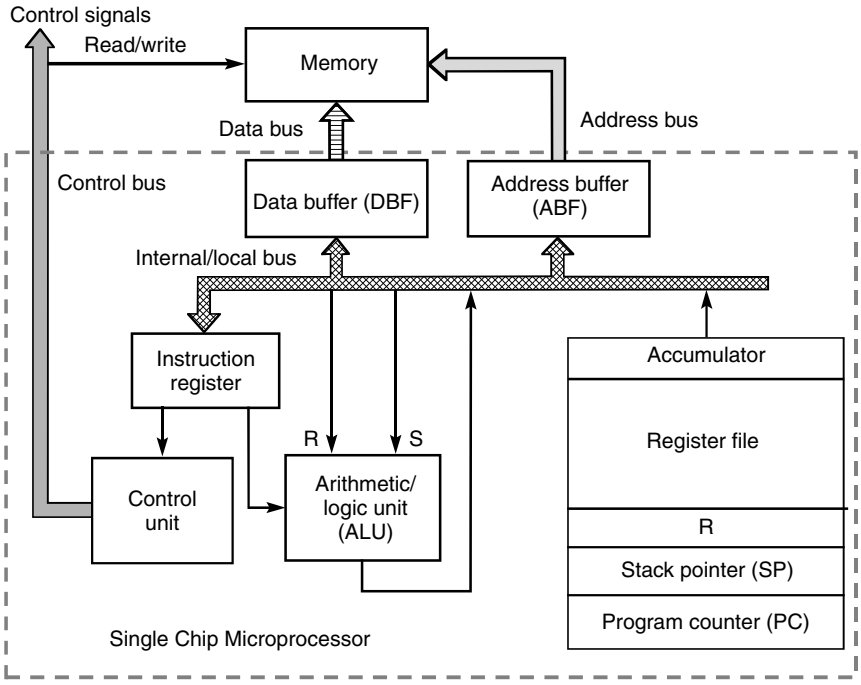


Figure 1. Operational organization of a microprocessor using a single address instruction format.

directed. For this reason, a second bus called the *address bus* is essential. The whole system needs to be controlled. A control bus carries the control signals utilized by the system to direct, time, and order the sequential proceedings of program execution.

The basic methods of information transfer use either SYNCHRONOUS bus transfers or ASYNCHRONOUS bus transfers. Reliable data transfer requires a data transfer protocol or defined set of rules. Three main problems must be overcome. When a master gates several parallel data or address lines with the signals, lack of synchronism causes problems. This is called *skew*. These problems are compounded by the differences in the rise times of each separate line signal. Moreover, the switching point of bus transceivers must be taken into account.

In synchronous bus transfers, synchronization of master and slave is timed as in Figure 2(a) using the circuit of Figure 2(b). Information on the direction of transfer, to or from the processor, the transfer timing, and whether it is a memory or input/output operation, is required. Data are placed on the parallel bus lines. After a delay, which is sufficient to cover the maximum skew and rise time differences, a stable strobe pulse is generated. The processor generates a request signal, together with data and address signals. When the appropriate device receives these signals, data is accepted or generated. The slave responds by completing a write operation to memory or some other interface. Although the timing is known in advance and taken into account by the processor, the processor must

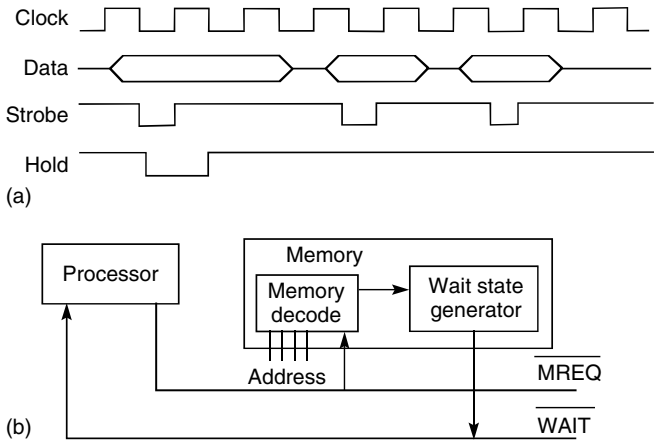


Figure 2. Synchronous protocol with hold and wait cycle: (a) signal timing diagram and (b) organization of the Z80.

not operate faster than the speed of the slowest device. However, since this can be restrictive, introducing a wait signal, or ready signal, can alleviate this problem. A wait state generator, contained within the slow memory being accessed, activates the wait line to the processor causing it to introduce dummy clock cycles in the middle of the data transfer period. Hence, when the slave is slow, it generates a hold pulse telling the master to insert wait cycles until the slave action is completed. Performance is limited by two main factors. In the first case, a clock pulse cannot be changed once it has been defined. This implies that

the system cannot take advantage of faster circuits as they become available. In the second case, slow devices need to be synchronized. Hence, a hold pulse has to be generated very quickly. This requires fast logic so that the master can decide whether to proceed with the next data cycle or to insert dummy wait cycles.

The above information is provided by control signals, which are usually active low or '0'. For example, in the Z80, the signals are $\overline{\text{RD}}$, $\overline{\text{WR}}$, $\overline{\text{MREQ}}$ (memory request), $\overline{\text{IORG}}$ (input/output request).

In the Motorola 6800 series, they are VMA or valid memory address, R/$\overline{\text{W}}$, E or ϕ. E or ϕ is the general processor clock with which timing is synchronized.

4 OVERCOMING PROBLEMS IN THE SYNCHRONOUS PROTOCOL

Problems that occur in the synchronous protocol are overcome by the inclusion of full handshaking signals: data strobe (DS*) and data acknowledge (DK*). The operations proceed in the sequence outlined in Figure 3(a). Two control lines are utilized. One from the processor initiates the start of data transfer, while the other, from the destination, which is the memory in this case, triggers completion of the transfer. Thus, the total transfer time is always dictated by the returning signal, whose absence would indicate a failure in the peripheral device. Compare this to WAIT state generation in which failure to generate is ignored by the processor. After first placing data on lines, the master waits until its own transmitter skew effects have settled and asserts DS*. When the slave sees DS* asserted, it waits until its own receiver skew effects have settled. It then waits for a long enough time either to accept the data or to place new data on the relevant lines before asserting DK*. Meantime, the master holds data steady until it sees DK* asserted. When this occurs, it then releases both data and DS* or waits until it has accepted the slave's data before releasing DS*. The slave holds its data until it sees DS* released. Finally, it then removes its data before releasing DK* in preparation for the next cycle. The data transfer rate is automatically adjusted to allow the fastest possible rate for the receiver and transmitter modules involved. The disadvantage of this method is that extra control signals are required. Propagation delays of these extra control signals may slow the speed of processor operation.

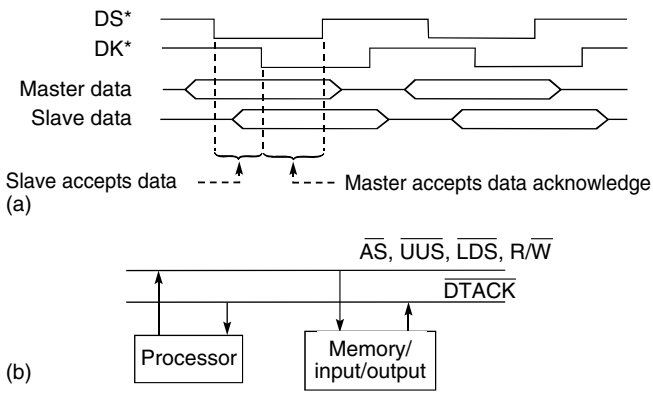


Figure 3. Asynchronous protocol: (a) signal timing diagram and (b) organization of the Motorola 6800.

In the case of the Motorola 6800, both synchronous and asynchronous signals are present. The asynchronous are $\overline{\text{AS}}$ (address strobe), $\overline{\text{UDS}}$ upper data strobe D08-D15, $\overline{\text{LDS}}$ lower data strobe D01-D07, R/$\overline{\text{W}}$, $\overline{\text{DTACK}}$ data acknowledge. The synchronous signals are similar to those of the 6800.

RELATED ARTICLES

Article 152, Fundamentals of the Stored Program Digital Computer, Volume 3; **Article 154, Internal Operation of the Microprocessor, Volume 3**.

REFERENCES

Holdsworth, B. (1987) *Microprocessor Engineering*, Butterworth and Co., London.

Lang, T.T. (1991) *Computerized Instrumentation*, John Wiley & Sons, UK.

Rangan, C.S., Sarama, G.R. and Mani, V.S. (1983) *Instrumentation Devices and Systems*, Tata McGraw-Hill, New Delhi.

Tocci, R.J. and Laskowski, L.P. (1987) *Microprocessors and Microcomputers*, Prentice Hall International.

Wilkinson, W. and Makki, R. (1992) *Digital System Design*, Prentice Hall, Europe.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

154: Internal Operation of the Microprocessor

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Fetching and Executing 1039
2 Register Transfer Logic, or RTL 1039
Related Articles 1040
References 1040

1 FETCHING AND EXECUTING

The operation of the processor can be divided into two cycles known as FETCH and EXECUTE (Conger, 1999; Gaonkar, 1999; Karalis, 1997; Heffer, King and Heath, 1981; (**Article 153, Single Address Instruction Microcomputer, Volume 3**); Wilkinson and Makki, 1992). It should be noted that all of the following operations are essentially register-to-register transfers. They are synchronized to the system clock. Figure 1 shows the control timing pulses, or T states, being derived from the system clock.

In the FETCH cycle, instructions are brought from the external primary memory to the internal instruction register of the processor. There are several distinct steps in this procedure. The output address is placed on the address bus, while appropriate control signals are established on the control bus. The instruction is transferred from the data bus to the instruction register after a suitable delay or alternatively after WAIT states or data acknowledge. When this is completed, the program counter is incremented by 1. Subsequently, other instruction bytes are fetched if required.

To illustrate the EXECUTE cycle, consider the implementation of the Z80 instruction LD A, (100). Firstly, the address 100 is placed onto the address lines, while the control signals are placed on the control bus. Data is loaded into the processor, before subsequent internal bus transfer to register A. If the instruction were LD (100), A the data would be transferred from register A via the internal bus to the data bus.

2 REGISTER TRANSFER LOGIC, OR RTL

The transfer of data between registers can be specified using a register transfer notation. For example, transferring the contents of register B to register A could be indicated by

$$\mathrm{A} \longleftarrow \mathrm{B}$$

Such a notation has found application in hardware description languages, or HDLs. They can be used to describe the operation of complex digital systems in which other techniques such as state diagrams are inappropriate.

Prefixing the transfer with a colon can include control signal operation

$$\mathrm{T1}\colon \mathrm{A} \longleftarrow \mathrm{B}$$

Thus, transfer takes place when T1 is high. Also

$$\mathrm{T1} \times \phi\colon \mathrm{A} \longleftarrow \mathrm{B}$$

indicates transfer when both T1 and ϕ are high.

Arithmetic operations in the ALU can also be specified. Thus, if register B goes to the ALU R-BUS and C to the L-BUS, then an ADD operation gives

$$\mathrm{T1}\colon \mathrm{A} \longleftarrow \mathrm{B} + C$$

Internal processor operation can be specified in terms of this notation. As an example, consider the Z80 processor. Assume that the control signals are generated by the loading of one bit registers corresponding to $\overline{MI}$ (Instruction fetch),

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

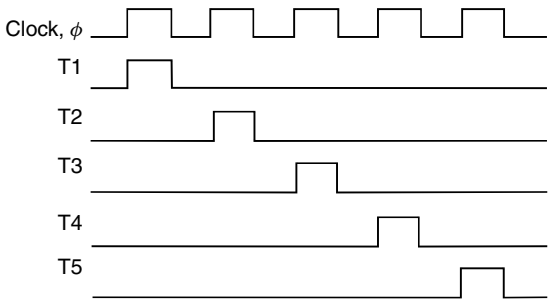


Figure 1. Control state timing for the instruction execute phase.

$\overline{MREQ}$, $\overline{RFSH}$ (dynamic RAM refresh), $\overline{RD}$, and $\overline{WR}$ Processor control inputs include ϕ (system clock) and $\overline{WAIT}$. Registers are loaded on the rising edge of the specified control signal, that is, T1, T3, T3, and so on.

For *instruction fetch*, the following transfers would be required;

FT1: ABUS ⟵ ABF ⟵ PC, 91- ⟵ 0

FT1: PC ⟵ PC + 1

FT$\overline{1}$: $\overline{MREQ}$ ⟵ 0, $\overline{RD}$ ⟵ 0

FT3: IR ⟵ DBUS

FT3: ABUS ⟵ RI, $\overline{MREQ}$ ⟵ 1, $\overline{RD}$ ⟵ 1, $\overline{M1}$ ⟵ 0, $\overline{RFSH}$ ⟵ 0

FT3: $\overline{MREQ}$ ⟵ 0, R ⟵ R + 1

FT4: $\overline{MREQ}$ ⟵ 1

FT1: $\overline{RFSH}$ ⟵ 1

For a *memory-to-register read* of the form LD A, (100)

ET1: ABUS ⟵ ABF

ET1: $\overline{MREQ}$ ⟵ 0, $\overline{RD}$ ⟵ 0

ET$\overline{3}$: $\overline{MREQ}$ ⟵ 0, $\overline{RD}$ ⟵ 1, A ⟵ DBUS

To obtain a *literal* of the form LD A, n from PROGRAM MEMORY, the following would occur after the op-code had been fetched.

ET1: ABUS ⟵ ABF ⟵ PC (PC points to next instruction byte)

ET1: PC ⟵ PC + 1

ET$\overline{3}$: $\overline{MREQ}$ ⟵ 0, $\overline{RD}$ ⟵ 0

ET$\overline{3}$: $\overline{MREQ}$ ⟵ 1, $\overline{RD}$ ⟵ 0, A ⟵ DBUS

A *write-from-a-register* such as LD (100), A

ET1: ABUS ⟵ ABF

ET$\overline{1}$: $\overline{MREQ}$ ⟵ 0, DBUS ⟵ DBF ⟵ A

ET2: $\overline{WR}$ ⟵ 0,

ET$\overline{3}$: $\overline{MREQ}$ ⟵ 1, $\overline{WR}$ ⟵ 1

T1′ DBUS ⟵ Zhigh (T1 of next cycle)

Arithmetic operations can also be described. These include ADD A, B, which uses the following operations.

ETl: ALUR ⟵ A

ET2: ALUS ⟵ B

ET3: A ⟵ ALU

Conditional microoperations might also be required, for example, interrogating the $\overline{WAIT}$ line until it goes to 1. Others are required for implementing conditional branch instructions depending on the STATUS of the ALU. In register transfer logic (RTL), a typical statement is

T2: IF $\overline{WAIT} = 0$ THEN T2 ELSE T3

T3:

The T2 cycle is repeated until $\overline{WAIT} = 0$.

RELATED ARTICLES

Article 152, Fundamentals of the Stored Program Digital Computer, Volume 3; **Article 153, Single Address Instruction Microcomputer, Volume 3**; **Article 155, External Operation of the Microprocessor, Volume 3**.

REFERENCES

Conger, D. (1999) *Fundamentals of Microcomputers for Technology Students*, Prentice Hall, NJ.

Gaonkar, R. (1999) *Microprocessor Architecture, Programming, and Applications with the 8085*, Prentice Hall, NJ.

Karalis, E. (1997) *Digital Design Principles and Computer Architecture*, Prentice Hall, NJ.

Heffer, D.E., King, G.A. and Heath, D.C. (1981) *Basic Principles and Practice of Microprocessors*, Edward Arnold, London.

Wilkinson, W. and Makki, R. (1992) *Digital System Design*, Prentice Hall, Europe.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

155: External Operation of the Microprocessor

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Input and Output–Programming and Types 1041
2 Interrupts and Interrupt Servicing 1042
3 Direct Memory Access 1043
Related Articles 1044
References 1044
Further Reading 1044

1 INPUT AND OUTPUT–PROGRAMMING AND TYPES

There are four parts to a microprocessor system, the microprocessor, which is an integrated circuit, or IC, central processing unit, or the microprogram, data memory, which is also usually fabricated in IC technology, and the input/output circuitry. The last two topics will now be considered.

Data transfer to and from the input/output devices, using processor machine instructions, requires that checks be made to ascertain that the I/O device, or input–output device, has data ready before a read, and is ready to accept data before data is sent. This is known as handshaking (Heffer, King and Heath, 1981; Holdsworth, 1987; (*see* **Article 153, Single Address Instruction Microcomputer, Volume 3**); Tocci and Laskowski, 1987; Wilkinson and Makki, 1992). A status register, which is polled, can be used to indicate the state of readiness of the interface (Doyle, 1999).

The two types of I/O are parallel I/O, or PIO, and serial I/O, or SIO. In PIO, data is sent in parallel at 1 bit per line. Common interfaces such as PIOs and versatile interface adapters (VIAs) have the necessary processor interfaces for controlling the I/O device, interrogating its status, interrupting the processor, and sending an interrupt vector. Handshake lines such as data strobe/acknowledge, or STB, and data ready, or RDY, ensure controlled data transfers at the port input.

For SIO, the information is sent along one wire. It consists of a number of bits separated in time. The following terms, which are described in more detail in **Article 165, Introduction to Networked Instrumentation, Volume 3**; **Article 166, Instrument Interconnection, Volume 3**; **Article 167, Asynchronous and Synchronous Interface Protocols, Volume 3**; **Article 168, RS 232 and EIA/TIA 232 Serial Interface, Volume 3**; and **Article 169, Voltage and Current Loop Transmission, Volume 3** relate to serial transmission.

- *Simplex*: data is sent in one direction only,
- *Half duplex*: data is sent in both directions one at a time,
- *Full duplex*: data is sent in both directions simultaneously.

Serial signals may also be transmitted synchronously or asynchronously. In the synchronous serial interface, a block of characters is transmitted one bit at a time without stop-bits. It is 20 to 27% faster than asynchronous transmission for the same bit rate. Synchronization characters, such as insert 3 characters are sent at the start and end of transmission, and periodically in between characters. The synchronizing characters must not be valid characters. In ASCII, or American Standard Code for Information Interchange, the SYN (0010110) character cannot be used. Parts available include USARTs, or Universal Synchronous/Asynchronous Receiver/Transmitters and SIOs.

The ASCII coded signal shown in Figure 1 is a standard asynchronous format. Transmission rates up to

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

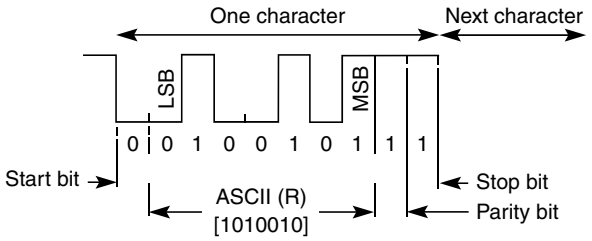


Figure 1. An ASCII coded signal.

9600 bps or 960 characters per second are common for video terminals. Parts used include universal asynchronous receiver/transmitters, or UARTs, or asynchronous communications interface adapters, or ACIAs, for microprocessors.

2 INTERRUPTS AND INTERRUPT SERVICING

Programmed I/O has the disadvantage that the processor must continually poll the I/O status register to see if data has to be sent or received. This slows down the processor. An alternative is to interrupt the processor during normal execution, causing it to service the I/O device. Special signals, such as *interrupt request* and *interrupt acknowledge*, are needed. Figure 2 shows the signaling lines and the interrupt handling scheme.

Interrupts are suited to unsynchronized I/O and randomly occurring I/O. It is especially suited to the multiprocessing environment, where several programs are interleaved. Interrupts can also be used to respond to internal processor hardware conditions such as ALU overflow.

An interrupt service routine, or ISR, is an integral part of every program, which allows the sequence of program operation to be suspended temporarily in an orderly manner. When one of the number of possible requests for interrupt servicing is acknowledged, the ISR may be initiated. Interrupt requests may be due to the following:

- An *unconditional jump* to a known location. This may be fixed in the processor design and activated as nonmaskable interrupts. Several locations may be available.
- In *vectored interrupts*, a jump address is provided to the processor by

  1. defined memory location *holding SR address,*
  2. interrupting device *providing SR address as a vector.*

Multiple interrupts may be caused if several devices are used to interrupt the processor. A priority interrupt number must be given to each so that simultaneous interrupts can be resolved. Priority can be established by the following:

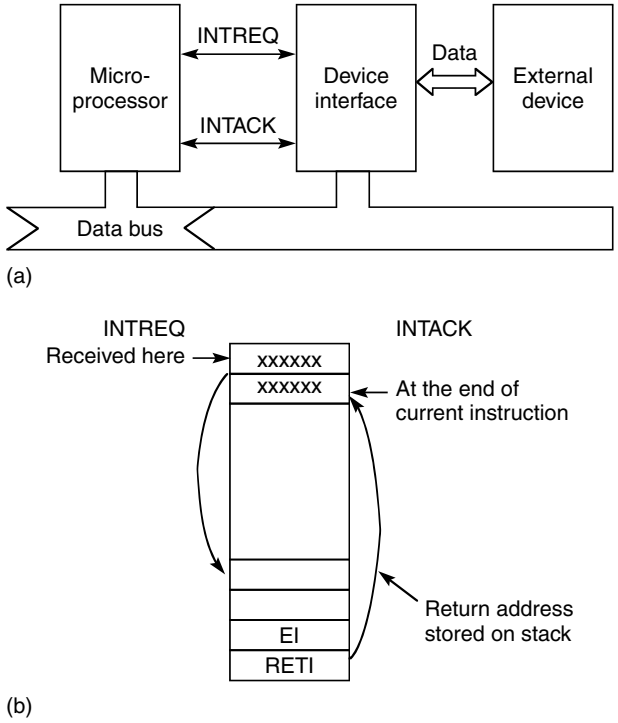


Figure 2. Interrupts in microprocessors: (a) interrupt request signals; and (b) interrupt handling.

- *Polled interrupts*: All interrupt requests are wire-OR'ed together for a single interrupt request, or INTREQ. The processor acknowledgment involves polling the devices to see which has sent INTREQ. This includes the following:

  1. The use of an INT FLAG on each device, which is set on INT. Processor checks flags in order of priority.
  2. The use of an INTERRUPT REQUEST REGISTER with bits set for each device. Set order of priority of bits in register.

- *Daisy chained interrupts*: One of the interrupt control lines passes through each device.

  1. *Daisy chained interrupt acknowledge*: As shown in the circuit of Figure 3, the highest priority is closest to the processor that receives the request directly. The interrupting device blocks the interrupt acknowledge, or INTACK, from the processor to devices with lower priority. Vectored ISRs can be used to identify the interrupting device.
  2. *Interrupt request*: The lowest priority device is closest to the processor. There is a common acknowledge to all devices.

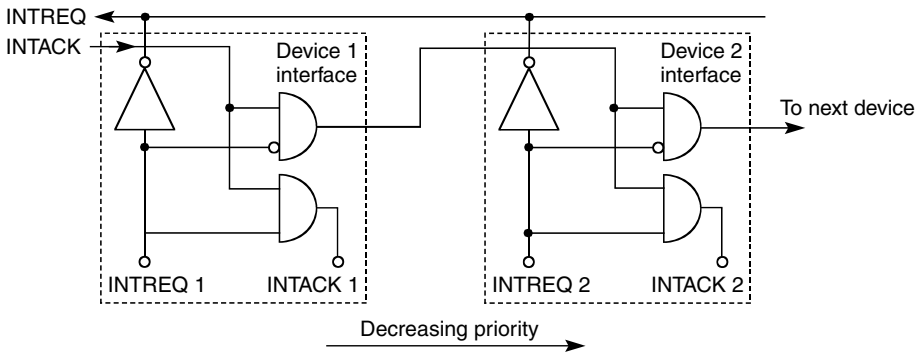


Figure 3. Handling interrupts using the daisy chained method.

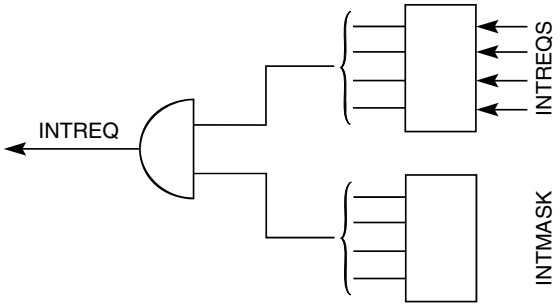


Figure 4. Added flexibility using an interrupt mask.

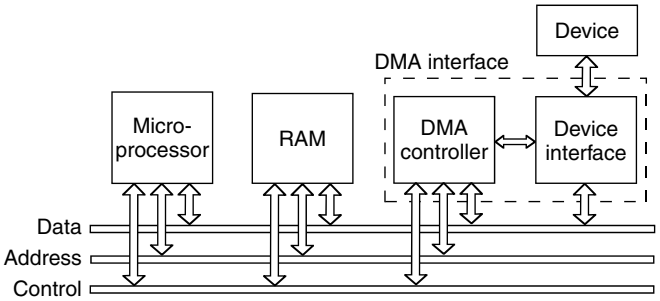


Figure 5. Use of a DMA controller.

Figure 4 illustrates the added flexibility of INTERRUPT MASKING. Higher priority devices can be selectively enabled immediately after accepting an INTREQ.

Interrupt controllers are devices that can

- handle multiple interrupts outside the processor,
- can support eight levels cascadable to more,
- can also supply interrupt vectors,
- can operate with fixed or rotating priority.

Rotating priority means that after the highest priority interrupt is serviced, it is then given the lowest priority. The next to highest becomes highest and so on. Priority therefore rotates thus preventing an interrupt dominating at the expense of others, which might not be serviced at all. This technique is also used in multiple bus-master systems.

3 DIRECT MEMORY ACCESS

Processor controlled transfers, such as memory to I/O, means that the processor cannot do other things. The speed of transfer is limited by the execution times of machine instructions. For large amounts of data, this is time consuming. A *direct memory access*, or DMA, controller, in the structural arrangement given in Figure 5, may be used as an attractive alternative. It is programmed from the processor by receipt of the starting address in memory plus the number of bytes to be transferred. The controller first captures the system data, address, and control buses, followed by transfer of the required number of bytes before signaling the processor that transfer is complete. The processor, which relinquishes the bus on receipt of a bus request, or BREQ, at the end of the current instruction, cannot use the bus again until transfer is complete. To allow the processor to continue operating, the DMA controller can 'steal' processor cycles, called *processor stealing*, to allow program execution.

The controller can operate in a variety of ways to

- fetch setup addresses and byte counts held in external memory,
- take data from source and compare it with the contents of a preloaded DMA register to find a data match,
- detect when a data match is not found. This is useful in memory testing where the same word is loaded into each location. If a different word is detected, an error is present at that location.

DMA is useful for devices such as magnetic disk memory units where a data block must be transferred as a constant stream and at a rate dictated by the shaft speed of the disk.

RELATED ARTICLES

Article 78, Embedded Systems, Volume 2; **Article 152, Fundamentals of the Stored Program Digital Computer, Volume 3**; **Article 154, Internal Operation of the Microprocessor, Volume 3**; **Article 158, Microcontroller Systems, Volume 3**.

REFERENCES

Doyle, L. (1999) *Computer Peripherals*, Prentice Hall, NJ, USA.

Heffer, D.E., King, G.A. and Heath, D.C. (1981) *Basic Principles and Practice of Microprocessors*, Edward Arnold, London.

Holdsworth, B. (1987) *Microprocessor Engineering*, Butterworth, London.

Tocci, R.J. and Laskowski, L.P. (1987) *Microprocessors and Microcomputers*, Prentice Hall International.

Wilkinson, W. and Makki, R. (1992) *Digital System Design*, Prentice Hall, Europe.

FURTHER READING

Lang, T.T. (1987) *Electronics of Measuring Systems*, John Wiley & Sons, UK.

Lang, T.T. (1991) *Computerized Instrumentation*, John Wiley & Sons, UK.

Rangan, C.S., Sarama, G.R. and Mani, V.S. (1983) *Instrumentation Devices and Systems*, Tata McGraw-Hill, New Delhi.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

156: Memory Management in the Microprocessor

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Introduction to Memory Management Systems 1045
2 Primary and Secondary Memory Word Block Transfer 1045
3 Replacement Algorithms 1046
4 Segmented System 1047
Related Articles 1048
Further Reading 1048

1 INTRODUCTION TO MEMORY MANAGEMENT SYSTEMS

When running large programs or when operating on data files that do not fit completely into the available primary memory, a memory management scheme must be implemented to control the transfer of information between the system's secondary, or disk memory, and primary or random access memory (RAM).

In simple systems such as microprocessors running MSDOS, a technique known as OVERLAYING is applied where existing programs are overwritten. A large software package is divided into a MAIN FILE and several OVERLAY FILES, which overwrite the main file. It may be necessary to restore the main file from time to time. Thus, a common problem is caused by interaction between program sections in different overlay files, which can cause disk threshing owing to excessive disk to memory transfers. Also, errors in different overlays can be difficult to detect.

2 PRIMARY AND SECONDARY MEMORY WORD BLOCK TRANSFER

In more sophisticated management systems, automatic replacement of memory blocks in primary memory with blocks obtained from secondary memory is desirable. Such systems require both the actual address used by the memory units and the program address produced by the processor. Translation between the two addresses is achieved by hardware and software. During program transfer, actual memory addresses may change but the program addresses used by the processor will remain the same. The two types of systems deal with memory in the form of pages of memory or segments of memory.

In paged or virtual memory systems, the actual address is the real address and the program address is the virtual address. A page is a memory block of fixed length of 128, 256, 512, 1024 words. An address is composed of a page number and a line number. Thus, there are two fields in the address as shown in Figure 1(a).

Memory management units, or MMUs, must provide hardware to translate the virtual page number into a real page number. This hardware works only on the primary memory. For secondary memory, software replacement algorithms are required to transfer pages from secondary to primary. The hardware checks to see if the page is in primary memory. If not, software must search secondary memory from which data is transferred to either a blank page or overwrites an existing one. Note the number of virtual addresses can exceed the actual number of available addresses in both memories as shown in Figure 1(b).

Address translation is achieved using one of two techniques. In direct mapping, as shown in Figure 2(a), all

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

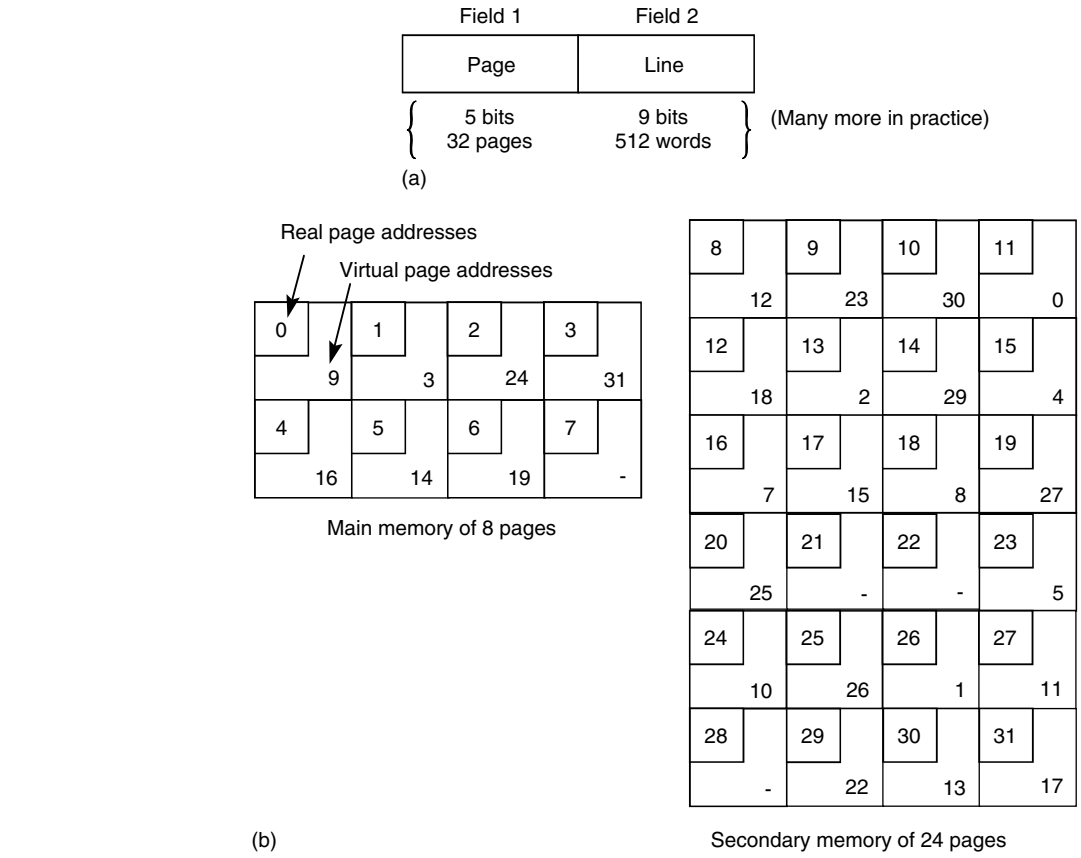


Figure 1. Memory management: (a) the two-field address structure and (b) translation of virtual to real page number.

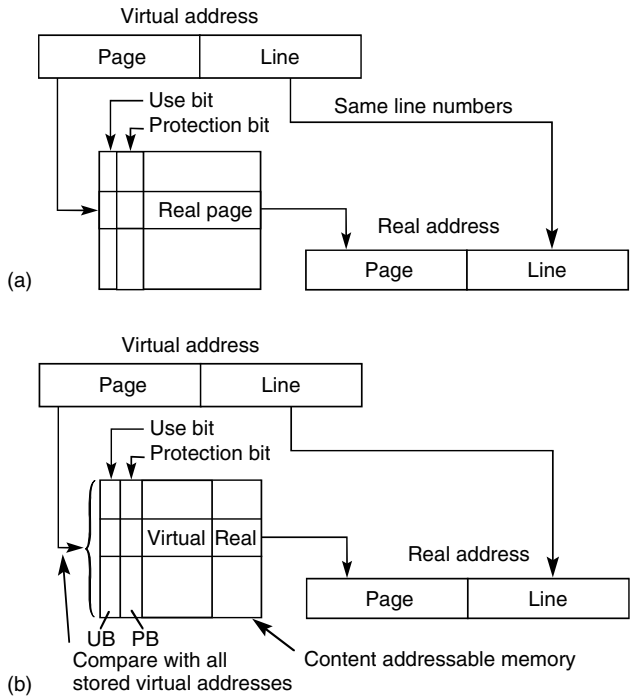


Figure 2. Address translation methods: (a) the direct mapping method and (b) the associative mapping method.

the real page addresses are stored in high speed RAM whose addresses correspond to the virtual page addresses and whose contents provide the real page addresses. The line numbers are the same in both the real and virtual page. Associative mapping uses a content addressable memory, or CAM, as shown in Figure 2(b). Here, a location is identified by its contents. The input data word is applied to the inputs of the CAM, which simultaneously compares data in all memory locations. The location containing the data is flagged as output. This memory, which uses bipolar technology, is fast, with access times of typically 30 ns. However, it is expensive. It is linked to a RAM, which contains the real addresses. Thus, the processor sends the virtual address; it is compared with all others, and if a match is found, the real address is accessed from the RAM.

3 REPLACEMENT ALGORITHMS

The transfer of data between primary and secondary memories can be achieved using a number of software algorithms. In random replacement, pages to be replaced in primary memory are chosen at random. This is not very good, since most programs access locations in proximity to each other.

First-in-first-out-replacement (FIFO) is an effective strategy for replacement. The incoming page replaces the page that is resident longest in memory. A rotating pointer identifies the real page in primary memory on a list where replacement has to take place. If a page fault occurs, which means the page is not in primary memory, the virtual page address at the pointer location is replaced by the new virtual page address. The counter is then incremented to the next location. Note that in a direct mapping scheme, the contents of the new virtual page address would be loaded with the real memory address of the pointer. In an associative scheme, the new virtual page address would be stored at the location corresponding to the real memory address, in essentially the same way.

First-in-not-used-first-out-replacement is a modification of the first-in-first-out-replacement, described above. It avoids unnecessary transfers of recently used data. A use bit is included, which is set if a page is referenced after initial loading (use bit = '0'). When a page fault occurs, the use bit at the current pointer location is checked. If it is 1, the bit is reset and the pointer incremented to the next location, where the bit is again checked. If the bit is zero, the location is loaded with the virtual address. Transfer of pages between primary and secondary memory then occurs and the counter is incremented.

Least recently used replacement, or LRU replacement, makes use of a list of the virtual page addresses held in primary memory. The list is updated by shifting every existing list line down by one when each new page is inserted at the top of the list as it is brought from primary. The page at the bottom of the list is then replaced. If, however, the page accessed is already on the list, it is moved to the top of the list, its preceding pages are moved one location down, while those after remain unchanged. Note that the actual pages are not reordered in primary memory. If many pages are accessed, then implementation is difficult. In one scheme, a counter is held for each page and regularly incremented. If the page is accessed, it is reset. The page with the highest count is replaced. It is expensive in hardware. In an alternative scheme, associative memory is used to record the primary memory pages. It is used for the top 16 or 32 entries while the rest are held in the RAM. This scheme essentially implements the list.

Some general points are worth emphasizing. The use bit can be incorporated into the virtual address translation table together with protection bits; for a read-only page, retain page in primary memory. The page size is typically 512 words and is a compromise. A small page means that (a) the transfer time is short and (b) there is a large selection of different pages in primary memory. Hence, a large page table is required. The minimum page size is constrained by disk sector size. For small pages, an increased proportion of information is required for sector identification. For a large page size, transfer times are long and there is an increase in unused space at the end of each page. This is called *internal fragmentation*.

4 SEGMENTED SYSTEM

As shown in Figure 3, the memory space is divided into different sizes of blocks, or segments, at contiguous locations. This technique may be better when programs and data are generated in various sizes. Segments can overlap, with the segment address consisting of a number plus a displacement, or offset, within the segment. The offset and segment are added together to form the real address. The logical address is the virtual address and the physical address is the real address.

Translation of logical addresses uses a segment table and direct mapping, as shown in Figure 4. The segment table can include protection bits in the form of a length field

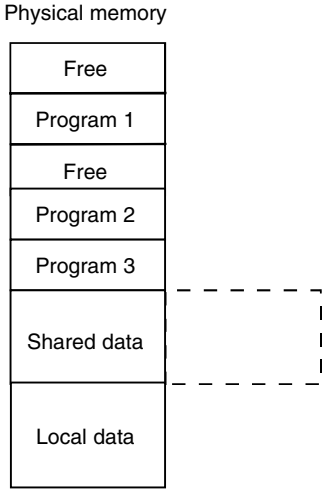


Figure 3. Division of memory into contiguous blocks.

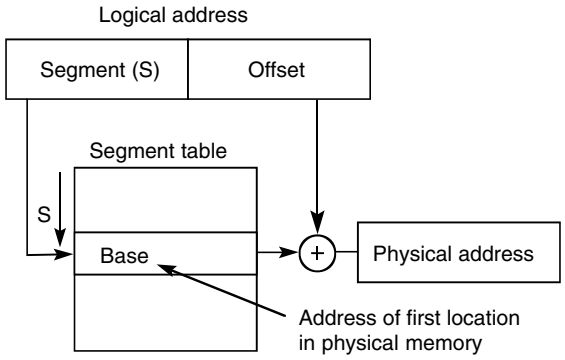


Figure 4. Addition of segment number to segment table.

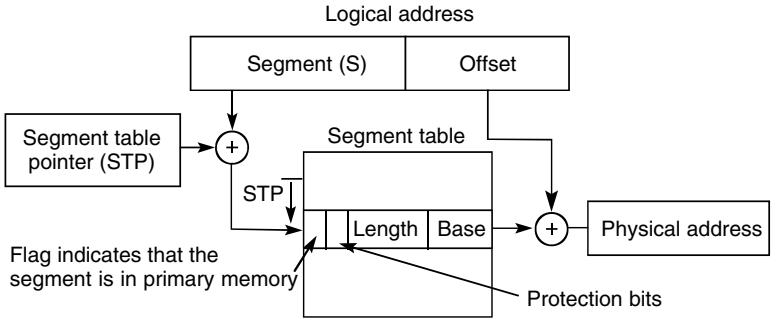


Figure 5. Translation of logical to real address.

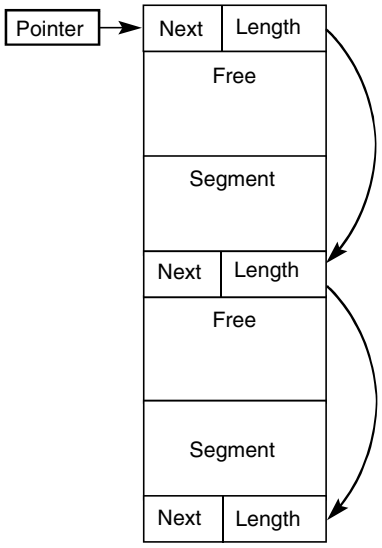


Figure 6. Implementation of the first fit algorithm.

so that referencing beyond a segment is prevented. Several segment tables can also be used.

In Figure 5, the segment number, S, is added to the segment table pointer (STP) to reach the correct address in the segment table. A flag is used to indicate that the segment is in primary memory. The offset is compared with the segment length and if the offset is greater, a segment error signal is generated. Otherwise, the segment base address is added to the offset to access the physical memory location.

Replacement algorithms, although similar to the pages system, must take account of the varying segment lengths. Care must also be taken to prevent external fragmentation by leaving a lot of unusable spaces.

For the first fit algorithm, a word that points to the next free space in memory and the length of the current space, as shown in Figure 6, is stored at the beginning of each free space in memory. A pointer, P, points to the first free space. If a replacement is required, the next/length word is read. If the length is greater than the segment to be inserted, transfer takes place with the next/length word inserted at the end. Meantime, the length is changed to indicate the remaining space. The next in the word of the previous free space is changed to point to the new next. If the length is less than the new segment, the space is ignored and the next space is checked until one of sufficient length is found. To prevent external fragmentation, a minimum residual free space can be specified.

RELATED ARTICLES

Article 78, Embedded Systems, Volume 2; **Article 152, Fundamentals of the Stored Program Digital Computer, Volume 3**; **Article 153, Single Address Instruction Microcomputer, Volume 3**; **Article 158, Microcontroller Systems, Volume 3**.

FURTHER READING

Gaonkar, R. (1999) *Microprocessor Architecture, Programming, and Applications with the 8085*, Prentice Hall, NJ.

Karalis, E. (1997) *Digital Design Principles and Computer Architecture*, Prentice Hall, NJ.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

157: Data Acceleration in Computers

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Single Thread Systems 1049
2 Acceleration of Main Memory at the Architectural Level 1050
3 Cache Memory 1051
4 Operation of the Cache 1052
5 Memory Fetch Policies 1053
6 Cache Mapping Schemes – Placement Policies 1053
7 Cache Mapping Schemes – Direct Mapping 1053
8 Cache Mapping Schemes – Associative and Set-associative Mapping 1053
9 Cache Mapping Schemes – Replacement Policies and Memory Update 1054
Related Articles 1055

1 SINGLE THREAD SYSTEMS

In many systems, where only one processor is used for program execution, there is only one program counter. Presently, there is a tendency to include coprocessors in dense program execution systems. When this is the case, the program execution time, T_{prog}, is then

$$T_{\text{prog}} = T_{\text{c}} \times \sum_{\text{i}=1}^{N} I_{\text{i}} \overline{C_{\text{i}}} \tag{1}$$

where T_{prog} is the program time, T_{c} is the cycle time, which depends on the architecture, I_{i} is the number of times instruction i is executed, which depends on both the program and the architecture, and $\overline{C_{\text{i}}}$ is the number of cycles or instructions, which depends upon the architecture. All of the variables determining T_{prog} depend upon the architecture of the system used while the indexing also depends upon the program.

Acceleration may be achieved at the manufacturing or realization level of the processors. The most commonly used manufacturing technique is the Very Large Scale Integration, or VLSI, process. For example, the Intel 80386 IC operating at 16 MHz has 275 000 transistors (100 mm^2), the MC68030 integrates an MC68020 CPU with an MC68851 MMU.

It is also possible to implement acceleration techniques at the system implementation level. The various elements involved in data manipulations together with their typical access times and capacities are shown in Figure 1. A number of terms should be explained first. Latency is the total time between the start of an operation and the availability of a result. The bandwidth is the number of operations per second, expressed as say 40 Mbytes/s over a bus or 2 MFlops/s for floating-point operations. The time taken to respond to a memory request is called the *memory latency time*.

Caches improve memory latency provided they are high-speed (∼50 ns). *Memory bandwidth* is also important. For example, a memory produces one byte in 100 ns and 8 bytes in 800 ns. Another produces the first 2 bytes in 200 ns and thereafter 2 bytes in 50 ns giving a total for the 8 bytes of 350 ns. Which is better? This introduces the concept of pipelining, or interleaving memory paths. Parallelism where wide data paths or multiple paths are used, is also important. Execution Latency is the delay associated with arithmetic operations. Having other processing units such as multipliers and coprocessors can solve this problem. Execution bandwidth is again effected by pipelining and

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

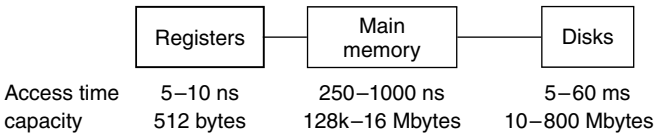


Figure 1. Illustration of memory latency for registers, main memory, and disks.

parallelism. Pipelining ensures that independent operations and instructions are processed simultaneously. This means that execution of the next instruction should start before the first has finished. For example, adding two floating-point numbers should be executed in the manner illustrated in Figure 2.

2 ACCELERATION OF MAIN MEMORY AT THE ARCHITECTURAL LEVEL

This is illustrated in Figure 3. It is based upon using special purpose processors for signal processing where the ALU and multiplier are on a single chip. Another approach used reduces the hardware complexity by using reduced instruction set computers, or RISC, so that the numbers of unique instructions, addressing modes, and data types are lower. Eliminate large microprogrammed controllers. Migrate complex functions to the compiler and away from execution time. It is possible to accelerate main memory using

- multiple memory paths,
- interleaving partitions,
  - low-level interleaving,
  - high-level interleaving.

Multiple memory paths use the Harvard architecture of Figure 3(a) instead of the von Neumann architecture, where instructions and data are mixed and share a common memory. The Harvard uses two memory paths, one for instructions and the other for data. It is usually implemented internally with data and instruction caches, while externally it appears as a von Neumann machine.

Memory may be divided into multiple physically independent modules, as shown in Figure 3(b) and (c). Memory interleaving partitions are then utilized with either low-order interleaving, illustrated in Figure 3(b), or high-order

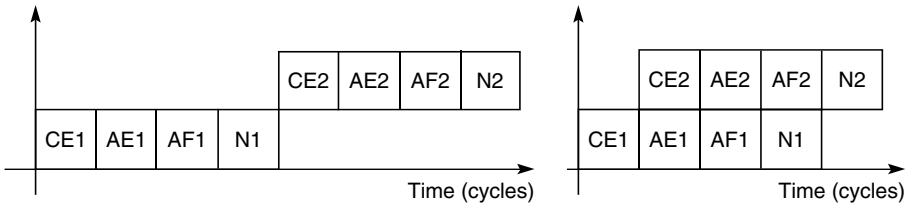


Figure 2. Principles of execution latency.

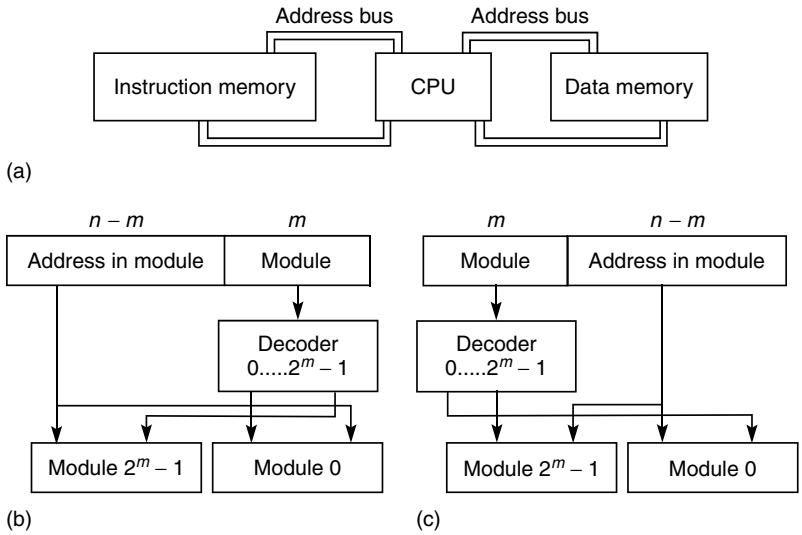


Figure 3. Acceleration of main memory using accelerating structures at the architectural level: (a) multiple memory paths, (b) low-order interleaving, and (c) high-order interleaving memory.

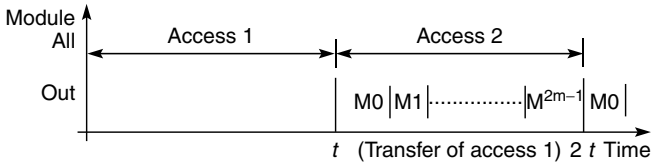


Figure 4. Speeding up memory access.

interleaving, illustrated in Figure 3(c). Low-order interleaving puts consecutive addresses in consecutive modules. It supports spatial locality and is good for vector elements and successive instructions. However, the failure of a module is catastrophic. In high-order interleaving, the data is in one module. In this method, shown in Figure 4, the address space is split into M contiguous parts. It does not support spatial locality since all are successive, easy to expand, and have fault tolerance. Low order is preferred in memories such as caches because it has spatial locality.

To speed up memory access, all the modules have to process memory requests simultaneously as in Figure 4. For low-order interleaving, simultaneous access is possible. High-order $n - m$ bits are sent to the modules at the same time with buffering of the output from each module. The low-order bits activate a multiplexer to select each module output in turn with the second memory access overlapping with the multiplexer accesses.

If all M words are accessed, memory access time is divided by M, the total number of modules. For nonsequential access to different modules, this scheme breaks down.

Concurrent Access overcomes the low efficiency of nonsequential access by offering to each module different high-order addresses, which are processed at the same time. Buffers now hold the memory addresses for each module allowing each to operate independently. A maximum speedup of M is again possible and the modules can be accessed at random. However, a memory control unit is required.

3 CACHE MEMORY

These are high-speed buffers between the main memory and the processor memory. Their success depends on locality. If the entire working set of a process can be placed into the cache, a substantial increase in performance can be achieved.

Purposes of Cache Memory are now discussed. There is usually a mismatch in the speed between transfers in primary and secondary memories, the latter being much slower. There is also a speed mismatch between primary memory access and processor operation times because the processor can operate faster. Having a cache of high-speed

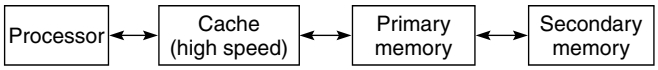


Figure 5. Structural principle of cache memory.

bipolar memory resident between the primary memory and the processor as shown in Figure 5 can reduce this problem.

The cache of limited capacity (512 words), which contains the current processor instructions and data, is updated as required, by reading from the primary memory. Outputs from the processor, which are then copied to the primary memory, can also update the cache. The cache works because the principle of locality applies, that is, instructions, and to a lesser extent data, are accessed locally to each other, and program loops are common. Thus, there are many cache references to each primary reference. If the first reference is obtained from primary memory and transferred to cache in time M s, and there then follows n references by the processor to cache in access time m s, the average access time, T_a, is

$$T_a = \frac{(M + n \times m)}{n} \tag{2}$$

Note the assumption that the output from primary goes to cache first before being accessed by the processor. In practice, output could go to both. Also, note that write operations to cache must be copied to primary memory at some stage. This will slow average transfer times.

For $M = 400\,\text{ns}$, $m = 50\,\text{ns}$ and

$n = 10$, then $T_a = 90\,\text{ns}$ rather than 400 ns.

To match the speed of data transfer to the cache, the primary memory is interleaved so that a number of words are accessed concurrently. The memory is divided into a number of modules, usually a power of 2, and the primary addresses are numbered across the module as shown in Figure 6. The number of modules is chosen to match the

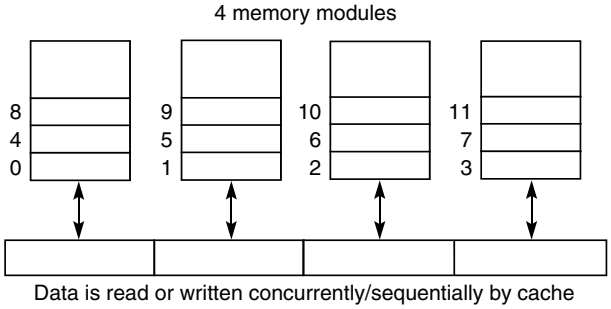


Figure 6. Numbering of primary addresses across the memory modules.

cache to which each word is transferred sequentially. Thus, for the example above, 8 memory blocks would be required having an access time per word to cache of $M = 50$ ns. For $n = 10$ the average access time, T_a, is now equal to 55 ns.

An important parameter in cache operations is the probability that the required word is in cache and is of the order of 0.8. A *hit* occurs when a cache location contains the required data, otherwise a miss results. The hit ratio, h, is

$$h = \frac{\text{the number of times the required word is found in cache}}{\text{the total number of references}} = \frac{\text{the number of hits}}{\text{the total number of references}} \quad (3)$$

Example A cache operates with a hit ratio of 0.85. Using the memory parameters as before with $M = 400$ ns and $m = 50$ ns, calculate the access time, T_a. Assume that the cache is loaded before the processor is accessed.

Solution:

$$T_a = \frac{(15 \times 400 + 15 \times 50 + 85 \times 50)}{100} = 100 \text{ ns} \quad (4)$$

4 OPERATION OF THE CACHE

A TAG is a memory address. Each entry in cache consists of a line of data, which can be several memory words, as represented in Figure 7. For an M-way interleaved memory, a cache line can be M words or $2M$ words or $4M$ words. TAGS are matched in the address field. If a hit, then select byte in line and update line status, which reflects reference, for example, write. If a miss, send the real address to main memory. During read from main memory, the cache selects a line in cache to be replaced. For virtual memory, a memory management unit, or MMU, is needed for address translation for cache and memory access.

The translation can be performed serially as shown in Figure 8(a). The alternative is to do it in parallel, as in Figure 8(b). This is called D-displacement. An extra field

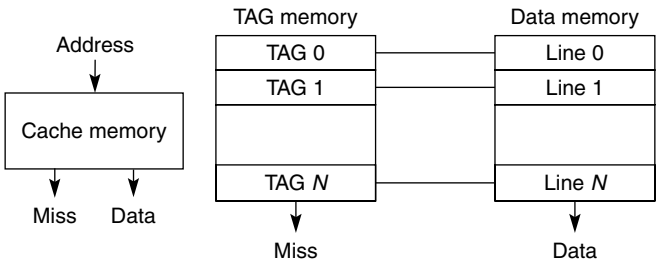


Figure 7. Cache memory TAG principles of cache operation.

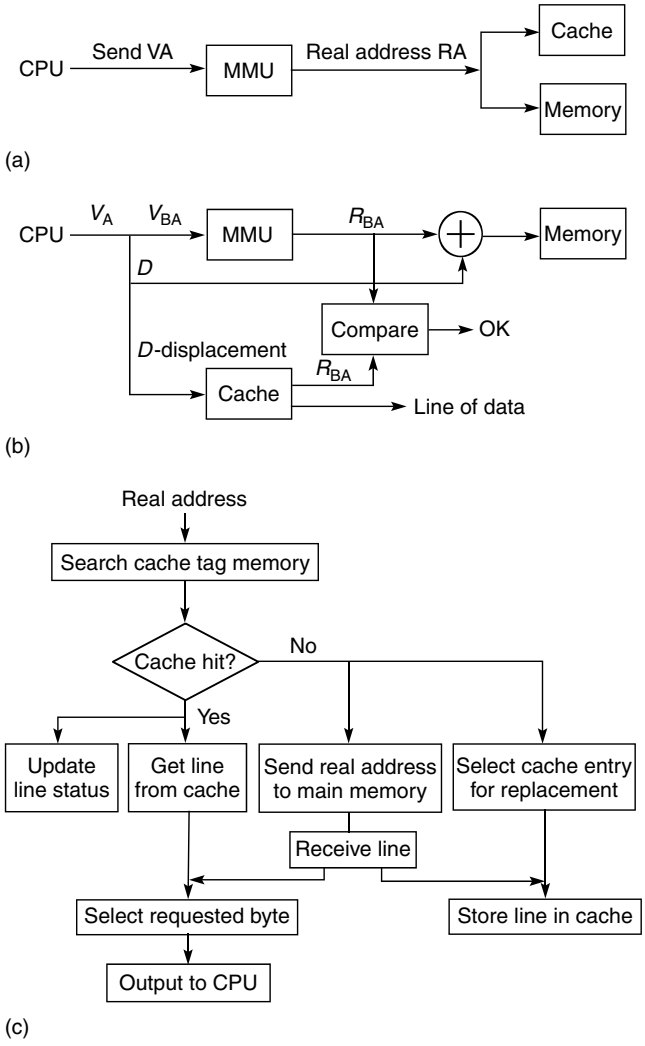


Figure 8. Operation of the cache: (a) serial, (b) parallel, and (c) typical flow of instructions.

is needed to hold R_{BA}. These assume that the cache is in the real address space, which is the most common.

The flow diagram for cache operation is given in Figure 8(c). When the real address is specified, the cache TAG memory is searched. If there is a hit, the line status is updated and the line is fetched from the cache. Subsequently, the operation, which is straightforward, selects the requested bytes and sends them to the CPU. If there is a cache miss, the real address is sent to main memory and the cache entry is selected for replacement. The main memory receives the real address, the receive line is activated and the line is stored in cache, and the requested bytes are selected.

In cache design, two hit ratios are of interest. The cold-start hit ratio applies to an initially empty cache, while the warm-start hit ratio applies when a process already has data in the cache.

Line size and cache size are the two most important factors regarding hit ratio. Long lines improve the spatial locality but it takes time to move them. For microprocessors, typical sizes are 16 to 64 bytes. A large number of lines improves the temporal locality up to a point. When greater than the working set, this will not result in a decrease in the miss ratio.

5 MEMORY FETCH POLICIES

There are various strategies in determining what lines should be fetched from main memory and loaded into the cache. Demand fetching fetches the line when it is needed. This is always implemented to some degree, but a measure of prefetching can improve matters. Prefetching tries to fetch lines, which are going to be needed in the near future. In high-speed caches, it is the line immediately following the fetched line i, that is, line $i + l$. Hence, fetch line 1 and prefetch line $i + l$. There are a number of policies that can be implemented. One is always prefetch, which tries to prefetch a line whenever line i is referenced. This increases memory traffic by about 80%. An alternative is to prefetch-on-a-miss, that is, when a cache miss occurs, line i is fetched and line $i + l$ is prefetched. Another more common method is tagged-prefetch, which prefetches on a miss and on the first reference to a prefetched line. This increases memory traffic up by 10 to 20% for the same miss ratio as always prefetch.

6 CACHE MAPPING SCHEMES – PLACEMENT POLICIES

Cache is smaller than main memory, so that a few lines map into each cache line. It is necessary to decide where to put data lines in the cache. Primary memory is hardware mapped into the cache by three main ways.

7 CACHE MAPPING SCHEMES – DIRECT MAPPING

Direct mapping is shown in Figure 9(a). A TAG field in the processor address is used to identify a page in primary memory, while an INDEX field finds the word within the page. Thus, words with different tags (pages) but the same index will share the same location in cache. To see if a word is in cache, the tags are compared. If the tags are the same, the word is read, otherwise the word is accessed from primary memory. For interleaved memories, a block of words can be transferred to cache as shown in Figure 9(b). Note that common blocks in different pages occupy the same cache locations. Take as an example a main memory

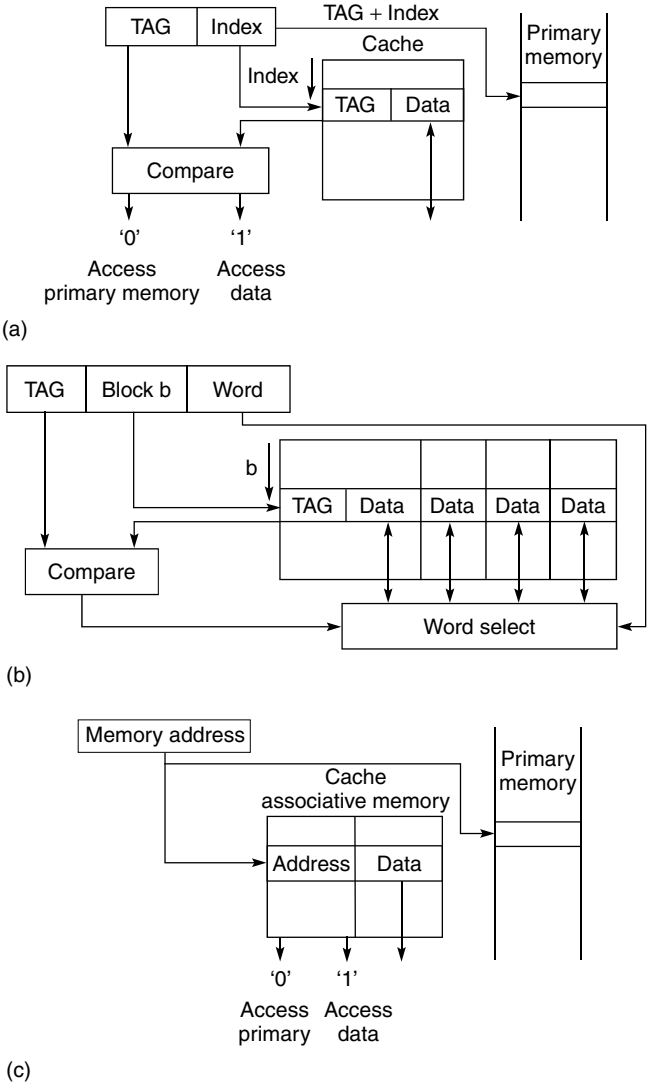


Figure 9. Mapping schemes for cache memory: (a) direct type for noninterleaved, (b) interleaved, and (c) associative type.

consisting of 1 MB with a 20-bit address, and a 16 kB cache with 16 bytes per line. Thus, there are 64 klines in main memory and 1024 lines in the cache. In direct mapping, line i is placed in line (i modulo 1024) in the cache. The advantage of this method is that associative comparison is not needed and the replacement of lines is trivial. However, the cache can only contain one of the 64, or 26, lines that map into each cache line one at a time.

8 CACHE MAPPING SCHEMES – ASSOCIATIVE AND SET-ASSOCIATIVE MAPPING

Associative mapping, shown in Figure 9(c), uses a content addressable memory, or CAM, to compare the incoming

address with all the cache addresses simultaneously. If the address is present, the data is read, otherwise it is accessed to the cache from primary memory and overwrites data. This scheme is expensive since words are stored in the CAM. An alternative is to store a block so that the block address is compared associatively reducing storage requirements. In both cases, during block or word replacement, the existing block or word must be chosen either on a random first-in-first-out or least recently used basis.

Fully associative cache is used when any line can reside at any location and is found by comparing the tag contained in the address field with the tags held in the cache. It has the highest hit ratio but may be slower due to the comparator hardware required for the matching. It is expensive to implement, but this could change with developments in VLSI.

Set-associative mapping, as illustrated in Figure 10, allows a block of data with the same index but different tags to be stored. It combines the advantages of direct mapping with fully associative placement. To find the required word, the tags are compared associatively and the required word is found. The cache is partitioned into a number of sets, S, containing a number of lines, L. The sets are directly mapped while the lines in each set are associatively mapped. Thus, for example, given that the number of lines per set is chosen to be $L = 4$, then the number of sets $S = M/L = 256$, where M is the number of lines in main memory. To find a line, the middle 8 bits of the address are used to find the set of line i such that it is placed in set (i modulo S) or (i modulo 256). The 8-bit tags within each set are then associatively mapped. Note that each set can contain any one of 256 lines. However, since each set has 4 lines, then 4 out of these 256 can be stored at the same time. This is better than direct mapping. Since the associative search is only over L elements, rather than M, it is much less costly than a full-associative cache. Note also that for $S = l$, the cache is fully associative, while for $L = l$ (one line per set) the cache is directly mapped.

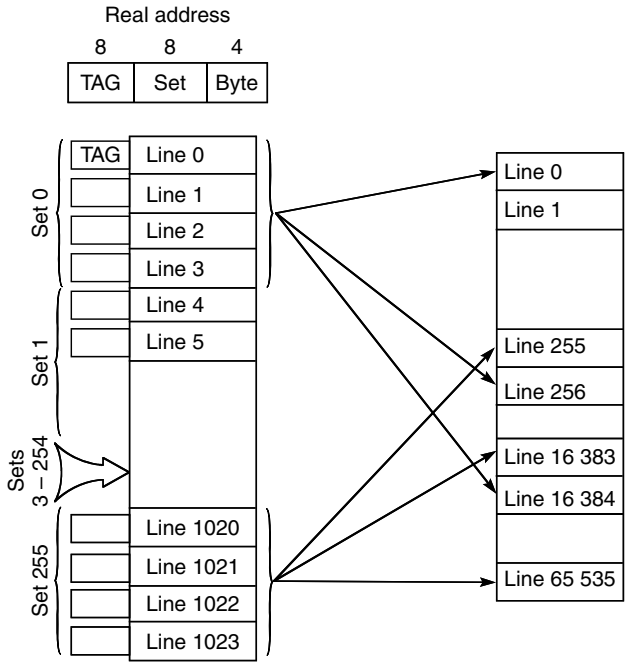


Figure 10. Cache mapping using the set-associative method.

9 CACHE MAPPING SCHEMES – REPLACEMENT POLICIES AND MEMORY UPDATE

Replacement policies: When a cache miss occurs and the cache is full, the question of which line to replace arises. This requires a cache replacement policy. The same techniques that were used for management of the main memory can be applied. However, since usage of the cache operates at maximum speed, such techniques must be implemented in hardware. Also, for simplicity, fixed sized partitions are used with each process occupying either the whole or a fixed area of cache. There are two approaches. One is based on the usage of line and the other ignoring the usage. LRU takes account of when the line was last referenced and has the highest hit ratio but is costly to implement. Random and FIFO policies are two nonusage based methods. Statistics show that the influence of placement is greater than that of replacement. For direct mapping, replacement is trivial. For a fully associative cache, replacement is difficult.

Main memory update: If data in cache is replaced by a write operation, information will be lost if the line has been modified and not updated in the main memory. It is necessary to ensure that any data written to cache is also written to main memory so that I/O operations such as DMA can operate on primary memory. There are two commonly used update policies.

Write-through: In this approach to memory update, every write operation from the processor always transmits changes to the cache and main memory. Although this takes time, there are usually some reads to cache. Hence, writes to primary memory can take place while the cache reads take place. This method is good for multiprocessor systems with multiple caches.

Copy-back also called write-back: There are three versions of this method. In the first two, a cache line is only transmitted to main memory if it is to be replaced in the cache. In always-copy-back, no account is taken of whether the line has been modified. In flagged-copy-back, a line is

copied only when it has been modified as indicated by a 'dirty flag'. This policy means that the main memory can contain 'stale' data, which will be lost if there is a processor or cache failure. The third method of writing to cache is known as *write-back*. Here, updating occurs during block replacement time. A tag bit indicates when a block has been updated. If a block has to be transferred into an existing block location, the old block is checked to see if it has been changed. If so, it is transferred to the primary memory.

Optimization of the cache is an important design consideration. Factors to be taken into account are hit ratio, access time on a hit, and delay time on a miss. A large cache gives a low miss ratio but a high cost. Note that if there is a write-miss, with the write-through policy, it is not specified whether the line is copied to cache. On copy-back the written line is usually copied to the cache on a write-miss.

RELATED ARTICLES

Article 152, Fundamentals of the Stored Program Digital Computer, Volume 3; **Article 153, Single Address Instruction Microcomputer, Volume 3**; **Article 154, Internal Operation of the Microprocessor, Volume 3**; **Article 155, External Operation of the Microprocessor, Volume 3**; **Article 156, Memory Management in the Microprocessor, Volume 3**.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

158: Microcontroller Systems

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Single Chip Microprocessors 1056
2 Microprocessor-based Data Acquisition 1056
3 Introduction to Special Purpose Digital Processors 1058
4 Bit-slice Devices and Elements 1059
5 The First Bit-slice Processor AMD 2901 1060
Related Articles 1061
Further Reading 1061

1 SINGLE CHIP MICROPROCESSORS

A typical single chip microprocessor has the architecture given in Figure 1. The instruction set is based on the standard 6800 instructions with additional functions. The central processing unit CPU has seven registers available to the programmer as shown in Figure 1. The two 8-bit accumulators, A and B, can be used by some instructions as a single 16-bit accumulator called the *D register*, which allows a set of 16-bit operations even though the CPU is technically an 8-bit processor.

Two index registers are available that may be used along with register D to perform fast 16-bit arithmetic operations. Two 16-bit by 16-bit divide instructions are also included.

2 MICROPROCESSOR-BASED DATA ACQUISITION

Microprocessors and microcontrollers form the heart of all modern acquisition systems. The microprocessor, which generally performs control, data handling, and data processing functions, requires external peripherals. Interfacing to these peripherals is achieved using a bus oriented connection consisting of a control bus, an address bus, and a data bus as shown in Figure 2. The external peripheral ICs (integrated circuits) handle input–output, I/O, operations, communications using Asynchronous Communications Interface Adapter, or ACIA, and external memory of both the read-only memory, or ROM, and random-access memory, or RAM, types.

The microcontroller unit, or MCU, which combines all these functions into one IC, thus simplifies the system design. This reduces costs and board space. A typical block diagram for an MCU is shown in Figure 3. The MCU example given in the figure, which is the Motorola 68HC11A8, is one version of the very popular 68HC family of microcontrollers. This controller, which is manufactured using the high density complementary metal oxide semiconductor, or HCMOS, IC technology, is an advanced 8-bit MCU with sophisticated on chip peripheral capabilities. Although the nominal bus speed is 2 MHz, it may operate at ultralow frequencies for reduced power applications. Note that the central subsystem is still a microprocessor.

The MCU I/O capability is grouped into ports. A port may be configured as a general I/O, similar to the VIA, or may have a specific function, such as serial communications, analog to digital converter, or ADC, input. Ports B and C may operate in single chip mode, where the MCU application requires no additional external memory, or in expanded mode, where external memory may be added. The latter requires the definition of an address bus, by combining ports B and C, and a data bus. Note that the data bus is multiplexed with low byte address lines to reduce the number of IC pins. The versatility of this configuration allows the device to interface, or be configured, to suit many applications.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

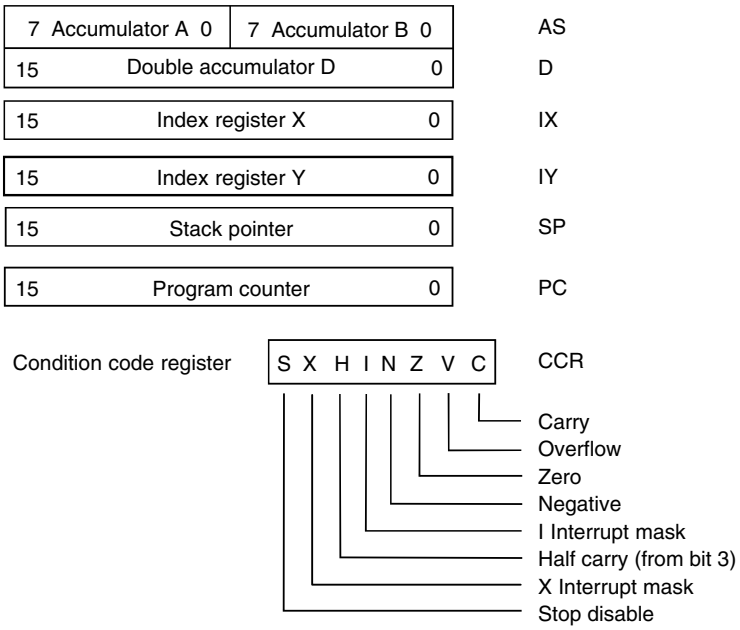


Figure 1. The programming registers in a microprocessor or microcontroller.

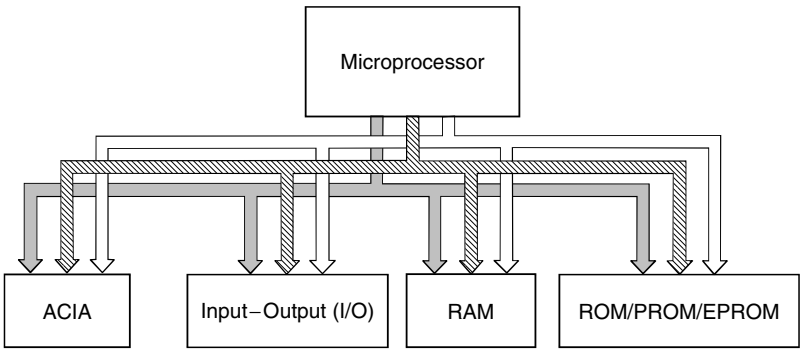


Figure 2. A microprocessor controlled data acquisition system.

The MCU also includes a 16-bit free running timer with three input capture lines, five output-compare lines, and a real time interrupt function. An 8-bit pulse accumulator subsystem can also count external events or measure external periods.

The version shown here also has an onboard 8-channel analog multiplexor, or MUX, feeding an 8-bit ADC. This allows up to eight individual analog sensors to be monitored by one MCU. Conversion is initiated under software control.

The device also includes an asynchronous serial communications interface, or SCI, and a separate synchronous serial peripheral interface, or SPI. Adding suitable line drivers allows a communication capability directly from the MCU for RS 232/RS 422, which are described in **Article 168, RS 232 and EIA/TIA 232 Serial Interface, Volume 3**. The SCI generates standard communication formats with 1 start bit, 8 or 9 data bits, and a stop bit, as previously illustrated in **Article 155, External Operation of the Microprocessor, Volume 3**. An on chip baud rate generator derives standard baud rate frequencies from the MCU oscillator. The SCI system can be used to connect a CRT or personal computer to the MCU. Alternatively, several widely distributed MCUs can use their SCI subsystems to form a serial communication network. This clearly allows remote data processing/information extraction as well as local data capture.

Self-monitoring circuitry is included on chip to protect against system errors. A computer operating properly watchdog system, or COP, protects against software failures. A clock monitor system generates a system reset in case the clock is lost or runs too slow. An illegal op-code detection circuit provides a nonmaskable interrupt if an illegal op-code is detected.

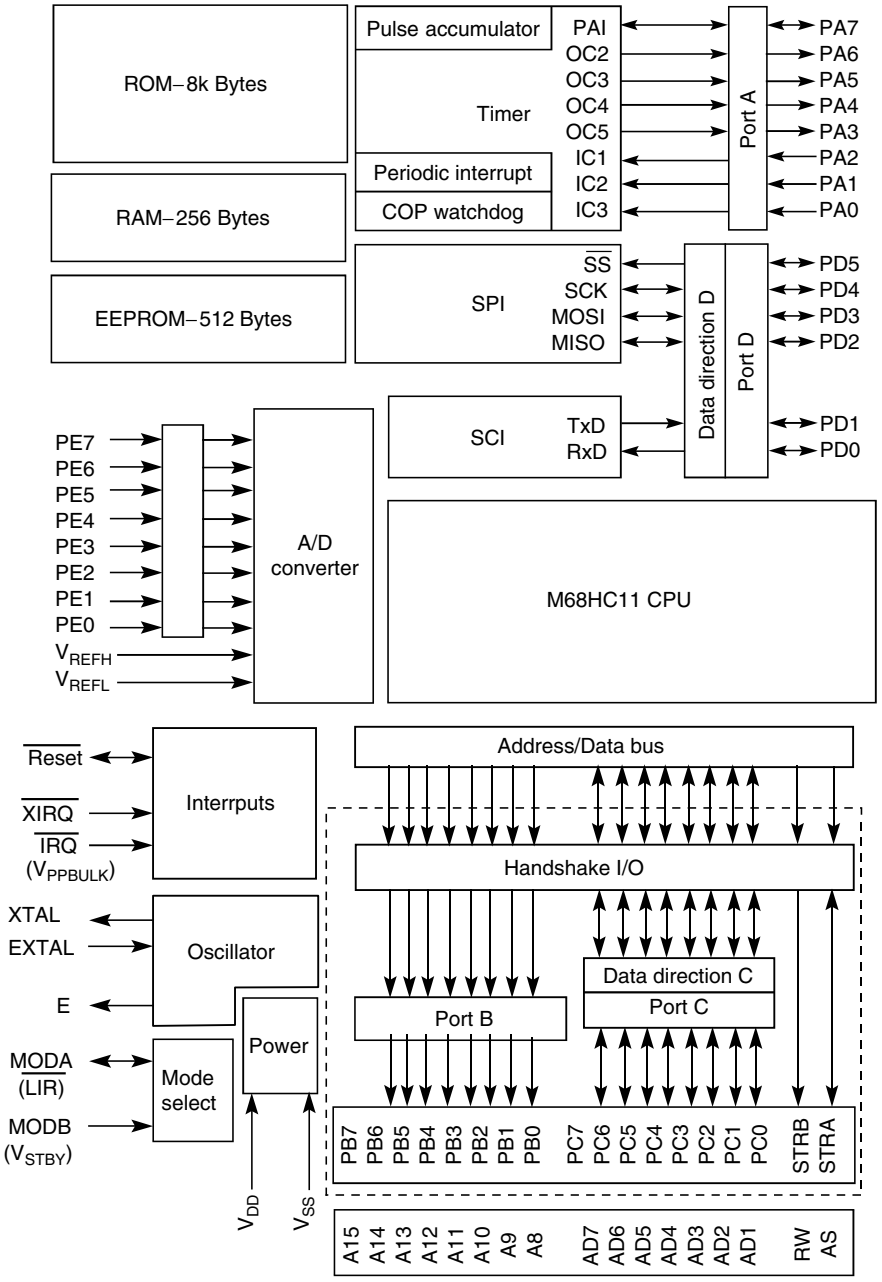


Figure 3. The Motorola 68HC11A8 microcontroller.

Two software controlled power saving modes of operation, WAIT and STOP, are available to conserve power. These modes make the 68HC11 family especially attractive for automotive and battery driven applications.

3 INTRODUCTION TO SPECIAL PURPOSE DIGITAL PROCESSORS

Digital systems of increasing complexity are playing a vital role in the processing of information. The reasons for this are as follows:

1. digital systems are less prone to error;
2. numerical accuracy depends on the number of variables, or bits, used;
3. information can be both numerical and nonnumerical;
4. a structured approach can be used by breaking the main task into a series of subtasks;
5. a general purpose system (computer) can be specified, which can be programmed for a particular task; and
6. VLSI technology has made the realization of complex systems cost-effective.

From an engineering point of view, the description and design of digital systems require the development of suitable models. For simple digital sequential networks, a systems model based on the state description of the system can be used with its concepts of output function and state transition function. For other systems, performing complex computations, an algorithmic model, which breaks down a complex digital computation into a sequence of primitive computations, becomes attractive. The algorithms themselves can be fixed by the hardware, or if storage is available they can be changed by programming. Such algorithmic models can be used for the specification, simulation, and implementation of complex digital systems.

The algorithmic model can be used at several levels of the system depending on the nature of the primitive computations. These range from (a) the hardware level, where logical operations are implemented by hardwiring networks of switching components; (b) the firmware level, where primitive computations involving hardware-like logical operations are implemented as microprograms; and (c) the software level, where language statements are executed in the computer.

A typical computer system consisting of memory, control, and processing elements has already been described in **Article 152, Fundamentals of the Stored Program Digital Computer, Volume 3**. The *data*, *control*, *processing*, *and I/O* structures of this system are all important. Arithmetic processing elements in the form of bit-slice components will be now introduced.

4 BIT-SLICE DEVICES AND ELEMENTS

Large scale integration, or LSI, of digital logic functions has led to higher switching speeds, due to shorter lead lengths, and higher reliability due to minimization of the number of interconnections. From an economic point of view, large scale use of a device is desirable (cf application specific integrated circuits or ASICs). This has led to microprocessors/computers on a chip and special purpose bit-slice elements, which are the building blocks of computer systems.

Bit-slice elements are fixed width computer hardware elements that range over 4, 8, 16, and 32 bits. A bit slice forms the basis of large computer systems, since they can be paralleled up. For speed, they are implemented in Bipolar Technology, which includes ECL, LSTTL, and ASTTL. Examples are 74LS481, MC10800, (ECL), and AMD 2901 (LSTTL). They can also be used for other applications such as signal processing. To be realized as a digital system, they require control logic. The use of bit slices allows partitioning of the system into a set of LSI building blocks of fixed words of width 4, 8, and so on to implement words in vertical slices of 8, 16, 32, and so on. This approach provides a reduction in the pin-out of the chip, a realization in LSI or VLSI, and a universal building block.

Bit-slice elements give the advantages of flexibility and speed when designing large systems with high performance. The trend has been to introduce higher levels of integration into bipolar families so that word length increases to 8, 16, and 32 bits. In the latter case, the ASS832 from TI (Texas Instruments) uses Advanced Schottky technology to obtain a 75-ns instruction execution. The structure of bit-slice elements may be organized to allow operation in either serial or parallel modes as in Figure 4. The serial modes are very similar to the conventional Von Neumann architecture. In the parallel units, many of the acceleration mechanisms already described in **Article 157, Data Acceleration in Computers, Volume 3** are used.

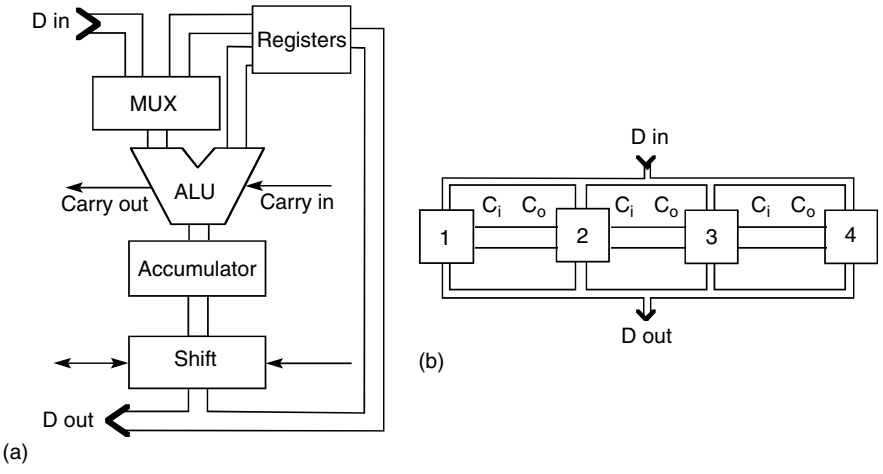


Figure 4. Architecture for typical bit-slice processors: (a) serial and (b) parallel.

A comparison with microprocessors is appropriate. In the first place, microprocessors are not bit-slice devices. They are LSI stored program computers, whose hardware consists of registers, ALU (arithmetic logic unit) and control unit, with fixed instruction sets. Since microprocessors are made in MOS and CMOS technology, they are generally slower than bit-slice elements, which are fabricated in bipolar technology.

5 THE FIRST BIT-SLICE PROCESSOR AMD 2901

Although this was one of the first *slices*, it is an important one since its general structure, which is shown in Figure 5, established the architectural structure of future devices. It was used in the DEC system 2020 and Data General Nova 4. It is a 40-pin LSI chip implemented in LSTTL with 500 gates. It has a 4-bit wide slice and an 18-word RAM. Any two of its registers are selectable on A and B data busses by a 4-bit control input. Note that the A register is read only, while the B register is read and write. Its ALU, which has three arithmetic and five logical instructions, R and S, has inputs for connection to R-A reg., D bus (external, 0 *S*-A reg, B reg, Q reg, 0).

The Q reg, which is designed for multiplication on the basis of shift and add, can be fed from the ALU or from itself with a shift up or down. It can shift the ALU output before loading into B reg. The ALU output can go to the external Y Bus, Q reg., or the Register Array with a shift, and can also output A reg contents to theY bus.

There are nine control lines for controlling (a) ALU Function (b) ALU Source Data (c) ALU Destination. Other control lines allow Y bus data enable ALU carry-in, ALU carry-out, and carry look ahead signals.

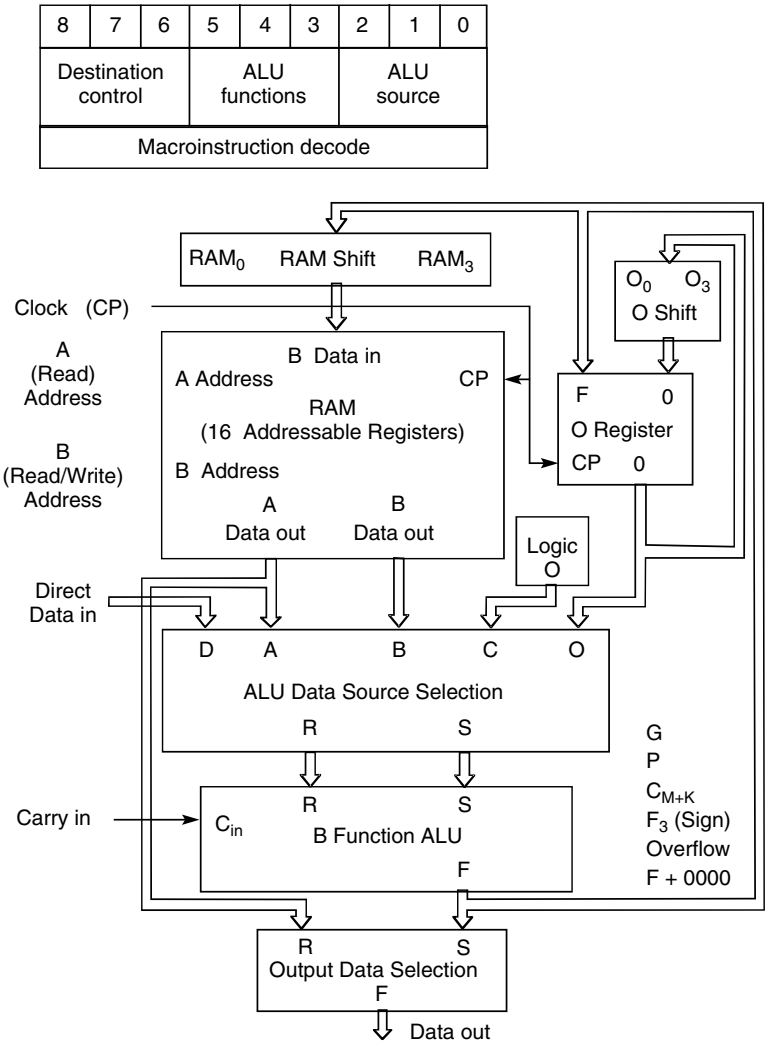


Figure 5. Functional block diagram of the AMD 2901.

The operation of the slice occurs within one cycle of the clock, which is sometimes referred to as a *microcycle*. Hence, data is taken from the RAM, the ALU operates on the data, which is then shifted and reloaded into the RAM registers: all within one microcycle. Different edges of the clock are used to prevent race hazards. Thus, the slice can be considered as a 'micromachine' that changes state every microcycle.

RELATED ARTICLES

Article 78, Embedded Systems, Volume 2; **Article 152, Fundamentals of the Stored Program Digital Computer, Volume 3**; **Article 159, Designing and Building Software for Measuring Systems, Volume 3**.

FURTHER READING

A relevant bibliography includes the following in addition to the references given in **Article 152, Fundamentals of the Stored Program Digital Computer, Volume 3**; **Article 153, Single Address Instruction Microcomputer, Volume 3**; **Article 154, Internal Operation of the Microprocessor, Volume 3**; **Article 155, External Operation of the Microprocessor, Volume 3**; **Article 156, Memory Management in the Microprocessor, Volume 3**; and **Article 157, Data Acceleration in Computers, Volume 3**.

Hintz, K.J. and Tabak, D. (1991) *Microcontrollers: Architecture, Implementation, & Programming*, McGraw-Hill.

Karalis, E. (1997) *Digital Design Principles and Computer Architecture*, Prentice Hall, NJ.

Kleitz, W. (1998) *Microprocessor and Microcontroller Fundamentals: The 8085 and 8051 Hardware and Software*, Prentice Hall, NJ.

Rigby, R. and Dalby, T. (1995) *Computer Interfacing: A Practical Approach to Data Acquisition and Control*, Prentice Hall, NJ.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

159: Designing and Building Software for Measuring Systems

Joseph E. Kasser
University of South Australia, Adelaide, South Australia, Australia

1 Software Development Life Cycles 1062
2 Developing the Software Requirements for a Measuring System 1063
3 Documenting Requirements 1064
4 Code Development 1064
5 Reviews 1065
6 Testing the Software 1066
7 Software Quality Characteristics 1066
8 Configuration Control 1067
References 1067

1 SOFTWARE DEVELOPMENT LIFE CYCLES

There are a number of software development life cycles (SDLC) or methodologies, which apply to both the functional and object-oriented approaches. The major ones used are the Waterfall, Spiral, Rapid Prototyping, and Cataract forms. These are little different to those used, in general, in the engineering of systems and they have been the basis of many of those developed; *see also* **Article 63, Outline of Systems Thinking, Volume 2**; **Article 64, Executing A Measuring System Design, Volume 2**; **Article 65, Life Cycle Concept, Volume 2**; and **Article 66, Phases of System Life Cycle, Volume 2**.

All of the forms of life cycle follow the following basic steps:

- Determine what has to be produced – this is generally the time in which the need is identified and the requirements written.
- Produce it – the design and implementation phase. This might also include developing prototypes and concept demonstration models.
- Test it – the test and evaluation step is to determine that the software meets its requirements.
- Deploy it – field it and use it.
- Operate and maintain it – this is the time in which latent defects show up and need to be corrected. Changes may also be suggested to improve or add functionality, which would then be incorporated in later releases of the software.

The alternative life cycles used in software development are now outlined to reveal their differences.

1.1 The waterfall methodology

The basic software development life cycle is known as the *Waterfall Methodology*. It follows the basic sequence of 'requirements, design, build, test, and operate'. It is known as the Waterfall, because when drawn, the process flow across the drawing looks like a waterfall. The waterfall methodology works very well if three conditions are met:

1. requirements are known before design begins;
2. requirements do not change while the design, build, and test phases take place; which generally means the
3. time frame is relatively short.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

These conditions are generally met when developing software for specific test and evaluation situations.

However, the waterfall methodology fails miserably if any of these three conditions are not met – which is usually the case if the project has substantially new issues involved.

As a consequence, the waterfall methodology has largely fallen out of favor, being replaced in software development by some sort of spiral or rapid prototyping methodologies. What is generally not noticed is that if the spiral is stretched out in a linear manner, it forms a series of small waterfalls.

1.2 Spiral and rapid prototyping methodologies

The spiral methodology Boehm (1988) implicitly added risk and risk mitigation to the basic waterfall methodology. Other rapid prototyping methodologies focus on converging the capabilities of the product to meet the requirements of the customer at the end of the product development cycle.

1.3 The cataract methodology

The Cataract Methodology (Kasser, 2002) combines the best of several methodologies and implicitly adds configuration management to the development process. It incorporates an initial Build version, Build Zero, which contains the same initial two phases, requirements and design, of the waterfall methodology with the exception that there is recognition that

- all requirements are not finalized at the Software Requirements Review (SRR) stage;
- additional requirements will become known as the project progresses;
- design and implementation decisions will be deferred and made in a *just-in-time manner* (Davies, 1998; Kasser, 2000). These decisions must also be made so as to maximize the 'don't care' situations (Kasser, 2001).

The work in Build Zero is to

- identify the highest priority requirements;
- baseline an initial set of user needs and corresponding system requirements;
- complete the first draft of the Test and Evaluation Management Plan (TEMP);
- document the assumptions driving the need for the software;
- develop the work breakdown structure (WBS) to level the workload across the future Builds and implement the highest priority requirements in the earlier Builds (Denzler and Kasser, 1995).

Build One, and each subsequent Build phase, is a miniature waterfall process in itself, hence the use of the term 'cataracts' to designate the methodology. The requirements for each Build are first frozen at the Build SRR. Then the design effort begins. Once the design is over, the Build is implemented and the software turned over for integration with the previous Builds.

While the design team does assist with the integration, their main effort is to start to work on the design of the next Build. Once the first Build has been built and is working, the 'requirements – design – integrate – test – transition and operate' stages of the SDLC commence for the second Build.

This cycle will continue through subsequent Builds until the software is decommissioned, although the contract may change from the development organization to the maintenance organization. Each Build is an identical process but time delayed with respect to the previous one. Each successive Build provides additional capabilities.

2 DEVELOPING THE SOFTWARE REQUIREMENTS FOR A MEASURING SYSTEM

Requirements articulate the implementation of a need. In this instance, the need is to measure something over some period of time with some degree of accuracy under certain conditions.

According to Juran (1988), the ideal unit of measure has the following characteristics:

- It provides an agreed basis for decision making. Different people view things differently and have different priorities. The metric must allow a meeting of minds between all interested parties.
- It is understandable. Technical metrics are normally understandable, but metrics of management or personal effectiveness are not. This may be because words do not have standardized meanings or may require an educational background that is lacking.
- It applies broadly in use to determine whether an improvement has occurred. Percentages and ratios apply to many situations and facilitate comparisons.
- It is susceptible to uniform interpretation. The units used and types of errors must have been defined with appropriate precision.
- It is economical to apply. There is a trade-off between the cost of making the measurements and the value of having them. The cost may depend on the precision, so care must be taken to specify the correct precision.

- It is compatible with existing designs of sensors. If you cannot measure it, there is little point in defining it as a metric.

Questions to be asked and answered when developing the requirements for a needed measuring system include:

- What is to be measured (temperature, pressure, voltage, current, etc.)?
- What is the degree of accuracy needed?
- Are the resources to make the measurement available (or can they be made available when needed)?
- What is the measurement data going to be used for? There is no need to measure something if the results are not going to be used.
- How will the software be tested to verify the requirements for sensitivity and accuracy are met?

It is imperative to develop good requirements; the penalty of not doing so can be considerable extra cost and time to change the software to correct errors appearing later when they are costly to fix. It is much better to find errors early. This, however, contrasts with the fact that the many requirements for novel systems can often be found only from a journey of discovery as the project develops. More on requirements is found in **Article 3, Measures and Metrics; Their Application, Volume 1**; **Article 67, Principle of Concept of Operations (ConOps), Volume 2**; and **Article 69, Requirements Allocation, Volume 2**.

3 DOCUMENTING REQUIREMENTS

Requirements can be documented in various ways including:

- *Text mode requirements*: the traditional statement of the form 'the software shall do X in response to Y within time T'.
- *Drawings and flowcharts*: process charts, timing or waveform diagrams, and Unified Modeling Language (UML) drawings.

Requirements are a means, not an end. There is nothing divine about requirements; they are just a convenient poorly used tool for translating user needs into the software that should be built.

They are generated as an intermediate work product in the software development process and are developed to provide formal communication between the stakeholders.

Writing text-based unambiguous requirements for combinatorial and sequential scenarios in the form of imperative construct statements is difficult. Timing and state diagrams are often used within the context of the Requirements Document (RD) to provide the necessary information. Sutcliffe, Galliers and Minocha (1999) proposed reducing human error in generating requirements by analyzing them using an approach of creating scenarios as threads of behavior through a *use case* and by adopting an object-oriented approach.

There is no single best way to document requirements. Requirements are a tool. In this instance, they are used to communicate the need to the developers and testers. Whatever format best acts as the communications medium is the 'right' way to use requirements.

Dedicated requirements extraction and management software tools are available, their use being, because they are costly and take considerable time and resources to deploy, more for the larger-size projects where requirements can number in the thousands and thus need a sound entry, modeling, recording, and tracking system in order to allow many people to work on them simultaneously.

4 CODE DEVELOPMENT

The IEEE 12207 Standard for code development describes the process for developing software. The process employed should be a tailored version of the standard appropriate to the size of the project developing the software. In general, the larger the project, the more people there will be working on it, bringing a greater need for the degree of formality applied. This work systematically converts the requirements into appropriate software computer code using whatever computer language(s) that has been selected for the project.

Selection of which language to adopt will usually be decided by the

- code already used in the project
- experience of the contractor doing the work
- dictates of the customer
- appropriateness of the language for the task at hand
- maintenance issues
- flexibility to interface to other software suites.

Efficient code building needs availability of suitable development tools for code generation from higher-level languages, with debugging tools and documentation aids all running on appropriate development-computing machines.

Experienced coders may appear to be overly expensive but are usually the most economic way to go as their speed and accuracy of development will be better than what can be obtained by the less experienced. The error rate will be lower with the consequence that the regression effort will be reduced as less repairs are needed.

5 REVIEWS

A key issue that brings control to code development is the 'work done' review process, for it flushes out potential problems at a time when they are cheap to fix.

According to Kasser (1998), reviews are not held to do work on the project but they too often degenerate into that form of meeting; that is best done in other types of meetings.

A review is a meeting held to provide visibility into the state of a product or process and to formalize approval of decisions previously made. It is not a time to present surprises or controversial design changes. As such, the general rules for running and chairing effective meetings apply. For example, there should be a prepublished agenda giving the issues and order of discussion. Well-prepared documentation should be made well before the meeting.

Any phase in the SDLC can be considered as a series of activities that take place between two milestone reviews – a 'start' milestone and an 'end' milestone. For example, the software design phase takes place between the SRR and the Software Design Review (SDR). Each milestone review covers two sets of work:

- work accomplished prior to the milestone (products completed, resources consumed);
- work to be accomplished before the next milestone review (products planned, resources available).

The products produced for the 'end' milestone are tested against the

- products-planned section of the 'start' milestone – the one before the current milestone;
- specifications for the specific products.

At each milestone, look at the work done now and at what is then planned to accomplish before the next review.

Take steps to ensure the necessary information is available to the meeting and that all concerned are not holding back on tardy issues in the hope that the difficulties they currently have will clear by the next meeting. A major source of project failure is lack of finding out problems early enough.

Reviews are an integral part of the SDLC. Each review focuses on the issues related to the particular phase. There are two kinds of reviews in the SDLC – formal and informal.

These reviews contribute to effective software development because they show progress. Customers want to see results quickly and equate results with software or hardware prototypes. Consequently, they tend not to allow enough time for analysis and design.

By taking part in the formal reviews and with the right team relationship in the informal reviews, the customers will see progress as the project progresses through the early stages (preproduction) of SDLC. They will also share the discovery process that is always associated with new projects and appreciate the difficulties as they arise.

Reviews serve as milestones to demonstrate progress in the completion of some work. The major formal reviews include the SRR, Preliminary Design Review (PDR), SDR, Operational Readiness Review (ORR), and so on, as are set up for the project. Informal reviews are mostly peer reviews when colleagues check each other's work. Reviews can also be thought of as being critical tests carried out in a way that lessens embarrassment over shortcomings.

Consider reviews from the perspectives of testing. The earlier the testing can be performed in the SDLC, the greater the reduction in the penalty costs of not doing it right the first time. Testing is done at well-established checkpoints through the SDLC. These check points include:

- *Concept reviews*: Verify that the conceptual software will meet the needs of the customer. Many times, people state problems in terms of particular solutions that may not meet their needs when subsequently researched and analyzed.
- *Implementation (management) plans*: Verify that adequate resources have been scheduled and sufficient planning has been performed to maximize the probability of successful completion of the tasks defined.
- *Requirements reviews*: Verify that the requirements have fully documented the customer's need as well as what can be done at that point.
- *Design reviews*: Verify that the design meets the requirements, namely, that nothing has been added or left out and that all interfaces will work correctly.
- *Code walk-throughs*: Verify that the code meets the module specifications, namely, that nothing has been added or left out.
- *Test plan reviews*: Verify that the tests, as planned, will exercise every requirement and will not waste effort testing for undocumented requirements.
- *Test procedure reviews*: Verify that the test procedures conform to the test plans without additions or subtractions.

Informal reviews are slightly different to formal reviews. However, to be effective, the following 'P' items need to be considered for formal reviews:

- *Prerequisite*: The product to be discussed in the review must be available for study before the review. The minimum is 24 h, perhaps longer, depending on people's schedules.

- *Peers*: The presenter, coordinator, and reviewers must be peers during informal reviews to allow free discussion and improve the tendency to discuss the problem, not the person. If management is absent, then, in theory, there will be no one to impress or be intimidated by.
- *Product item*: This must be small enough to be completely reviewed within one hour.
- *Period of time*: The time spent discussing each subitem must be limited to a few minutes.
- *Pact*: Consensus is to be reached by the team during the review as to the disposition of discussion items. A healthy coalition of activity is essential as each kind of skill integrates to produce the whole.

For effective formal reviews, only the 'Prerequisite' and 'Product item' issues are essential.

These reviews shown here, in general, cover items that are not discussed at the Configuration Control Board level. However, they do take place in a formal or informal manner in the process of implementing change requests – see Section 8.

6 TESTING THE SOFTWARE

The code should be built and tested in phases. It should not be left to be tested at too long a period.

The ideal implementation for test and evaluation of software is to build it from the user interface to the measurement software. This allows the users/operators to use the completed parts of the software to test sections developed later. This approach ensures that the software is 'user friendly'. There are enough instances of software being released that is so clumsy to use that it is obvious that the people who wrote the code never used it.

Software is a partial system product and needs to be tested just like the whole system product that will be tested using the software. The following are examples of the attributes of software that need to be tested (Hughes, 2000):

- *Correctness*: Extent to which a program satisfies its specifications and fulfills the customer's requirements.
- *Reliability*: Extent to which a program can be expected to perform its intended function with the required precision when called upon to operate in stated conditions.
- *Efficiency*: Amount of computer resources required by the software when executed.
- *Integrity*: Extent to which access to the software or data by unauthorized persons can be controlled.
- *Usability*: Effort required for learning, operating, preparing input, and interpreting output.
- *Maintainability*: Effort required for locating and fixing an error in an operational program.
- *Testability*: Effort and expediency required for testing a program to ensure that it performs its intended function.
- *Flexibility*: Effort required for modifying an operational program.

Depending on the circumstances, not all parameters for testing may be required. For example, if the program is going to run on a personal computer in a stand-alone environment, with no other program running at the same time, there is little point in testing to determine the actual resources used by the program as long as it performs in a satisfactory manner.

All effective meetings need a suitable set of metrics to gauge success and progress – *see* **Article 3, Measures and Metrics; Their Application, Volume 1** – consider now metrics used for reviews. Reviews form an important part of the communications process between the developer and the customer. Both parties participate to ensure that the work is proceeding correctly, that is, the contractor is producing the system that the customer wants.

One metric for reviews could be the number of Review Item Discrepancies (RID) generated during or following a review. Since reviews with the same name (i.e. SRR) have different levels of complexity depending on the size of the system being reviewed, comparing numbers may not be useful. RIDs are commonly used, but while they do point out deficiencies in the products delivered at the time of the review, the absence of RIDs is an ambiguous condition. It is possible to produce an operations concept review that will receive zero RIDs. This is possible by discussing the concept with all the stakeholders during the period of performance before the review. Everybody's comments and wants can be incorporated into the concept. After the review, when zero RIDs are received, the issue is whether it is because a good job has been done or because nobody cared.

A better metric is based on the amount of rework to be performed following the review or the amount of work that has to be scrapped and replaced (as a percentage of the whole). Measuring the number of RIDs does provide a metric for the degree of customer involvement in the process, but as explained, these kinds of metrics can be manipulated by skilled project managers.

7 SOFTWARE QUALITY CHARACTERISTICS

It is useful to consider the considerable body of knowledge on software quality. ISO 9126 (1991) identifies the following six software quality characteristics:

- *Functionality*: Covers the functions that the software product provides to satisfy user needs.
- *Reliability*: Relates to the capability of the software to maintain its level of performance.
- *Usability*: Relates to the effort needed to use the software.
- *Efficiency*: Relates to the physical resources used when the software is run.
- *Maintainability*: Relates to the effort needed to make changes to the software.
- *Portability*: Relates to the ability of the software to be transferred to a different environment.

The applicable attributes and quality characteristics that apply to the software must be noted as requirements and the test program for the software must verify that all stated requirements have been met.

At all times, it is also necessary to ensure no new requirements have crept in without being allowed for or that any no-longer-needed requirements are still being worked upon.

8 CONFIGURATION CONTROL

All software should be placed under configuration control using a change management process. Changes should be approved before being implemented. The actual change control process will vary depending on the situation but the key elements are constant. These are:

- No changes are made without prior approval by an authorized body.
- Estimates of the effort and schedule time needed must be made before the change is implemented.
- Changes are to be made in an orderly and recorded manner.

Two types of changes occur:

- planned
- unplanned.

These can each be allocated into one of three degrees of urgency:

- stop the software from working: change now;
- can live with it: change soon;
- would be nice to have this addition: change whenever.

A serious problem with software changes is that due to the high degree of interdependence of all software instructions, there is a distinct risk that a change not only fixes known defects but can also add new features that are usually unwanted. Each change has to be subject to a regression analysis to reveal the impact of the intended change. The effort needed to carry this out can be major. Sound design partitions the modules to lessen the degree of interaction and makes it easier to investigate effects of changes.

REFERENCES

Boehm, B. (1988) A Spiral Model of Software Development and Enhancement. *IEEE Computer*, **21**(5), 61–72.

Davies, J. (1998) Making Choices at the Right Time: Producing Better Systems, in *Fourth Annual Symposium of the INCOSE-UK*, 1998, Hendon, UK.

Denzler, D. and Kasser, J.E. (1995) Designing Budget Tolerant Systems, in *5th Annual International Symposium of The National Council of Systems Engineering (NCOSE)*, St. Louis, Missouri.

Hughes, B. (2000) *Practical Software Measurement*, McGraw Hill, New York.

ISO 9216 (1991) *Information Technology – Software Product Evaluation – Quality Characteristics and Guidelines for their use*.

Juran, J.M. (1988) *Juran on Planning for Quality*, The Free Press, New York, pp. 76–78.

Kasser, J.E. (1998) *Course Notes – CSMN 648 Software Maintenance*, University of Maryland University College, College Park, Maryland.

Kasser, J.E. (2000) The WebConference: A Case Study, in *The INCOSE Mid-Atlantic Regional Conference*, Reston, VA.

Kasser, J.E. (2001) Writing Requirements for Flexible Systems, in *INCOSE-UK 5th Annual Symposium*, Daventry.

Kasser, J.E. (2002) The Cataract Methodology for Systems and Software Acquisition, in *Systems Engineering Test and Evaluation (SETE) Conference 2002*, October, Sydney, Australia.

Sutcliffe, A., Galliers, J. and Minocha, S. (1999) Human Errors and System Requirements, in *IEEE International Symposium on Requirements Engineering*, Limerick, Ireland.

160: Smart Sensor System Features

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Introduction to Smart Sensing Systems 1071
2 The Capability Aspect 1072
3 The Suitability Aspect 1072
4 Architectures of Smart Systems 1073
5 Methods for Automating Smart Features 1074
6 Presets and Balances 1075
7 Autocalibration and Selftest 1076
8 Implementation of Smart Sensor Systems 1077
9 Pros and Cons of Smart Systems 1077
10 Operational Aspects of Smart Sensor Systems 1078
Related Articles 1078
References 1078

1 INTRODUCTION TO SMART SENSING SYSTEMS

Today, users want sensors that are easy to select, install, and use on the 'fit and forget' basis.

Basic measurement sensor systems, however, have many potential failure modes that can degrade the metrological performance – the key performance characteristic of a sensor. Users also seek easy-to-install systems that can be used without the need to appreciate the theory and detailed practice of their operation in order to set them up and maintain them.

To make a basic sensor 'smart' has been a need for a long time. Numerous things can go wrong with a sensor; and it is often the key to sound overall systems success. A much more robust sensor unit is required.

Producing smart sensors is a matter of designing and making a sophisticated system in its own right. What was not so long ago a vision has now become widespread reality. Today, they are either designed as additional boards in the system's computer or in microcircuit form – the so-called microsmart sensor that can allow packaging with the sensor element.

Combination of sensor experience and research, advanced microelectronic technology and sophisticated signal processing (Tränkler and Kanoun, 2001) has led to many reliable sensors now being used routinely that can be truly defined as suited to fit and forget use. They are found in many places – automobiles, process control, aviation, health, manufacturing, agriculture, and more. The application of smart sensors in robotics has been discussed (Siegel, 2001).

So what is it that makes a simple dumb sensor become a 'smart' one?

A high-level system engineering definition of a smart measuring sensor would be:

> A sensor is smart when it detects, corrects or warns about any mechanism that will, is, or is about to degrade the metrological integrity of the sensor system during operation, this being with respect to both its capability and suitability regimes.

A word of caution: the name 'smart' is sometimes used in association with sensors without reasonable justification, one instance having only the smart feature of possessing a stabilized power supply!

A sound appreciation of smart sensor systems is needed in:

- creation of smart sensors
- appreciating the features of given systems

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

- critiquing operating systems
- maintenance
- making a sensor system more competitive.

Being relatively new, the literature on smart systems is still found much in the primary journals, or on manufacture's releases, where schemes and ideas are reported. Several books that survey aspects of the smart sensor field include Frank (2000), Gardner, Varadan and Awadelkarim (2001), Howell and Hamilton (1992), Inaudi (2002), Kirianaki *et al*. (2002), Pallás-Areny and Webster (2000), Zobrist and Ho (2000).

This article provides an appreciation of the basic issues of smart sensors. To this end, it first discusses the requirements of its two main regimes – capability and suitability. It then looks at the architectures of the basic sensor and how it changes to become smart. An overview of mechanisms that may need automation shows how the smart features rapidly make a simple sensor a complex system. Presets and balances are covered before the issue of autocalibration is dealt with. Implementation of smart systems and their pros and cons are discussed before finishing with operational aspects.

2 THE CAPABILITY ASPECT

In the engineering of systems, the completed design has to provide the user with the desired capability – *see* **Article 63, Outline of Systems Thinking, Volume 2**. A sensor then has to sense a measurand, satisfactory with respect to its metrological task – *see* **Article 57, Static Considerations of General Instrumentation, Volume 1** and **Article 58, Description of Accuracy, Linearity, and Drift, Volume 1** for explanations of the many features of a measurement determination.

The necessary capability aspects of any sensor are that it is:

- **Calibrated**

At any time of use to determine a measure, it must provide the required measurement functions of discrimination, precision, and accuracy. These are each different characteristics needing ongoing testing and possibly automatic correction to data. A word of caution is needed here; autocalibration is a misunderstood, and sometimes misrepresented feature of many smart systems – see Section 5.

- **Linear in operation**

Many sensors are not linear in their operation. Linearity is desired as the signal moves over its valid dynamic range. (The actual function may be of other laws such as exponential.) Linearity is vital to correct application and it is not covered as such in a sensor calibration that is usually only being performed for one or two spot values. Linearity needs periodic automatic determination of several points over its dynamic range so that correct ongoing signal linearization correction can be provided.

- **Giving adequately precise values**

A sensor signal needs to possess sufficient short-term grouping for repeated measurements of a given value (called *repeatability*), and also for its long-term reproducibility. Automation can speed up initial commissioning and long-term verification of these parameters.

3 THE SUITABILITY ASPECT

As well as the above-mentioned features needed for success, many others need to be correct for a sensor to be deemed fully capable. They are concerned with the suitability of the sensor at providing that capability over its working life. This aspect is its suitability for use in a specified task.

In general, the reasons for failure under operation are that parameters of the sensor system go too far over, or under, the design limits in which the sensor will give faithful metrological performance.

The main areas in which this can happen are

- set up conditions of the sensor system elements
- connections
- measurement range selection
- amplifier gain settings
- automatic data processing settings
- fault conditions.

Fault conditions provide numerous risk situations many of which can be given automated attention for their detection and possible correction within a smart sensor system. It is not only a case of total failure taking place but also the measurement being sufficiently corrupted from its correct value by a small amount such that it is not producing faithful metrological data.

Check lists of likely sources that might need to be considered follow.

External conditions

- excessive influence by an external parameter such as humidity, pressure, temperature, vibration, lightning, and so on causing severe degradation;
- external influence parameter altering working levels of key voltages, currents, and so on;
- power excesses overheating key components, causing metrological degradation;
- protective packaging failures;

- incorrect connection being made, such as DC power lines reversed or wrong connections inserted;
- connection to other systems needing selection of parameters, as in autoranging wireless links;
- leakage of electric currents, fluids, gases into the system.

Internal conditions

- stoppage of cooling fan;
- failure of drive motors or actuators;
- degradation of resistors, capacitors, inductors, and integrated circuits over time;
- rising of hot spots from dust, preventing cooling;
- occlusion in optical systems from dust and insect webs;
- chemical growth.

Detection and correction of numerous features could be included in a smart sensor. Obviously, the more that are needed, the higher will be the cost in terms of money and space.

Those features easily provided in the electronic domain have become routine. Their size of manufacturing scale and development tool availability has made them inexpensive to offer. However, addition of special needs outside an electronic-only solution can be very costly.

4 ARCHITECTURES OF SMART SYSTEMS

Designing a well-developed smart sensor with many features can be a major systems development task requiring application of the systems engineering process and detailed design, all coupled with a sound appreciation of the metrological aspects. The latter aspect is not always so well appreciated; implementation will usually be by persons not that familiar with sound measuring systems theory and practice. Overall, their design can be particularly difficult, for the system has to operate correctly within small measurement limits at all times.

The usually encountered basic architecture of a sensor unit that provides a processed signal ready for transmission or display is given in Figure 1. A sensor will need either, or both, auxiliary excitation and an offset bias.

For this classical structure, the sensor will need adjustment for linearity; the offsets and balances of the initial analog blocks will need adjustment. Gains will need setting. To make this structure handle automatic operation of such issues needs, in addition, several kinds of support subsystems in a systemic and systematic manner.

Figure 2, a simple schematic, shows how some basic smart features are added to the basic system to allow for sensor calibration in the short and long term; offset gain and zero adjustment; and linearization.

To effect a particular smart operation, as will be discussed in the following sections, additions will require extra circuitry and/or suitable physical devices that are coupled into nodal points, or are placed in series with signal lines. While the basic system of Figure 1 has already become quite complex, the extras shown are only a few of the additions that need to be added into a really smart sensor. The main difference shown here is that switches, test points, and a calibrator have been added under processor control to carry out the smart functions.

Software is used to sequence the various activities. Some need the sensor output to be held at a suitable value while they are implemented. Corrective actions, such as voltage regulation, can be continuous and be operating at all times. The order in which the functions are switched in can be critical, as is discussed in Section 6.

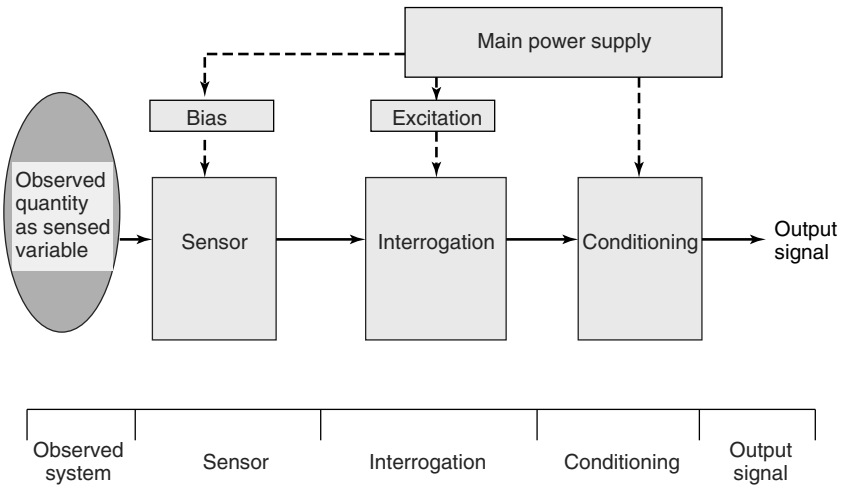


Figure 1. Architecture of basic 'dumb' sensor system.

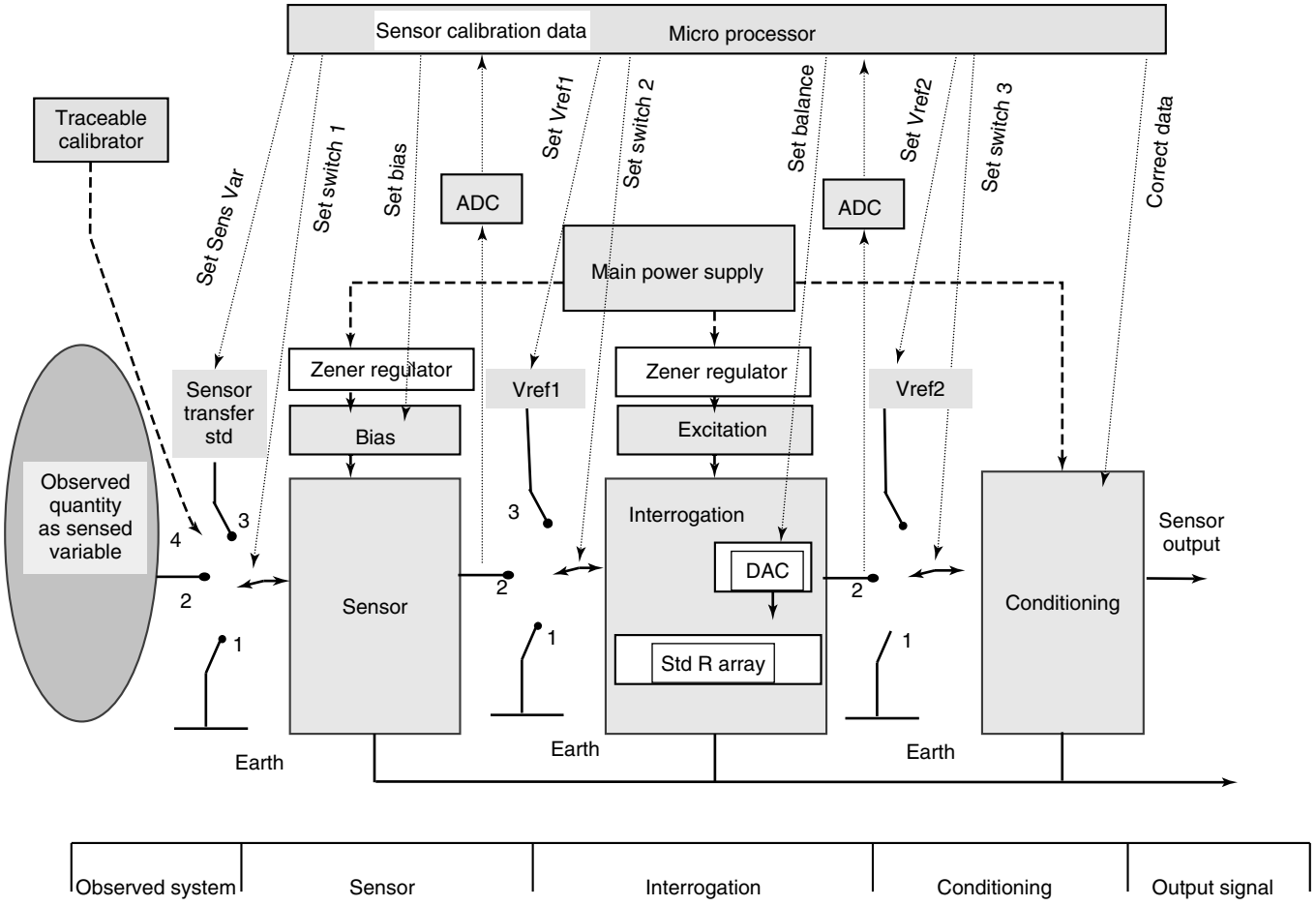


Figure 2. Smarts added to basic sensor system.

Some smart functions are best achieved in the early analog stage of the sensor signal chain. Others are a case of mathematical correction of the data stream. It is important to consider where they are best done to keep sensor performance high.

Advanced smart systems will also log sensor performance data for use in trend calculations that can be employed to make near to real-time corrections and for future design reference.

The rapid uptake of miniature wireless communication systems has called for smart sensors that can operate in the 'plug and play' mode.

5 METHODS FOR AUTOMATING SMART FEATURES

A collection of smart effects and their implementation is now outlined. Offsets, zero settings, and balances are covered in Section 6. Autocalibration is left for discussion in Section 7. The main features are now discussed.

- **Use of common-mode rejection of unwanted signals**

It is common practice in sensitive measuring systems to use a second similar unit, connected in opposition, to balance out unwanted effects. This is done routinely in the Wheatstone bridge – *see* **Article 126, Electrical Bridge Circuits – Basic Information, Volume 2** and **Article 127, Unbalanced DC Bridges, Volume 2**, in capacitance sensing – *see* **Article 204, Capacitance and Inductance Measurement, Volume 3**, in strain gauges – *see* **Article 192, Strain Sensors, Volume 3**, and in instrumentation amplifiers – *see* **Article 122, Operational Amplifiers, Volume 2**. This is the first issue to smarten, as there is need to reduce unwanted effects caused by not exactly matching elements.

- **Compensation by feedback from in-built auxiliary sensor**

Where a common-mode operation is not practical to implement, an auxiliary sensor can be used to sense the level of the unwanted effect and be used to produce a compensating signal.

An example of this is an auxiliary temperature sensor being used to correct strain gauges for the influence effects of temperature; the strain gauge unit is fitted with a metal film or semiconductor-sensing element.

Another instance arises in use of precision amplifiers; a solid-state temperature sensor is often used to measure the silicon substrate temperature in order to compensate for thermal errors.

- **Amplifier gain**

Circuitry is set up temporarily to create the network conditions needed to make error calculations or system component adjustments. Detail is given in Offsets and Balances – see Section 6. In array systems, such as photo detectors stages and amplifiers, this is usually performed in sequence of each gain stage as this allows use of the same test transfer standard for each channel test.

- **Incorrect voltage effecting metrological aspects**

Whereas the stability of the main power supply voltage is not that critical, it is vital that voltage applied as excitation to common-mode circuits, such as bridges, is sufficiently stable.

Not always appreciated is that common-mode arrangement such as the DC bridge configuration, only balance out unwanted effects when in the balanced null state. As unbalance increases, the compensation effect reduces. Thus, an unbalanced bridge will 'see' some effect from an unstable excitation, depending on the extent of unbalance and the voltage stability factor of its excitation source. This need also applies to amplifier power supplies for preamp stages, reference bias voltages, and when dealing with the measurement of critical currents.

The smart mechanism here is to test the stability of the voltage by comparing it with a sufficiently stable voltage reference source. Zener diodes set up with low power loss, that is, low through currents, to give more stability are often used. They have a wide plateau range wherein voltage across the diode is reasonably constant for a range of through currents. A more superior method is to use the reference Zener to control an amplifier, thus further reducing the through current in the Zener diode. Two Zener diodes used in piggyback connection can give superbly stable transfer reference voltage sources. These units are available in integrated form as standard voltage reference sources over a wide range of voltages.

Having detected an error voltage, closed loop digital methods are used to change the used voltage source until agreement is reached with the reference source.

- **Out-of-range detection in bridge and sensor outputs**

Fault conditions for specific sensor systems can be identified from a critical analysis of the system and the mechanisms set up to sense their presence.

For example, a Wheatstone bridge will have a normal output swing for inputs – such as when used to interrogate a strain gauge or resistance temperature sensor – that only swings over some $\pm 10\%$ of the magnitude of the excitation voltage. A common error state is for one arm of the bridge to go open circuit owing to resistance element failure. This state causes the output to swing to the maximum excitation voltage. A test for this magnitude is a sure sign of bridge failure and could be used to bring up an alarm. Auto replacement of failed components is done sometimes by replacement with a set of redundant components using a suitable switching arrangement.

- **Model-based operation**

All the above-mentioned methods make use of knowledge of the correct state of the sensing system by reference to some known factor. The idea can be extended by forming an executable model of the critical sensing parts of the sensor in the processing computer. If the model is white, that is, its detail models the interior of the system, then its nodal values at any given overall state can be used as a reference source for correct operation. Development of models can take time to develop – *see* **Article 71, Modeling Methodology, Volume 2** – but with manufacturing volume, or in supercritical applications, the expense can be justified.

These models can take the form of formal mathematical descriptions or be AI based using Knowledge-based Systems, KBS methods – *see* **Article 128, Nature and Scope of AI Techniques, Volume 2**.

6 PRESETS AND BALANCES

The main elements of a sensor – its sensing element, interrogation stage, excitation source, and preamplifier, usually, cannot be manufactured precisely enough not to need adjustment of such parameters as common-mode balances, gain, and presets adjustments.

- **Bridges**

These need the resistance of one or more of the arms to be trimmed to obtain an adequately zero output. If not balanced, the null point will be a minimum, having a value never equal to zero. A computer-controlled resistor network is switched into the bridge circuit using detection of the balance point to select the correct resistance value.

A trick here is to trim the bridge by adding parallel resistors, not by putting resistance in series. Their resistance will then need to be large, making the required resistance of the connecting switch more easily accommodated.

- **Instrument amplifier**

The gain of an electronic amplifier varies with ambient temperature and time. As the output of a sensor is decided by the gain factor of the processing chain, it is necessary to adjust the gain when first commissioned, and henceforth periodically.

Gain is usually adjusted by digitally altering a stable resistance network added to the circuitry, this often being incorporated into the amplifier chip. A similar setup as is shown in Figure 2 shows how this can be done in practice.

To detect a gain error, the amplifier input is first clamped to zero, using a suitably low resistance switch placed from the input to earth. The zero offset of the chip is then adjusted electronically to zero and that adjustment value held.

The next step is to replace the amplifier input with a suitable known voltage. The output of the amplifier is measured automatically and stored in the computer's memory. Gain is then calculated as the output/input ratio. Gain is then adjusted on the chip. Not setting the offset first can lead to a second order error in the above procedure.

To obtain linearity figures, the process is repeated at several positions across the working range. These values are then used to correct, in close to real time, the sensor.

7 AUTOCALIBRATION AND SELFTEST

It is common practice for measuring systems, when started up, to run through a suite of tasks that test for many fault conditions. This is called selftest. This process identifies areas where a system is already out of specification or, perhaps, trending toward it. The general principles and practices of selftest and auto correction are described above.

Special mention is needed about autocalibration. While many systems claim they provide automatic calibration, it needs to be understood as to what level and traceability the claims are valid.

Like selftest, autocalibration is a matter of comparing the present value of a parameter important to correct operation with one that is held to be sufficiently correct from direct calibration.

First let us clarify what calibration is? It is finding out how far a measured value lies from the value that is defined as being the true one by a method that is traceable to an accepted standard. This means the method used to declare true autocalibration has to have a traceable link used at a calibration interval – *see* **Article 46, Calibration Interval, Volume 1** suited to the application.

It is only possible to conduct calibrations without reference to the agreed standard in two situations:

- where the sensor itself is inherently calibrated in a traceable manner;
- where the sensor can, from past experience, refer to a replacement transfer standard for a sufficient time of operation before external calibration is needed again.

It is a fact that there is only one measuring apparatus that is inherently calibrated. The frequency-stabilized laser interferometer, if operating within design limits, has been declared to be a traceable length standard to stated limits.

The semiconductor diode/transistor has been described as also being self-calibrating and not needing reference to the Practical Temperature Scale. This is because it can be shown that its temperature current is related to only Boltzmann's constant, the charge of the electron and its temperature in K, thus being independent of geometry or specific manufacturing details.

Thus, at first sight, this appears to be able to provide a standard temperature reference source of high precision and accuracy. Analysis, however, of its practical use shows that the temperature variation needed to be measured is in the $\mu V K^{-1}$ range, yet the standing voltage for room temperature is nominally 600 mv. To gain adequate sensitivity, it requires the nominal value to be compared with a standard voltage source in a voltage comparator. The measurement then is not traceable to the metric System of Units for the standard voltage source needed is not inherently traceable.

This means all self-calibration sensors must periodically be compared to an external reference to be traceably calibrated. This situation is often overlooked.

Autocalibration in sensor systems needs the availability of a cheap and small enough reference source. A list of likely transfer standard sources used with measuring sensors is as follows:

- voltage (Zener diode, and the older Weston cells)
- radiation (Black body held at controlled temperature)
- chemicals (Standard Reference Materials)
- temperature (Standard Temperature Points, such as ice and boiling point and limited use of the semiconductor diode/transistor temperature sensing effect)
- length (Gauge blocks, interferometers)
- resistance (Calibrated stable resistor network with adequately low resistance connections).

Many other calibrators can be designed to order; careful analysis is always needed to determine that their accuracy and long-term repeatability are sufficiently stable.

For example, a standard pressure can be produced by the known sweep of a piston containing a fluid, or a relative humidity sensor can be calibrated in a suitable

chamber into which water is injected. In both cases, the physics of the chamber conditions needs to be well understood. Many special calibrators will, however, usually be too large to be practical in microsensor autocalibration use.

When conducting an autocalibration, a choice of calibration test signal and profile will need to be considered.

The input calibration signal format needs to match the practical situation. Where there exists considerable inertia in the object under test, such as when attempting to move mechanical systems, the ramp test signal is more realistic than trying to produce a step movement. Use of an impulse or sine wave can sometimes be used with domain Fourier transformation to obtain the needed response coefficients.

Some sensors will need special waveforms to be generated such as a heartbeat monitor signal. In cases where the signal arises in the electrical domain, the smart system digital processor can generate the test signals.

Some tests need noise sources. Again the digital processor can generate the type needed, be it white, grey, pink, or some other specific distribution.

8 IMPLEMENTATION OF SMART SENSOR SYSTEMS

Today, the list of suppliers of smart sensors is ever growing, IRN (2004). Growth of the smart sensor market has attracted the preparation of reports, Frost & Sullivan (2003).

Their design is often supported by excellent database libraries. These, coupled with application of best practice systems engineering, makes the creation of sophisticated sensor systems more routine by the specialist sensor makers, Tränkler and Kanoun (2001).

There are also available specialized computer-based tool kits for building the microsmart sensors. The *NI Plug & Play*® kit facilitates the integration of a smart design into a virtual instrument formed in the LabView tool system – *see* **Article 105, Modeling with LabVIEW™, Volume 2** and **Article 106, Virtual Instrumentation in Physics, Volume 2**. It allows rapid integration of the design of IEEE P1451.4 smart analog sensors into computer-based data acquisition applications. It also assists conversion into embedded computer chips. This tool kit works with a broad range of sensors, including accelerometers, temperature sensors, load cells, pressure sensors, displacement sensors, and others.

Judgment will be needed as to how much is automated, for special features can be expensive to design and make.

A microprocessor is used to pace and control most of the features explained above, although some will be done continuously with special circuitry. Processors often provide features that support smart operation. They need some nonvolatile memory to retain the settings in the case of a power down.

The first smart sensors used additional circuitry boards to carry the extra components. Progression into cheaper, better systems has taken much of supply into the integrated chip form wherein the sensor and its basic signal conditioning and additional smart features are all integrated into one unit. Some are still built in printed circuit and hybrid forms, being necessary for special cases with low volume demand.

During selftest and autocalibration operations, it is usually necessary to hold the output of the sensor to the most recent value, for during these operations the output can run to extremes, causing glitches in the output.

Analog signal correction loops can be replaced with digital Expert System methods, for example, a simple rule set might be:

If Voltage V exceeds 10.0 V (a limit of safe operation, for example),
Then reduce V by 0.1 V
Repeat until V is equal to 6 V

A wide range of smart sensors are now on offer, so it is wise to consider already available products before launching into a new design.

9 PROS AND CONS OF SMART SYSTEMS

The main advantage of smart sensors is that they can provide fit and forget operation, thereby reducing maintenance costs and risks of total or gradual failure.

One very common problem seen with basic sensor systems is that they often make erroneous indications. For example, they can indicate states in equipment that cause unnecessary shutdowns. Smart systems should be more reliable in operation, for a good design exercise is needed to thoroughly analyze the sensor systems for its various fault conditions and set up corrections for many.

On the other hand, a smart sensor can be very complex and that condition can involve poor reliability unless very well designed and tested. The smart sensor can be much more expensive than its basic unit, but the extra cost should be easy to justify in terms of cost avoidance.

The increased facility for taking a basic sensor to the smart state carries penalties:

- need for more access connection and test points in the sensor system;

- need for additional auxiliary components and subsystems;
- more system administration and control processing power and its supporting software is needed;
- additional complexity exists compared to the simple basic sensor;
- can be more difficult to maintain – need training and special test sets.

Assessing their capability and suitability, however, is a costly issue. This can be validated by three strategies:

- where sufficient detail is available, using assessment of the design features and the design execution (called *verification*);
- from experience with a given design, this being usually from another's use (validation);
- by submitting the sensor system for an instrument evaluation exercise – *see* **Article 217, Instrument Evaluation, Volume 3** (Full evaluation and maybe accreditation).

All take time and cost to perform and thus there can be the tendency to accept claims of a manufacturer without question. Instrument evaluation experience – **Article 217, Instrument Evaluation, Volume 3** – has repeatedly shown, over several decades, that instrument performance claims are often not all met. Smart sensor systems are little different to many complex instrument systems so it is reasonable to assume that they might also sometimes fall short of their stated specifications. Any area of difficulty will usually lie in their metrological performance. Caution is needed in their selection.

10 OPERATIONAL ASPECTS OF SMART SENSOR SYSTEMS

The sequence in which the initial set up, and then in-service calibration, is performed is most important.

As a general guide the order of tests for a start-up situation is as follows:

1. Ascertain that the nonautomatic settings needed are correctly set.
2. Check if system alarm states are all within the 'OK to go' state.
3. Check if power supply availability is within range and with sufficient endurance.
4. Ensure that environmental conditions are within acceptable bounds.
5. Generally, test for 'out of normal range' operation of internal operations, such as wildly incorrect open and short circuit input states; and out of range values for bridges and sensors.
6. Determine and adjust any zero states that need to be balanced at their device source.
7. Determine and set offsets to correct value.
8. Calibrate the internal electronic subsystems with standard electrical components, holding the raw sensor input first to zero and then at the 100% range value by switching in suitable shunts and sources.
9. Calibrate the sensor against a suitable transfer calibrator, using the now set up electronic amplification. It is usually not possible to sense the low-level sensor output signals without this gain being added.
10. Calibrate the correction table for linearity over the range from a suitable set of test points.
11. Record all values that have later use, such as trends, maintenance needs, and audit of performance. These help to build up confidence in the system.

Not doing these steps in this order can result in an apparently correctly set system. However, there are some traps for learners here. For instance, balancing out the offset in common-mode rejection systems at the post-block stage can be used with signal processing, but it could leave one of the two compared variables at an extreme of the range, thus, offering little working range before saturation is reached.

Overall the reliability of well-tried varieties of smart sensors, as are used in industrial control and process plants, is such that they get close to being a 'fit and forget' application. Smart sensors are not a fad but an essential part of sound measuring practice.

RELATED ARTICLES

Article 64, Executing A Measuring System Design, Volume 2; **Article 71, Modeling Methodology, Volume 2**; **Article 99, Safety Organization, Volume 2**; **Article 107, Principles of Sensor Science, Volume 2**; **Article 108, Transducer Fundamentals, Volume 2**; **Article 109, Structure and Energy in Sensor Systems, Volume 2**; **Article 117, Materials in Measuring Systems, Volume 2**.

REFERENCES

Frank, R. (2000) *Understanding Smart Sensors*, Artech House, Boston, MA.

Frost & Sullivan. (2003) *Smart Sensors-global Developments and Growth Opportunities*, Frost & Sullivan Consultants, London.

Gardner, J.W., Varadan, V.K. and Awadelkarim, O.O. (2001) *Microsensors, MEMS and Smart Devices*, John Wiley, Chichester.

Howell, S. and Hamilton, S. (1992) Intelligent Instruments, in *Handbook of Measurement Science*, (eds P.H. Sydenham and R. Thorn), Wiley, Chichester.

Inaudi, D. (2002) *Smart Sensor Technology and Measurement Systems*, SPIE-International Society for Optical Engineering.

IRN. (2004) *Directory of Suppliers*, Industrial Resource Network, http://www.powersourcing.com/se/smartsensors.htm.

Kirianaki, N.V., Yurish, S.Y., Shpak, N.O. and Deynega, V.P. (2002) *Data Acquisition and Signal Processing for Smart Sensors*, John Wiley, Chichester.

Pallás-Areny, R. and Webster, J.G. (2000) *Sensors and Signal Conditioning*, 2nd edn, John Wiley, Chichester.

Siegel, M. (2001) Smart Sensors and Robots, Recent Advances in Sensor Technology, in *Proceeding of Conference IMTc/2001*, 21–23 May, Vol. 1, Budapest, (pp. 303–308).

Tränkler, H. and Kanoun, O. (2001) Recent Advances in Sensor Technology, in *Proceeding of Conference IMTc/2001*, 21–23 May, Vol. 1, Budapest, (pp. 309–316).

Zobrist, G.W. and Ho, C.Y. (2000) *Intelligent Systems and Robotics*, T&F STM.

161: Knowledge-based Systems

Dietrich Hofmann
Steinbeis Transfer Centre for Quality Assurance & Quality Measurement, Jena, Germany

1 Choosing Methods from the KBS Range 1080
2 Setting Up the Sensor Signals and Processing 1080
3 Example for Measurement Signal Processing Situations 1081
4 Measuring Instruments in Knowledge-based Society 1082
5 Knowledge-based Simple and Complex Measurements 1083
6 Knowledge-based Legal Metrology 1084
7 Knowledge-based Industrial Metrology 1085
References 1086

1 CHOOSING METHODS FROM THE KBS RANGE

In Information societies, people desire free access to information any where, any time, and on demand with every device. This ensures success within a knowledge-based society (KBS). The dividing lines between professional and private life are shrinking.

Knowledge-based systems (KBS) or, in short, knowledge systems comprise knowledge servers; multiuse ontologies, knowledge bases and knowledge system modules; computational environments for modeling structure, behavior, and functionality of technical (physical) devices; compositional modeling; architectures for adaptive intelligent systems; and knowledge-based systems for science, engineering, and defense applications (http://www-ksl.stanford.edu).

Ontologies are definitions of terms and their relationships to other terms with the aim of building semantic webs (http://ftp.ksl.stanford.edu/pub/KSL_Reports/KSL-01-07.doc).

Intelligent measurement is an experimental process for acquiring new knowledge about a product with the help of measurement devices, computers, and specialized software.

Ambient intelligence and ubiquitous computing holds a high potential of applications for micro system technology (MST), micro electromechanical systems (MEMS), and micro-opto–electromechanical systems (MOEMS) (http://www.mstnews.de/download_0202.htm).

These paradigms are due to become available in all kinds of products and services; within their software, hardware, and processed material and processes (see ISO 9000:2000, point 3.4.2) (http://www.iso.ch).

Many products comprise elements belonging to different generic product categories. Whether the product is then called a service, software, hardware, or processed material depends on the dominant element.

Process is a set of interrelated or interacting activities that transform inputs into outputs (see ISO 9000 : 2000, point 3.4.1) (http://www.iso.ch). Inputs to a process are generally outputs to other processes.

Knowledge-based or intelligent measuring instruments or measuring equipments are known as virtual instruments or VIs (Figure 1) (http://www.ni.com)

2 SETTING UP THE SENSOR SIGNALS AND PROCESSING

Virtual instrumentation systems (VIS) have developed with improving sophistication in measurement resolution and frequency bandwidth. In some five years, virtual instrumentation systems have improved in measurement performance

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

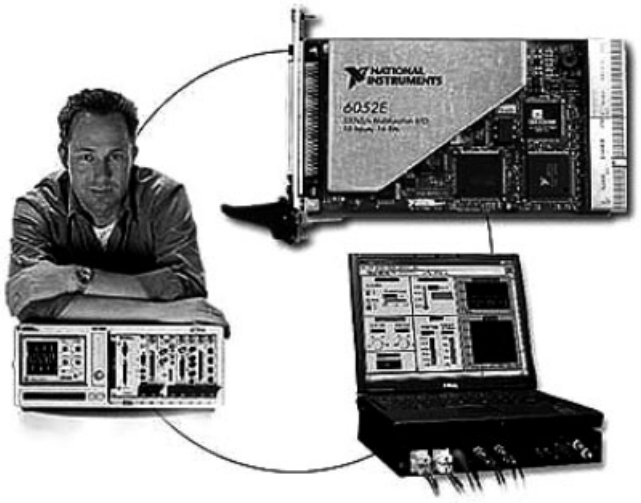


Figure 1. Knowledge-based intelligent measurement systems (http://www.sine.ni.com/apps/we/nioc.vp?cid=1036&lang=us).

to meet, and sometimes exceed, the capabilities of traditional analog instruments. With the introduction of the FlexDMM and RF signal analyzer, virtual instrumentation systems have taken an order of magnitude leap in performance.

The key to the innovative improvements in virtual instrumentation is the leverage they derive from standard PCs; and off-the-shelf analog-to-digital (ADC) and digital-to-analog (DAC) converter technologies – *see* **Article 132, Data Acquisition Systems (DAS) in General, Volume 3**; **Article 133, Amplifiers and Filters for DAS, Volume 3**; **Article 134, Analog Multiplexers, Volume 3**; **Article 135, Sample-hold Circuits, Volume 3**; **Article 136, Quantizing Theory Relevant to DAS, Volume 3**; **Article 137, Coding for Data Converters, Volume 3**; **Article 138, Sampling Theory Relevant to DAS, Volume 3**; **Article 139, Analog-to-Digital (A/D) Converters, Volume 3**; **Article 140, Integrating Type A/D Converters, Volume 3**; and **Article 141, Digital-to-Analog (D/A) Converters, Volume 3**. In contrast, stand-alone instruments often use custom ADCs and proprietary computing architectures. Several years ago, these proprietary architectures offered superior performance to their mainstream counterparts. However, advances in semiconductor technologies and the explosive demand in the last decade for computing and telecommunications capabilities have created an abundance of high-performance, commercial ADCs. Consequently, gaps between virtual instrumentation only appear at the extremes of frequency and resolution (Figure 2) (http://digital.ni.com/worldwide/bwcontent.nsf/web/all/721E2E2BB3AD847B86256C25005890BC)

3 EXAMPLE FOR MEASUREMENT SIGNAL PROCESSING SITUATIONS

The ever-increasing processing power of PCs, combined with the advanced analysis tools available in software packages like LabVIEW – *see* **Article 105, Modeling with LabVIEW™, Volume 2**, provide a more recent example of the progress of virtual instrumentation. For example, new Lock-In Amplifier start-up kit for LabVIEW, offers a FREE downloadable VI that brings the functionality of a lock-in amplifier – *see* **Article 184, Photon Counting, Volume 3**, to virtual instrumentation (Figure 3).

The LabVIEW Lock-In Amplifier start-up kit provides for precise measurement of AC signals that are partially or completely buried in noise. An example of this includes

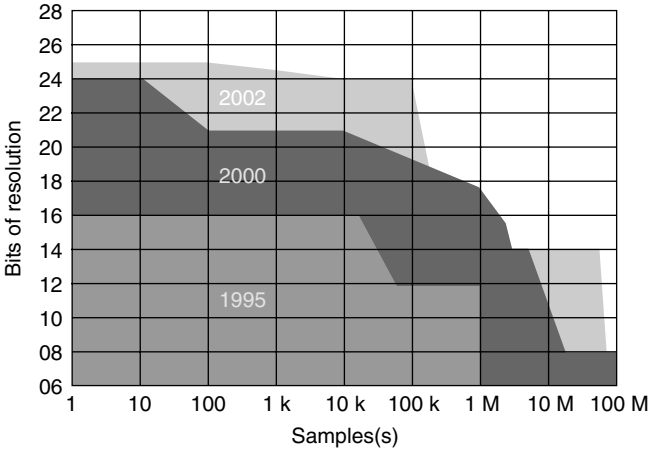


Figure 2. Performance development of digital measuring instruments (http://digital.ni.com/express.nsf/bycode/nsi2401?opendocument&lang=en&node=seminar_US).

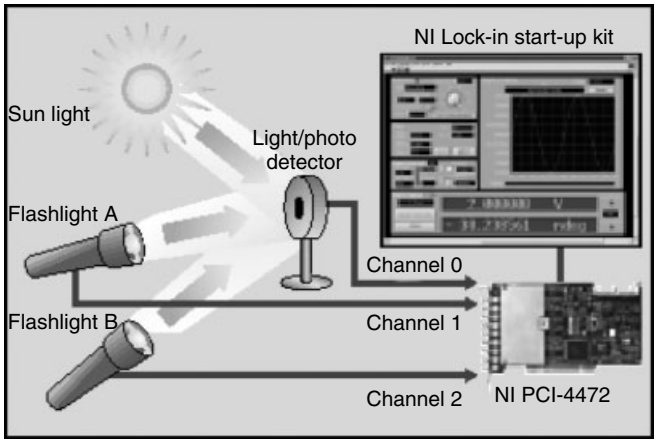


Figure 3. Knowledge-based intelligent measurement of noise-buried measuring signals (http://www.ni.com/pdf/newsletters/us/inl_q4_2002.pdf).

extracting a 50-μV sine wave from a white noise signal with 5-V amplitude.

The lock-in amplifier removes the noise by performing a Fast Fourier Transform (FFT) on the input signal at the frequency and phase the reference signal carries – *see* **Article 143, DFT and FFTs, Volume 3**.

In Figure 3, a flashlight A modulated at 10 kHz is pointing to a photodetector. Sunlight acts as unwanted noise in the system. The opto sensor converts incoming light from the sun and flashlight A into an electrical signal. The electrical signal from the sensor is connected to channel 0. The reference signal coming from the on/off switch of flashlight A is connected to channel 1. The Personal Computer Instrument PCI-4472 digitizes the data and the Lock-In Amplifier start-up kit processes it. The flashlight A signal is extracted from the sunlight noise. Flashlight B is modulated at 1 kHz and connected to channel 2. Both signals can be measured with the same opto sensor (photodetector) (http://www.ni.com/pdf/newsletters/us/inl_q4_2002.pdf) page 12.

4 MEASURING INSTRUMENTS IN KNOWLEDGE-BASED SOCIETY

Owing to the triumphal development of microelectronics and microcomputing in the last 30 years, increasing measurement processes are designed and accomplished with electrical and electronic instruments (Figure 4). Typical examples are

1. time measurement with electrical watches;
2. temperature measurement with electrical thermometers;
3. weighing with electrical balances;
4. energy measurement with electrical Watt meters;
5. health care with electrical pacemakers;
6. security supervision with digital cameras;
7. pattern recognition with electrical microphones;
8. channel hopping with electrical remote control;
9. automatic light switching with opto sensors;
10. computer games with electrical joy sticks;
11. wireless cooking with microwave oven;
12. wireless communication with cell phones.



Figure 4. Measuring instruments in knowledge-based society.

Modern algorithms for pattern recognition – another kind of measurement – distinguish automatically between normal and dangerous situations (Figure 4).

Components of modern embedded technologies are

1. Personal Computer Instrumentation (PCI) (http://www.ni.com)
2. Pocket Personal Computers (PPC) (http://www.pocket.at)
3. MS Windows Operation Systems (http://www.microsoft.com)
4. Local Networking with Ethernet (http://www.gigabit-ethernet.org)
5. Global Networking with Internet (http://www.internet.com)
6. Wireless Communication with Bluetooth (http://www.bluetooth.com).

Modern embedded technologies have turned the character of conventional measurements to intelligent or knowledge-based measurements with hardware and software having equal rights. Microelectronics and microcomputing bring low-cost measurement tools to scientists and engineers. Global remote access to knowledge (http://www.kmbook.com) and especially measurement knowledge via Internet is state of the art (http://www.messweb.de and http://www.globalspec.com).

The main information and knowledge source for measurements is the Internet. From 1996 till now, the Web archive has collected 100 Terabyte data on more than 10 billion Web pages (http://web.archive.org).

The availability of highly accurate, reliable, and robust measurement technologies is very important to research and development, manufacturing quality control, and marketplace exchange. Measurements of electrical quantities are particularly important.

Electrical quantities are the common currency of measurements; that is, many non-electrical quantities are converted into electrical quantities to facilitate measurement by electronic instrumentation. Further, the foundational International System of Units (SI), and all research based on it,

rely on a time-invariant, internationally consistent realization of the electrical units

(http://www.eeel.nist.gov/811/nat-elec-stds.html).

5 KNOWLEDGE-BASED SIMPLE AND COMPLEX MEASUREMENTS

Simple (plain) measurement is defined as the direct or indirect comparison of a measurement object with a known metrological measurement standard (reference standard) (Figure 5) and a known measurement method. In practice, it is a standardized measurement.

Metrological standards based on the International System of Units (SI) are, for example, etalons for m, kg, s, A, K, mol, cd (http://www.bipm.fr).

Complex measurement (Figure 6) is defined as direct or indirect comparison of a measurement object with different measurement standards (reference standards) like

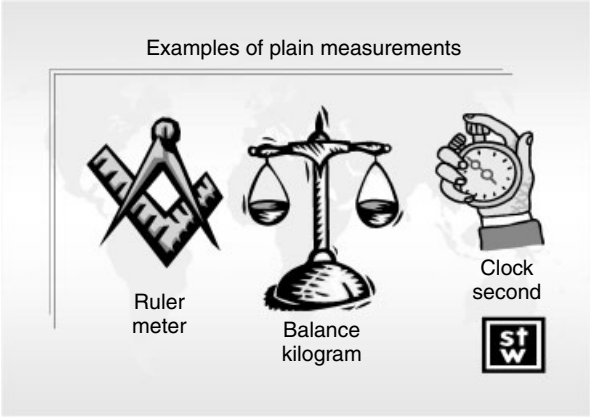


Figure 5. Simple (plain) measurements clipart (see http://www.google.de).

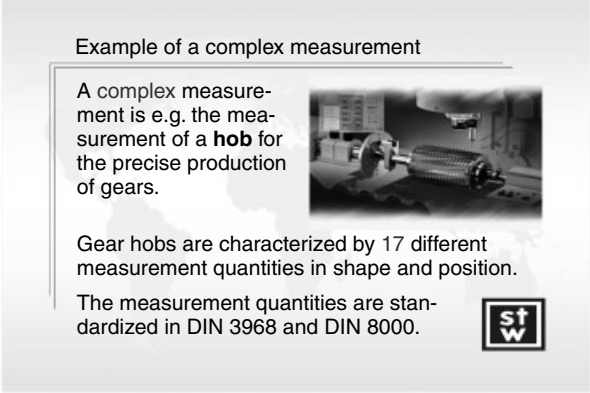


Figure 6. Complex measurement Hob (see http://www.okm-jena.de).

- metrological (etalons)
- particular (material pattern, master tools, gauges)
- virtual objective (simulation programs)
- virtual subjective (experience)

measurement standards and appropriate measurement methods.

Complex measurements are not yet standardized, or cannot readily be standardized.

Complex measurements are now accomplished as knowledge-based measurements in the framework of Internet Metrology – *see also* **Article 47, Internet Calibration, Volume 1**.

A *particular* measurement standard (Figure 7) is an objective material pattern intended to

- define,
- represent material,
- conserve, and
- reproduce

the identification marks of an object or an event in order to transmit it to other measuring objects by comparison.

A *virtual objective* measurement standard (Figure 8) is a nonmaterial pattern, for example, a numerical, textual or graphical one like a software program or a database intended to

- define,
- represent nonmaterial,
- conserve, and
- reproduce

the definition signs of an object or an event in order to transmit it to other measuring objects by comparison.

The application of virtual objective measurement standards (reference standards) is increasing in Net-environments due to the fact that they are more

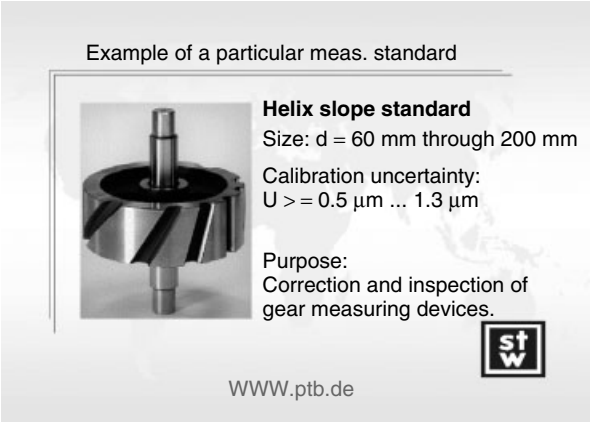


Figure 7. Particular measurement standard Helix slope standard (see http://www.ptb.de).

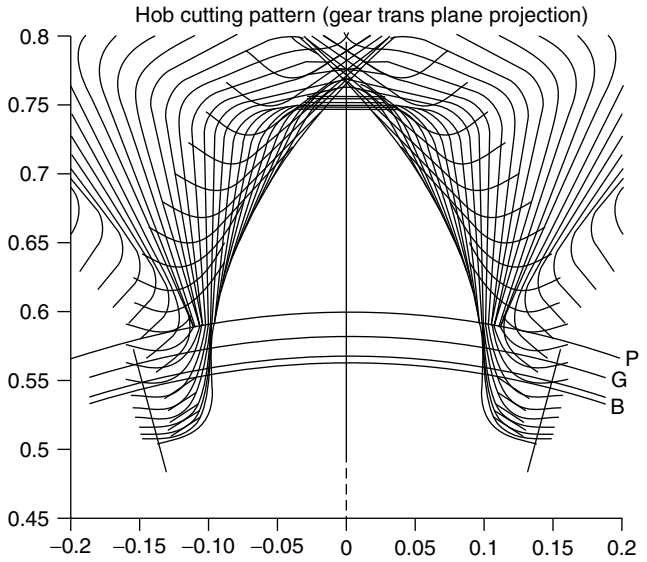


Figure 8. Virtual objective measurement standard hob cutting pattern (see http://www.uts.com/gear/411ex1.html).

- individually appropriate,
- functionally powerful, and
- financially affordable

in comparison with material patterns.

The knowledge about *virtual subjective* measurement standards is not yet adequately introduced and accepted in the body of knowledge on metrology. Nevertheless, a huge number of measurements are daily supported by the individual experience of the user. People are able to 'measure' mechanical, electrical, optical, and thermal quantities by using their individual sensual apparatus with many millions of natural sensors (receptors), nerve strings, and brain. (Figure 9) (http://www.berlin.ptb.de/8/_indexe.html).

6 KNOWLEDGE-BASED LEGAL METROLOGY

The electrical current I is a base measurement quantity of the metrical International System of Units SI (http://www.bipm.fr). The base unit of the electrical current is the Ampere A.

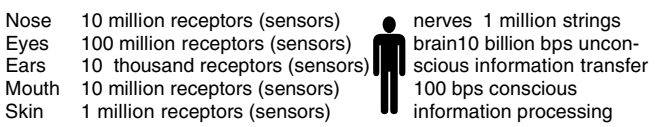


Figure 9. Information acquisition and processing capabilities of man.

The formal definition is

*The **Ampere** is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross section, and placed 1 meter apart in vacuum, would produce between these conductors a force equal to* 2×10^{-7} *Newton per meter of length.*

From the standpoint of a knowledge-based metrology, it is critical that the terms 'infinite', 'negligible', and 'would' in the definition of Ampere make its practical realization from a technological standpoint impossible. Therefore, the Ampere, since 1994, is realized by indirect measurements (http://www.ptb.de). Voltages and resistances are therefore used in the practical realization of governmental etalons in the institutions of legal metrology since January 1, 1990. Bases used are Ohm's law and the von Klitzing constant.

The electrical voltage U, current I, and resistance R relationship follow, after George Simon Ohm's (1787–1854) work, the characteristic interrelation given in Ohm's law (Figure 10):

$$U = I \times R \tag{1}$$

The National Physical Laboratory NPL, for example, adopted on January 1, 1990 (http://www.npl.co.uk/npl/cem/dclf/publications/booklet/resistancelc.html) a new value for its reference value of the ohm based on the new SI value of the von Klitzing constant

$$R_{\mathrm{K}-90} = 25\,812.807\,\Omega \pm 0.2\,\mathrm{ppm} \tag{2}$$

Measurements of resistance are referred to the quantum Hall effect using the conventional value $R_{\mathrm{K}-90} = 25\,812.807\,\Omega$ exactly (Figure 11). The uncertainty of this value in the International System of Units SI is not included in the uncertainty assessments.

The first step in the calibration chain of resistance measurements is to relate a room temperature wire-wound resistor to R_{K}. This is done, for example, at NPL using an automated cryogenic current comparator bridge (CCC) to measure a 100-ohm resistor against a quantized Hall resistance (QHR) sample.

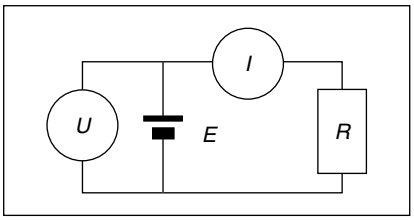


Figure 10. Fundamental electrical circuit with E energy source; U voltage meter (Volt V); I current meter (ampere A); R resistance (Ω); $R_{\mathrm{Ui}} = \infty$; $R_{\mathrm{Ii}} = 0$; $_{\mathrm{i}}$ = internal.

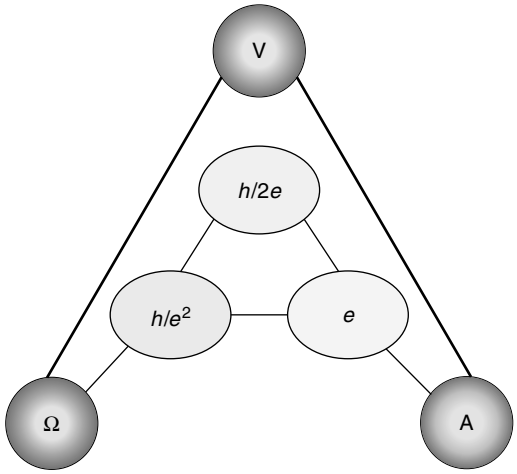


Figure 11. Interrelation between electrical units and natural constants (http://www.ptb.de Bildstelle) with V Volt voltage U; A Ampere current I; Ω Ohm resistance R; h Planck's constant; $h = 6.626\,068\,76 \bullet 10^{-34}$ J•s; e elementary charge of the electron; $e = 1.602\,176\,462 \bullet 10^{-19}$ C; R_K von Klitzing constant; $R_K = h/e^2$, R_H Hall resistance; $R_H = 1/i\,R_K$ with $i = 1, 2, 3, \ldots$; K_J Josephson constant; $K_J = h/2e$ $U = n K_J f$ with $n = 1, 2, 3, \ldots$; $i = nef$ with $n = 1, 2, 3, \ldots$; f, frequency of the irradiated microwave; J joule = W•s = V•A•s; C Coulomb = A•s; W Watt = V•A; s Second.

The high ratio accuracy of a CCC overcomes the difficulty of comparing the awkward value of the QHR with a decade value of resistance. Turns ratios of 4130 : 32 and 4130 : 64 on the CCC are used to measure the 100-ohm resistor against the i = 2 and i = 4 plateaux respectively. The measurements are estimated to have an accuracy of 0.01 ppm at the 1 sigma uncertainty level.

For the physical reproduction of Volt and Ohm measurement standards, microelectronic devices are used (Figure 12).

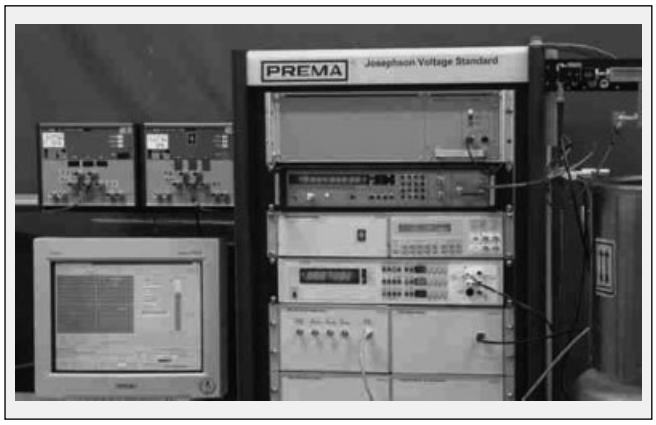


Figure 13. Josephson voltage standard rack with drive frequencies in the 70 GHz range (http://www.ipht-jena.de/BEREICH_1/abt13_cryo_electronics/cryo_start_b1.htm) and (http://www.prema.com/D/JVSe.htm).

Paradoxically, the small dimensions of the central tools of the voltage and resistance standards, contrast sharply with the very large cost of the instrumental apparatus, laboratory, and operation that is needed to make determinations. (Figure 13).

7 KNOWLEDGE-BASED INDUSTRIAL METROLOGY

Under the influence of modern technologies with standardized affordable microelectronic components for analog, digital, and mixed signal processing, the paradigm of instrument designers and manufacturers for electrical instrumentation is changing. The functionality and artificial intelligence of new instruments is high. The price of offering this extension to capability is low.

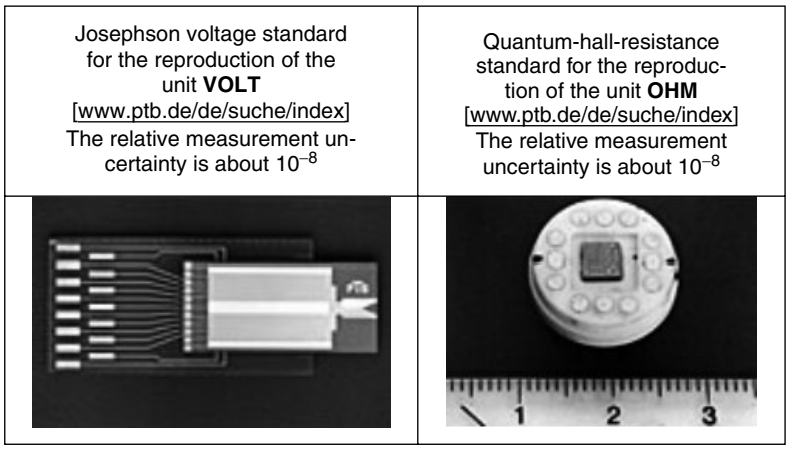


Figure 12. Microtechnological electronic standards for the units volt and ohm (http://www.ptb.de/de/suche/index).

The number of different instrument makers in the market place is decreasing. The qualification of people who use the instruments might be lower than before and individual repair of the instruments is usually not economic.

A standard statement on modern electrical and electronic instruments is



Do not attempt to disassemble your product. The product does not contain consumer serviceable components.

Owing to their high technological reliability, electrical measurement devices are less likely to age in their functionality than do their esteem and modernity factors. Users are less interested in *what* to do, than in *how to* do, in real time, in easy and affordable ways. Enlightened firms follow this thinking.

All-in-one (AIO) intelligent electrical instruments for knowledge-based industrial measurements are seen, for example on (http://www.fluke.com/about_fluke/corporation.asp?AGID=11&SID=0):

Electrical AIO instruments for industrial service and maintenance

- digital multimeters,
- electrical power controls and accessories,
- integrated ScopeMeter hand-held test tools.

Electrical AIO instruments for measurement of electrical and thermal quantities

- power harmonics meters,
- electrical testers,
- digital multimeters (DMMs) and accessories,
- clamp meters,
- ScopeMeter test tools,
- current clamps,
- digital thermometers.

Electrical AIO instruments for precision measurements

- calibrators,
- standards,
- waveform generators,
- calibration software products.

Electrical AIO instruments for process measurements

- test tools for field calibration and maintenance,
- simple loop calibrators,

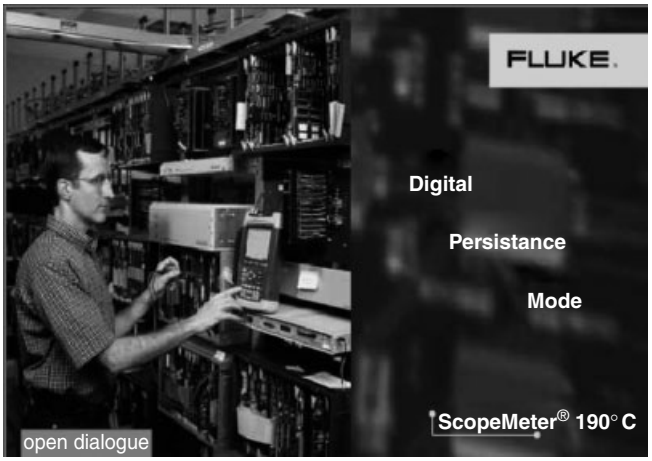


Figure 14. Multimedia sequences of ScopeMeter tests (http://www.fluke.com/products/home.asp?SID=0&AGID=0&PID=31768).

- sophisticated documenting process calibrators for current, voltage, resistance and frequency, temperature and pressure.

In addition, many intelligent instruments offer computer connectivity and documentation software to help report and analyze data results for rigorous maintenance documentation.

Market leaders provide exacting standards for companies and government organizations who rely on tightly calibrated instruments for their quality and standards programs as well as to meet strict international quality requirements.

The full functionalities of, for example, ScopeMeters are shown in multimedia sequences (Figure 14) at (http://www.fluke.com/products/home.asp?SID=0&AGID=0&PID=31768).

REFERENCES

At present there appear to be no titles available in KBS measuring systems.

Primary papers can be searched in sources such as the *KBS Journal*. However, the content of papers is often, despite the article titles, not that closely associated with KBS methods. Only a few papers will address use of KBS in measurements.

162: Principles of MEMS

Janusz Bryzek
BN Ventures, Fremont, CA, USA

1 Definition of MEMS 1091
2 Why a High Level of Interest in MEMS? 1091
3 Problems with MEMS Commercialization 1092
4 MEMS Processes 1092
5 Role of MEMS in Measurement Systems 1094
6 Key Principles used for MEMS Sensors 1095
References 1096

1 DEFINITION OF MEMS

MEMS is an acronym for Microelectro Mechanical Systems, it defines mechanical structures fabricated with IC processing on (most often) silicon wafers. In Europe, MEMS is labeled Microsystems and in Japan it is labeled Micromachines.

The term MEMS evolved in the United States in the 1990s. Prior to that period the technology was labeled *silicon micromachining*.

MEMS defines the technology; not specific products. This technology encompasses a collection of a variety of processes enabling three-dimensional shaping of wafers or stacks of wafers. While most of the applications use silicon wafers, many other materials have been used, including glass and quartz wafers.

As a result of batch manufacturing technology (fabrication of multiple devices photolithographically defined on the wafer, Figure 1), the cost of single device depends on its size; wafer processing cost is fixed for a given process. The cost difference between a 1×1 mm device and a 10×10 mm device is 100 times, as the first device would yield about 16 000 devices on a 6-in diameter wafer, and the larger device would yield only about 160 devices on the same wafer.

The beginning of MEMS technology dates back to the discovery of semiconductors at Bell Laboratories in the early 1950s. Many consider their 1954 paper, announcing the discovery of piezoresistive effect in silicon and germanium, as the birth date of MEMS.

A related acronym MOEMS stands for Micro Opto Electro Mechanical Systems and defines a subset of MEMS, that is, devices performing optical functions.

2 WHY A HIGH LEVEL OF INTEREST IN MEMS?

A high level of interest in MEMS technology results from both business and technical factors.

Factors on the business side are given below:

- Multiple emerging markets for MEMS devices promise large financial gains. The cumulative venture capital industry investments into MEMS-based companies is estimated at over $1 billion as of 2003.
- In 1999, Business Week selected MEMS as one of the three technologies expected to fuel the growth of the economy in the twenty-first century. (The other two were information technology and biotechnology.)
- The IC industry created a solid technology infrastructure immediately available for MEMS.

On the technical side, some of the multiple factors that contributed to making MEMS attractive are as follows:

- Potential for integration of devices with IC circuitry, to create integrated systems on a chip.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

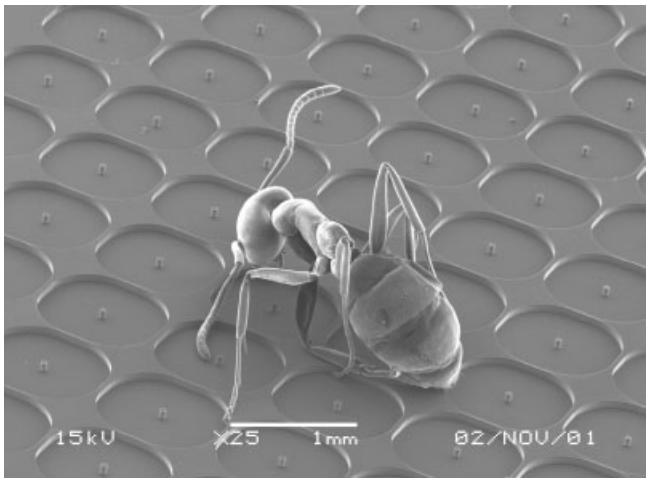


Figure 1. Scanning electron microscope (SEM) close-up of a wafer showing photolithographically defined and etched pattern forming part of the mirror array (1 mm each), contrasted in size with that of an ant. (Copyright © Transparent Networks, Inc.)

- Excellent mechanical properties resulting from the extremely pure crystalline structure with a silicon content of 99.999% or better, resulted in no material fatigue or mechanical hysteresis. This makes silicon almost a perfect material for sensors. The mechanical properties are comparable to steel, see Table 1.
- Batch wafer processing technology, enabling low-cost, high-volume production.
- Excellent lateral dimension control to submicron level.
- Available cutting edge IC processing equipment, offering easy transition to volume production.
- Available ultrapure (no mechanical fatigue) low-cost materials.
- Available sophisticated diagnostic and test equipment.
- Available design and simulation tools (software).
- Available high-volume IC packaging technologies.
- Available pool of educated silicon processing technologists.

Another aspect of MEMS that adds to its attractiveness is its synergy with nanotechnology, which receives a high level of government funding worldwide. In many cases, MEMS can be used as a packaging vehicle for nano devices.

Table 1. Mechanical properties of steel and silicon.

	Steel	Silicon	Units
Yield strength	<4.2	7.0	$1010\,\mathrm{dyn\,cm^{-2}}$
Young's modulus	2.1	1.9	$1012\,\mathrm{dyn\,cm^{-2}}$
Density	7.9	2.3	$\mathrm{g\,cm^{-3}}$
Thermal conductivity	0.97	1.57	$\mathrm{W\,cm^{-1}\,^{\circ}C}$
Mechanical hysteresis	Yes	No	
Sensitivity to stress	Low	High	
Fabrication resolution	10	0.1	μm
Fatigue failures	Yes	No	

3 PROBLEMS WITH MEMS COMMERCIALIZATION

While MEMS technology has been very promising, commercialization efforts encountered multiple stumbling blocks thereby significantly delaying the availability of commercial devices as summarized below:

- IC industry has not provided many process modules required to process mechanical devices, such as wafer-deep etching of wafers, double sided wafer alignment, multiple wafer bonding, and so on. Development of such unique processes was long and expensive, as they were usually funded by a small company with limited funding and a pool of technologists dedicated to a particular process.
- Mechanical properties of many IC materials are unknown and vary with the process selection.
- Structure thickness control was underdeveloped by the IC industry, which typically was satisfied with ±10% accuracy for ICs; but often required better control for mechanical structures.
- Mechanical and electrical stability of structures were considered black magic for a long time, as scientists did not understand its origin for a long time.
- Hermetic wafer lamination technologies critical to many devices had to be developed entirely by the MEMS industry.
- High-volume, low-cost MEMS dedicated testing, especially with the application of a mechanical stimulant (such as pressure) had to be developed entirely by the MEMS industry.
- Compatibility with the real world, such as fluids, was a major task to be developed by the MEMS industry.

As a result of these problems, volume implementations of initial devices was slow, as shown in Table 2.

These incubation periods may seem long, but are similar to the incubation times of other emerging technologies, including VCRs, cell phones, and satellite television.

4 MEMS PROCESSES

Many processes for fabricating mechanical structures were developed over time. Three major categories of such processes are as follows:

Table 2. Development of MEMS.

Product	First prototype	Cum volume 1 M	Incubation (years)
Pressure sensor	1961	1984	23
Acceleration sensor	1970 piezoresistive	1995	25
	1977 capacitive	1995	18
Ink jet printers	1977	1996	19
Displays (DLP)	1979	2001	22

- bulk micromachining of silicon
- surface micromachining
- deep reactive ion etching.

4.1 Isotropic etching

Bulk micromachining makes use of the entire wafer thickness, which typically may range from 500 to 1000 μm.

The first bulk micromachining process made use of the anisotropic wet etch developed for pressure sensors in 1970 by Kulite, New Jersey. This etch could remove silicon in all directions, independently, according to the crystallographic orientation of the substrate. The shape was defined by a masking process, Figure 2. The accuracy of the structure definition was not very good, as it depended on many factors, such as time, temperature, and etchant concentration. No good etch-stop technology enabling termination of the etch at a predefined depth was developed.

4.2 Anisotropic etching

A major improvement in accuracy of etching was achieved with the development of anisotropic wet etching, which relied on crystallographic orientation. This was first demonstrated by Kulite, in New Jersey, in 1976 for pressure sensors. Different crystallographic directions of silicon exhibited different etch rates for selected etchants, such as KOH. The etched rates could differ by 2 orders of magnitudes (depending on the direction); thus, the mask was defining the shape of the etched cavity, with the only variable being the depth. This depth still depended on temperature and concentration and could be controlled to about ±10% of etched depth in the best cases. If the starting wafer thickness was 600 μm and etch depth was 500 μm, this meant the variation of thickness of unetched silicon would be 100 ± 100 μm, or 0 to 200 μm. This was not good enough for the first application, that is, in pressure sensors, as pressure sensitivity varies as the square (for piezoresistive sensors) or cube (for capacitive sensors) of the diaphragm thickness.

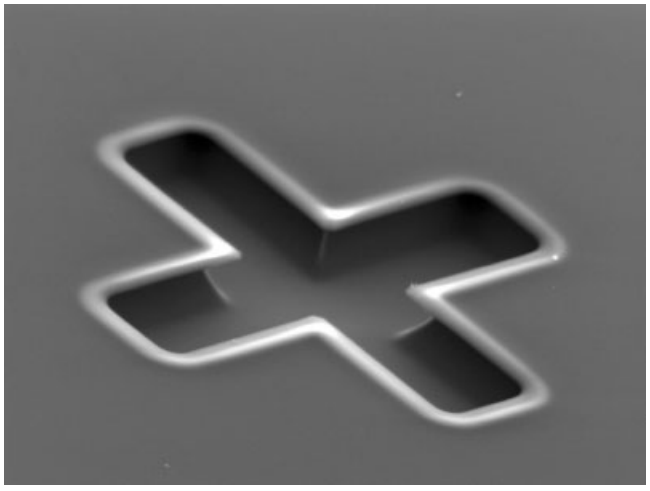


Figure 2. Example of isotropic etching of silicon through a cross-shaped mask and with nonetched surface layer of silicon oxide. (Copyright © Transparent Networks, Inc.)

Several techniques for stopping the etch on this layer were later developed. One of them took the advantage of the high differential etch rate as a function of the silicon-doping level: for high doping, the etch rate was significantly slower as compared to the low doping level. Using diffusion or ion implantation, a thin (submicron to tens of microns) highly doped (carrier concentration $>10^{19}\,\mathrm{cm}^{-3}$) surface layer could be formed, which would terminate the etch. The depth accuracy of such a layer was still about ±10%, but the tolerance of the intended feature was not dependent on the etched thickness. If 5-μm thickness was programmed, it would have ±0.5-μm tolerance, which was not achievable without the etch stop. This type of etch stop was deployed (among others) in General Motors pressure sensors, Figure 3.

Another etch-stop technique was based on the p-n junction, biased by the voltage. During the initial etch, the p-n junction was blocking the flow of the current. As soon as one of the layers (p or n) was etched, the current flow caused oxidation of the exposed area, stopping the etch. This type of etch stop was deployed (among others) by NovaSensor, California.

4.3 Sacrificial etching

Surface micromachining focused on using the few micron thick surface layer of the wafer, to simplify cointegration

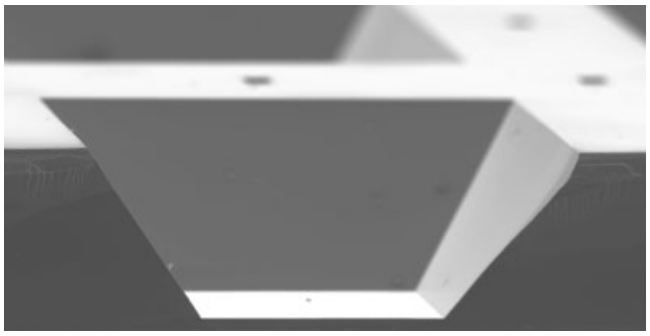


Figure 3. Anisotropic etching of silicon follows crystallographic planes, enabling precision control of the rectangular patterns and formation of thin membranes. With etch stop, the thickness of the thin membrane (on the bottom of the chip) could be controlled to less than 10%. (Copyright © Transparent Networks, Inc.)

with electronics. The process emerged through the 1980s, with major progress in its development at the UC Berkeley and Sandia National Laboratories (2004).

The principle of sacrificial etching is based on silicon oxide etchants that do not etch polysilicon. The structures are formed by patterning a layer of oxide into the required shape by photolithography, covering it with patterned layers of polysilicon, and then etching away (sacrificing) the oxide through small holes etched in polysilicon.

The thickness of structures fabricated with surface micromachining is typically under 4 μm, due to limitations in polysilicon deposition. One of the highest volume products based on this technology is the family acceleration sensor from Analog Devices, Massachusetts.

A well-advanced process is Summit V, developed at Sandia National Labs, enabling five layers of polysilicon to be independently shaped to create complex mechanical systems in the chip, such as gears, pumps, and mirrors – see Figure 4.

4.4 Deep reactive ion etching (DRIE)

None of the deep wet etch technologies are friendly to the processing of integrated circuits. Furthermore, sloping silicon in an anisotropic etch unproductively consumes the chip area, thus, increasing its cost.

Development of plasma-based deep etching was a dream tool for the MEMS industry. This tool emerged in the 1990s, with major progress contributed by Bosch. Their process was based on alternating two cycles: plasma etching and vertical polymer wall coating to prevent lateral etching.

Initial etch speed was about 1 $\mu m\,min^{-1}$. As of 2004, the fastest machines can deliver up to 10 $\mu m\,min^{-1}$ rough etch (dimensions are not very well controlled).

While this deep reactive ion etching (DRIE) etches silicon, its etch rate of silicon oxide is 100 to 1000 times slower

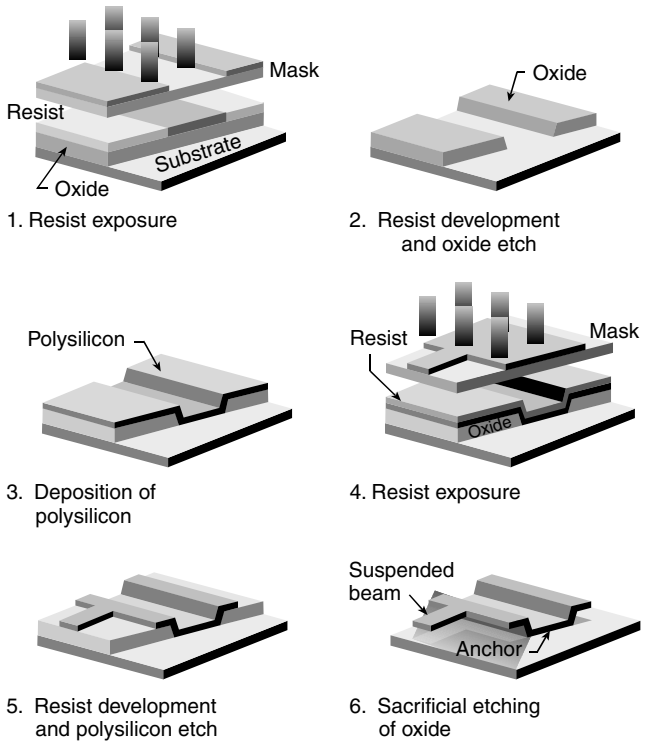


Figure 4. Foundation for surface micromachining is sacrificial etching of silicon oxide underneath of polysilicon, to create suspended mechanical structures. (Copyright © Nadim Maluf, New Focus.)

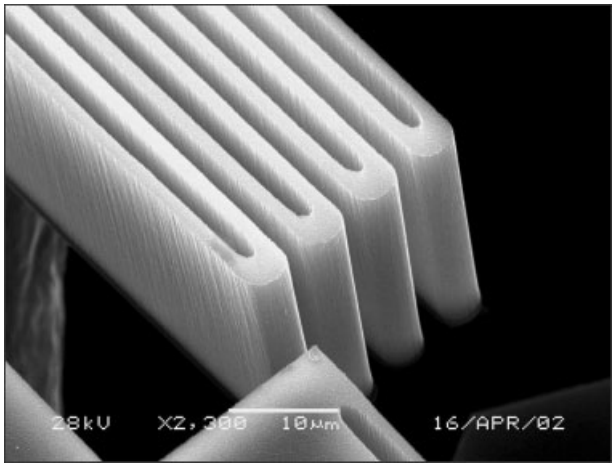
(depending on settings), enabling creation of the etch stop on the oxide. One of the easiest approaches is the use of silicon on insulator (SOI) wafers, containing an embedded layer of oxide.

While the etch speed for DRIE is now comparable to anisotropic etch, the throughput is not. DRIE equipment can etch only one wafer at a time, while wet etching can simultaneously process a large number of wafers (e.g. 100), limited only by inexpensive tooling.

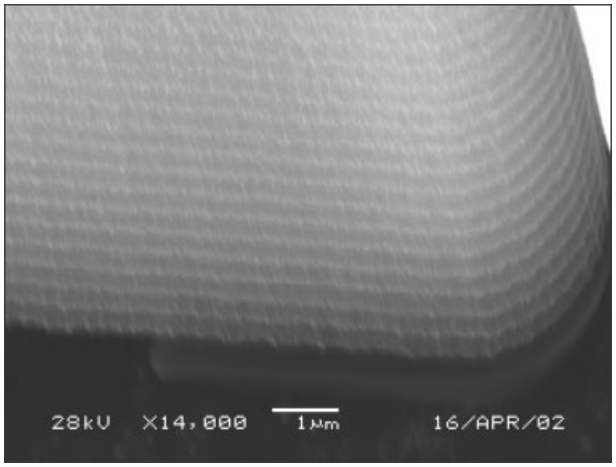
If the design needs several DRIE etches with a cumulative depth of 800 μm (typical for many sensors), and the effective (accuracy limited) etch speed of 4 $\mu m\,min^{-1}$, each wafer will use an etcher for 200 min. This allows etching of only 7 wafers per day using three shifts, or about 1800 wafers per year, in a 5-day week. This makes DRIE a significantly more expensive process than wet etching. Examples are shown in Figure 5.

5 ROLE OF MEMS IN MEASUREMENT SYSTEMS

MEMS-based devices play a valuable and instrumental role in the sensor world, in their application in the measurement of physical variables. Key MEMS sensors include the following:



(a)



(b)

Figure 5. Deep reactive ion etching (DRIE) enables high aspect ratio vertical etching of silicon. (a) SEM picture shows 1-μm wide, 30-μm deep silicon springs with 1-μm spacing between springs. (b) Close-up of the vertical wall, showing the surface scalloping resulting from the alternative etching and polymer deposition during the DRIE process. As shown, each cycle etches approximately 0.3 μm of silicon, resulting in highly vertical etch. (Copyright © Transparent Networks, Inc.)

- *Pressure*: with MEMS sensors representing over 90% of units and 50% of revenues.
- *Acceleration*: with MEMS sensors representing over 95% of units and 25% of revenues.
- *Rate*: with MEMS sensors representing about 95% of units and 75% of revenues.
- *Force*: with MEMS sensors representing about 20% of units and 5% of revenues.

Pressure sensors were the first devices based on silicon technology introduced to the market in the 1960s, for military and aerospace applications. In the year 2003, the volume of pressure sensors produced approached 200 million units, with automotive (most models of advanced cars use over 10 pressure sensors) and medical (primarily disposable blood pressure sensors) markets dominating the unit volume.

Acceleration and gyro MEMS-based sensors were developed for military and aerospace market in the 1970s. In the year 2003, over 100 million units were produced and it was dominated by the automotive airbag applications.

MEMS rate sensors were developed in the 1990s, first for military, and then for automotive applications. In 2003, their unit volume was about 2 million units. The largest volume markets are automotive (stability control and GPS) and consumer (camera shake stabilization and cell phone GPS), with the potential for 10 million rate sensors by 2010.

The major market for silicon force sensors are home scales, absorbing millions of devices.

The emerging MEMS sensing applications include the following:

- infrared
- flow
- chemical
- biological.

6 KEY PRINCIPLES USED FOR MEMS SENSORS

Major sensing technologies that can be applied in the MEMS form include the following:

- piezoresistive
- capacitive
- resonant
- thermoelectric.

The piezoresistive effect (Smith, 1954) in silicon produces resistance change that is approximately 2 orders of magnitude larger when compared to the change that would result from a dimension change under applied stress (typical for metals). For example, if the material is elongated 0.1% by stretching, typical metallic resistors used for strain gauges would change by about 0.2%, but the resistance of silicon would change by about 10%. Piezoresistive sensors dominate pressure, acceleration, and force applications. Typically, four piezoresistors are connected into a Wheatstone bridge configuration to reduce temperature errors. Figure 6 is a simple example of a piezoresistance sensor.

Capacitive sensors have at least one electrode moving under the input variable, which typically would be pressure, acceleration, or rate. While the simplest configuration is two flat electrode capacitors, the interdigitated silicon fingers (Figure 7), gained wide acceptance as an inertial sensor, as it allows for larger sensing capacitance.

Resonant sensors are based on resonating beams or diaphragms. Applied strain changes the resonant frequency

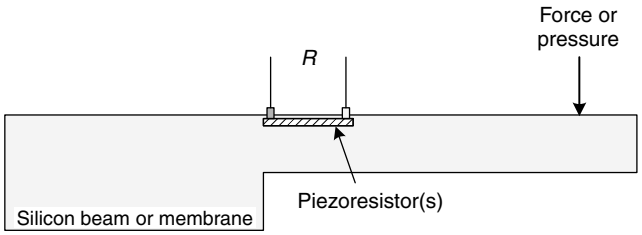


Figure 6. Sensing piezoresistor schematic.

(similar to a guitar string), enabling measurement of input variables such as pressure, acceleration, rate, and temperature. Resonant frequency pickup may use any of the sensing technologies.

MEMS thermoelectric sensors are based on a large number of tiny thermocouples connected in series. These are primarily used to measure IR radiation. One end of the thermocouple is located over a thermally isolated region of the chip, the other end over conducting silicon. In response to radiation, the thermally isolated end of the thermocouple gets warmer, enabling measurement of IR power.

The emerging approach for both chemical and biological MEMS sensors is based on carbon nanotube technology, demonstrated by Silicon Valley companies; it has great potential for creating new markets for yet more MEMS sensors – *see* **Article 163, Uses and Benefits of MEMS, Volume 3** and **Article 164, Principles of MEMS Actuators, Volume 3**.

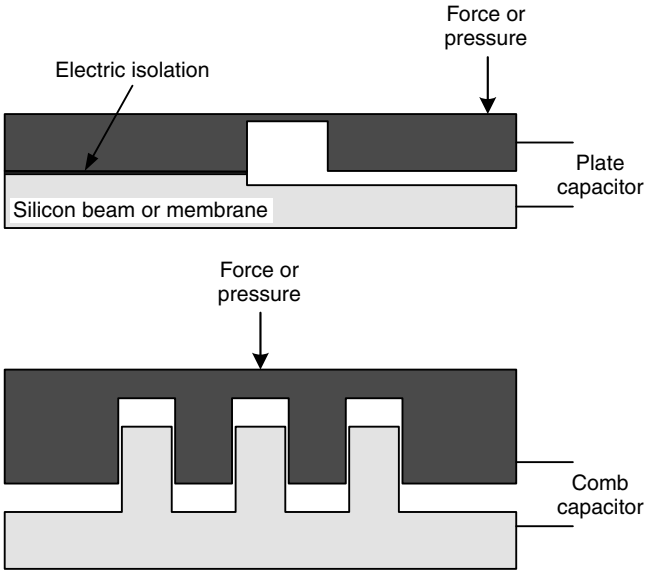


Figure 7. Capacitive sensing schematic.

REFERENCES

Sandia National Labs, (2004) http://www.sandia.gov/mstc/technologies/micromachines/tech-info/technologies/summit5.html.

Smith, C.S. (1954) Piezoresistive Effect in Germanium and Silicon. *Physical Review*, **94**, 42–49.

163: Uses and Benefits of MEMS

Janusz Bryzek
BN Ventures, Fremont, CA, USA

1 Introduction 1097
2 Benefits of Using MEMS Devices 1097
3 MEMS Market Overview 1098
4 Pressure Sensors 1098
5 Acceleration Sensors 1099
6 Displays 1100
7 Ink-jet Printers 1101
8 Lab on the Chip 1101
Related Articles 1102
References 1102

1 INTRODUCTION

MEMS (Microelectro Mechanical Systems) products, limited in their first decades to pressure and acceleration sensors, significantly lagged in making an impact on international economy and wealth. In the 1980s and 1990s, many governments heavily funded the development of MEMS technology, with the United States, Japan, and Germany giving the highest investment levels.

As a result, a large number of MEMS devices were demonstrated and some were commercially deployed in high volumes.

A general introduction to the principles of MEMS devices is given in **Article 162, Principles of MEMS, Volume 3** and their use in actuators is covered in **Article 164, Principles of MEMS Actuators, Volume 3**. This article discusses their application in products.

During the period from 1998 to 2001 ('optical bubble'), over US $1 billion was invested into MEMS-based companies by the venture capital industry, with the main focus on optical networking. This greatly helped accelerate the development of MEMS products, and the design and manufacturing infrastructure for their fabrication.

Automotive pressure and acceleration sensor markets were the first to become a multibillion-industry sector, with high-end value cars using up to 14 pressure and 11 inertial (acceleration and rate) sensors on the 2003 model cars.

Several other killer applications based on MEMS also emerged. The most visible examples of these new billion-dollar markets include ink-jet printers and projector displays. Several more applications are getting ready to create other major areas, including multiple biotech, medical, RF, aerospace, and consumer applications, each capable of creating their own billion dollar market.

As a spin-out of MEMS, nanotechnology started to receive significant attention, culminating in over US $3.6 billion of funding in 2004 from the US Government. While most of the nanotechnology applications are still decades away from commercialization, selected developments promise high volume applications.

2 BENEFITS OF USING MEMS DEVICES

MEMS is labeled an enabling technology: it enables development of products that would not otherwise be possible.

In some cases, it could provide a feature that is not possible with other technology, such as a single chip optical display with two million tiny mirrors enabling wall-size projection of movies with high definition.

It could provide a breakthrough in cost reduction, such as disposable blood pressure sensors.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

It could also be a combination of multiple factors similar to the benefits seen by the introduction of integrated circuits, such as:

- low cost of manufacture per item due to batch replication manufacturing of multiple wafers, each with a large number of devices per wafer, as compared to one device at a time in the classical manufacturing technology;
- improved mechanical performance due to the excellent mechanical properties of silicon in its pure state;
- capability to deliver very large volume of products, the result of adopting IC processing technology;
- significant reduction in size and weight of the structures, often enabling new applications in which fast response and small size are needed;
- improved reliability because of the use of an integrated structure with a small number of external connection points;
- possibility to cointegrate the MEMS structure with that of the processing electronics, thus enabling new system architectures to be made that were not possible with discrete electronics; and,
- feasibility of building single chip complex electro-mechano-optico multisystems.

As the MEMS technology has entered a high volume production phase for only several applications out of the demonstrated several hundred, it is to be expected that many new benefits will be discovered along the way.

3 MEMS MARKET OVERVIEW

In the 1990s, MEMS programs were established at all major universities around the world. Several hundred MEMS fabrication facilities (MEMS foundries) have been funded, as of 2004, in the academic and commercial environment. Hundreds of MEMS applications have been demonstrated during the last two decades. Only selected applications, however, reached volume production, this not being unusual for any new technology.

One of the problems with a widespread introduction of manufacturable MEMS technology has been its multidisciplinary nature: the integrated circuits designed primarily by electrical engineers who also required a deep background in other disciplines, such as mechanical engineering, material science, optics, biotechnology, and measuring systems.

The largest applications for MEMS technology in the near future (Mancef, 2002) are characterized in Table 1.

Some of these applications are characterized below.

Table 1. Forecast applications for MEMS, Mancef (2002).

Application	2005 (US $ in billion)	2015 (US $ in billion)
Pressure sensors	3.0	6.0
In-vitro diagnostics	0.01	5.0
Read/write heads for hard disks	2.0	4.0
Ink-jet print heads	2.0	3.5
Optical displays	1.0	3.0
Gyroscopes	0.1	2.0
Lab-on-a-chip	0.01	2.0
Drug delivery systems	0.0	1.5
Inertial sensors	0.2	1.5
Chemical sensors	0.1	1.0
Optical switches	0.1	1.0
RF devices	1.0	0.1
Microspectrometers	0.02	0.4

Copyright (2002) © Mancef.

4 PRESSURE SENSORS

The first MEMS-based pressure sensors were developed in the early 1960s in the form of silicon strain gages bonded to metal diaphragms. In the 1970s, silicon strain gages were integrated into silicon diaphragms, and in the 1980s, signal-conditioning electronics had been cointegrated.

Applications for silicon pressure sensors were initially in aerospace and military markets because of their low weight and small size. The first high accuracy silicon pressure sensors were used on commercial DC-9 airplanes in 1967 for airspeed and altitude measurement.

The next market to adopt MEMS pressure sensors was the process control market. Honeywell introduced its Class 41 pressure transmitters in this market in 1970.

As soon as prices started to move toward the US $100 mark, and below, industrial applications started to reach the market in 1970s. Currently, thousands of products use pressure sensors, including compressors, paint sprayers, air conditioning systems and hydraulic controllers.

The medical market adopted MEMS pressure sensors in the 1980s. The world's first disposable blood pressure sensor (developed by the author) reached the market in 1982, with the volume in 2004 being about 20 million units per year. Other medical applications include infusion pumps, catheter tip pressure measurement, and intrauterine pressure measurement during the delivery of babies that was able to prevent unnecessary cesarean sections. An example of a blood pressure sensor system is shown in Figure 1.

Automotive applications for pressure sensors received a major boost from the US government. In 1974, the United States created its Corporate Average Fuel Economy (CAFE) standard in response to the energy and pollution crisis. This targeted an increase of fuel economy from 12 mpg (miles per US gallon) in 1974 to 27.5 mpg in

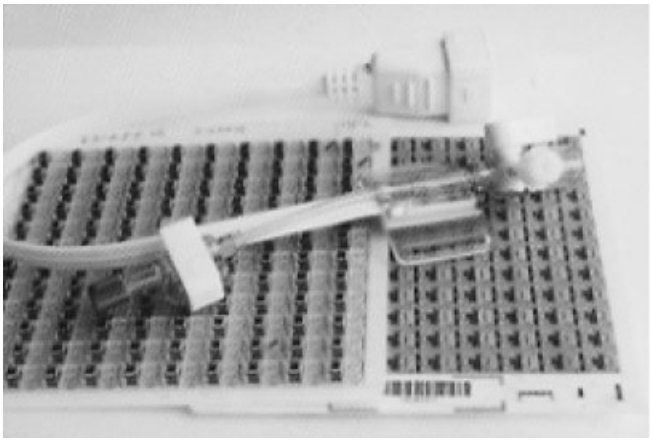


Figure 1. Disposable blood pressure sensor based on MEMS chip bonded to ceramic substrate with laser trimmed calibrating resistors. (Copyright © NovaSensor/Abbott.)

1997 also with a dramatic reduction in exhaust pollution. The goal was achieved by a closed loop gasoline combustion control using a pressure sensor enabling stoichiometric air-to-fuel ratio control, independently of altitude and barometric pressure. The world followed this development with some delay.

In 2003, high-end value cars used up to 15 pressure sensors, including fuel injection, barometric pressure measurement, suspension pressure control, electrohydraulic brake actuation, tire pressure sensing, air-conditioning compressor control, exhaust gas recirculation control, side air bags actuation, oil pressure measurement and many other measurements. In 2004, the total usage of pressure sensors exceeded 100 million units, with a potential for 500 million pressure sensors by 2015.

5 ACCELERATION SENSORS

The acceleration sensor was the second MEMS device to reach the market in large numbers. It was initially developed for the military and aerospace market, with piezoresistive devices being the first to reach the market in 1970s. The first capacitive sensors reached the market in 1980s.

In the 1990s, the US government mandated implementation of air bags in cars. Initially air bag sensors were based on mechanical sensing of a crash using a metal ball moving inside of a sleeve, forming an acceleration activated switch. While such a device was sufficient for the first air bags, typically five of such sensors had to be wired and mounted in the front of the car.

To lower the cost, an improved acceleration sensor was required that would open up multiple options for air bag system design. These include

- Production line reprogrammability that would reflect the car configuration affecting crash response (e.g. engine weight and geometry).
- Elimination of sensor wiring cost; the sensor could be mounted inside of the cabin or engine compartment, and need not be in the bumper unit.
- Reduction of number of required sensors, as the car crash pattern signature, in frequency domain, could be stored in the memory through use of the fast Fourier transform of acceleration.

In the early 1990s, a massive effort was initiated throughout the world to develop a low-cost silicon acceleration sensor. NovaSensor (Silicon Valley) started shipments of advanced sensors in 1992, but exited in 1995 due to dropping prices. Another Silicon Valley startup, ICSensors, started shipments in 1993, but exited in 1998 when it failed to introduce a lower cost device. Sensonor in Norway started shipments of a simple but cheap sensor in 1993, becoming the industry leader in 1997. Failure to introduce a low cost, more advanced model almost destroyed the Company in late 1990s.

Analog Devices, USA, started shipments of the industry's first, surface micromachined, sensors integrated with its signal conditioning electronics in 1994, after a rumored US $200 million investment into the program. It shipped an estimated 50 million units in 2003, becoming one of the key players in this market.

Bosch, Germany, unsuccessfully tried to buy NovaSensor technology in 1990. It introduced its own version, see Figure 2, in 1997 shipping about 50 million units in 2003, primarily to captive applications.

Nippon Denso, Japan, developed a sensor for captive applications in 1995 and shipped an estimated 30 million units in 2003.

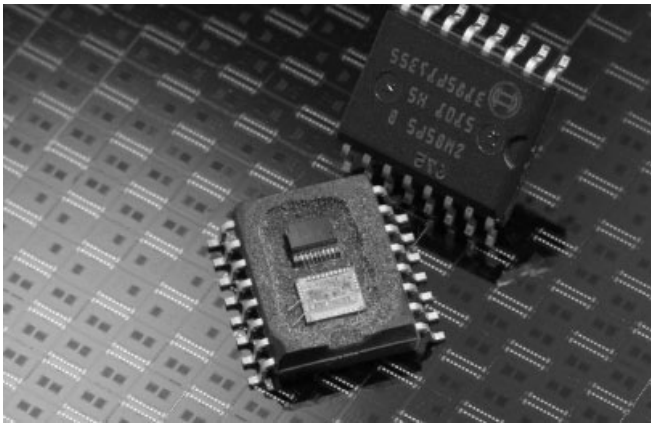


Figure 2. Air bag acceleration sensor built as a two-chip configuration: hermetically sealed sensor chip (upper part of picture) wirebonded to sensor signal processing IC (bottom) are molded in a plastic package for low cost. (Courtesy of Bosch.)

Finally, Motorola, USA, introduced an integrated sensor in 1997 and shipped an estimated 20 million units in 2003.

Clearly, the market for MEMS devices is major and is still in its infancy.

6 DISPLAYS

Developments of MEMS light modulators started at IBM in late 1970s (Petersen, 1977). Texas Instruments, TI, started development of MEMS optical scanners in 1977 and in 1987, demonstrated the digital micromirror. In 1994, a prototype of the digital light projector (DLP), based on digital mirror display chip (DMD) was shown (Figure 3).

The first DMD chip consisted of one small ($17 \times 17\,\mu m$) mirror per display pixel. For their initial 800×600 pixel displays, the chip had 480 000 mirrors. Each mirror could direct the light either to the screen, or the light absorber by tilting $\pm 10^{\circ}$ on one axis. For forthcoming HDTV2 displays, the DMD chip will have about two million mirrors.

Projection was based on time division multiplexing, when each mirror was sequentially controlled to pass light to the screen or to the absorber. One cycle consisted of 256 intensity levels for three colors. If during the cycle the color light was directed towards the screen all the time, it represented the brightest pixel for this color, and if light was directed during the entire cycle to light absorber, it was the darkest pixel.

In a single DMD chip, the color switching (RGB) was accomplished using a rotating color wheel.

DLP™ technology-enabled projectors for very high image quality or high brightness applications such as cinema and large home-theater displays use a three-DMD-chip configuration to produce stunning images, whether moving or still.

In a three-chip system, the white light generated by the lamp passes through a prism that divides it into red, green, and blue. Each DMD chip is dedicated to one of these three colors; the colored light that the micromirrors reflect is then combined and passed through the projection lens to form an image.

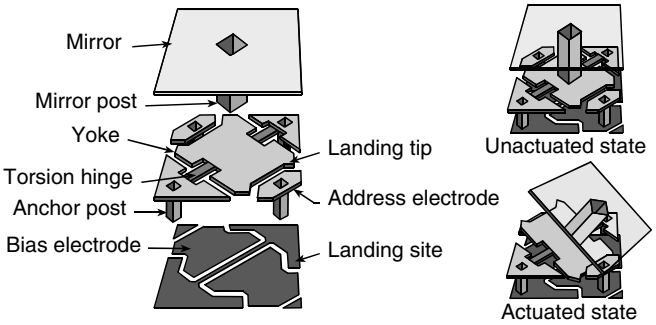


Figure 3. Mirror light switch is electrostatically driven to tilt mirror by about 10° in approximately $10\,\mu s$. (Courtesy of TI. (2004) Texas Instruments DLPs, http://www.dlp.com/dlp_technology/dlp_technology_overview.asp.)

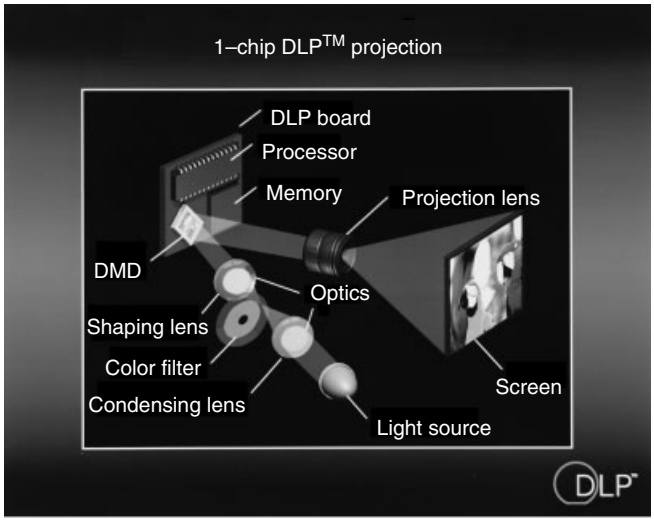


Figure 4. Single-chip DLP Projection system uses rotating color filter to sequentially modulate color on a display, relying on human eye to integrate the resulting picture. (Copyright © Texas Instruments.)

The first projector based on the DLP technology, diagrammatic view shown in Figure 4, was shipped in 1996 following US $1 billion cumulative of funding by Texas Instruments. The first cumulative million units of the DLP™ subsystem were shipped in 2001, the second million in August 2003, and the third million in April 2004, TI (2004).

A promising, emerging application of this digital mirror technology is the pixilated headlight for cars, Figure 5. The left headlight is a xenon type that blinds oncoming traffic. The right one, also energized, includes a unit with 477 000 MEMS mirrors, which individually direct the light to prevent the blinding effect and improve road lighting.



Figure 5. Pixilated headlight for cars using digital mirrors. (Copyright © BMW.)

7 INK-JET PRINTERS

Ink-jet printers emerged in the 1990s, creating one of the largest volume MEMS applications. The first companies to develop this technology were Xerox and Hewlett Packard.

While several different approaches for ink-jet printers are used, one of the more popular printers (the bubble jet form) is based on the formation of tiny bubbles through the on-chip heater and ejection of the bubble through a nozzle. As of 2004, consumer ink-jet printers reached the lower size of 1.5-pL (10^{-12}) droplets and print density 5760 × 1440 dots per inch having 180 nozzles on the chip (for the Epson Stylus Photo R800), delivering professional picture quality for home printers.

The printer application became the largest volume MEMS device in production, with hundreds of millions of printer cartridges being shipped yearly (see Figure 6).

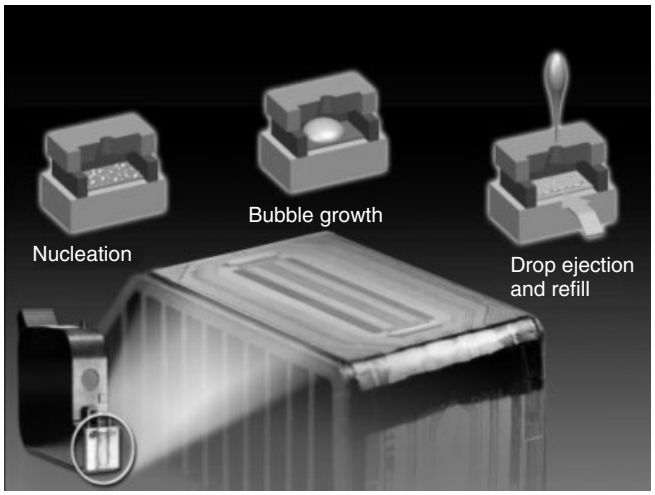


Figure 6. Print head of ink-jet print cartridge includes MEMS chip enabling nucleation of the droplet, bubble growth and drop ejection with micron precision to deliver high resolution photo quality prints. (Copyright © ST Microelectronics.)

8 LAB ON THE CHIP

The most promising and potentially the largest MEMS applications are foreshadowed to be in the biotechnology and medical markets. Single chip integration of a variety of sensors and actuators on the single chip makes possible development of systems replacing entire laboratories. This offers multiple benefits:

- spectacular reduction of size and weight, enabling the laboratory to be taken to the site (i.e. a patient);
- dramatic reduction of cost making the laboratory disposable, if needed;
- stunning reduction of sample volume needed from milliliters to nano- and picoliters;
- major reduction of test time from hours or days to minutes.

The functionality that can be provided on-chip can integrate pumps, channels, fluid mixers, and separators, and optical, biological, and chemical detectors (Figure 7).

The idea of a microminiature, mass replicated, chromatograph-style analyzer is not new and its various parts

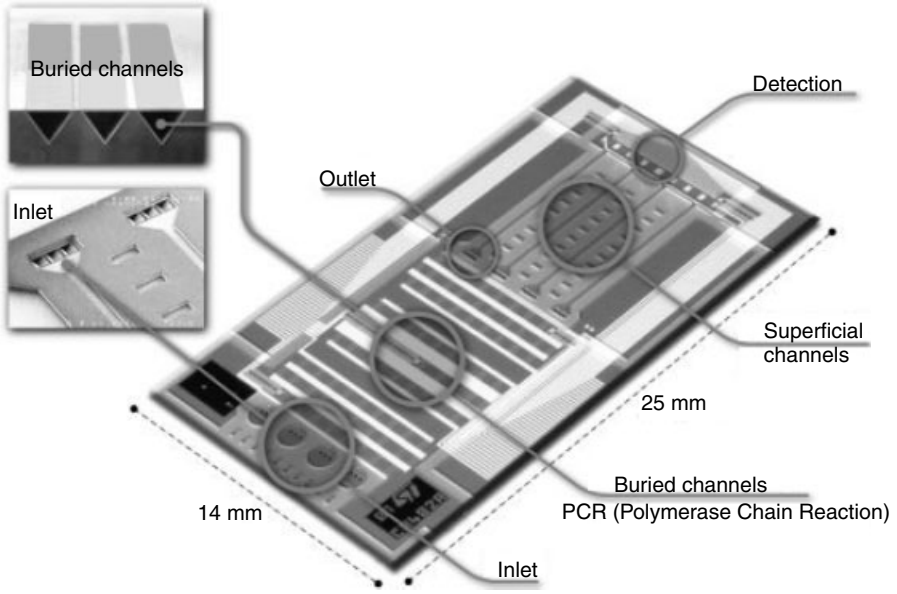


Figure 7. Lab-on-chip example includes micromachined inlets for sample introduction, buried channels for polymerase chain reaction processing DNA and detectors. (Copyright © ST Microelectronics.)

have been demonstrated. This is clearly another area of sensing that will take on many a new application once these microminiature forms are available at low price with satisfactory performance.

RELATED ARTICLES

Article 80, Principles of Fine Mechanics – Kinematic and Elastic Designs, Volume 2; **Article 81, Principles of Fine Mechanics – Systems Considerations, Volume 2**; **Article 113, Force-feedback Sensors, Volume 2**; **Article 162, Principles of MEMS, Volume 3**; **Article 164, Principles of MEMS Actuators, Volume 3**.

REFERENCES

Mancef. (2002) *MEMS Roadmap*, Micro and Nanotechnology Commercialization Education Foundation MANCEF, Albuquerque, www.mancef.org.

Petersen, K. (1977) Micromechanical Light Modulator Array Fabricated on Silicon. *Applied Physics Letters*, **31**(8), 521–523.

TI. (2004) Texas Instruments DLPs: http://www.dlp.com/dlp_technology/dlp_technology_overview.asp.

164: Principles of MEMS Actuators

Janusz Bryzek
BN Ventures, Fremont, CA, USA

1 Definition of an Mems Actuator 1103
2 Electrostatic Actuation 1103
3 Electromagnetic Actuation 1105
4 Piezoelectric Actuation 1105
5 Thermal Actuation 1106
6 Optical Actuation 1106
Related Articles 1107
References 1107

1 DEFINITION OF AN MEMS ACTUATOR

An actuator converts one of many forms of energy into a motion or force. MEMS actuators are microminiature integrated devices, for example, converting electrical energy into a deflection.

Actuators find use in measuring systems as devices that can balance a known force against an unknown force, thus enabling a measurement to be obtained from the input force voltage or current – *see* **Article 113, Force-feedback Sensors, Volume 2**. They also are found as driver mechanisms in printer heads, plotters, and printers. This article describes the microminiature forms used in MEMS devices; *see also* **Article 162, Principles of MEMS, Volume 3**; and **Article 163, Uses and Benefits of MEMS, Volume 3**.

Five basic control actuators have been developed for MEMS devices:

1. Electrostatic
2. Magnetic
3. Piezoelectric
4. Thermal
5. Optical

These are discussed in turn.

2 ELECTROSTATIC ACTUATION

These are formed in the archetype version as two electrically conducting plates held apart with insulating supports. Voltage is applied across the plates. It is essentially a capacitor.

The forces of electric fields are inversely proportional to the square of the distance between the plates forming the capacitor. As the device is made smaller, this force becomes significant enough to provide displacement. Electrostatic actuators, however, still provide very low drive energy; they became a dominant approach for MEMS actuation as they involve electric fields. Such drives require a relatively high drive voltage (20 to 500 V) to be effective. Predominately, electrostatic actuators are implemented in two types: parallel plate and comb drive.

2.1 Parallel-plate electrostatic drive

The parallel-plate actuator uses parallel electrodes to form the capacitor structure. At least one electrode is suspended on a spring support such that it can change position in response to the applied voltage differential connected across the insulated plates.

Figure 1 illustrates this concept. The electrode with voltage $V1$ is suspended on a spring allowing its rotation. Stationary electrodes with voltages $V2$ and $V3$ create differential forces F_{1-2} and F_{1-3} that try to rotate the upper

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

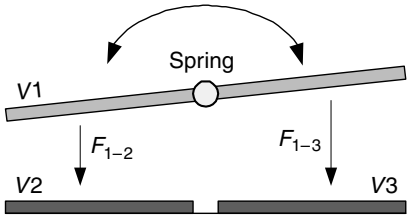


Figure 1. Concept of parallel-plate electrostatic actuator.

electrode. As electrostatic force is only positive, the rotation will be in a direction of the larger of the $V2, V3$ voltages.

The equilibrium of a simple system can be found by equating the electrostatic force F_e with the restoring force F_r provided by the spring constant of the system:

$$F_e = \frac{\varepsilon V^2 A}{2d^2} = F_r \tag{1}$$

where

- ε is the dielectric constant for the substance between the electrodes,
- A is the area of the plates,
- d is the gap height between the electrodes at the zero position, and
- V is the difference voltage applied to the electrode set.

The electrostatic force F_e varies as the voltage squared and the inverse of the gap squared.

However, the restoring force F_r of the spring varies linearly with the change in gap. Because the electrostatic force increases much faster than the restoring force, it can be shown that this system is unstable in the open-loop control mode at less than one-third of the gap height. After that point, the parallel-plate actuator will pull in, or snap down to the electrode.

Parallel-plate actuators have the advantage of being simple to fabricate. They require only an electrode (typically metal) and the conductive actuator. Many applications of electrostatic actuators have been demonstrated including microgrippers, microrelays, micromotors, cantilevers, optical shutters, variable optical attenuators (VOAs), and micromirrors (Figure 2).

2.2 Comb drive

Comb drives were developed to reduce the drive voltage, improve drive linearity to voltage and eliminate the snap-down effect in plate actuators.

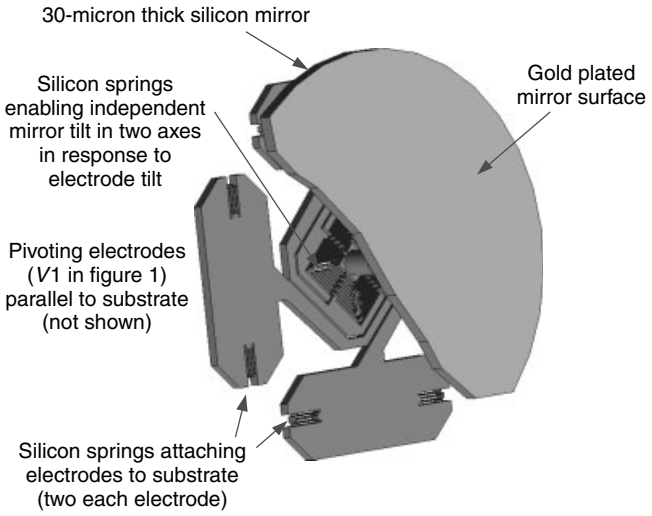


Figure 2. Integrated micromirror with parallel-plate electrostatic actuator. (*Source*: Transparent Networks, Milpitas, CA.)

The force F_x generated by applying a voltage across a comb structure can be expressed as

$$F_x = \frac{\varepsilon V^2 h}{2g} \tag{2}$$

where

- ε is the dielectric constant for the substance between the electrodes pairs,
- h is the height of the comb,
- g is the gap between fingers, and
- V is the applied voltage.

Unlike the parallel-plate capacitor, the gap between comb fingers does not change as the combs slide together. The force generated is solely a function of the fabricated geometry and the applied voltage, not the position of the combs. This simplifies the actuator response and control.

Interdigitated finger or comb drives were first introduced as lateral comb drives in surface micromachined polysilicon. The comb drive pictured in Figure 3 was produced at Sandia National Labs and is now available through the standard SUMMiT™ V process. Comb drives have also been made in bulk silicon by deep reactive ion etching (DRIE).

A second variation on the comb drive design is vertically asymmetric combs where differences in height and lateral restriction cause the comb structures to move vertically rather than laterally. This is less common because it is much more difficult to fabricate.

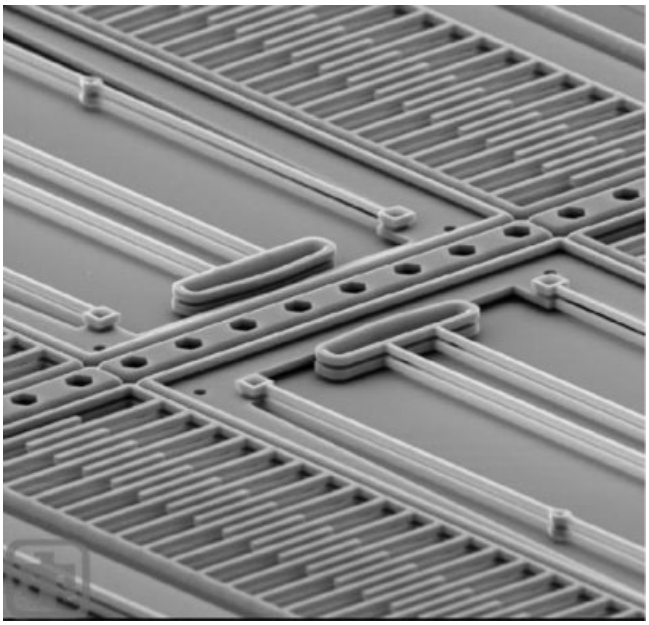


Figure 3. Shown is the lateral comb drives of a three-layer polysilicon microengine. The springs in the center provide the restoring force. (Courtesy Sandia National Laboratories, SUMMiT™ Technologies, www.mems.sandia.gov.)

3 ELECTROMAGNETIC ACTUATION

In the electromagnetic form of actuator, an electromagnetic field generates force that causes motion. This is done by passing current through a conductor to generate a magnetic field.

These can be formed in many styles in traditional designs but they are more challenging to implement in the MEMS form, as very few MEMS foundries offers a suitable fabrication process. A typical implementation involves a blanket magnetic field across an entire MEMS device with a patterned loop-generating field locally. Unlike electrostatic devices, electromagnetic devices typically require a constant power dissipation to hold a position.

For a current carrying coil in a uniform magnetic field, two important aspects of the electromagnetic form of actuator exist.

The first is that the torque of a rotational form is linear with input control coil current and decreases as the dipole from the coil and the magnetic field move into alignment (Figure 4). Consequently, unlike the capacitor form, there is no instability point in controlling the rotation of a magnetic actuator.

There is, however, zero net force in a homogeneous magnetic field. Producing a net force on the actuator requires more complex engineering of both the mechanics of the actuator and the gradients of the field.

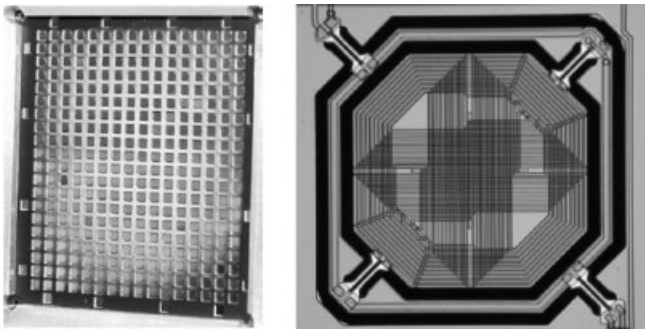


Figure 4. 320-mirror array and patterned control coils under each mirror. (Copyright Corning.)

4 PIEZOELECTRIC ACTUATION

Piezoelectric actuators rely on the piezoelectric transducer (PZT) effect whereby the application of an electric field causes a material to exert a force along the axis of the electric field (Figure 5).

Displacements obtained from piezoelectric materials are typically very small. Changes in dimension range from 10^{-10} to 10^{-7} cm/V and several hundred volts are usually required to achieve adequate displacement. They are often used as part of bimorphs or other leveraging structures to increase the effective travel.

The force generated by applying a voltage across a piezoelectric material can be derived from

$$F = \frac{V\varepsilon_0\varepsilon_r A}{\ell d} \tag{3}$$

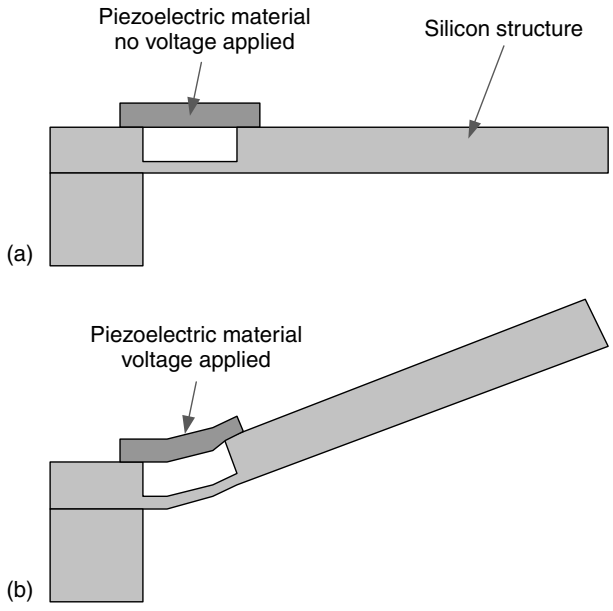


Figure 5. Principle of piezoelectric actuator.

where V is the applied voltage,

ε_0 is the permittivity of free space,
ε_r is relative permittivity for the substance between the electrodes,
A is the area of the electrodes formed on the PZT material,
ℓ is the length of the material (along the axis of the field), and
d is the piezoelectric constant.

The piezoelectric constant d, and hence the amount of force generated, is dependent upon the crystalline axis of the material. To give an idea of the size of parameters involved, for the commonly used lead zirconate titanate PZT material, d in one crystal direction of cut is 370 pC/N. For more detail on PZT materials and their application, *see* **Article 118, Ultrasonic Sensors, Volume 2**; and **Article 119, Ultrasonic Instrumentation Principles, Volume 2**.

One application where piezoelectric microactuators have been used extensively is for precise positioning of hard disk drive heads (Evans, Griesbach and Messner, 1999).

While the small displacement makes them difficult to use for some applications, the precision achievable is ideally suited for positioning within tracks, especially as hard disk drives reach for submicron track width. Typically, piezoelectric actuators are the second stage of a multistage head where the gross positioning is done with a voice coil actuator. The control requirements for this application are very demanding and extensive research has been done in this area (Oboe, Beghi and Murari, 1999).

One of the newest areas for piezoelectric actuators is energy scavenging for wireless sensors. With a proper design, vibration can be on-chip converted into voltage with a power on the level of single μ W to mW, depending on the chip size.

5 THERMAL ACTUATION

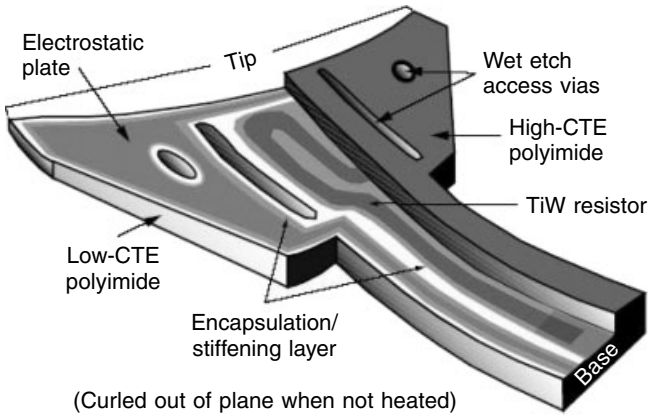
Thermal actuators take advantage of material expansion in a material due to heating, generating the force for actuation. While it would be possible to use this simply, a more effective use of thermal actuation involves pairing two materials of dissimilar coefficient of thermal expansion (CTE) into a thermal bimorph. As the structure is heated, the CTE mismatch causes strain between the layers, which if unconstrained along an axis, will cause motion. The force exerted by a temperature change in a material can be quite considerable for the small displacements generated.

An example of this implementation is the ciliary array produced by Suh *et al*. (1999) (Figure 6). A TiW heater was sandwiched between two layers of polyimide selected to have severe CTE's mismatch (50 and 2 ppm/C respectively). The chip achieved a 95-μm vertical displacement, generating 23 μN of force. In addition to the thermal actuator, each cilia element had a plate that was used for an electrostatic hold down.

Thermal actuators have the disadvantage of high static power dissipation. They also typically exhibit relatively slow time constants.

6 OPTICAL ACTUATION

Optical actuation is a relatively new approach. One of the first approaches was presented by Zook *et al*. (1995), wherein optical energy was used to power the resonant beam with a very high Q (Figure 7). The beam located on a diaphragm was stressed by pressure, generating a high accuracy conversion of pressure to frequency.



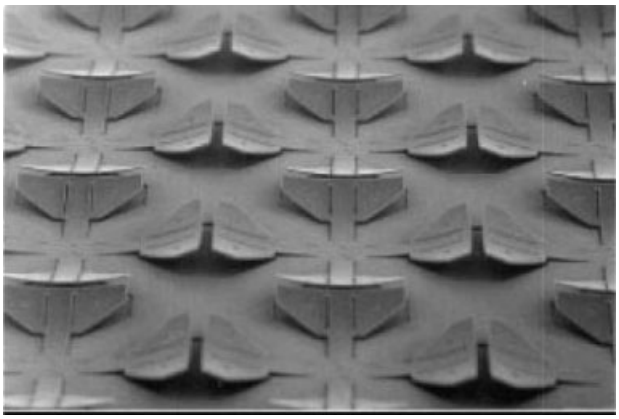


Figure 6. Thermally actuated ciliary array. (Reproduced from *Journal of Microelectromechanical Systems* Vol. 8(4), pp. 483–496, CMOS Integrated Array as a General-purpose Micromanipulation Tool for Small Objects, Suh J.W. *et al*., © (1999) IEEE.)

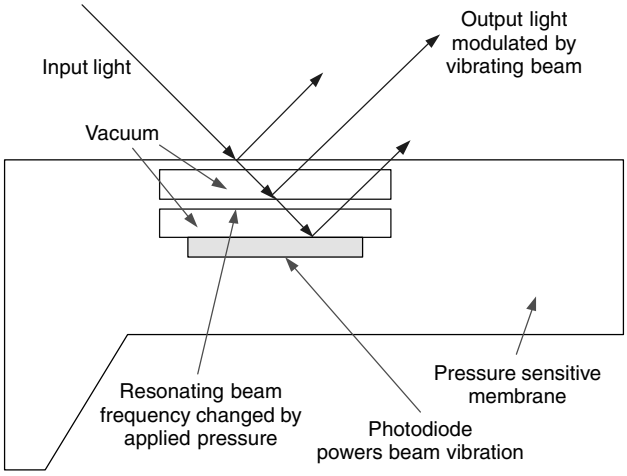


Figure 7. Light actuated and read resonant beam pressure sensor. (Reproduced from *Conference Proceedings: International Solid-State Sensors and Actuators Conference – TRANSDUCERS '95*, June 25–29, Vol. 2, Stockholm (pp. 600–603), Light Actuated & Read Resonant Beam Pressure Sensor, Burns J.D. *et al*., © (1995) IEEE.)

RELATED ARTICLES

Article 113, Force-feedback Sensors, Volume 2; **Article 118, Ultrasonic Sensors, Volume 2**; **Article 119, Ultrasonic Instrumentation Principles, Volume 2**; **Article 162, Principles of MEMS, Volume 3**; **Article 163, Uses and Benefits of MEMS, Volume 3**.

REFERENCES

Evans, R.B., Griesbach, J.S. and Messner, W.C. (1999) Piezoelectric Microactuator for dual stage control. *IEEE Transactions on Magnetics*, **35**(2), 977–982.

Oboe, R., Beghi, A. and Murari, B. (1999) Modeling and Control of a Dual Stage Actuator Hard Disk Drive with Piezoelectric Secondary Actuator, in *Proceedings of the 1999 IEEE/ASME, International Conference on Advanced Mechatronics*, September 19–23, Atlanta, GA (pp. 138–143).

Suh, J.W., Darling, R.B., Bohringer, K.-F., Donald, B.R., Baltes, H. and Kovacs, G.T.A. (1999) CMOS Integrated Ciliary Actuator Array as a General-purpose Micromanipulation Tool for Small Objects, *Journal of Microelectromechanical Systems*, **8**(4), 483–496.

Zook, J.D., Burns, D.W., Herb, W.R., Guckel, H., Kang, J.-W. and Ahn, Y. (1995) Light Actuated and Read Resonant Beam Pressure Sensor, in *Conference Proceedings: International Solid-State Sensors and Actuators Conference – TRANSDUCERS '95*, June 25–29, Vol. 2, Stockholm (pp. 600–603).

165: Introduction to Networked Instrumentation

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Computer-based Instrumentation 1113
2 Data Communication Standards 1113
3 Data Communication Equipment 1115
4 Data Communication Methods 1115
5 Communication at Close Proximity 1116
6 Intermediate Distance Communication 1117
7 Large Distance Communications 1117
Related Articles 1117
References 1117
Further Reading 1118

1 COMPUTER-BASED INSTRUMENTATION

Instrumentation using machine information processing improves the versatility of the equipment so that it enhances the abilities of humans to perform information-handling operations.

Although all instrumentation involves some processing of information by machine, intelligent instrumentation involves some machine functions beyond basic information processing. Hence, processing operations such as interpretation of complex data, reasoning, learning, inference, or any operation associated with Artificial Intelligence or Knowledge Expertise must be included before instruments may be termed *intelligent – see also* **Article 160, Smart Sensor System Features, Volume 3**.

Computerized measurement systems (Lang, 1987; Lang, 1991) do not differ essentially from other measuring instruments. Using special instrument interfaces for communicating with a computer is the main difference (Iotech Inc., 1991; Maine, 1986; Senbon and Hanabuchi, 1987). The task of the computer is to schedule and command the instruments and then to perform any further processing of the measurement data. As a computer has a large processing capacity, several instruments can be connected to the computer simultaneously using different connection structures (*see* **Article 166, Instrument Interconnection, Volume 3**). In the same way as inside a computer or microprocessor, a bus structure is used to reduce the wiring and the number of interface circuits. These buses also use asynchronous or synchronous transfer protocols (*see* **Article 167, Asynchronous and Synchronous Interface Protocols, Volume 3**). The instruments are all connected to the bus in parallel. To transport the data in an orderly fashion, each instrument is equipped with a more or less intelligent interface.

Communication of data is an essential enabling component of all measurement systems. It provides the essential core of technology, which permits both local and remote distribution of information. It is only possible to give a general background of this topic, as it is relevant to measurement systems.

2 DATA COMMUNICATION STANDARDS

In the application of computers to measurement, the measurement system must quickly convert input data to processed output results. Hence, a high priority is placed upon real-time processing. Progress in microprocessors, as well as the technologies applied to them, has advanced the intelligence and decentralization of measurement systems. These advances are responsible for the increased importance in measurement of data communication technologies (Duck, Bishop and Read, 1996; Marshall, 1980)

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

that allows high-fidelity communication of the required data in real time.

Data communication can be described as the linking of equipment such as measuring instruments and control devices by means of communication lines, through which digital data is mutually transferred with good efficiency and with fidelity according to certain predefined procedures. To send correct data to the destination system, the communication circuits in both the data source and the data sink systems require the same methods and specifications. These include the types of communication cables and connectors, the electrical characteristics of the signal, the signal delivery procedure and the method of expressing the data. Simplification of the overall aims of the interconnection is achieved by the following standardized methods and specifications for hardware, software, signal definitions and communication protocols.

One of the most frequently applied approaches is the generalized set of standards known as the *open systems interconnection*, or OSI, basic reference model (Iotech Inc., 1991; Duck, Bishop and Read, 1996) created by the International Organization for Standardization, or ISO. As shown in Table 1, the OSI model classifies all the functions that are indispensable for data communications into seven layers, ranging from the physically implemented hardware layer, which includes the function of converting data to an electrical signal for the transmission line, to the software implemented application layer. This layer includes application-oriented functions such as methods of expressing messages and file transfer procedures. In addition, the functions of each layer are defined and serve as guidelines for standardization. Nevertheless, the boundaries between these layers are sometimes not clear. A system that implements all of these layers is called a *high-level interface*. Many other interface definitions omit many layers.

Table 1. OSI reference model.

Layers	Label	Function
Layer 7	Application	Provides all services directly comprehensible to application programs
Layer 6	Presentation	Restructures data to/from standardized format used within the network
Layer 5	Session	Synchronizes and manages data
Layer 4	Transport	Provides transparent, reliable data transfer from end node to end node
Layer 3	Network	Performs packet routing for data transfer between nodes
Layer 2	Data link	Improves error rate for frames moved between adjacent nodes
Layer 1	Physical	Encodes and physically transfers bits between adjacent nodes
Physical medium		Transmission line

There are several defined standards relating to both serial and parallel transmission. Serial signal transmission, which can be used for much more distant communication than parallel transmission, is very commonly applied in telemetry systems. Factors to be taken into account are

1. data rate
2. transmission distance
3. immunity to interference and noise and
4. the number of multiplexed signals over one link.

Well-known serial communication specifications are

1. EIA/TIA 232 E, also known as RS 232, RS 449, RS 422 and RS 423 (Axelson, 1996; Campbell, 1989; Iotech Inc., 1991; Lang, 1987; Maine, 1986; Seyer, 1991), *see* **Article 168, RS 232 and EIA/TIA 232 Serial Interface, Volume 3**;
2. 4–20 mA current loop (Jordan, 1995; Lang, 1987; Michalski *et al*., 2001), *see* **Article 169, Voltage and Current Loop Transmission, Volume 3**;
3. CCITT Consultative Council on Information Technology Transfer (Europe) V24;
4. Fieldbus (Burton, 1995; ISA, 1997; Verhappen and Pereira 2002) and its automation version PROFIBUS (Bender, 1993).

Equally well-known parallel communication specifications are

1. IEEE-488 (Caristi, 1989; Colloms, 1983; IEEE, 1987; Iotech Inc., 1991; Lang, 1987; Senbon and Hanabuchi, 1991)
2. VMEbus/VXIbus (Heath, 1994)
3. SCSI, IDE and ATAPI. (Schmidt, 1998).

The RS 232 serial communication standard is considered in **Article 168, RS 232 and EIA/TIA 232 Serial Interface, Volume 3**. The IEEE-488 standard is described in **Article 170, IEEE-488 Instrumentation Bus, Volume 3**. Voltage and current loop transmission is described in **Article 169, Voltage and Current Loop Transmission, Volume 3**.

On the other hand, there are also many widely used communication networks and protocols, which do not entirely conform to international standards. One such example is the Systems Network Architecture, or SNA, which has been the

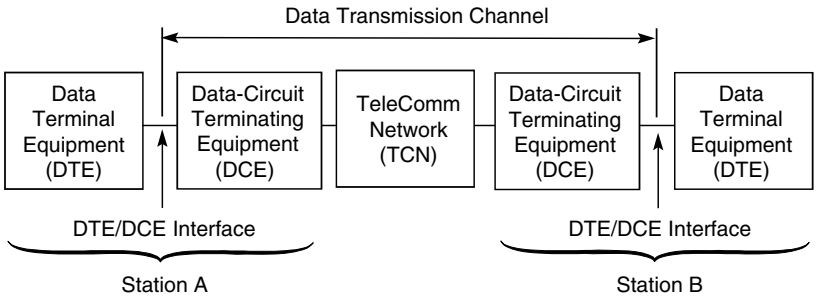


Figure 1. A data transmission channel showing the various subsystem blocks.

IBM Corporation standard network since 1974. It is also used worldwide.

There are many examples of practical optical communications systems. This is another area in which standardization is proceeding slowly. In any case, when the use of nonstandard protocols is unavoidable, it is necessary to make detailed preparations and carefully design for stability and reliability.

3 DATA COMMUNICATION EQUIPMENT

Figure 1 illustrates the various parts of a channel for data transmission. There are three main elements of such a channel. Data terminal equipment, or DTE, includes terminals or computers. Data circuit terminating equipment, or DCE, interfaces between the DTE and the telecommunications network, or TCN. DCE includes modems, packet switched networks, connection pads, and statistical multiplexers among others. The TCN may be circuit switched, packet switched, or private leased line. Circuit switched networks route connections through fixed paths selected by switches. In packet switched networks, each data message is subdivided into a number of separate blocks, or packets, of data.

Some terminology associated with data communications requires explanation. When data can only be transmitted or received, the exchange of data is called *simplex* operation. *Half duplex* is an extension of the simplex method, which allows both transmission and reception but not simultaneously. When the data transmission channel can transmit and receive simultaneously, it is said to be operating in *full duplex*.

The *modem* (Duck, Bishop and Read, 1996) is a hardware unit that can modulate and demodulate signals. It is often used as a simple and convenient way to make a fixed connection between two, probably remote, devices. Introducing a modem between devices equipped with the RS 232 interface, described in **Article 168, RS 232 and EIA/TIA 232 Serial Interface, Volume 3**, allows extension of the transmission distance. There are different types of modems. The telephone protocol of the modem is governed by the following special signals:

1. REQUEST TO SEND, or RTS, and CLEAR TO SEND, or CTS, are used in *half-duplex* telephone lines. In this operating mode, traffic can be bidirectional, but the telephone line can only send in one direction at a time. When the computer is ready to transmit a character, it signals the modem with RTS. Transmission will be achieved only when the computer receives a CTS signal coming from the modem.
2. The CTS and RTS lines are held at a constant voltage in *full-duplex operation.* In this mode, transmit and receive functions can be performed simultaneously on independent wires.
3. DATA SET READY, or DSR, indicates that the modem is operational.
4. DATA TERMINAL READY, or DTR, indicates that the equipment on the opposite end, such as a computer or terminal, is operational.
5. CARRIER DETECT and RING INDICATOR are related to the telephone network functions. CARRIER DETECT indicates that the remote connection is normally active.

4 DATA COMMUNICATION METHODS

Representing the many data communication methods, in terms of communication distance versus communication speed results in Figure 2 (Senbon and Hanabuchi, 1991). Although the grouping looks clearly defined in the diagram, it is much looser. Nevertheless, the following rough classes, or categories, can be distinguished:

1. communication between devices in close proximity;
2. communication between devices at intermediate distances;
3. communication between devices at large distances.

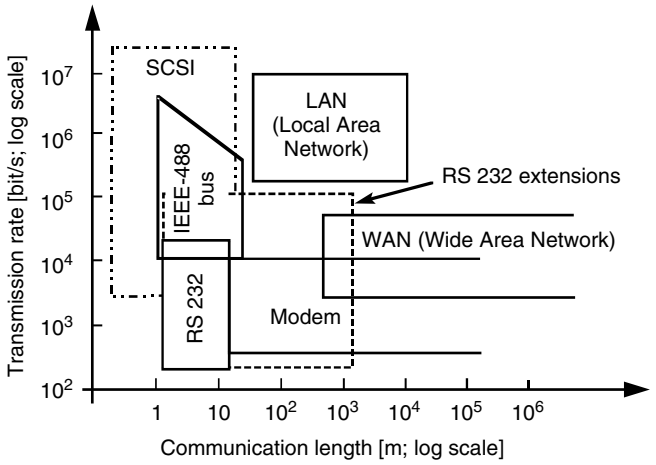


Figure 2. Performance ranges of different instrument interfaces.

It is not possible to give precisely defined divisions between the three classes given above. However, it can be seen that the SCSI, RS 232, and IEE-488 interfaces can be regarded as in close proximity over distances up to about 50 m. Distances from 50 m up to about 10 km, which normally use local area networks (Anderson and Minasi, 1999; Hodson, 2001; Thompson, 2000) and extensions to RS 232, can be considered as intermediate distance communication systems. Wide area networks, whose lower distance range lies within the upper limit of the intermediate distance range, are generally considered as most suitable for distances between about 800 m and more than 1000 km.

5 COMMUNICATION AT CLOSE PROXIMITY

Within testing laboratories and other closely integrated locations, automatic measurement and data processing is often carried out with various measuring instruments connected to a computer. The IEEE-488 parallel instrument bus, specified by the Institute of Electrical and Electronics Engineers (USA), is widely used for such purposes. It is compatible with the specification for the IEC-625 bus, specified by the International Electrotechnical Commission, or IEC. IEEE-488 allows easy reconfiguration of the system for various experiments because instruments can be easily connected to and disconnected from the system. The Computer Automated Measurement and Control Equipment, or CAMAC standard, which is used for roughly the same purposes as IEEE-488 and IEC-625, has been established in the atomic power field.

Another possibility is to use the VERSAModule Eurocard, or VMEbus, which was originally specified by Motorola as the VERSAbus. Later, extensions to the VMEbus were incorporated in the 1987 specification for VXIbus, which is an acronym for "VMEbus eXtensions for Instrumentation". Although VXIbus expands on VMEbus, it has an open architecture, including features of IEEE-488. It is useful for automated test systems and data collection. The issue of electromagnetic radiation is part of the VXIbus specification.

Small computer system interface, or *SCSI*: This interface contains specific design features to permit communication between small computer systems and their peripherals. It was first standardized by the American National Standards Institute, or ANSI, in 1986, although it was originally developed by Shugart as Shugart Associates System Interface, or SASI. It has now reached the third revised version.

All IBM compatible personal computers possess what is commonly called the *AT-Attachment*, or ATA interface. It is also widely known as the *Integrated Disk Electronics*, or IDE interface.

Its development, which began in 1984 with the support of Compaq, grew out of the desire to embed the hard disk controller of IBM AT compatibles on the hard disk itself. By 1988, ANSI began its involvement, which led to the approval of a standard. This resulted in the IDE interface being renamed as the ATA in 1994. Subsequent developments resulted in ATA-2 in 1995. The ATA Packet Interface, ATAPI standard, which was developed to allow components other than hard disks to operate under the IDE interface, was developed in parallel with ATA-2. Although it will be well known to many users of CD-ROM technology, it may also be used for other peripherals. It is a combination of SCSI and IDE. Lately, ATA-3 has evolved to offer new commands as part of the ATA interface. It is intended that ATA-4 and ATAPI will be integrated.

SCSI, IDE, and ATAPI will not be considered further.

EIA/TIA 232 E, or *RS232*: This very widely used serial interface, which was originally supervised by the Electronic Industries Association, or EIA, later was supervised jointly by EIA and Telecommunication Industries Association, or TIA, in the United States. Since 1997, EIA changed its name to Electronic Industries Alliance, an alliance of five trade associations representing the most dynamic and fastest-growing industry in the world. It is commonly used for connection of computers to terminal equipment such as CRT (cathode ray tube) terminals and printers, connection of various types of measurement instruments and automation machinery such as robots and numerically controlled or NC, machines. To overcome the limitations of the RS 232, it is now common practice to use the RS 422, RS 423 and RS 485 standards as an extension of the RS 232 for longer distance communication at higher data rates. These

RS 232 extensions are described in **Article 168, RS 232 and EIA/TIA 232 Serial Interface, Volume 3**.

6 INTERMEDIATE DISTANCE COMMUNICATION

In the field of measurement and control, local area networks (Anderson and Minasi, 1999; Duck, Bishop and Read, 1996; Hodson, 2001; Thompson, 2000), or LANs, are the most important method for data exchange. PROWAY, or Process Data-way, is a widely used standard, specified by the IEC. It is a means of highly reliable, high-speed data exchange between the operator console and multiple dispersed field measurement and control units. In a distributed control system, these may be computers, control devices, terminals and other distributed equipment at distances up to 10 km.

Recently, the demand for total network integration of production information systems, at both the line and corporate enterprise levels, has become high. For this reason, the Manufacturing Automation Protocol, or MAP, which is a factory LAN, is being used increasingly by many enterprises in the United States, Europe, and Japan. MAP, which also conforms to OSI, is under the primary leadership of the General Motors Corporation. MAP and PROWAY are both receiving wide attention as a means of integrating dispersed equipment in various environments such as offices, research centers, and factories.

7 LARGE DISTANCE COMMUNICATIONS

When instruments are to be connected over large distances through commercial communication systems, the network is frequently referred to as a *wide area network* or WAN (Duck, Bishop and Read, 1996). Data is transmitted over the communication lines in the form of data blocks called *packets*. Information on the destination and packet sequence number are included in the packet. Thus, the communication circuit utilization efficiency is very high because many users can share a circuit. This method can be used for on-line transmission of orders, production plans, and other information from the head office to distant factories.

RELATED ARTICLES

Article 78, Embedded Systems, Volume 2; **Article 155, External Operation of the Microprocessor, Volume 3**; **Article 158, Microcontroller Systems, Volume 3**; **Article 168, RS 232 and EIA/TIA 232 Serial Interface, Volume 3**; **Article 170, IEEE-488 Instrumentation Bus, Volume 3**; **Article 171, Local Area (LANs) and Wide Area Networks (WANs), Volume 3**.

REFERENCES

Anderson, C. and Minasi, M. (1999) *Mastering Local Area Networks*, Sybex International.

Axelson, J.L. (1996) *Serial Port Complete: Programming and Circuits for RS-232 and RS-485 Links and Networks*, Lakeview Research, (Digital Download).

Bender, K. (1993) *PROFIBUS: The Fieldbus for Industrial Automation*, Prentice Hall.

Burton, P. (1995) *Fieldbus for Industrial Control Systems*, Kluwer Academic Publishers.

Campbell, J. (1989) *The RS-232 Solution*, Sybex International.

Caristi, A.J. (1989) *IEEE-488 General Purpose Instrumentation Bus Manual*, Academic Press.

Colloms, M. (1983) *Computer Controlled Testing and Instrumentation: An Introduction to the IEC-625 : IEEE-488 Bus*, John Wiley & Sons.

Duck, M., Bishop, P. and Read, R. (1996) *Data Communications for Engineers*, Addison-Wesley, Harlow, UK.

Heath, S. (1994) *VMEbus: A Practical Guide*, Butterworth-Heinemann.

Hodson, P. (2001) *Local Area Networks*, Continuum International Publishing Group.

IEEE (1987) *IEEE Standard Codes, Formats, Protocols, and Common Commands for Use with IEEE Std 488.1-1987*, IEEE Standard Digital Interface for Programmable Instrumentation, IEEE Publications.

Iotech Inc. (1991) *Instrument Communications Handbook*.

ISA (1997) *Standards and Practice for Instrumentation: Fieldbus Standard for Use in Industrial Control Systems*, Instrument Society of America.

Jordan, J.R. (1995) *Serial Networked Field Instrumentation*, John Wiley & Sons, Chichester, UK.

Lang, T.T. (1987) *Electronics of Measuring Systems*, John Wiley & Sons, Chichester, UK.

Lang, T.T. (1991) *Computerised Instrumentation*, John Wiley & Sons, Chichester, UK.

Maine, A.C. (1986) *Interfacing Standards for Computers*, IEETE, London, UK.

Marshall, G.J. (1980) *Principles of Digital Communications*, McGraw-Hill, Maidenhead, UK.

Michalski, L., Eckersdorf, K., Kucharski, J. and McGhee, J. (2001) *Temperature Measurement*, John Wiley & Sons, Chichester, UK.

Schmidt, F. (1998) *The SCSI Bus and IDE Interface*, Addison-Wesley, Harlow, UK.

Senbon, T. and Hanabuchi, F. (eds) (1991) *Instrumentation Systems: Fundamentals and Applications*, Springer-Verlag, Berlin, Germany.

Seyer, M.D. (1991) *RS 232 Made Easy: Connecting Computers, Printers, Terminals and Modems*, Prentice Hall.

Thompson, A. (2000) *Understanding Local Area Networks: A Practical Approach*, Prentice Hall, Englewood Cliffs, NJ.

Verhappen, I. and Pereira, A. (2002) *Foundation Fieldbus: A Pocket Guide*, American Society of Civil Engineers (ASCE).

FURTHER READING

ISA (1997) *The ISA Fieldbus Guide*, Instrument Society of America, Research Triangle Park, NC.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

166: Instrument Interconnection

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 General Issues of Instrument Interconnection 1119
2 Parallel and Serial Transfer 1119
3 Bus and Loop Configured Instrument Interfaces 1120
Related Articles 1120
Reference 1120

1 GENERAL ISSUES OF INSTRUMENT INTERCONNECTION

Main forms of interfacing allow serial or parallel information distribution. A typical minimum configuration set for computerized instrumentation is shown in Figure 1. These systems consist of components, boards, and peripherals. Components within the computer have direct access to the processor bus of the system. Extension boards, which are located within the computer housing, have access to the back plane or system bus of the computer. Bus extension means that an external box may also be connected to the PC by only one slot. Bus length is usually limited to a maximum of 2 m.

Versatility is achieved by three main kinds of communicating boards. General Purpose Input/Output (GPIO) boards acquire signals of diverse types such as status, alarm, control, analog, and set points. The IEC bus (International Electrotechnical Commission), also known as IEEE-488 or General Purpose Interface Bus (GPIB), is a peripheral bus system. It allows up to 15 devices to operate together at distances not exceeding about 20 m. It is highly suitable for laboratory automation. Serial connections allow operation over a wider distance. Hence, it is suitable for single buildings or factory floors. It is also versatile in data communications for Local Area Networks, or LANs, and Wide Area Networks, or WANs. Front-end subsystems are flexible, versatile, and economical for measurement systems.

2 PARALLEL AND SERIAL TRANSFER

Instruments may be connected for either parallel transfer of data or serial transfer of data (Maine, 1986). These two possibilities are illustrated in Figure 2. Serial data is passed between devices by using either asynchronous data transfer method or the synchronous data transfer method – *see*

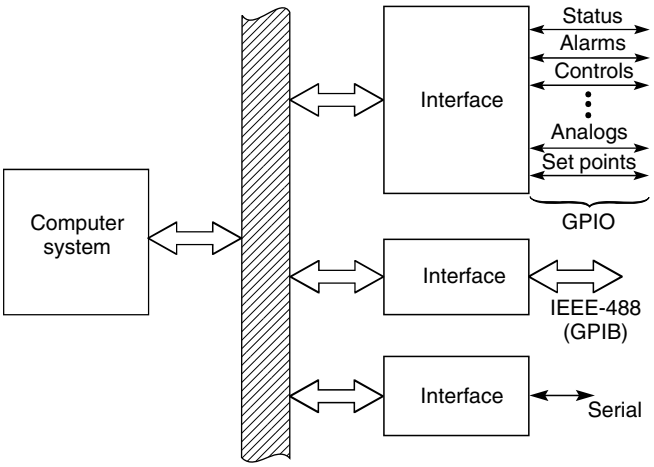


Figure 1. Microcomputer with a minimum equipment of interfaces for intelligent measurements. (GPIO: General Purpose Input/Output; GPIB: General Purpose Interface Bus, identical with IEEE-488).

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

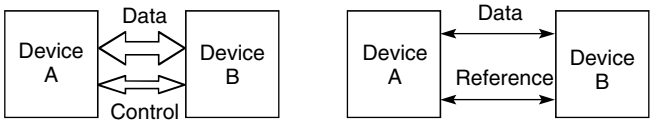


Figure 2. Parallel and serial instrument interconnection.

Article 167, Asynchronous and Synchronous Interface Protocols, Volume 3.

3 BUS AND LOOP CONFIGURED INSTRUMENT INTERFACES

When more than two instruments are interconnected, the configuration may be either a bus connection or a loop connection as in Figure 3. The data transfer is usually bit parallel byte serial in the bus system, where the limiting factor is the signal path length. Each device has equal access to the data. In the loop structure, where the data transfer is normally serial, the limiting factor is the loop transmit time. Information passes from device to device. The system is low power, with a higher reliability, as there are normally only two wires.

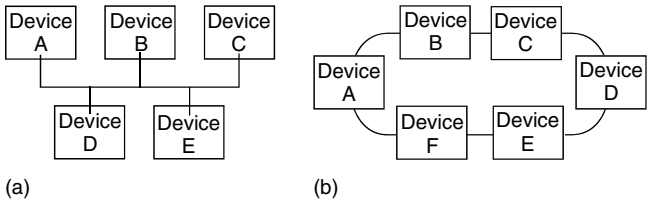


Figure 3. Instrument configuration: (a) bus structure and (b) loop structure.

Some systems employ analog voltage or current loop systems for communication. These have been in use for decades and are still good choices in some circumstances. The life of these is being extended by gradual take-up of digital communication methods. They are covered in **Article 169, Voltage and Current Loop Transmission, Volume 3**.

Designers of communicating links usually simply specify proprietary circuits boards for this. In use, they need setting with some hardware toggles and within the operation software.

RELATED ARTICLES

Article 78, Embedded Systems, Volume 2; **Article 158, Microcontroller Systems, Volume 3**; **Article 165, Introduction to Networked Instrumentation, Volume 3**; **Article 167, Asynchronous and Synchronous Interface Protocols, Volume 3**; **Article 170, IEEE-488 Instrumentation Bus, Volume 3**.

REFERENCE

Maine, A.C. (1986) *Interfacing Standards for Computers*, IEETE, London, UK.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

167: Asynchronous and Synchronous Interface Protocols

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Asynchronous Protocols 1121
2 American Standard Code for Information Interchange (ASCII) 1121
3 Synchronous Interface Protocols 1122
4 Binary Synchronous Communications Protocol 1123
5 High-Level Data Link Control Protocol 1123
Related Articles 1123
References 1123

1 ASYNCHRONOUS PROTOCOLS

A protocol is a mutually agreed arrangement for initiating and maintaining a connection in a communication network (Duck, Bishop and Read, 1996; Marshall, 1980). There are asynchronous protocols or synchronous protocols. The asynchronous protocol is mainly used in serial communication systems. Three hardware connections commonly used in serial communication (**Article 165, Introduction to Networked Instrumentation, Volume 3**) are

- EIA/TIA 232 E, previously known as RS 232 and still widely referred to as such, with its extension possibilities embodied in the RS 422 A, the RS 423 A, and the RS 449 – *see* **Article 168, RS 232 and EIA/TIA 232 Serial Interface, Volume 3**;
- 4–20 mA current loop – *see* **Article 169, Voltage and Current Loop Transmission, Volume 3**;
- CCITT V24, which is more common in Europe;
- Fieldbus.

The Recommended Standard, or RS, which refers to the published standards of the Electronic Industries Alliance, or EIA, of the United States, can use asynchronous signals with the general timing for the character structure illustrated in Figure 1(a). In this figure, the idle state is assumed to be in a high state or '1'. Data is handled in units of one character, which may have 6, 7, or 8 bits.

A character is delimited with an appended starting '0' level bit, followed by 8 data bits, then followed by one and a half or two successive '1' bits. Note that data are always transmitted with the least significant bit first. One of the 8 data bits is a simple error-checking bit, termed the *parity bit*. It is used to make the total number of '1' bits in the character always odd or even, depending on whether odd parity or even parity is required. The bit interval is defined with a fixed period controlled by a local clock in the transmitter and receiver.

The receiving device recognizes the beginning of the character by detecting the start bit. From that time it reads each bit making up the character at fixed time intervals, determined by the transmission speed.

2 AMERICAN STANDARD CODE FOR INFORMATION INTERCHANGE (ASCII)

The seven character bits in Figure 1 are assigned according to the American Standard Code for Information Interchange, or ASCII system, which is also known as *American National Standards Institute*, or ANSI, standard X3.4-1977.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

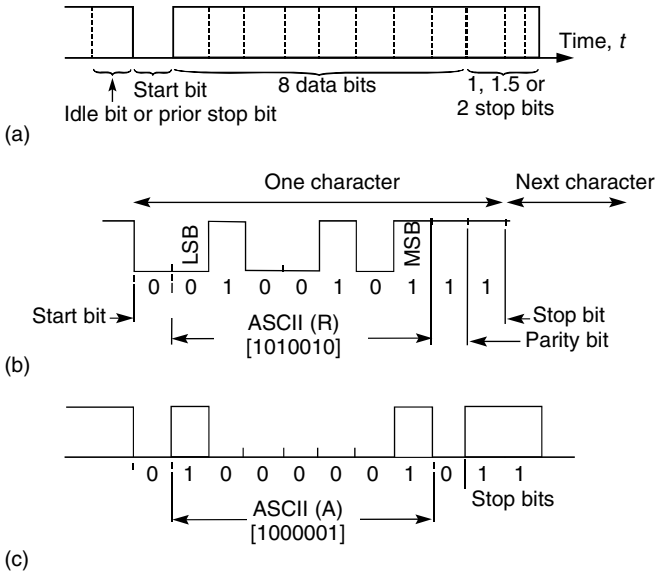


Figure 1. Asynchronous data: (a) character structure showing location of 'idle', start, data, and stop bits; (b) representation of letter 'R' in ASCII; and (c) representation of letter 'A' in ASCII.

This code, which uses 7 bits/characters, covers a range of 128 characters. Thus, the code represents all of the upper and lower case letters, the numerals 0 to 9, punctuation marks, and 32 nonprintable, or control, characters. Another encoding scheme applies to information other than text. For example, the IBM PC extends the ASCII character set from 128 to 256 characters using the binary encoding scheme. This allows representation of non-English alphabets and special graphic characters.

The ASCII code for the letters R and A are shown for illustrative purposes in Figure 1(b) and (c) respectively. Note that the time patterns are written with the least significant bit, or LSB, first in both cases. The *start* and *stop* bits, which identify the beginning and the end of each data character, also allow the receiver to synchronize a local clock to every character. Signaling frequencies can be 300 Hz, 1200 Hz up to 19.2 kHz.

3 SYNCHRONOUS INTERFACE PROTOCOLS

Synchronous protocols, which are intended to allow communication at very high speed, are characterized by the presence of a synchronous clock designed to control character-to-character timing within a stream of characters. Thus, they present numerous advantages over the asynchronous standards. As start and stop bits are eliminated, the bandwidth for data transmission is widened. Data are made up of a number of characters into units of character

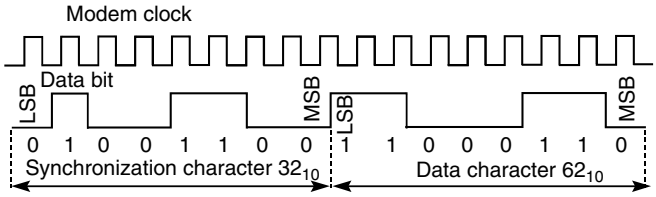


Figure 2. Synchronous data character structure.

strings. As shown in Figure 2, no start bits or stop bits are added as is done when the transfer unit is one character. Bit position is detected by synchronization with a special clock signal. Characters are distinguished by detecting a special synchronization character at the head or in the middle of the character string. Bits are then counted as the data is received.

Communications can be achieved at higher bit rates over longer distance than in the asynchronous protocol described above. Asynchronous links run at a maximum speed of 19.2 kHz, while a synchronous link can easily run at about 500 kHz or higher. However, proper electrical connections are required to attain very high data rates. For instance, the use of a double-ended transmitter/receiver such as an RS 422 interface is recommended.

Block-oriented synchronous protocols allow the inclusion of control and redundancy information in the communication link. The receiver locks onto the frequency of the received data through a phase-locked loop located at the input of the synchronous I/O port. This extracts the received clock from the received data stream. The receiver is synchronized to the beginning of a character by an additional logic, which may be different from protocol to protocol.

After character synchronization, the receiver then divides data bits into 8-bit characters, with character-to-character synchronization established in the same time with block-to-block synchronization. The interface consists of shift registers and buffer registers to achieve conversion between parallel and serial data. Typical synchronous interfaces use three registers in a queue known as a *first-in-first-out*, or FIFO, structure.

The three widely used protocols are the Binary Synchronous Communications Protocol, or BSC, the Digital Data Communication Message Protocol, or DDCMP, and the High-Level Data Link Control, or HDLC (Duck, Bishop and Read, 1996).

Only the BSC protocol (Senbon and Hanabuchi, 1991) and the HDLC protocol (Duck, Bishop and Read, 1996; Senbon and Hanabuchi, 1991) will be considered here because they are easily realizable.

4 BINARY SYNCHRONOUS COMMUNICATIONS PROTOCOL

This protocol, which has been primarily used in IBM equipment since 1960, operates in half-duplex transmission. In this protocol, one or two successive synchronizing characters are used to identify character and block synchronization. Character synchronization is achieved by a special pattern providing three operating modes with respect to character synchronization. These are external synchronizing mode, one-synchronizing character mode, and two-synchronizing character mode.

5 HIGH-LEVEL DATA LINK CONTROL PROTOCOL

A synchronous transfer method called the *High-Level Data Link Control Protocol*, or HDLC, has been standardized. This protocol, which was initially proposed by IBM, operates in full duplex. HDLC protocols are frequently used for point-to-point links. It is also used in wide area networks for connecting computers, where successive throws of transmissions may be interspersed by idle periods. As their serial bandwidth is high enough, it is possible to connect several synchronous devices together on one line in order to share the bandwidth.

The method uses frames of variable bit length as the data transmission unit rather than single characters or character strings. Character synchronization is maintained by a unique pattern at the beginning of a block.

The typical structure of the HDLC protocol, which appears in Figure 3, indicates that the *flag sequence*, or synchronizing pattern, 01111110, is placed at the beginning and end of the frame. These flags are used to detect the frame and the bit positions. Address and control information follows the opening flag. In the figure, the address represents the address of the secondary station corresponding to the main station, while the control information specifies the data frame type.

Initial flag	Address	Control	Information field	Frame check	Ending flag
01111110	1 byte	1 or 2 bytes		2 bytes	01111110

Figure 3. Typical structure of the data frame of the HDLC protocol.

When the data sequence contains many successive '1's, it is important to distinguish the *data sequence* from the *synchronizing sequence*. Hence, whenever the transmitter has sent five successive '1's, it inserts a false '0'. The receiver then has to remove the '0' located after the sequence of five '1's. The address and control information, consisting of 8 or 16 bits, generally contains a source and destination address and a sequence number.

After the address and control information, the next block can use the ending flag as its initial flag. The information block may have variable length. It is followed by a 2 bytes check, which has to determine if any errors have occurred in the previous transmission. When the transmitter runs out of data, an abort code, consisting of a sequence of eight '1's, is used.

When the transmission line is long, or in factory environments where there is much electrical interference, noise errors may be generated in the transmitted data. Thus, a means of detecting and correcting these errors is needed to allow transmission of correct data. In asynchronous transfer, a parity bit is added to each character unit of data. In the synchronous method, a check code that has a high error-detection power is appended to the end of each data unit. Using these features, the receiving side can detect whether there is an error in the data. If an error is detected, the receiving side issues a request to the transmitting side for retransmission. Coding, as this is, is covered in **Article 33, Coding Theory and its Application to Measurement, Volume 1**.

RELATED ARTICLES

Article 78, Embedded Systems, Volume 2; **Article 155, External Operation of the Microprocessor, Volume 3**; **Article 158, Microcontroller Systems, Volume 3**; **Article 165, Introduction to Networked Instrumentation, Volume 3**; **Article 166, Instrument Interconnection, Volume 3**; **Article 170, IEEE-488 Instrumentation Bus, Volume 3**.

REFERENCES

Duck, M., Bishop, P. and Read, R. (1996) *Data Communications for Engineers*, Addison-Wesley, Harlow.

Marshall, G.J. (1980) *Principles of Digital Communications*, McGraw-Hill, Maidenhead.

Senbon, T. and Hanabuchi, F. (eds) (1991) *Instrumentation Systems: Fundamentals and Applications*, Springer-Verlag, Berlin.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

168: RS 232 and EIA/TIA 232 Serial Interface

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Introduction to the Serial Data Interface 1124
2 RS 232 Interface and Modems 1125
3 Specifications in RS 232 1125
4 What RS 232 Does Not Cover 1127
5 Extension of RS 232 to RS 422, RS 423 and RS 485 1128
Related Articles 1129
Further Reading 1129

1 INTRODUCTION TO THE SERIAL DATA INTERFACE

Serial communication is the simplest, cheapest, and probably the most widespread type of communication method used between computers and other measuring instruments. Unfortunately, it is also the slowest because data are transmitted using one data line in a serial form.

This uses either an asynchronous protocol, or one of the synchronous protocols described in **Article 167, Asynchronous and Synchronous Interface Protocols, Volume 3**. Since computers handle their internal data in parallel, it is necessary to include some means of transforming any parallel data into serial form for subsequent serial transmission. It may also be necessary for any measurement instrument to reconvert the received serial data back to a parallel format.

A widely used serial data communications protocol is the *recommended standard*, or RS interfaces, which are specified by the Electronic Industries Alliance, or EIA, of the United States. The EIA was known as the *Electrical Industries Association* until 1997, when it joined with five other organizations to form the Electrical Industries Alliance.

The RS 232 standard was originally used for data communications on telephone lines with interfacing through a modem. Since the original development of this standard more than 30 years ago, the EIA has published three modifications. Initially, it provided a means for connecting remote systems together by the use of a telephone line as a transmission channel. In 1987, the EIA 232 D, also widely referred to as the RS 232 D, was issued as a revision of the 1969 specification for RS 232 C.

A little later, in 1991, the EIA collaborated with the Telecommunications Industry Association, or the TIA, in issuing a new version of the standard called EIA/TIA 232 E. Besides changing the name, some signal lines were renamed and various new ones were defined, including a shield conductor. In spite of the name change, it is still widely referred to simply as RS 232. It now broadly conforms to CCITT V24, V28, and ISO.

Nowadays, it is widely used in both terminal and computer systems. It may be used at distances up to about 15 m with data rates up to 9600 baud, or 20 000 bps. Terminals can be connected directly to computers through EIA/TIA 232 E ports. For remote connections, modems must be used through a telecommunications network similar in structure to that in Figure 1 of **Article 165, Introduction to Networked Instrumentation, Volume 3**.

Electrical conventions for RS 232, RS 422, RS 423, and RS 485 circuits, which are given in Table 1, show that the voltage levels are largely different from TTL and CMOS voltage levels. The RS 422 A, RS 423 A and RS 485 interfaces, which are extensions of the RS 232, are considered in more detail later in this article. For example, voltages in the RS 232 standard, which use negative, bipolar logic devices,

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

Table 1. Typical electric specifications of RS 232, RS 422, RS 423, and RS 485.

Driver configuration	RS 232	RS 422	RS 423	RS 485
Connection mode	1 : 1	1 : 10	1 : 1	*n : n*
Txd–Rxd circuits				
Transmitter	Single ended	Differential	Single ended	Differential
Receiver	Single ended	Differential	Differential	Differential
Txd signal	±5 to ±15 V	±2 to ±6 V	±4 to ±6 V	±1.5 to ±6 V
Rxd signal threshold	±3 V	±200 mV	±200 mV	±400 mV
Transmission distance	15 m	1.2 km (100 kbps or less) (0.2 mm^2 twisted pair)	1.2 km (6 kbps or less) (0.2 mm^2 twisted pair)	1.2 km (10 mbps) (special cable)

are symmetrical with respect to ground. In this logic, a negative voltage signal represents logic '1', and a positive voltage represents logic '0'. This use of negative logic can be traced to the use of teletype machines in the 1950s, where a flowing current represented logic '1'. It should be noted that this negative logic used in RS 232, is different from the modern digital circuit convention. The standard also uses the terms 'Mark' and 'Space' to replace logic '1' and logic '0' respectively. Table 1 indicates 3 V for a logic '0' (or Space) and −3 V for logic '1' (or Mark). Practically, the voltage levels are ±15 V and ±5 V (TTL). It is necessary to use translator circuits for changing the voltage levels. More details appear later. These circuits must be able to accept TTL input level (logic '1' > 2 V, logic '0' < 0.8 V) and to produce RS 232 output levels. They are located in the transmitter output. Similar circuits also exist in the receiver input to transform RS 232 input levels into TTL output level.

2 RS 232 INTERFACE AND MODEMS

The word modem is an abbreviation for modulator–demodulator. These devices are used to extend the distance for remote communication routes. They perform different functions at the sending and receiving ends. Specialized modems for on-site communication lines are used to extend the distance between connected devices within a facility. Public communication line modems and private communication line modems can connect devices that are far apart on a one-to-one basis through lines provided by commercial communications enterprises. At the sending end, a serial digital signal is used to modulate an analog signal, using frequency modulation, for example, which is sent out on the communication circuit. The analog signal is detected and demodulated at the receiving end to reconstruct the original serial digital signal. Modems for connection to private or public communication lines must be authorized, approved, and must conform to standard frequency band and transmission power specifications. They must also be designed to transmit signals while avoiding external interference from adjacent communication lines. Modulation is covered in **Article 34, Modulation Theory, Volume 1**.

The interface between the modem and the devices it is connected to is standardized in terms of connectors, signal lines, electrical specifications, data exchange method, and timing. One such standard is the RS 232 interface.

The RS 232 interface can also be used for direct connection of devices within 15 m of each other, as well as for the widely used standard modem connection. For these reasons, it is standard equipment on many computers and control devices. As a substitute for the RS 232 interface for transmission without using a modem at relatively longer distances of the order of several hundred meters, depending on the cable diameter, the RS 422, differential type, the RS 423, single-ended type, and the RS 485, differential multidrop type, interfaces have been standardized, and their use is spreading. The electrical specifications of these interfaces are given in Table 1.

3 SPECIFICATIONS IN RS 232

RS 232 only defines specific characteristics of an interface between the DTE (Data Terminal Equipment), such as a computer or VDU (Visual Display Unit), and some sort of DCE (Data Circuit Terminating Equipment), such as a modem, packet switched network pad, statistical multiplexer, and so on. All of the different aspects of the standard are covered in four parts describing

- electrical signal characteristics,
- mechanical interface specifications,
- functional description of interchange circuits,
- standard interfaces for selected communication system configurations.

Electrical specification: The signal definitions of RS 232, illustrating pass and fail types of signal, are shown in Figure 1. These important characteristics mainly concern the signal levels, impedances, and rates of change, which are also covered in CCITT V28. RS 232 signals are characterized by their very large voltage swings, which are required for noise immunity on the transmission wire. For the maximum recommended cable length of about 15 m and

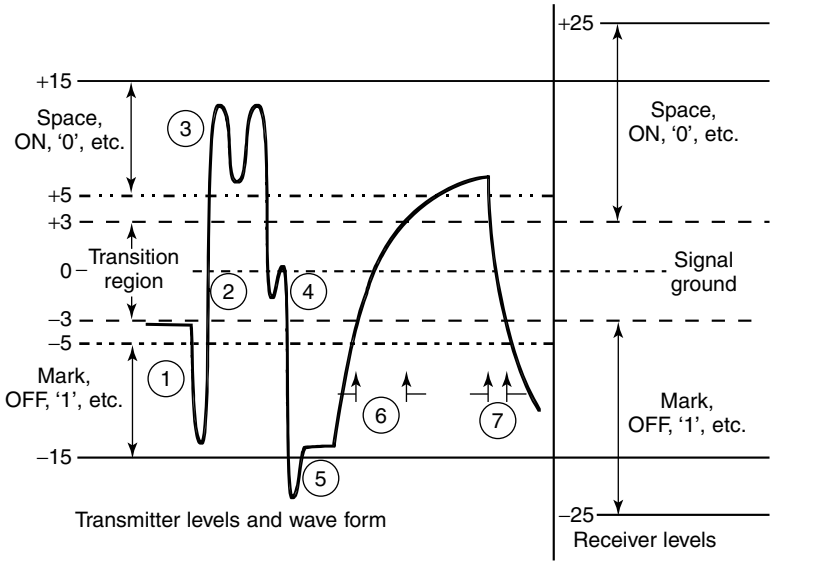


Figure 1. Signal definitions for RS 232 interface.

for these voltages, the standard is able to cover signaling rates up to 20 kHz.

Essentially, any voltage above +3 V is treated by a receiver as a logic '0', a space condition or an ON condition. A signal below −3 V is treated as a logic '1', a mark or an OFF condition. The band between −3 V and +3 V is an undefined transition region.

The generator, or driver, must provide a minimum of +5 V for logic '0', or a maximum of −5 V for a logic '1' at the interface point. Any level between these two, such as at point 1, is a 'fail' because the steady state is out of range. During a pass through the transition region, the dV/dt must not exceed 30 V μs^{-1}, otherwise, the condition, similar to point 2, is a 'fail'. Provided the level of interference, such as power supply hum in point 3, does not cause the signal to go outside the permitted steady state band, the interference level is acceptable.

An unacceptable level of interference is shown at point 5, where a voltage spike drives the signal outside its steady state band. In less than a certain time, the slope must not reverse during the transition as it does in the illustration at point 4, otherwise the transition is a 'fail'. This situation could arise if the duration of a transmitted bit is too short.

Other 'fail' and 'pass' conditions are shown at points 6 and 7 respectively. At point 6, the transition occurs too slowly, while at point 7 the transition, although slow, is still within the accepted tolerance on dV/dt.

Mechanical interface specifications: The specified connector in RS 232 is defined as a 25-way 'D' connector, with dimensions specified in the standard. The DTE connector shall have Male pin contacts and a Female shell, and the DCE shall have Female socket contacts and a Male shell as shown in Figure 2(a) and (b). The signal functions, which enable a fully interlocked handshake exchange of data between equipment at opposite ends of a communications channel, are also covered in CCITT V24. However, V24 defines many more signal functions than RS 232. Those that are common are compatible. Designated functions are associated with 22 of the 25 connector pins, although few, if any, practical implementations use all of them. A detailed discussion of all these signals is beyond the scope of this summary.

The most commonly used signals with the connector pin number in parenthesis, corresponding to Figure 2(c), are as follows:

- *Signal Ground* (7) – interface ground reference.
- *Transmitted Data* (2) from the DTE to the DCE. Data must be transmitted only when the modem compatible signals *Request to Send*, or RTS, *Clear to Send*, or CTS, DCE ready, or DCR, and DTE ready, or DTR, are all in the ON state. A common reason *standard* RS 232 devices fail to work together is that one implements only some of these four handshake signals, but the other is expecting them all.

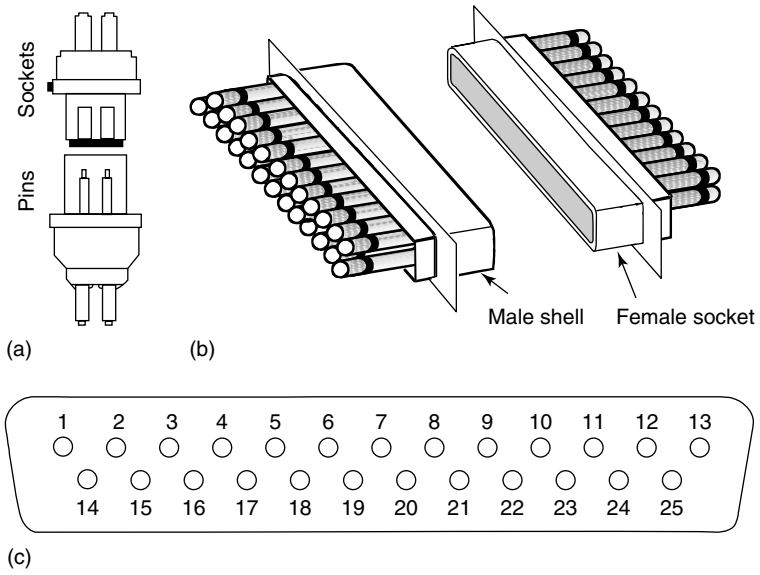


Figure 2. Mechanical specification for RS 232.

- *Received Data* (3) from the DCE to the DTE.
- *Request to Send, or RTS*, (4) – the DTE telling the DCE it wants to *transmit*. This is also used to control direction of communication in a half-duplex system.
- *Clear to Send, or CTS*, (5) – the DCE telling the DTE, in response to RTS and its own ready condition, that the DTE may transmit.
- *DCE Ready, or DCR* (6) – the DCE telling the DTE that it is connected to a communications channel and all dialing, talking, testing, and other operations are finished. This used to be Data Set Ready.
- *DTE Ready, or DTR* (20) – the DTE telling the DCE that it is ready to transmit or receive data.
- *Received Line Signal Detector* (8) – the DCE telling the DTE that it is receiving valid signals over the channel. Sometimes called *Carrier Detect*.
- *Ring Indicator* (22) – the DCE telling the DTE that a ringing signal is being received on the communication channel. Used in auto-answer systems. There are several other, less commonly used handshake functions, including lines to control synchronous communication timing.

All serial protocols operate on the basis that the receiver is ready to receive and identify transmitted data. The handshake timing must always ensure that the transmitter knows when the data set is ready to receive and also when successful reception has been completed. Serial communication uses asynchronous and synchronous protocols as in microcomputer buses.

Figure 3(a) shows that a minimal system only requires two lines for transmission and reception and one signal return line. In asynchronous communication, two lines are used for data transmission and reception, four for control

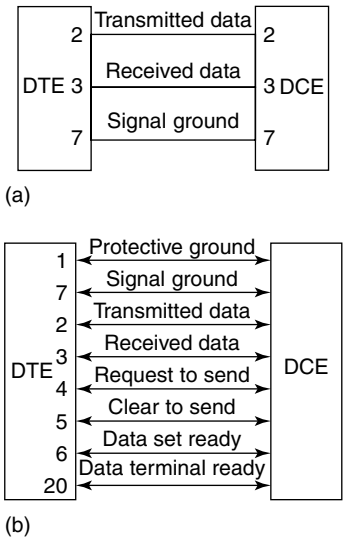


Figure 3. Connections for RS 232: (a) minimal and (b) typical.

signals and one for common return as shown in Figure 3(b). This communication interface is effective at rates of up to 20 kbps over short distances of up to 15 m.

4 WHAT RS 232 DOES NOT COVER

The RS 232 standard covers many aspects of serial communication. However, it should be emphasized that it is not primarily an interface between computers and printers, though it is often used as such.

Moreover, it is not a standard for asynchronous communication. In fact, it contains specific provisions for

synchronous communication if desired. The familiar asynchronous character structure of *one* start bit, eight *data* bits, *two stop bits* is widely used in conjunction with RS 232. However, this is just the most common application of the standard. As it does not appear in the specification of the standard, RS 232 is not a communication format standard.

There is no provision for tri-stating communication lines in the definition of RS 232. Hence, since it is not defined as a local area network, it cannot be used for multiple access to a single line without modification. Networking of multiple lines is covered in **Article 171, Local Area (LANs) and Wide Area Networks (WANs), Volume 3**.

5 EXTENSION OF RS 232 TO RS 422, RS 423 AND RS 485

The two most significant disadvantages of the RS 232 serial communications interface are the comparatively short transmission length of up to 15 m and the maximum transmission rate of 20 kbps. These limitations are mainly due to the fact that the transmission system is an economical space division multiplexed system for transmit and receive lines that share the same common signal. Consequently, they are susceptible to externally induced series mode interference. Moreover, the resistance/capacitance effects of the lines cause attenuation of voltage pulses.

Over longer distances, up to about 1.2 km, and for higher data transmission rates, up to 100 kbps, the RS 449 standard, which uses a 37-pin connector, is gradually replacing RS 232. RS 449 is the mechanical specification for the RS 422 and RS 423 electrical specifications of the extensions. RS 422, which is a balanced double-ended transmission specification, allows the higher data rate with the RS 232 compatible signal. Both this system and the RS 423 have higher transmission rates because they each may use twisted pair cables due to their possessing a unique signal return line. RS 423 is a single-ended unbalanced system. This allows connection to RS 232 and RS 422, but with the RS 449 mechanical specification.

The electrical characteristics of the RS 422 and the RS 423: These standards, which are specially designed for a high signaling rate bandwidth and greater immunity to electrical interference over long distances, of the order of 1.2 km, are now replacing the older RS 232. They have the electrical characteristics given in Table 1. Full duplex connections for both RS 422 and RS 423 are shown in Figure 4. A single-ended RS 423 standard, given in Figure 4(b), is intended to achieve compatibility with other standards. In particular, the RS 232 and RS 422 links can be connected through an RS 423 link, which is characterized by the electrical characteristics of unbalanced digital interface circuits.

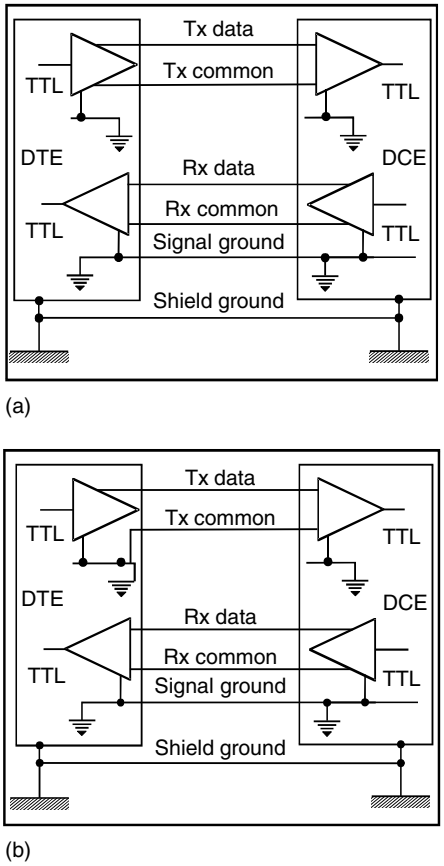


Figure 4. Full-duplex connections for extended RS 232: (a) RS 422 connection and (b) RS 423 connection.

Note that the electrical characteristics of RS 232 are nearly identical to those of the RS 423 standard. The principal objective of RS 422 and RS 423 is to move to higher speed communications.

RS 422 supports multipoint connections, whereas RS 423 supports only point-to-point connections. RS 422 defines a balanced electrical interface, while the RS 423 standard is characterized by its unbalanced interface. This is emphasized by the differential nature of RS 422, which is specified to have two lines of opposite polarity for data transmission. This pair of lines is such that one is positive, in the range 2 to 6 V, while the other is negative, within the range -2 to -6 V, for a logic '1'. Logic '0' corresponds to a reversal of the polarity of each of these lines. Compatibility between these different interfaces is achieved through adequate differential receivers. Proper precautions must be taken when an RS 423 interface is connected to RS 232 interfaces. For example, the SIGNAL GROUND of RS 232 must be tied to the SEND COMMON connection of RS 423. To go beyond 20 kHz, the double-ended signaling, defined by the RS 422 standard is recommended.

RS 485 standard: This is a specialized interface that would not be considered standard equipment on today's home PC but is very common in data acquisition. RS 485 will support 32 drivers and 32 receivers, which means it is a bidirectional, half duplex, multidrop communications interface over a single twisted pair cable or a dual twisted pair cable. An RS 485 network can be connected in a 2-wire or a 4-wire mode. Maximum cable length can be as much as 1.2 km because of the differential voltage transmission system used. The typical use for RS 485 is a single PC connected to several addressable devices that share the same cable. RS 485 can be viewed as a *party-lined* communications system in which the addressing is handled by the remote computer unit. RS 232 may be converted to RS 485 with a simple interface converter. This can have optical isolation and surge suppression.

When choosing a transmission line for RS 485, it is necessary to examine the required distance of the cable and the data rate of the system. Losses in a transmission line are a combination of AC losses, due to skin effect, DC conductor loss, leakage, and AC losses in the dielectric. In high quality cable, the conductor losses and the dielectric losses are of the same order of magnitude.

RELATED ARTICLES

Article 78, Embedded Systems, Volume 2; **Article 158, Microcontroller Systems, Volume 3**; **Article 165, Introduction to Networked Instrumentation, Volume 3**; **Article 166, Instrument Interconnection, Volume 3**.

FURTHER READING

Axelson, J.L. (1996) *Serial Port Complete: Programming and Circuits for RS-232 and RS-485 Links and Networks*, Lakeview Research, (Digital Download), Madison, WI.

Campbell, J. (1989) *The RS-232 Solution*, Sybex International, Berkeley, CA.

Iotech Inc. (1991) *Instrument Communications Handbook*, Cleveland, OH.

Jordan, J.R. (1995) *Serial Networked Field Instrumentation*, John Wiley & Sons, Chichester, UK.

Maine, A.C. (1986) *Interfacing Standards for Computers*, IEETE, London, UK.

Seyer, M.D. (1991) *RS 232 Made Easy: Connecting Computers, Printers, Terminals and Modems*, Prentice Hall, NJ.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

169: Voltage and Current Loop Transmission

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Voltage Loop Transmission System 1130
2 Current Transmission System 1130
3 Principles of the Current Loop 1131
4 Variable Current Potentiometers 1131
5 Current Loop Transmitters 1131
6 Intelligent Transmitters 1133
Related Articles 1134
Further Reading 1134

1 VOLTAGE LOOP TRANSMISSION SYSTEM

Communicating signals over cables requires means of accepting and delivering signals with accuracy. The simple electrical circuit loop can be used for this by varying the voltage or current in the loop. These two methods are discussed here, starting with the voltage option.

Consider the circuit shown in Figure 1. The voltage at the input of the receiving instrument, v_I, contains a component, v_{ISV}, due to the transmitted signal, v_s, given by

$$v_{ISV} = \frac{R_I}{R_I + R_s + R_w} v_s \tag{1}$$

Another component appearing at the input of the receiving instrument is due to the interfering signal, v_m. Interfering signals may be caused by coupling to nearby power lines or from cross talk in adjacent transmission channels. The interfering component at the input of the receiving instrument is

$$v_{ImV} = \frac{R_I}{R_I + R_s + R_w} v_s \tag{2}$$

Assuming that $R_s \ll R_w$ gives

$$v_{ISV} \approx \frac{R_I}{R_I + R_w} v_s \tag{3}$$

and

$$v_{ImV} \approx \frac{R_I}{R_I + R_w} v_m \tag{4}$$

Hence, if the voltage source is ideal, then only cabling attenuation due to R_w affects the received signals. Note that both the transmitted information bearing signal and the interfering signal are attenuated in the same proportion so the receiving end signal-to-noise ratio is unaffected by either the cabling or whether the source is ideal or not.

2 CURRENT TRANSMISSION SYSTEM

The current loop option is now explained.

Suppose the signal is a current, i_s, generated in a nonideal current source with an output resistance, R_g, as shown in Figure 2. In this case, the signal generated at the receiver by the current, i_s, is

$$v_{ISC} = \frac{R_g(R_I + R_w)}{R_g + R_I + R_w} \times \frac{R_I}{R_I + R_w} i_s \tag{5}$$

whilst

$$v_{ImC} = \frac{R_I}{R_g + R_I + R_w} v_m \tag{6}$$

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

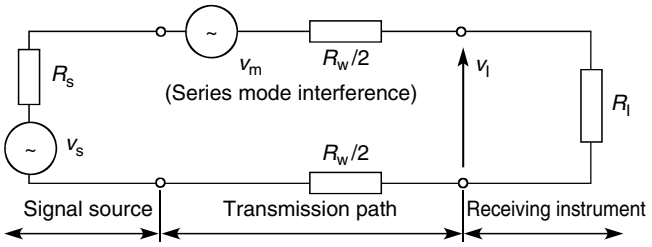


Figure 1. Voltage transmission with series mode interference.

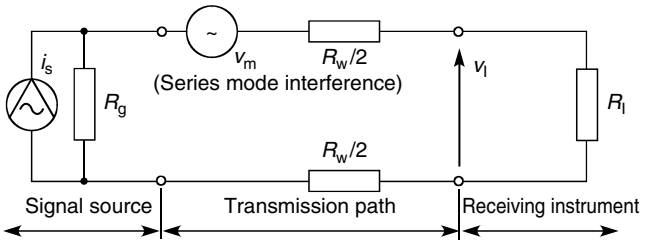


Figure 2. Current transmission with series mode interference.

Substituting for the relation between v_{ImV} and v_m gives

$$v_{ImC} = \frac{R_I + R_s + R_w}{R_g + R_I + R_w} v_{ImV} \ll v_{ImV} \tag{7}$$

Thus, the received voltage due to series mode interference in the current transmission system is very much less than the received voltage due to series mode interference when the information signal is transmitted as a voltage.

3 PRINCIPLES OF THE CURRENT LOOP

In harsh industrial and process system environments, even the RS 449 protocol is not impervious to severe series mode interference. Consequently, it cannot be used. A current-transmitting loop, which can be used for transmitting either analog or digital signals, has a limited digital rate of 4800 bits s^{-1}. When the signal is digital, the transmitter is switched between two distinctly different current levels that indicate the binary logic '11111' and '110011' levels. If the signals have to be transmitted/received between computers or other instruments, it is normal to include a converter to restore the loop current to a voltage signal compatible with RS 232 – *see* **Article 168, RS 232 and EIA/TIA 232 Serial Interface, Volume 3**. By using frequency modulation – *see* **Article 34, Modulation Theory, Volume 1** – it is possible to transmit high-fidelity current signals with low interference over longer distances. In the current control, a logic '1' is encoded by a current of 20 mA and a logic '0' by zero current.

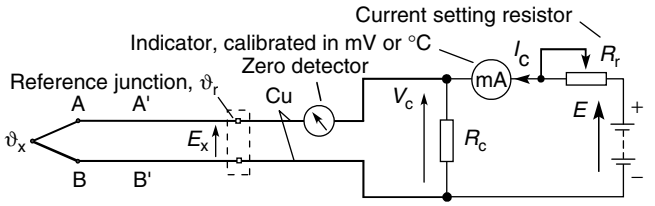


Figure 3. Basic circuit of a variable current potentiometer.

4 VARIABLE CURRENT POTENTIOMETERS

Process transmitters, similar to those based upon BS 5863, which still represent the most widely used forms, are usually based on the principle of variable current potentiometers. BS 5836 has two parts. Part 1 (1984), which replaces BS 3586, is equivalent to IEC 60381-1 for CENELEC HD 452.1 signals. Part 2 (1980) is equivalent to IEC 60381-2 for its specification of direct voltage signals. Variable current potentiometers have the basic circuit given in Figure 3, which uses a thermocouple sensor as a specific example. The temperature to be measured is ϑ_x, and ϑ_r is the temperature of the reference junction. The voltage to be measured, E_x, which may also be due to any other type of sensor, represents the thermoelectric emf. It is opposed by an adjustable voltage, V_c, which is generated as a voltage drop across a constant resistor, R_c. This voltage drop is adjusted by changing the compensating current I_c, using a slide wire resistor, R_r, to satisfy the condition

$$E_x = I_c R_c \tag{8}$$

The value of the current, I_c, is a measure of the unknown voltage, E_x to be determined. A zero-detector, which is used to indicate the null balance condition, may be a current indicating milliammeter actually measuring in mA, but calibrated in mV. The restricted accuracy of milliammeters is compensated for by the simplicity of the construction of the potentiometer. It is easy to assemble such a potentiometer from typical laboratory instruments. Their accuracy is about 0.2 to 0.5% of the measuring range. When they are configured as self-balancing potentiometers on the basis of an operational amplifier with feedback, they become what are now known as *current transmitters*. More detail is available on amplifiers for data acquisition systems (DAS) in **Article 133, Amplifiers and Filters for DAS, Volume 3**.

5 CURRENT LOOP TRANSMITTERS

Current transmitters, which may be configured as two-wire or four-wire transmitter units, are widely used. Figure 4

shows typical block diagrams for each, once again using a thermocouple as the sensor.

In the schematic diagram of a four-wire current transmitter, given in Figure 5(a), it can be seen that the output signal is proportional to the measured emf, E_x. The voltage drop across the resistor, R_c, caused by the current, I_c, compensates the measured emf, E_x, so that the input voltage, e, of the operational amplifier is

$$e = R_c I_c + E_x + V_{off} \tag{9}$$

where V_{off} is the input offset voltage of the amplifier.

Terms used in measurement are covered in **Article 41, Terminology and Classification of Measurement Systems, Volume 1**; and **Article 57, Static Considerations of General Instrumentation, Volume 1**.

Define the amplifier gain as $K = -I_c/e$. Inserting this into (8) and (9) gives

$$I_c = -\frac{K}{1 + K R_c}(E_x + V_{off}) \tag{10}$$

Assuming a high amplifier gain, K, (10) can be rewritten as

$$I_c \approx -\frac{E_x + V_{off}}{R_c} \tag{11}$$

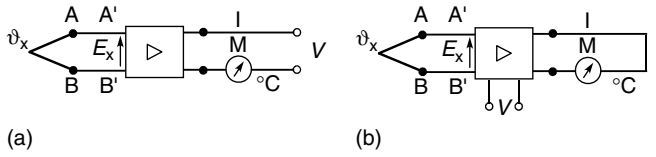


Figure 4. Current transmitter types: (a) two wire and (b) four wire.

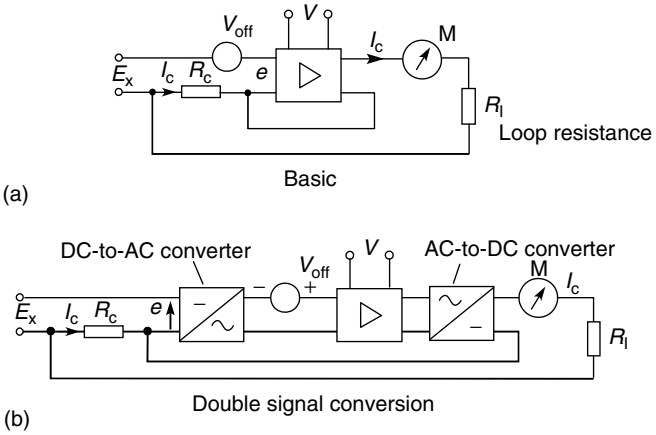


Figure 5. Four-wire thermocouple transmitters.

The compensating output current, I_c, is thus directly proportional to the measured emf, E_x. The effects of low frequency drift and offset in the amplifier are eliminated using double signal conversion transmitters, as shown in Figure 5(b). The two converters and the amplifier form one integrated circuit.

The two-wire transmitter, with the block diagram shown in Figure 6, is primarily intended for use with thermocouples. Amplification of the thermocouple signal in 6 (see Figure 6), is followed by conversion in 7 into an output current, proportional to the measured emf. The transmitter's measuring current, I_c, is composed of two components so that

$$I_c = I_o + I_m \tag{12}$$

The current, I_o, in most cases 4 mA in 4 to 20 mA systems, which is used to supply the necessary power to the transmitter of two-wire systems, is kept constant by a current stabilizer, 8. The component, I_m, is a function of the measured temperature. An advantage of 4 to 20 mA systems is that a transmitter failure, corresponding to the occurrence of zero output current, cannot be mistaken for a zero-measured temperature value. The reference junction corrector, 1, adjusted to match the thermocouple characteristic, simultaneously feeds the thermocouple break-safe circuit, 2. In the case of a thermocouple break, the output current either falls to zero or increases over the 20-mA high limit, to a value of 35 mA in most transmitters. The voltage stabilizer, 3, and the resistors of the circuit, 4, determine the minimum measuring range, whereas the circuit, 5, is used to set its maximum value. Both output current components flow through the diode, D, protecting it against any reversal of supply voltage, while protection for the transmitter against any voltage peaks is provided by the other Zener diode, 9.

A transmitter for resistance thermometers is shown in Figure 7. As for the thermoelectric emf in the thermocouple transmitter of Figure 6, the temperature-dependent resistance variations of the RTD (resistance temperature detector) are converted into an output current, linearized in current source, 1. The input signal is fed through the temperature range module 2 and the amplifier, 3, to the output current converter, 4. The current stabilizer, 5, sets a component of the current, I_o, smaller than 4 mA. Two-wire, three-wire, or four-wire connections are used for circuit completion between the RTD and the transmitter.

Although using a two-wire connection between the RTD and the transmitter is a convenient method of circuit completion for head mounted transmitters, it does not give precise measurements. It also imposes a limit upon the loop resistance variations. As a three-wire connection or a four-wire connection between the RTD and the transmitter

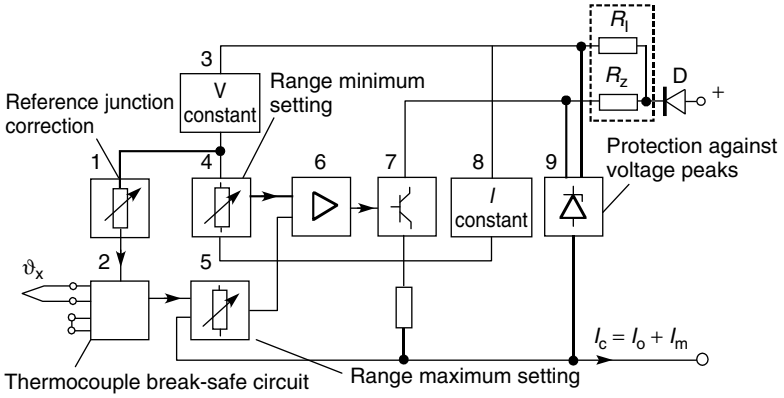


Figure 6. A two-wire transmitter.

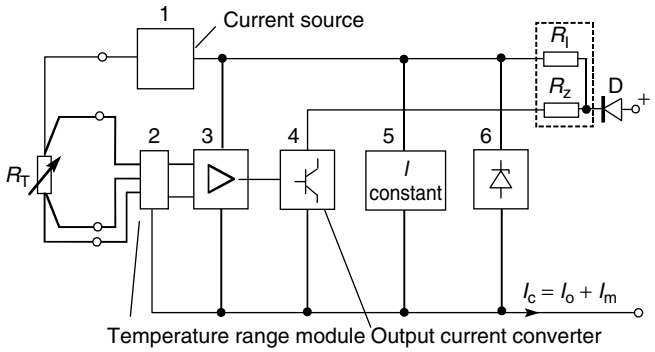


Figure 7. A two-wire transmitter for RTDs.

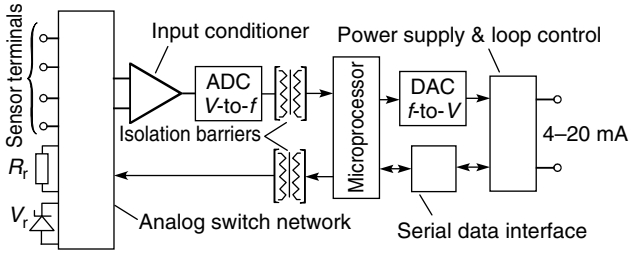


Figure 8. An intelligent current transmitter.

ensure the compensation of loop resistance variations, they are normally used in the case of larger distances between the sensor and the transmitter.

6 INTELLIGENT TRANSMITTERS

Further development of transmitters has led to this generic type. For example, the transmitter, shown in Figure 8, can accept inputs from a variety of temperature sensors such as Pt-RTD or different thermocouples without changing any modules or altering any links or switches. This microprocessor-based transmitter can be programmed by a remote transmitter interface, or RTI, which also enables the measuring ranges to be changed. The output signal can be linear or any other function of measured temperature. Galvanic isolation between input, output, and earth is an important feature of the transmitter. These transmitters usually have a high accuracy over their entire working range as well as high immunity to any airborne radio-frequency interference, or RFI, due to the included effective filtering. This last quality is very important in view of the increasing use of radio transceivers for on-plant personnel communication.

Selection of a sensor type causes the analog switch network in Figure 8 to gate the relevant sensor signal. The same network also gates the reference resistor, R_r, or the reference voltage, V_r, to the output circuit, which conditions the sensor signal. After conversion by a voltage/frequency converter, the frequency signal is passed through an isolating transformer to the microprocessor. Between samples from the sensor, the resistor, R_r, voltage, V_r, and short circuit are checked to calculate the circuit drift due to ambient temperature variations or aging of elements. The microprocessor, which then corrects these influences, also compensates for the influence of changing reference junction temperature, if a thermocouple is used. Either a digital or an analog 4 to 20 mA DC output signal is provided by the microprocessor. For computer interfacing, these transmitters can also communicate digitally over the 4 to 20 mA lines using an appropriate protocol.

Modern current transmitters are very reliable, reaching up to 12 000 working hours, which is about eight years.

It is to be expected that digitally based transmitters, where the current transmitted will be in digital form, will continually replace the 4 to 20 mA analog standard units as their performance deteriorates with age.

Further information on these still popular traditional analog loop transmission systems is found in the classical process control texts given in the bibliography and on the

web sites of major manufacturers by searching with 'current loop transmission'.

Intelligence in measuring systems is covered in **Article 160, Smart Sensor System Features, Volume 3**.

RELATED ARTICLES

Article 17, EMC and EMI, Volume 1; **Article 76, Electronic Systems Design, Volume 2**; **Article 121, Signals in the Presence of Noise, Volume 2**; **Article 166, Instrument Interconnection, Volume 3**; **Article 167, Asynchronous and Synchronous Interface Protocols, Volume 3**.

FURTHER READING

Anderson, N.A. (1997) *Instrumentation for Process Measurement and Control*, CRC Press.

Considine, D.M. and McMillan, G.K. (1999) *Process/Industrial Instruments and Controls Handbook*, McGraw-Hill Professional.

Liptak, B.G. (1985) *Instrument Engineers' Handbook: Process Control*, Chilton Book Company.

Shinskey, F.G. (1988) *Process Control Systems*, McGraw Hill.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

170: IEEE-488 Instrumentation Bus

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 General Description of IEEE-488 Instrumentation Bus 1135
2 Mechanical Specification of the IEEE-488 Bus 1137
3 Electrical Specification of the IEEE-488 Bus 1137
4 General Format of the IEEE-488 Bus 1138
5 IEEE-488 Bus Operation–Controllers, Talkers and Listeners 1139
6 IEEE-488 Bus Operation–Bus Messages 1139
7 IEEE-488 Bus Operation–Bus Management Protocol 1140
8 IEEE-488 Bus Operation–Bus Handshake Protocol 1140
Related Articles 1141
Further Reading 1141

1 GENERAL DESCRIPTION OF IEEE-488 INSTRUMENTATION BUS

Several attempts to standardize similar bus systems that were suitable for interconnecting measurement instruments from the specifications of various manufacturers were made during the 1970s. The International Electrotechnical Commission, or IEC, and the Institute of Electrical and Electronics Engineers, or IEEE, have respectively succeeded in producing the IEC-625 bus and the IEEE-488 bus, which are the most widely adopted and generally compatible, if not identical, instrument bus standards.

The IEEE-488 bus or General Purpose Interface Bus (GPIB) is an internationally accepted specification, which defines the mechanical and electrical interface between all IEEE-488 compatible instruments. The mechanical, electrical, functional, and operational features are shown in Table 1. It defines a standard communication link for data and commands between instruments. This leads to significant simplification of the combined use of a variety of devices from different manufacturers. However, it should be noted that only the operational aspects of this bus depend upon the device or system. It does not define *device-dependent programming*, which is the way any instrument will respond to a given command. One analogy is the international telephone system, where the hardware and hookups are all compatible, but the language spoken over the system is not defined.

The bus, which is restricted to a maximum of 15 measuring instruments, can consist of up to 31 in certain circumstances. The IEEE-488 instrument bus is most often used to connect computers and personal computers to various kinds of laboratory testing and measuring instruments, such as multimeters, frequency counters and analyzers, and other peripheral devices such as plotters and signal generators among others. In this way, automatic measurement, analysis, data processing, and recording can be accomplished.

Controllers, *talkers*, and *listeners* are the three classes of functional devices that are connected to the bus. Each device must be capable of executing one of these three functions. An instrument that functions as a talker can only transmit messages to the bus. A listener can only receive messages from the bus. However, most instruments are designed for both transmitting and receiving messages,

Table 1. Some features of the IEEE-488 bus.

Mechanical	Electrical	Functional	Operational
Device or system independent specification			Device or system dependent specification

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

although not simultaneously. Examples could be for the control of the measurement range and for the transmission of the measured data. Only one instrument can talk at a time, although several instruments can function simultaneously as listeners.

To regulate the information stream along the bus, one of the connected instruments, usually a computer, acts as a controller, or supervisor. The controller determines which instrument is the talker and which is the listener. Each instrument has a unique address, which is set with a series of small switches, usually located on the back plane of the instrument. Only the instrument with the address that corresponds to the address transmitted by the computer receives the message. It is even possible to let two instruments communicate without the intervention of the computer.

The controller performs management functions such as releasing a connection or specifying the talker and listener. Whereas a talker loads data onto the bus, a listener downloads data from the bus. Computers are usually equipped with all three functions, while measurement instruments are talkers, and plotters and other such devices are listeners. As an example of a simple system, a computer performs the required controller functions, specifies an instrument as a talker and itself as the listener after which it then receives the measurement data. After completion of data transmission, the designations for both talker and listener are released. When data analysis is complete, the computer specifies itself as the talker and a plotter as the listener before sending the analysis data to the plotter for recording.

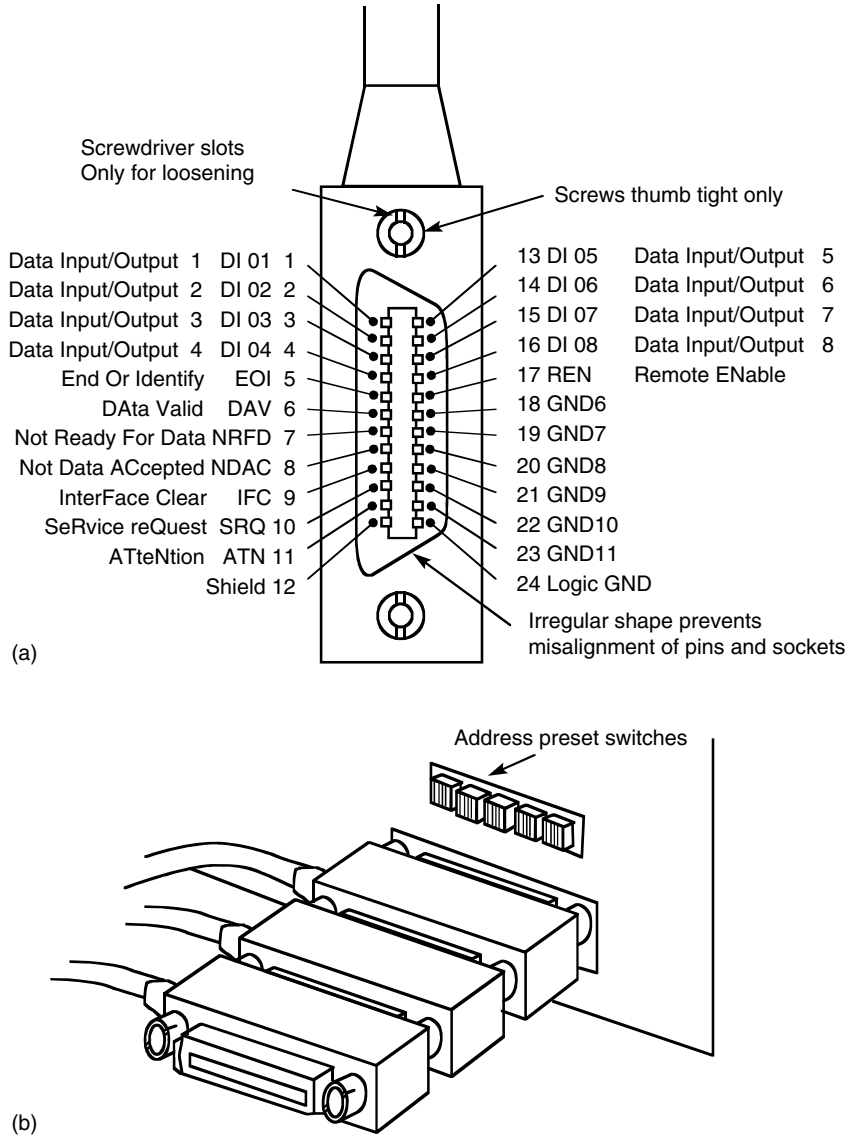


Figure 1. Mechanical and electrical specification for IEEE-488 bus.

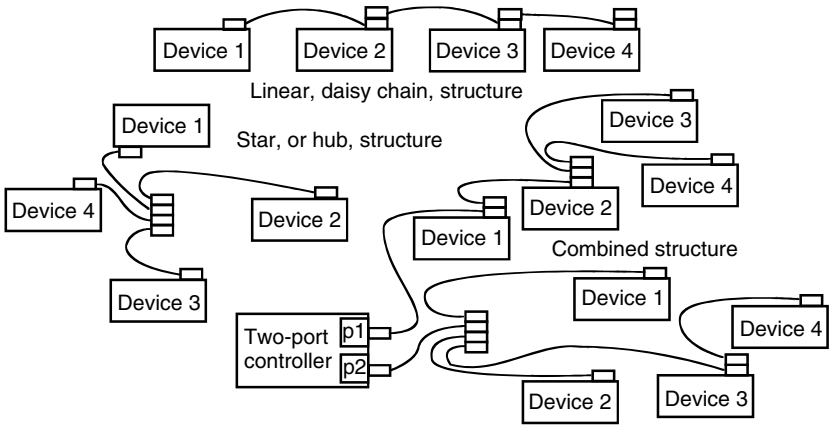


Figure 2. Network connections for the IEEE-488 bus.

2 MECHANICAL SPECIFICATION OF THE IEEE-488 BUS

The 16-line highway cables between devices are terminated in 24-way connectors of the *Amphenol micro ribbon* type with the pin assignments shown in Figure 1(a). These connectors, usually located at the instrument rear, can have further *piggy-backed* connectors, as shown in Figure 1(b), where the address setting switches are also conveniently located. The piggy-backed structure allows *linear systems*, *star systems* or *combined systems* to be configured as shown in Figure 2, where the address setting switches are also conveniently located.

3 ELECTRICAL SPECIFICATION OF THE IEEE-488 BUS

The highway is terminated within each device in a specific manner as in Figure 3. The bus, which is TTL compatible, uses a negative logic convention so that logic *high* is $> +2\,\text{V}$ to indicate a *false* state, whilst '*low*' is $< 0.8\,\text{V}$ to indicate *true*. The line drivers may be either of the open collector or the tristate variety. Their specified drive capability permits bus lengths of up to 20 m or 2 m times the number of instruments connected to the bus, whichever is less. The specification also covers details of screening, cable resistance and capacitance, and so on.

Physically the bus is a 16-line bidirectional passive highway to which up to 15 instruments, including a controller may be connected in parallel to form a *party line* structure. All lines are functionally grouped into the three component buses summarized in Table 2. These include 3 lines for data transfer timing, also called *handshake lines*, 5 internal control signal lines and 8 data lines for 8-bit parallel data. The 8 lines, forming the *data bus*, allow data

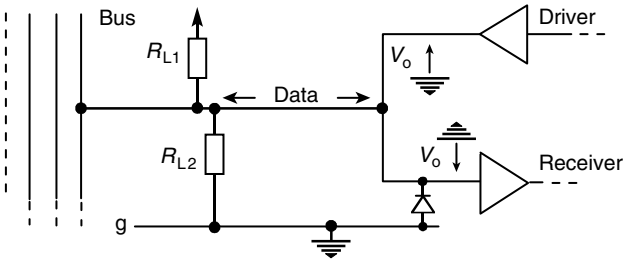


Figure 3. TTL compatible connection between the IEEE-488 bus and the devices input/output circuit.

Table 2. Summary of bus signals for IEEE-488 bus.

Classification	Signal line		Function
Handshake signal	DAV	DAta Valid	Data is valid
	NRFD	Not Ready For Data	Not ready for data
	NDAC	Not Data ACcepted	Data is not accepted
Control signal	ATN	ATteNtion	Command designation
	REN	Remote ENable	Remote control designation
	IFC	InterFace Clear	Reset of interface
	SRQ	SeRvice reQuest	Interrupt notice to controller
	EOI	End Or Identify	End-of-data
Data	DIO 01 to 08	Data input/output	Data

transport in bit parallel word serial form with the data being transmitted byte after byte. This may be between talkers, which are transmitting devices, and listeners, which are receiving devices, or between the system controller and subordinate devices. The controller, which talks and listens, designates the status of controlled instruments. The overall system direction is controlled by commands, to or from the controller, via the 8 data lines in conjunction with the 5 management lines. There may be a number of listeners operating on the bus at the same time, but only one talker operating at a time. The data may be commands for the selected instruments or measurement data. One message may comprise several bytes. The remaining 3-byte transfer lines control the *handshake* routine by which data is transferred from one device to another. This process is necessary since the rates at which the various bus devices can transmit or receive data may vary considerably. These 3 lines permit a talker to synchronize its data transmission with the readiness of the appropriate listeners to receive.

4 GENERAL FORMAT OF THE IEEE-488 BUS

Figure 4 represents the general format of the bus. Although only four devices are shown, up to five can be used on a single bus system. If more than five instruments are required in the system, additional bus loops may be used.

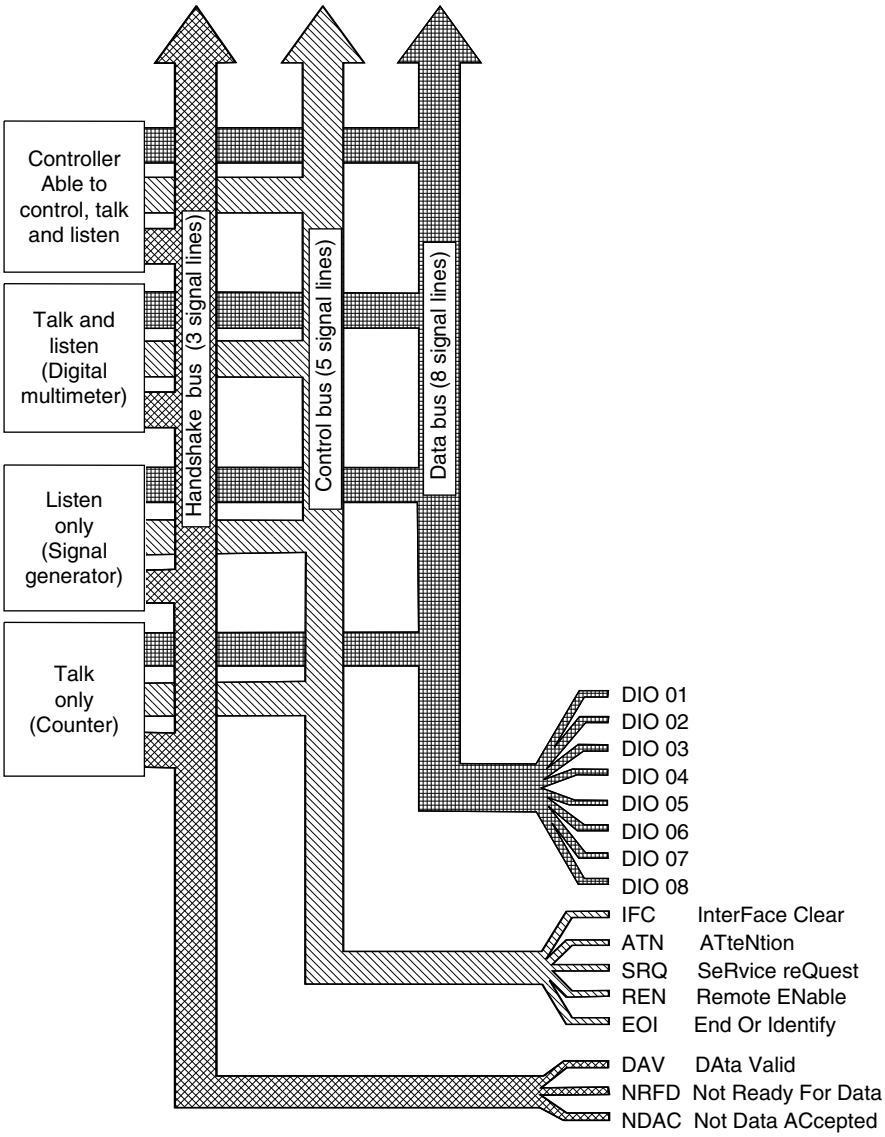


Figure 4. General format and structure of the IEEE-488 bus lines.

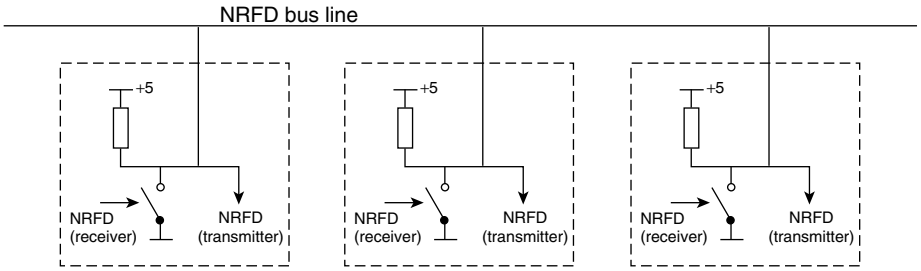


Figure 5. Wired OR lines for the NRFD lines. When the listener is not ready to accept data, the switch is off, the line voltage is low (0) and the talker is waiting.

The data transport over the bus is asynchronous. Due to the handshake procedure, instruments with diverse processing times can be connected to the bus, so the transport speed is determined by the slowest of the connected instruments.

Some bus lines are connected in a wired OR manner as in Figure 5. The bus lines for the *Not Ready For Data*, or NRFD, line in Figure 5 are all connected to the positive power supply voltage via a resistor. Each instrument contains a switch that can connect the lines to ground. The line voltage is only *high*, or '1', if none of the instruments has short circuited the line. The line is *low*, or '0', if at least one instrument has asserted the switch on. In this way, the OR function between the NRFD lines is realized without the necessity of separate OR gates so that introducing complex wiring is avoided.

5 IEEE-488 BUS OPERATION–CONTROLLERS, TALKERS AND LISTENERS

Each of the instruments in an IEEE-488 based system must be capable of operating in one or more of the modes as a Controller, a Talker, or a Listener.

Controller: The Controller addresses other devices to listen or talk. It can send interface messages as well as device independent commands or device-dependent commands. It will generally hold a *control program*, which schedules the system devices and also captures data under predetermined conditions. The data may be stored, processed, and then displayed by the controller using its own output devices, such as a CRT display or plotter, or by using a dedicated peripheral connected to the bus such as a high-speed printer. There is normally one controller on the bus directing data between talkers and listeners using interface messages. These can be multiline messages using the 8 data lines and/or '*line messages*, using the 5 management lines. Each device on the bus has its own address code, which is usually set up on rear panel switches as shown in Figure 1(b). This allows device-dependent messages to be addressed to that particular instrument. Addressing, which schedules instruments as talkers or listeners, may involve multiline talk or listen commands. Universal interface commands, which can be multiline or uniline, such as InterFace Clear, or IFC, Remote ENable, or REN, ATteNtion, or ATN, and so on, affect all interfaces on the bus.

Talkers and listeners: A *Talker* is a device, which can transmit data on the bus but can not receive it, while a *Listener* can receive data but cannot transmit. Typical examples might be a simple DVM as a talker and a printer as a listener. Some instruments can function as talkers and listeners. For example, most DVMs can *Listen* to remote programming instructions sent via the bus. Subsequently, having taken measurements, the DVM then *Talks* by transmitting that data to other listeners on the bus. Clearly, the simplest possible system consists of one as a *Talker* and another as a *Listener*. For example, a device taking measurements and capable of, or manually set to, a *talk only mode* where data is transmitted over the bus to a device in *listen only mode*. An expansion of this may be with one talker transmitting to a number of listening devices.

However, it is not until an intelligent controller is included in a system that the real value of the bus is realized. An intelligent controller, such as a programmable calculator, can both talk and listen as well as act on data to process it and subsequently direct the overall measurement operation to produce a genuinely intelligent system.

6 IEEE-488 BUS OPERATION–BUS MESSAGES

Messages sent over the bus, which are generally based on the ASCII code, are all transmitted via the 8 data lines in conjunction with the 5 management lines. The messages may be classified as either *interface messages* or *device-dependent messages*.

Interface messages: These are commands sent over the bus via the 8 data lines in association with appropriate management line status. For these commands, the ATN management line is always held true so that data is directed to the *instrument interface circuitry*. Contrastingly, *device-dependent commands* direct the operation of the instrument. These *commands* are specified by IEEE-488 and therefore have the same meaning to all bus instruments. If a device does not have a particular bus facility, associated commands are ignored completely. In this mode, the system controller is the 'talker' and all other devices on the bus are awaiting instructions. These commands include

Talk/Untalk; *Listen/Unlisten*; *Remote/Local*; *Clear/Trigger*; and so on.

Device-dependent messages: These messages are transmitted to *command* a particular instrument on the bus to adopt a specific measurement mode or function. Since the functions and capabilities of different units, even of the same general type, vary considerably, it is necessary for instrument designers to use the most appropriate message coding. The IEEE-488 standard therefore specifies only the general format for command transmission. It does not define the exact instrument response to any given command. Consequently, instruments are rarely interchangeable without reprogramming. Moreover, some of the available manual user functions to a unit are not necessarily available on the bus. It is a significant feature of most new DVMs that virtually all the functions available to the manual user can also be controlled remotely on the bus. First addressing the DVM as a listener and then sending a simple alphanumeric ASCII string achieves programming. An alpha character designates the required program area, such as range, function, or even Autocal. A numeric character to specify a particular range or function follows this.

7 IEEE-488 BUS OPERATION–BUS MANAGEMENT PROTOCOL

Five control lines perform the following bus management functions:

- *IFC, or InterFace Clear*: The controller can reset the bus. All connected instruments are set in an initial state and wait for further commands.
- *ATN, or ATteNtion*: The controller can switch from command mode to data mode. In the command mode, instrument control signals are transported over the data lines. In the data mode, the data lines are used to transport measurement data.
- *REN, or Remote ENable*: Here, the controller imposes its command of the instruments. Manually controlled front panel elements are disabled.
- *SRQ, or SeRvice reQuest*: One or more instruments can interrupt the normal operation by signaling the controller for the execution of a particular program.
- *EOI, or End Or Identify*: In the data mode (ATN = false), the EOI means that the current byte is the last byte of the message. In the command mode, the EOI refers to the execution of a service request.

Normally, the user of the bus system has nothing to do with the bus interfaces. Special ICs take care of the communication, including the handshake procedure. The user, however, has to write a program for the controller to execute the desired measurement sequence.

8 IEEE-488 BUS OPERATION–BUS HANDSHAKE PROTOCOL

Handshake lines: To guarantee proper data transport, irrespective of the differences in response times of the instruments, the bus contains three special control lines, called the handshake lines. These three *handshake*, or Data Byte Transfer Control Lines, control the transfer of data bits between talkers and listeners. This allows devices with differing input/output data rates to be connected together. The data transfer rate is always, automatically adjusted to that of the slowest device by a wired OR process. Maximum Transfer rate is one megabyte per second. These three lines, which are NRFD, or Not Ready For Data, DAV, or Data Valid, and NDAC or Data Not Accepted, have the following functions:

- *NRFD (Not Ready For Data)*: All instruments use this line to indicate whether they are ready to accept data. It goes high, or *false*, to indicate that the listener(s) are now ready for data. If any device is not ready, it forces the line to be held low. A talker on the bus puts data on the bus, after it has used this line to determine when all listeners are ready. The talker must wait until all listeners are ready, which is indicated by a false NRFD. Logically speaking, this signal is the OR operation on all connected NRFD lines: $\text{NRFD} = \text{NRFD}(1) + \text{NRFD}(2) + \cdots + \cdots$.
- *DAV (Data Valid)*: This signal is asserted true by the instrument assigned as talker, just when all active listeners are ready to receive data. This is indicated with NRFD false. It goes low when the talker has placed data on the bus, but only after a short time interval to allow for settling. The DAV line thus tells the listeners that data is on the bus and is fit to be read. Thus, all

receiving instruments know that the data on the data lines contain relevant information.

- *NDAC (No Data Accepted)*: Each listener attempts to assert this signal false as soon as data transmission takes place. However, it is held low, that is, remains *true* corresponding to NDAC, by the slowest listener until it has received and accepted the data from the data lines, or has completed data transfer. The talker must maintain the data on the data lines at least until that moment. Once this is pulled high again, the talker may change the data so that a new byte of data can be sent over the bus. Because a possible next byte can only be put on the data lines when NDAC is false, NDAC is a logic OR function. In other words:

$$\text{NDAC} = \text{NDAC}(1) + \text{NDAC}(2) + \cdots.$$

Handshake timing: Figure 6 shows a handshake routine with the relative timing of each line, although not to scale. The dotted lines show the state of faster devices and the solid lines the change of state, which results when the slowest device on the bus finally reacts:

- A – As each listener device becomes ready to accept data, NRFD is set false. By 'A', the slowest device is ready to accept data and NRFD is set false.
- B – DAV is set true by the talker to indicate to the listener(s) that data is on the lines, has settled and is fit to be read.
- C – As each listener accepts data, its NDAC is set high. When the slowest device has accepted data, the NDAC line goes false to indicate to the talker that all the listeners have accepted data.

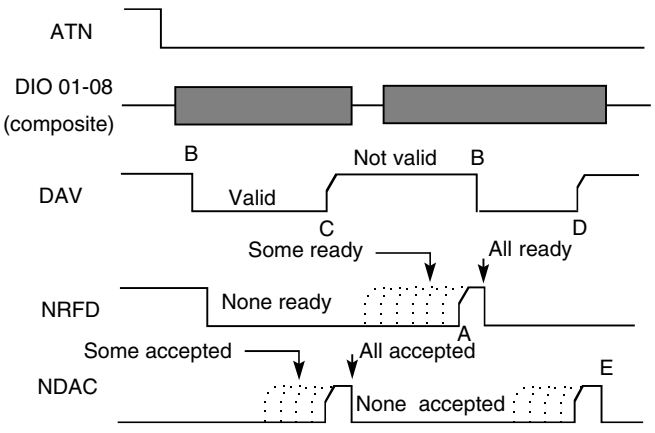


Figure 6. Handshake routine with relative timing of each line in IEEE-488 bus.

- D – The talker now sets DAV false to indicate that the data on the lines is no longer valid.
- E – All listener(s) respond to DAV going false and set their NDACs true. After 'E', NRFD can be reset high so that a further handshake can occur.

The IEEE-488 standard has recently been upgraded (IEEE 2003), enabling buses transfer data at rates of up to 8 Mbps.

RELATED ARTICLES

Article 132, Data Acquisition Systems (DAS) in General, Volume 3; **Article 158, Microcontroller Systems, Volume 3**; **Article 165, Introduction to Networked Instrumentation, Volume 3**; **Article 166, Instrument Interconnection, Volume 3**; **Article 172, Fieldbus Systems, Volume 3**.

FURTHER READING

There does not appear to be any recently published books specifically on this IEEE-488 bus. The technology is mature and earlier books are still relevant today.

Web sites of suppliers offer information of use – search on 'General Purpose Instrumentation Bus' to obtain links to catalogs.

Caristi, A.J. (1989) *IEEE-488 General Purpose Instrumentation Bus Manual*, Academic Press, San Diego, CA.

Colloms, M. (1983) *Computer Controlled Testing and Instrumentation: An Introduction to the IEC-625 : IEEE-488 Bus*, John Wiley & Sons.

IEEE (1987) *IEEE Standard Codes, Formats, Protocols, and Common Commands for Use with IEEE Std 488.1–1987, IEEE Standard Digital Interface for Programmable Instrumentation*, IEEE Publications.

IEEE (2003) *Standard for Higher Performance Protocol for the Standard Digital Interface for Programmable Instrumentation, IEEE Standard 488.1–2003*, IEEE Publications.

Iotech Inc. (1991) *Instrument Communications Handbook*, Iotech Inc, Cleveland, OH.

Lang, T.T. (1987) *Electronics of Measuring Systems*, John Wiley & Sons, Chichester.

Lang, T.T. (1991) *Computerised Instrumentation*, John Wiley & Sons, Chichester.

Senbon, T. and Hanabuchi, F. (eds) (1991) *Instrumentation Systems: Fundamentals and Applications*, Springer-Verlag, Berlin.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

171: Local Area (LANs) and Wide Area Networks (WANs)

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Introduction to Local Area Networks 1142
2 Optical Communications 1143
3 Star Topology 1143
4 Ring Topology or Token Ring Network 1144
5 Bus Topology 1144
6 Hub Topology 1145
7 Wide Area Network Principles 1145
Related Articles 1145
Further Reading 1145

1 INTRODUCTION TO LOCAL AREA NETWORKS

Compact, high-speed data exchange through a single medium places special requirements on distributed measurement systems where peer-to-peer communication may be necessary. Peer-to-peer communication between each station does not require the intervention of a separate bus controller to intervene and manage the communication process as occurs in the IEEE.488 system. Local area networks, or LANs, which were developed to meet this need, are generally used to interconnect computers on the same floor of a building, within a building of an organization, or between local buildings as in a factory complex. They may also be used for process-control interfacing with intelligent nodes on the network. The LAN provides a high-speed data link over a limited working area. Many types of LAN have been developed for various purposes, but they can be simply classified according to their physical configuration, or topology. The four main topologies are the star, the ring, the bus, and the hub configurations. More details are given later.

The access methods or techniques for transferring data to the transmission medium, include the token passing or baton passing scheme and the carrier wave multiple access with collision detection, or CSMA/CD method. In the token pass method, a frame, or token, makes rounds to all the devices on the LAN. Only the device that is to receive the token can send a signal to the medium. In the CSMA/CD method, any device can send a signal at any time if there is no signal on the medium, thus signal collisions may occur. At such times, each device is made to wait for a set time, which is different for each device.

The CSMA/CD method is efficient when the traffic load is light. However, when the load increases, multiple conflicts develop regardless of the waiting. As a result, efficiency is lost and response time is lengthened. Degradation of response time beyond this limit is unacceptable in a network for factory control and management. Thus, the token passing method is most often used for factory networks. The Manufacturing Automation Protocol, or MAP, and the Process Data-way, or PROWAY, are examples. The ETHERNET method, which is a bus-type connection used extensively for light-load situations, such as in offices and in research and design facilities, is a typical example of CSMA/CD.

The two main standards commonly used are the IEEE-802.4 and the PROWAY. Brief specifications for the IEEE-802.4 method, which is intended for factory use and is also

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

Table 1. Specification summary of IEEE-802.4 and PROWAY.

Item	IEEE 802.4	PROWAY
Physical		
Structure	Bus	Bus
Transmission medium	75 Ω coaxial cable	75 Ω coaxial cable
Transmission distance	–	2 km max.
Node numbers	–	100 max.
Signals		
Modulation method	Phase continuous FSK Phase coherent FSK Broadband	Phase continuous FSK
Control method	Token passing	Token passing (Baton passing)
Error control	No data error monitoring	Retransmission and recovery by immediate acknowledge monitoring
Timing		
Transmission rate	1, 5, 10 Mbps	1 Mbps
Access time (worst case)	In the case of no transmission error, it is decided by station numbers and working parameter value	50 m or less

employed in MAP, are listed in Table 1, together with the specifications for PROWAY.

2 OPTICAL COMMUNICATIONS

Optical fibers are often used as the transmission medium for LANs. When used as a transmission medium, optical fibers have advantages such as insensitivity to electrical noise, the capability for high-speed transmission over great distances, and safety because they do not transmit electric power. For these reasons, they have great potential for widespread use as a transmission medium in factories.

In the recent past, optical connections were difficult because of their high cost and their poor compatibility among the products of different manufacturers. Previously, optical media have been used for special environments and high-performance communications. However, these difficulties are being steadily overcome. It is inevitable that the use of optical communication media will spread.

When it was difficult to split, or branch, the optical fibers used in communications, they were almost exclusively used in one-to-one communications or in ring-type communication configurations. Nowadays, the production technology has advanced sufficiently to broaden their area of application. Optical fibers have now been introduced to bus LANs and star LANs by using a multiple-branch element called the *star coupler*. Furthermore, RS 232 connectors, with built-in optical modems for a duplex communication, have been developed. Low-cost, convenient optical communication is now becoming more widespread.

3 STAR TOPOLOGY

The best-known star type is the *private automatic branch exchange* or the PABX system shown in Figure 1(a). Many large companies such as electrical generating and gas distribution and processing enterprises, as well as water and sewerage, and petroleum and crude products companies use it. The PABX unit, which uses switching to direct communications between points on the star, is an analog communication system. It was initially designed to route telephone calls. PABXs are being replaced by digital switching exchange units called *private digital exchange* or PDX systems as represented in Figure 1(b). An advantage of this replacement method is that much of the wiring is already in place. The inclusion of computers, ADCs, and DACs – *see* **Article 139, Analog-to-Digital (A/D) Converters, Volume 3** and **Article 141, Digital-to-Analog (D/A) Converters, Volume 3** – in the PDX allows both data and voice signals to be handled. PABXs at different sites may also be linked to form a *wide area network*, or WAN.

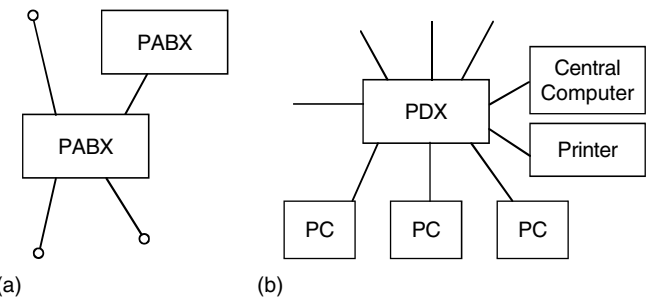


Figure 1. The PABX and PDX local area networks.

Connections between sensors and networks often follow the conventions and standards of the field-bus methods – *see* **Article 172, Fieldbus Systems, Volume 3**.

4 RING TOPOLOGY OR TOKEN RING NETWORK

In the ring topology, shown in Figure 2, the signal is intercepted and retransmitted at the connecting point, or node, for each device. In this network, the communication cable is connected from one node on the network to the next to form a loop or ring. Data passing round the network is seen and processed by each node to determine whether the message is addressed to itself or should be retransmitted to the next node in the loop. One node on the network acts as master, performing management functions, such as token generation. Thus, any node is allowed to communicate directly with any other node on the network.

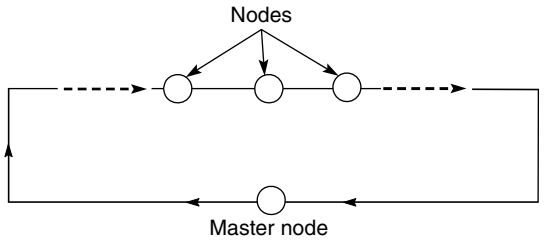


Figure 2. A ring LAN.

The network traffic is unidirectional. Data rates are typically 1 to 10 Mbps.

5 BUS TOPOLOGY

Here, a single cable is passed through nodes on the network so that equipment located in various rooms and floors of an organization can be serially connected. The equipment taps into the cable to receive data on the bus. Data rates are typically 1 to 10 Mbps. The bus topology has high reliability because the transmission route contains no active elements like transistors, LSI circuits, or others in series, unlike the ring and star configurations. Thus, the bus topology is suitable for factory online networks.

A good example of a bus topology is the multifunction vehicle bus, or MVB, which is schematically illustrated in Figure 3. The physical layer of this bus is specified to support a short-distance copper medium and a longer distance optical medium. The copper bus is divided into *bus segments*, with a *bus terminator* at each end. The bus segments may be divided into *bus sections*, which are not terminated. The devices are attached to a segment by passive taps. To extend the bus beyond the limit imposed by electrical constraints, a repeater containing a signal shaper or regenerator may connect segments. Remote bus segments may be connected by an optical repeater link consisting of a pair of point-to-point optical fibers.

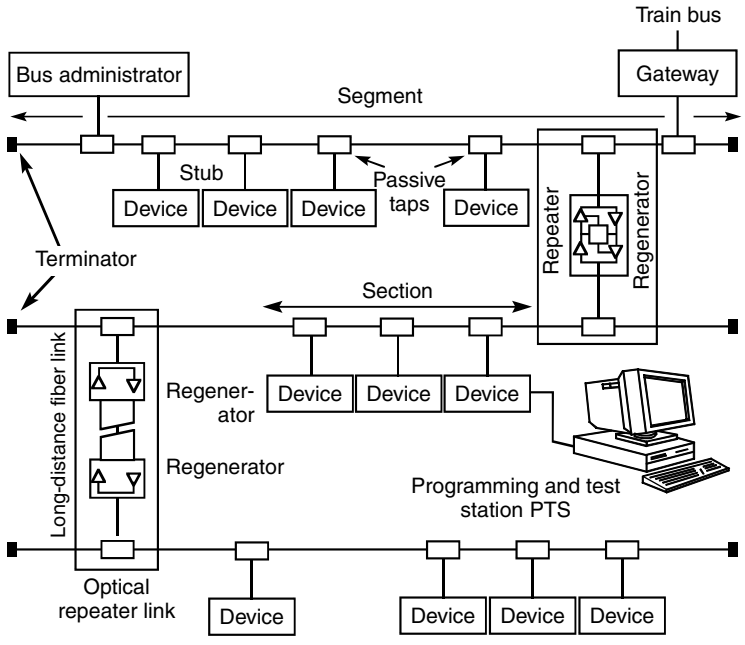


Figure 3. The multifunction vehicle bus used to illustrate a bus configured instrumentation system.

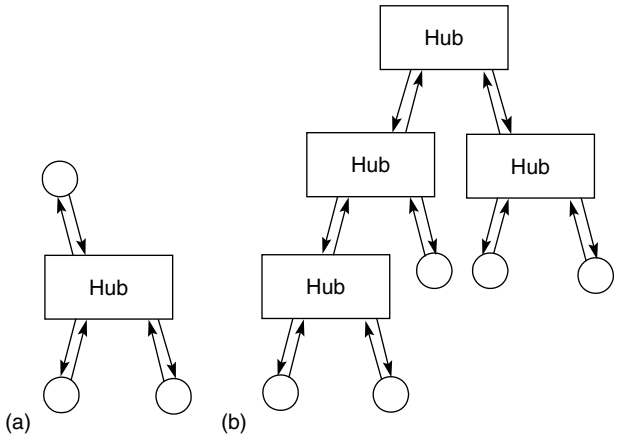


Figure 4. A hub topology LAN.

6 HUB TOPOLOGY

Hubs, which are configured, as shown in Figure 4(a), are simply multiple repeater units for signals received from all the nodes on the network. They perform no switching function as in the star topology. The formation still uses the basic bus and ring topologies. Hub trees, as in Figure 4(b), may also be formed.

7 WIDE AREA NETWORK PRINCIPLES

Local area networks may be connected via gateways to public or private communication networks to form *wide area networks*, or WANs, connecting different business sites of one company across a large city. A gateway, as in Figure 5, allows two systems having different methods or protocols, to communicate with each other. The gateway handles differences in speed, coding, and protocols.

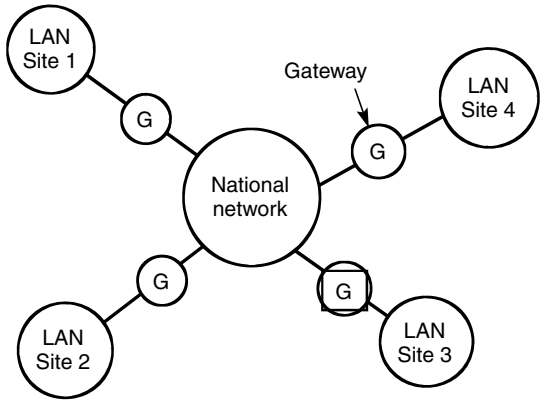


Figure 5. A wide area network, or WAN, topology.

A WAN may be either a publicly or a privately operated network or possibly a mixture of the two. In the United Kingdom, the main network is private. To control and protect the passing of data, the Council of Europe issued a Convention on Data Protection in 1981. This restricts the free flow of data within the EU.

Networks are costly to install and use because of telecommunication charges. Careful thought must be given to the extent to which communication and information via networks is necessary. IBM is an example where instant communication and information is considered important by management. They installed a network in 1987 to satisfy the five point '*any*' rule for their employees. The five '*anys*' are

> *That* **any** *employee at* **any** *IBM location should be able to go to* **any** *terminal and log onto* **any** *IBM application running in* **any** *IBM computer centre*.

This system caters for 20 000 IBM employees in the United Kingdom. The high cost and security aspects have prevented other companies having such an open policy. Is such a system necessary and do all employees need to have access to such a wide range of applications and information? Company IT costs can be very high due to the information demands of staff who do not have a real need for such information. Strategic use of IT funds is necessary.

RELATED ARTICLES

Article 15, Typical Measurement Systems Architectures, Volume 1; **Article 165, Introduction to Networked Instrumentation, Volume 3**; **Article 166, Instrument Interconnection, Volume 3**; **Article 172, Fieldbus Systems, Volume 3**.

FURTHER READING

Anderson, C. and Minasi, M. (1999) *Mastering Local Area Networks*, Sybex International, London.

Brooks, K. (2001) *Networking Complete*, Sybex International, London.

Dennis, A. (2002) *Networking in the Internet Age*, Wiley, Chichester.

Duck, M., Bishop, P. and Read, R. (1996) *Data Communications for Engineers*, Addison-Wesley, Harlow.

Hodson, P. (2001) *Local Area Networks*, Continuum International Publishing Group, New York.

Iotech Inc. (1991) *Instrument Communications Handbook*, Iotech Inc, Cleveland, NE.

Senbon, T. and Hanabuchi, F. (eds) (1991) *Instrumentation Systems: Fundamentals and Applications*, Springer-Verlag, Berlin.

Stallings, W. (2000) *Local and Metropolitan Area Networks*, Prentice Hall, Englewood Cliffs, NJ.

Thompson, A. (2000) *Understanding Local Area Networks: A Practical Approach*, Prentice Hall, Englewood Cliffs, NJ.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

172: Fieldbus Systems

Halit Eren
Curtin University of Technology, Perth, Western Australia, Australia

1 Introduction to the Fieldbus 1147
2 Communicating Sensors to Networks 1147
3 Elements of Connecting Sensors 1148
4 Network Protocols 1149
5 The Fieldbus Concept and Systems 1150
6 Characteristics of Fieldbus Implementation 1151
7 Case Study of Fieldbus Application for a Measurement System 1151
8 Designing, Implementation, and Testing of Fieldbuses 1152
9 Conclusions 1153
Further Reading 1153

1 INTRODUCTION TO THE FIELDBUS

The fieldbus is a digital means of integrating field devices into the information and control system of the entire plant. Thus, the fieldbus communication systems have access to the physical medium by the use of Data Link Layer (DLL) and the communication takes place by the protocols. Fieldbus protocols are largely developed for distributed processes with centralized control philosophy (as in the IEC/ISA fieldbus), or distributed processes with distributed control philosophy (as in the Profibus).

Many versions of the fieldbus are offered by a diverse range of vendors driven by commercial interests or application-specific reasons. There is a worldwide attempt to standardize fieldbus protocols so that it becomes independent of vendors or application specific requirements, a situation that has existed for many years.

Although this had been realized by the IEC-61158 standards in the year 2000, its effect and wide acceptability is yet to be seen. A common fieldbus standard will allow the implementation of control strategies universally, thus giving advantages such as improving intrinsic safety, giving better security, enabling easy installation, maintenance, monitoring and diagnostics of the equipment resulting in interoperability, and yielding cost reduction for the vendors, and ultimately for the end users.

This article discusses the fieldbus as a part of a digital control system, provides a short market survey of the existing fieldbuses, and analyzes future trends.

2 COMMUNICATING SENSORS TO NETWORKS

The fieldbuses interconnect field devices. A typical interconnection of field devices is illustrated in Figure 1. In a complete fieldbus configuration, a variety of components such as sensors, instruments, analog and digital components, and controllers are connected to the system for measuring, monitoring, and control purposes. These interconnected field components can be grouped into three major categories:

1. traditional analog and discrete input/outputs (I/O)
2. hybrid analog and digital devices, and
3. purely digital devices.

Analog devices are usually connected to controllers by means of dedicated 4 to 20 mA analog current loops. The hybrid components are capable of using both analog and digital communication techniques. For example, a protocol called *Highway Addressable Remote Transducer* (*HART*)

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

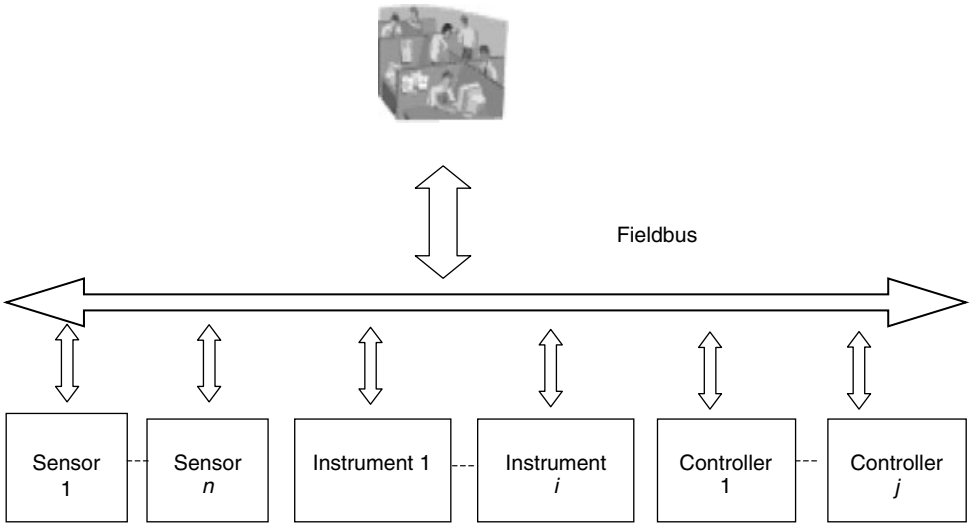


Figure 1. Connection of sensors, instruments, and controllers to a fieldbus.

uses digital communication signals onto a conventional 4 to 20 mA analog signal. The third type, purely digital components, usually requires digital interface hardware and software such as the RS-232, USB, and IEEE 488 parallel buses.

In the wake of relatively new technology on intelligent or smart sensors – *see* **Article 160, Smart Sensor System Features, Volume 3** – the communication may be realized by new wired-in techniques, such as I^2C or I^2S, Integrated Smart Sensor buses, and many others. Some of these buses have the following features:

- they are simple, requiring only two lines for communication (clock line and data);
- the required supporting electronic componentry is at a minimum;
- the transmission can be terminated either by the master or the sensor, an example being the Motorola MCM20027 color image sensor.

Nowadays, many instruments and intelligent sensors communicate by wireless techniques using protocols such as Bluetooth, UWB; 802.11a/b/g (Wi-Fi); GPRS; ZigBee; 802.15.4; IEEE P1451.5, and so on.

Communication by the use of radio frequencies (RF) is now a well-established field. It can be realized by standardized or nonstandard methods. The choice of these methods depends on the application requirements and complexity of the instruments, and on the sensors and capabilities of the host fieldbus system. Some RF techniques have a high level of power consumption so it may not be convenient in some applications such as in portable wireless instrumentation systems.

3 ELEMENTS OF CONNECTING SENSORS

Almost all modern fieldbus systems operate on digital and computerized systems. As yet, the majority of sensors still produce analog signals thus requiring analog-to-digital (A/D) conversion by means of suitable converters. Similarly, many controllers require analog signals; hence, digital signals from the bus need to be converted back into analog forms – *see* **Article 132, Data Acquisition Systems (DAS) in General, Volume 3**.

Typical elements needed for the connection of sensors to the digital environment are illustrated in Figure 2. Front-end analog-signal processing is necessary for amplification or attenuation and filtering of signals generated by the sensors. In addition to analog-signal processing, the A/D conversion process takes place with the aid of appropriate sample-and-hold devices and multiplexers. Multiplexers allow the connection of many sensors and transducers to the same signal-processing media.

Analog-to-digital conversion involves three stages: sampling, quantization, and encoding. In all-digital instruments, the Nyquist sampling theorem must be observed; that is, 'the number of samples per second must be at least twice of the highest frequency present in the continuous signal'. As a rule of thumb, depending on the significance of the high frequencies, the sampling must be about 5 to 10 times of the highest frequency component in the signal. The next stage is the quantization, which determines the resolution of the sampled signals. Quantization involves errors that decrease as the number of bits increases. In the encoding stage, the quantized values are converted to binary numbers to be processed digitally.

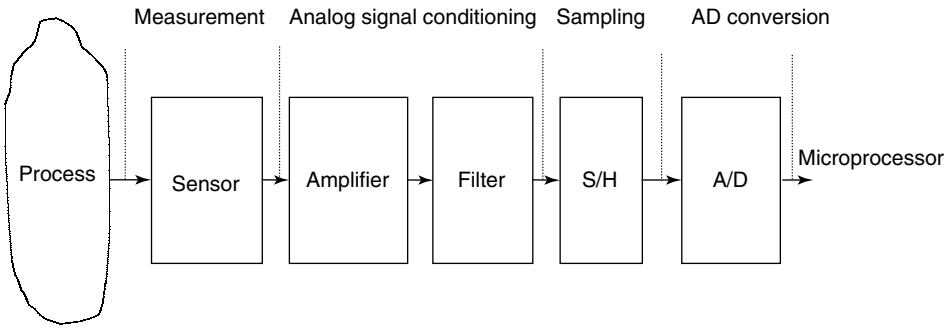


Figure 2. Basic elements of a sensor linking to a fieldbus.

For the process actuators and controllers, the digital signals are converted to analog waveforms by D/A converters and then filtered and amplified before being applied to the controllers.

Nowadays, due to easy and cost-effective availability of advanced microprocessors and the supporting components, many sensors and instruments are based on powerful digital systems integrated at various stages, starting right from the sensing elements through complex communication capabilities. Therefore, microprocessors, microcontrollers, and digital signal processors constitute the other essential elements for interconnecting sensors to the fieldbuses. The use of microprocessors makes the sensors and controllers self-supporting allowing them to take local decisions under the instruction of the fieldbus system. Microprocessors also allow convenient man–machine interface, diagnostic capabilities, self-configuration, and self-identification.

4 NETWORK PROTOCOLS

Field instruments are connected to digital systems and computers, thus forming networks such as Direct Digital Control (DDC) systems, Supervisory Control (SC) systems, Distributed Control Systems (DCS), Hybrid Control Systems (HCS), Supervisory Control and Data Acquisition (SCADA) systems, and so on.

The interconnected devices communicate through a particular common bus network using communication protocols (defined ways of doing things) that are developed to guarantee precise intended information flows between components.

In industrial systems, different control units may need to share the data collected by a sensor, and disseminate the processed data to the various actuation units. Thus, transactions involving one producer on the field and several consumers of a real-time variable have to be supported. In this respect, a focus on the group communication on the shop floor bears particular importance. These types of communications are realized by the data link layer protocols or the Fieldbus protocols. These protocols provide a series of facilities, which help the application process to check that the real time operation–requirements are met.

The messages that sensors generate can be divided into short and long messages. While the short messages are usually processed locally, the long messages are processed by the central computers or remotely located computers. The length and type of messages is fairly important in the configuration of fieldbus. In this respect, the fieldbus forms the first stages of information gathering, data flow and decision making about the process under control. Some examples are the integration of the fieldbus with Computer-aided Facility Management (CAFM) systems and the integration of fieldbus systems into Distributed Object-oriented Systems (DOOS).

As mentioned earlier, the information flow between the nodes, individual instruments, and the computers are regulated by protocols. According to the IEEE, a networking protocol is ‘a set of conventions or rules that must be adhered to by both communicating parties to ensure that information being exchanged between two parties is received and interpreted correctly’. The protocol defines the following issues:

1. Network topology supported: star, ring, or bus (tree), and so on.
2. ISO reference model layers implemented: physical, datalink, network, transport, session, presentation, and application.
3. Data communication modes: simplex, half-duplex, or duplex.
4. Signal types: digital or analog.
5. Data transmission modes: synchronous (include bit-oriented and character-oriented), or asynchronous, and so on.
6. Data rate supported: several bps (bits per second) to several Gbps (gigabytes per second), it depends on both oscillator frequency and transmission medium.
7. Transmission medium supported: twisted pair, coaxial cable, optical or microwave, and so on.

Table 1. ISO 7-layer reference model for communication and networks.

No	Layer	Application	Protocols
7	Application	Common application service elements (CASE); Manufacturing message services (MMS); File transfer and management (FTAM); Network management.	ISO 8650/2 (DP), RS-511, ISO 8571 (DP), IEEE802.1.
6	Presentation	Transformation of information such as file transfer. Data interpretation, format and code transformation.	Null/MAP transfer. ISO 8823 (DP).
5	Session	Communication and transaction management, synchronization, administration of control sessions between two or more entities.	ISO Session Kernel. ISO 8237 (IS).
4	Transport	Transparent data transfer, mapping, multiplexing, end-to-end control, movement of data elements.	ISO Transport, Class 4. ISO8073 (IS).
3	Network	Routing, switching, segmenting, blocking, error recovery, flow control. Wide area relaying.	ISO DIS 8473, Network services, ISO DAD 8073 (IS).
2	Link	Transmission of data in local network. Establish, maintain and release data links, error and flow.	IEEE 802.4 Token Bus. IEEE 802.2 Type 1 Connection services.
1	Physical	Electrical, mechanical and packaging specifications. Functional control of data circuits.	ISO/IEEE802.4, phase Coherent Carrier, Broadband, etc.

8. Medium access control methods: Carrier Sense Multiple Access with Collision Detection (CSMA/CD), or control token, and so on.
9. Data format: mainly based on data transmission modes and individual protocol specifications.
10. Error detection methods: parity, block sum check, or Cyclic Redundancy Check (CRC), and so on.
11. Error control methods: echo checking, Automatic Repeat Request (ARQ), or sequence number I, and so on.
12. Flow control methods: X-ON/X-OFF, window mechanisms, or sequence number II, and so on.

Table 1 shows ISO seven-layer reference model layout for the protocols and some of the protocols that are relevant to that particular layer. The ISO reference model has seven layers of which each is an independent functional unit. Each one uses functions from the layer below it and provides functions for the layer above it. The lowest three layers are network-dependent layers, the highest three layers are network-independent layers (application-oriented) while the middle layer (transport layer) is the interface between the two.

Most fieldbuses use only a few of these layers, being usually the physical, data link, network, and application layers.

5 THE FIELDBUS CONCEPT AND SYSTEMS

The fieldbus is an industrial communication system that is configured and supported by single companies or groups of companies or private and government agencies. In the course of development of the fieldbus, unfortunate proliferation took place because many different technology approaches existed among the vendors serving their commercial desire for market dominance.

Implementation of a fieldbus involves hardware as well as software, with a trade-off being necessary between the two. Most fieldbuses have associated chips for implementing all or portions of system specifications. For example, in LonWorks protocol, all layers can be implemented by chips and associated software. The software defines all the objects and libraries that are readily available for the customer. But, for another CAN, the lower two layers are implemented by protocol chip (82C200); all the software has to be done by the user.

There are now more than 50 different fieldbus products or standards in the marketplace being reported by manufacturers in conferences, magazines, and exhibitions. Table 2 lists some of the most commonly used fieldbuses. Other fieldbus products not included in the table are the I/O-lightbus, Foundation Fieldbus, ICE 61 158, P-NET, INSTA

Table 2. Some examples of commonly available fieldbuses.

Research	
MPS (Michigan Parallel Std., MSS	University of Michigan
I^2S	Delf University of Technology
Time-Triggered Protocol	University of Wien
Industrial	
DeviceNet	Allen-Bradley
SP50 Fieldbus	Fieldbus Foundation
LonTalk/LonWorks	Echeleon Corp
Profibus DP/PA	Siemens
ASI-Bus	ASI
InterBus-S	InterBus-S club, Phoenix
SERCOS	VDW
IEEE-488	HP
ArcNet	Datapoint
WorldFIP	WorldFIP

(EIB), Modbus, Bitbus, ArcNet, HART, and so on. The list continues to grow with movement toward any form of Open Systems Interconnect (OSI) methodology, like the technology that made computers easy to use and is still to come. In the mean time, users will need to cover the extra cost of customizing applications with additional middle-ware code and hardware to connect devices to communication channels, where they are not all made to the same standards.

6 CHARACTERISTICS OF FIELDBUS IMPLEMENTATION

Implementations of fieldbuses vary from system to system as vendors offer different forms of supporting hardware and software. Also, the nature and requirements of processes differ from each other. In order to explain characteristics of fieldbus implementation, Foundation Fieldbus will be taken as an example.

The Foundation Fieldbus is an all digital, serial, two-way communication system, which interconnects field equipment such as sensors, actuators, controllers, and computers. It is essentially a Local Area Network (LAN) for instruments used in the manufacturing, automation, and process industries, which has built-in capability to distribute the control application across the network. It is supported by an independent nonprofit corporation, known as the *Fieldbus Foundation* consisting of over 100 of the world's leading suppliers, manufacturers, and end users. Foundation Fieldbus, created in 1994, is a joint effort of two major organizations; the Interoperable Systems Project (ISP) and WorldFIP/North America. Its committee is formed by the members of the IEC and ISA, backed up by Profibus of Germany and FIP of France.

Foundation Fieldbus is particularly suitable in systems where the communication traffic involved is the periodic exchange of short information frames. However, purely periodic exchange of short information frames is not sufficient to meet all the communication requirements of the process. In this respect, the implementation of the communication models offered by the application layer through the access medium mechanisms are made available at the data link level. Therefore, asynchronous message (e.g. alarms, operator commands) transmissions are allowed as well as synchronous transmission. This approach is the key point in the definition of this particular protocol, as only correct mapping between the application layer communication models and data link layer mechanisms can guarantee to meet the time-critical requirements of the processes.

An important characteristic of the Foundation Fieldbus is the function block that defines the inputs, outputs, mode structures, and event subsystems for common control systems. These blocks provide a consistent and easy-to-use method for modeling standard control system functions. As far as the control computers are concerned, the system is seen as a collection of function blocks rather than as a collection of instruments. An important contribution of the Foundation Fieldbus is the flexibility of creating field devices that are capable of storing and executing the software routines in their own way. With the use of such intelligent devices, the function blocks permit the distribution and integrated execution of process functions among the field devices themselves.

The Foundation Fieldbus topology comprises two basic parts: a lower speed fieldbus called $H1$ and a higher speed fieldbus, the $H2$. These two fieldbuses can be connected together via a device called a *bridge*. The device descriptions are a standard mechanism that permits a control system to acquire the definitions of messages from a field device. Foundation Fieldbus is equipped with information on the device characteristics such as manufacturer's details, supporting software capabilities, available function blocks, and diagnostic capabilities. Such information is developed in Device Description Language (DDL), which originated from the HART protocol. This approach is expected to enable vendors to describe the functions of their devices in a standard way. They can also add new capabilities and novel features to the devices that can easily be accessed by the existing systems and users.

7 CASE STUDY OF FIELDBUS APPLICATION FOR A MEASUREMENT SYSTEM

Fieldbuses are applied across all cross sections of industry ranging from under the bonnet in cars to complex manufacturing plants consisting of tens of thousands of sensors, transducers, and controllers. Many companies offer interoperable products with a particular fieldbus. For example, the System Integration Specialist Company (SISCO) of Michigan, USA, offers a broad range of fieldbus products including adapter cards, servers, interface boards, programming tools, bus analyzers, protocol source codes, and so on. The company also offers fieldbus pressure transmitters, temperature transmitters, programmable logic controllers, fieldbus-to-current converters, fieldbus-to-pneumatic-signal converters, valve petitioners, and other accessories. Also, many companies are producing wireless connections for accessing distant sensors and actuators.

As an example of implementation of fieldbuses, Fuji Electric of Japan has developed open systems in the field of instrumentation control and promoted the establishment of Foundation Fieldbus standards. The company has been working on fieldbus specifications, hardware developments,

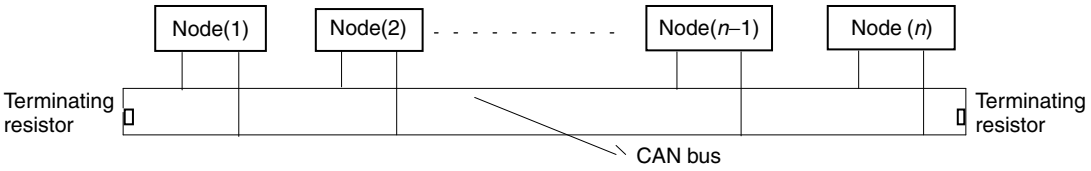


Figure 3. A typical CAN network layout in a laboratory environment.

promotions for increased usage, and public relations. Fuji Electric has started the fiber optic fieldbus working group, and the results of the work have been finalized as a final specification in the Fieldbus Foundation.

In a laboratory environment, the author implemented the CAN protocol, as illustrated in Figure 3. This particular system contains a number of sensors and actuators. From amongst many other CAN controllers, the 82C200 chip from PHILIPS was selected.

In this project, the implementation was realized by three sublayers. These were the Logical Link Control (LLC) and Medium Access Control (MAC) sublayers for the data link layer. This CAN implementation allowed messaging between sensors, actuators, man–machine interfaces, controllers, and other control activities. This architecture provided a communications environment upon which a high-speed, real-time centralized or distributed control platform could be created.

8 DESIGNING, IMPLEMENTATION, AND TESTING OF FIELDBUSES

Designing: In the design of a fieldbus system, the process characteristics must be well understood. Once the measuring, monitoring, and control objectives are identified, selection of the appropriate fieldbus will be dependent on costs of the system and integration of the selected fieldbus to the existing instrumentation in the plant. As indicated in Table 2, a suitable fieldbus must be selected for a particular application, such as BACnet for intelligent buildings, Profibus for industrial applications, and so on. Once the decision is made, manufacturer's instructions for installation, commissioning, and maintenance must be followed closely. It is important that in the selection process the interoperability of sensors and instruments offered by different manufacturers must be observed.

Implementation: There are many methods of implementing fieldbuses, some of which will be described here. Apart from the hard-wired connections, wireless implementations of fieldbuses are becoming popular as many of today's instruments often have wireless communication features built in. The most popular means of transmission using wireless systems are the LAN adapters, radio modems, and packet controllers. The difference between the three is in the way that the data is processed between and during the transmission. Wireless transmission devices are used to create complete wireless networks or partial networks with wireless segments of wired networks. Simple fieldbuses such as those based on RS-232C standard (e.g. Modbus) can easily be realized with use of radio modems or packet controllers.

If the network is more complicated or the required transmission speed is higher, LAN cards and more powerful processors are used. A number of reports describe solutions for wireless transmission in the networks such as LonWorks, CAN, or FIP.

The possibility of having wireless transmission is also included in the Foundation Fieldbus standard. At the moment, only low speed radio networks are standardized, but inclusions of higher speed radio networks are planned to be implemented in the near future.

Another implementation uses fuzzy component networks such as fuzzy sensors, fuzzy actuators, and fuzzy inference on fieldbuses. These fuzzy components are formed in groups to form fuzzy cells. Depending on its configuration, each fuzzy cell is able to perform the functions of one or several fuzzy components. Fuzzy cells are characterized by their ability to handle and exchange fuzzy symbolic information. Their configuration is performed via the fieldbus using the language Prototype of Integrated Language for Symbolic Actuators and Sensors (PLICAS), which is specially designed for this purpose. A PLICAS compiler is integrated inside each cell, thus conferring interoperability properties to the cell.

In time-critical communication systems, the Hopfield Neural Network (HNN) is extensively used. In time-critical systems, the correct implementation of proper scheduling sequence is very important. The implementation of the sequence must cater for the instants at which each transmission requirement must be fulfilled within the time constraints. Generally, the length of the scheduling sequence is equal to the minimum common multiple of the periods of the processes and needs to be memorized for later implementation. During this process, some problems arise if the sequence is large. The computational complexities arising in the sequences are overcome by adopting an HNN integrated with the fieldbus.

One important implementation is the integration of fieldbuses with the Internet. There are many examples of such systems that combine the benefits of the two to create synergy. It appears that the protocol Simple Network Management Protocol (SNMP), originally designed for the Internet, can be suitable for this purpose. This combination provides access to the fieldbus system from local or global computer networks by the use of common LAN management tools.

Testing: Testing of the operations of a fieldbus is conducted during the commissioning stage. Manufacturers supply self-testing and diagnostic capabilities under normal operating conditions. For testing and troubleshooting, manufacturer's recommendations must be adhered to.

9 CONCLUSIONS

Fieldbuses are industrial communication protocols that come with appropriate hardware and software for interconnecting field devices to computer or digital networks. There are many types of fieldbuses offered by a diverse range of manufacturers. In the implementation of fieldbuses, the interoperability of devices is important. Fieldbuses can be integrated with wireless instrumentation systems and the Internet.

FURTHER READING

Cena, G., Demartini, C. and Valenzano, A. (1997) On the Performances of Two Popular Fieldbuses, in *Proceedings of the IEEE International Workshop on Factory Communication Systems WFCS'97*. (Cat. No. 97TH8313) IEEE, New York, 177–186,.

Demartini, C. and Valenzano, A. (1998) The EN50170 Standard for a European FieldBus. *Computer-Standards-&-Interfaces*, **19**(5–6), 257–273.

Eren, H. (2004) *Electronic Portable Instruments-Design and Applications*, CRC Press, LLC, Boca Raton, FL.

Eren, H. and Fung, C.C. (1999) Automation and control equipment and implementation, in *Encyclopedia of Electrical and Electronics Engineering*, (ed. J.G., Webster), Wiley, New York.

Jordan, J.R. (1996) *Serial Networked Field Instrumentation*, Wiley, Chichester.

Mahalik, N.P. (2003) *Fieldbus Technology: Industrial Network Standards for Real-Time Distributed Control*, Springer Verlag, Berlin.

Santori, M. and Zech, K. (1996) Fieldbus Brings Protocol to Process Control. *IEEE Spectrum*, **33**(3), 60–64.

Schumny, H. (1998) Fieldbuses in Measurement and Control. *Computer Standards & Interfaces*, **19**(5–6), 295–304.

Stallings, W. (1994) *Data and Computer Communications*, 4th edn, Maxwell Macmillan International, New York.

Vitturi, S., Rivaroli, L., Pasqualini, M. and Zannotti, S. (1997) Communication in Industry: A Profibus Application. *Automazione-e-Strumentazione*, **45**(4), 167–171.

173: Scheduling Systems

Emil Michta
University of Zielona Gora, Zielona Gora, Poland

1 Timing Analysis in Networked Measurement–Control Systems 1154
2 Introduction to Scheduling Systems 1155
3 Scheduling Theory 1155
4 Functional Model of NMCS 1157
5 Tasks Modeling 1158
6 Scheduling Tasks on Node 1159
7 Scheduling Messages Across a Network 1160
8 Trends 1160
Related Articles 1161
References 1161

1 TIMING ANALYSIS IN NETWORKED MEASUREMENT–CONTROL SYSTEMS

NMCS (networked measurement–control systems) are widely used in applications ranging from difficult process control to simple discrete manufacturing.

Usually, these systems impose real-time requirements on the nodes and communication networks (Kopetz, 1997; Sha and Sathaye, 1995). Stankovic (1988) defines real-time computing systems as those systems in which the correctness of the system depends not only on the logical result of computation but also on the time at which the results are produced.

There are many examples of real-time computing systems, such as control systems, flight control systems, and robotics. A typical real-time computing system has a real-time program running in the system, which reads inputs from input devices, processes these inputs, and produces outputs to be sent to output devices – *see also* **Article 78, Embedded Systems, Volume 2**.

The time between the arrival of an input from a device and the completion of the processing for that input is called the *response time* for the device (Joseph and Pandya, 1986).

The relative deadline for the device can be defined as the maximum interval between the instant of the input arrival and the completion of the processing for that input. The required response time for a device must be smaller or equal to its relative deadline.

In NMCS, the relative deadline is often assumed to be equal to the period. The task deadline value is set by object requirements but the value of the worst-case response time is a result of simulation methods or analytical techniques used to do the timing analysis.

It is assumed that each input and output device is assigned a task of the application program and that the tasks share the same processor. The problem of determining whether the processor will meet its peak-processing load becomes one of schedulability analysis of tasks in NMCS (Burns, 1991).

If the input (sensor node) and output (actuator node) devices are network components, then the input device processes events and generates messages to the output device. The problem of determining whether the communication system will meet its peak load is one of schedulability analysis of tasks.

Consideration of priority levels is crucial to a real-time system. If different inputs have different response time requirements, we need to consider different priority levels to schedule the related processing tasks. The problem is to determine whether for a given assignment of priority levels, the system will meet its peak load – *see also* **Article 165,**

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

Introduction to Networked Instrumentation, Volume 3 and **Article 166, Instrument Interconnection, Volume 3**.

2 INTRODUCTION TO SCHEDULING SYSTEMS

Developers and integrators of NMCS face the challenge of meeting end-to-end real-time requirements. This requires techniques to ensure timing correctness both in embedded nodes and in the communication network. Schedulability analysis can be applied to prove that time requirements are satisfied.

Scheduling is defined as a resource (processor, network) allocation required to task (program, message) execution and its termination in the required time.

The general problems are to find a sequence in which tasks pass between the resources, that there is a feasible schedule, and that it is optimal with respect to some decided economic and/or performance criterion.

Reliable timing behavior is crucial in NMCS with real-time requirements. In many cases, designers and/or developers have had to rely on testing to establish timing reliability. Using suitable scheduling methods, system developers can verify timing behavior at the design stage and implement the required timing behavior into the application right from the start. Scheduling allows estimate resource utilization and establishes relationships between the task's worst-case response times and their deadlines. It determines whether a designed system is schedulable and indicates the margin by which a deadline is met or might be missed.

A large number of approaches to modeling and solving scheduling problems have been reported in the literature for this topic. This article focuses on the static approach because this methodology best fits the NMCS. Scheduling theory provides a well-founded way of predicting whether a designed system having real-time requirements will meet their deadlines or not.

3 SCHEDULING THEORY

3.1 Classification

The schedulability analysis can be performed online or offline as it is shown in Figure 1. In the online case, the schedulability of the task set is analyzed at run time, whereas in the offline case, it is performed prior to run time.

Offline scheduling requires little run-time overhead and the schedulability of the task set is guaranteed before execution. However, it requires prior knowledge of the tasks timing characteristics. If the tasks characteristics are not known prior to the run time, schedulability analysis must be performed online.

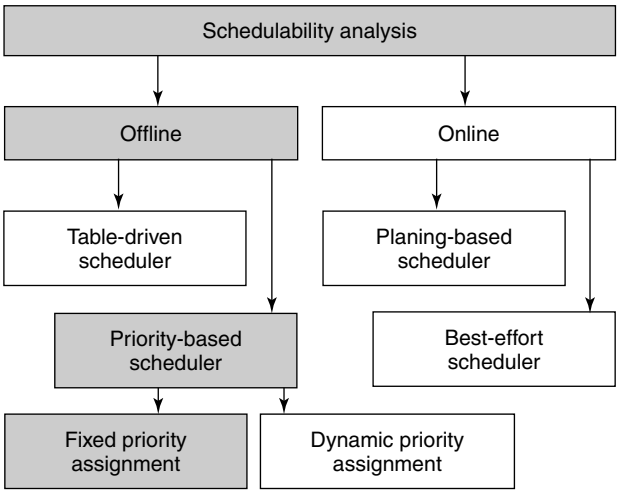


Figure 1. Basic types of schedulability analysis.

Ramamritham and Stankovic (1994) distinguish two basic types of online schedulers: planning-based and best-effort schedulers. In the planning-based type, when a new task arrives, the scheduler tries to redefine a new schedule, which is able to comply with both the requirements of the new task and the requirements of the previously scheduled tasks. The new task is only accepted for execution if the schedule is found feasible.

In the best-effort type, when a new task arrives, the scheduler does not try to perform a new schedule. The new task is accepted for execution; the system tries to do its best to meet deadlines.

Offline scheduling paradigms depending on whether the schedulability analysis produces a schedule according to which tasks are dispatched at run time.

The table-driven approach is the best-known example of an offline scheduling that produces a schedule.

The priority-based approach is an example of offline scheduling where no explicit schedule is constructed.

At run time, tasks are executed in a highest-priority-first basis. Priority-based approaches are much more flexible and accommodating than table-driven approaches.

3.2 Priority assignment schemes

The most used priority assignment scheme is to give the tasks a priority level based on its period: the smaller the period, the higher the priority. Thus, if they have smaller periods, their worst-case response time must also be smaller. This type of priority assignment is known as the RM (rate monotonic assignment).

RM scheduling theory ensures that, for a given fixed set of tasks with fixed and known priority ordering, as long as the system utilization of all tasks lies below a certain bound and appropriate scheduling algorithms are used, all tasks meet their deadlines. This puts the development and maintenance of real-time systems on an analytical basis, making these systems easier to develop and maintain. RM theory was introduced by Liu and Layland (1973) for scheduling independent, periodic tasks with end-of-period deadlines. RM is an optimal static priority scheduling algorithm for independent periodic tasks with end-of-period deadlines.

If some of the tasks are sporadic, it may not be reasonable to consider the relative deadline equal to the period. A different priority assignment can then be to give the tasks a priority level based on its relative deadline: the smaller the relative deadline, the higher the priority. This type of priority assignment is known as the DM (deadline monotonic) assignment (Leung and Whitehead, 1982).

In both RM and DM priority assignments, priorities are fixed, in the sense that they do not vary along time. At run time highest priority tasks are dispatched first. A similar dispatching policy can be used if the task, which is chosen to run, is the one with the earliest deadline. This also corresponds to a priority-driven scheduling, where the priorities of the tasks vary with time.

Thus, the EDF (earliest deadline first) method is a dynamic priority assignment scheme. Pre-run time–schedulability analysis of tasks dispatched according to the EDF assignment scheme was introduced by Liu and Layland (1973).

In all three cases, the dispatching phase will take place either when a new task is released or when the execution of the running task ends.

3.3 Preemptive and non-preemptive systems

A common implementation for embedded node software and for a Master/Slave communication system is to use a static cyclic system.

This involves the static creation of the schedule that typically consists of a number of tasks running one after another to form an overall schedule. The cyclic system does not make effective use of the CPU and worst-case response time for each task exceeds its period.

In a cyclic system, all tasks must run at harmonic processing rates. Such a solution can be used for nodes with a small number of tasks and for implementations without real-time requirements. To improve effective use of the node CPU, and to optimize worst-case response time, a task's priority levels with preemptive or non-preemptive scheduling policy can be used.

In a priority-based scheduler, a higher-priority task may be released during the execution of a lower-priority one. If the tasks are being executed in a preemptive context, the higher-priority task will preempt the lower-priority one.

In a non-preemptive context, the lower-priority task will be allowed to complete its execution before the higher-priority task starts execution.

This situation can be described as a priority inversion because a lower-priority task delays a higher-priority task.

3.4 Offline schedulability analysis

There are two basic types of analytical methods to perform pre-run time–schedulability analysis.

One is based on the analysis of the processor or network utilization. The other is based on the response time analysis for each individual task.

In Liu and Layland (1973), the authors demonstrated that by considering only the processor or network utilization of task set, a test for pre-run time–schedulability analysis could be obtained.

An analytical approach is used to predict the worst-case response time for each task. The obtained values are then compared with the relative deadlines of the tasks. Lechoczky (1990) and Audsley *et al.* (1993) improved the schedulability analysis introduced by Liu and Layland (1973).

The utilization-based test needs a simple computation procedure, which is applied to the overall task set. For this reason, they are very useful for implementing schedulers that check the feasibility online. However, utilization-based tests do not give task's response time values. It constitutes sufficient, but not necessary, conditions. For more complicated task models, the utilization-based tests cannot be used, see Tindell (1992).

3.5 Basic utilization-based test

For the RM priority assignment, Liu and Layland (1973) proved that a set of N independent periodic tasks characterized by a worst-case computation time C and a period T scheduled by the rate monotonic algorithm will always meet their deadlines for all tasks, if a utilization-based pre-run time–schedulability test is following.

$$U = \sum_{i=1}^{N} \frac{C_i}{T_i} \leq N \times (2^{1/N} - 1) \tag{1}$$

where: C_i is the worst-case computation time of task τ_i, and T_i is the minimum time between task τ_i releases (period).

Utilization-based tests are sufficient but not necessary conditions. This utilization-based test is valid for periodic independent tasks having relative deadlines equal to the period and for preemptive systems.

For the non-preemptive case Sha, Klein and Goodenough (1991) shows that similar analysis can be adapted to include task τ_i blocking time B_i, during which high priority tasks are blocked by low-priority tasks:

$$U = \sum_{i=1}^{i} \left(\frac{C_i}{T_i} \right) + \frac{B_i}{T_i} \leq i \times (2^{1/i} - 1), \quad \forall_{i,1 \leq i \leq N} \qquad (2)$$

where B_i is the maximum, blocking a task τ_i by lower-priority tasks than task τ_i.

3.6 Response time tests

The task response time test, for determining the interval between its minimum and worst-case values, is a very important time parameter to determine in difficult and safety-critical real-time projects.

To confirm that each task meets its deadline, the worst-case completion time of each task is calculated taking into account the influence of the other tasks. It was proven by Joseph and Pandya (1986) that the worst-case response time R_i, of task τ_i is found when all tasks are synchronously released at their maximum rate. For that instance, R_i can be computed using the following recursive equation:

$$R_i^{n+1} = \sum_{j \in hp(i)} \left(\left\lceil \frac{R_i^n}{T_j} \right\rceil \times C_j \right) + C_i \qquad (3)$$

where *hp(i)* denotes the set of tasks of higher priority than task i priority. The symbol $\lceil \rceil$ is the ceiling function: it is a roundup function.

Consider the initial value for R_i is zero. Recursion ends when $R^{n+1} = R^n = R_i$. If the worst-case response time R_i exceeds T_t (in the case of RM priority assignment) or D_i, (in the case of DM priority assignment) the task τ_i is not schedulable. This result is valid in the preemptive context.

For the case of non-preemptive tasks, the utilization-based test using (3) was updated by Sha, Klein and Goodenough (1991) to include a blocking factor, during which higher-priority tasks can be blocked by low-priority tasks.

Taking into account the blocking factor B_i the worst-case response time for non-preemptive approach can be derived from the following recursive equation:

$$R_i^{n+1} = B_i + \sum_{j \in hp(i)} \left(\left\lceil \frac{R_i^n}{T_j} \right\rceil \times C_j \right) + C_i \qquad (4)$$

where $B_i = \max C_j$ for *lp(i), lp(i)* denotes tasks with lower priority than task τ_i.

4 FUNCTIONAL MODEL OF NMCS

An NMCS is implemented by a set of embedded nodes. Each node runs a number of tasks. These tasks communicate with each other by passing messages between nodes across the shared network. In order to meet the timing requirements, the end-to-end communication delay, composed of sender node generation delay, queuing delay during accessing network, transmission delay, and delivery delay in receiver node, must be bounded.

Now consider the techniques that can be applied both to software within nodes, and to messages passing on shared buses, to determine that deadlines will be met.

To analyze the time constraint of a networked, distributed system, we can distinguish three functional layers: node application layer, communication stack layer and network layer shown in Figure 2 (Michta, 2002).

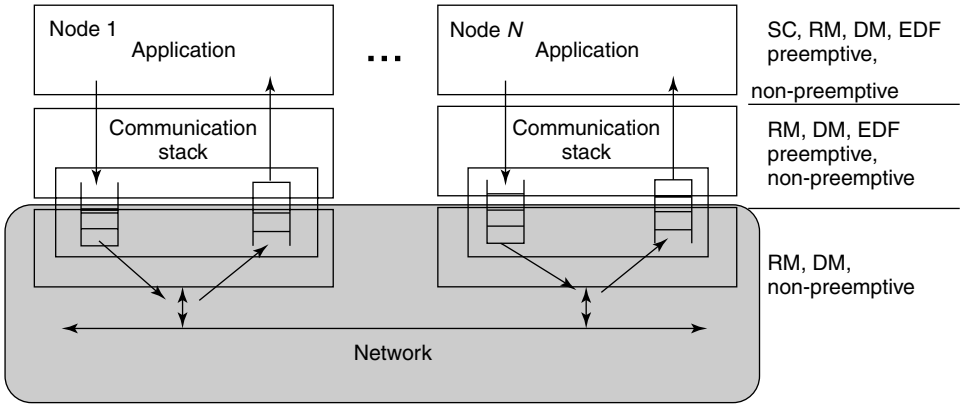


Figure 2. Functional model of NMCS and scheduling approaches.

During analysis of each layer, a combination of different scheduling policies can be used. For example, the network layer may be priority driven with use of nonpreemptive RM or DM polices: nodes may be priority or time driven with use in both preemptive and nonpreemptive approaches. Different nodes may have different scheduling policies due to manufacturer's preferences; thus, during the NMCS designing phase, there exist many influences on the choice of a communication system.

RM, DM, and EDF are a priority-based scheduling algorithm, so the system must have an adequate number of priority levels that can be assigned to tasks on each layer.

The primary objective during designing phase is to decouple the scheduling of resources and analyze the scheduling of each node processor and each network. As we can see in Figure 2, the same techniques of analysis for task scheduling on the processor node and message scheduling across the network can be used. Decoupling resources allow us to separate the scheduling problems and simplify the development and maintenance of nodes and NMCS.

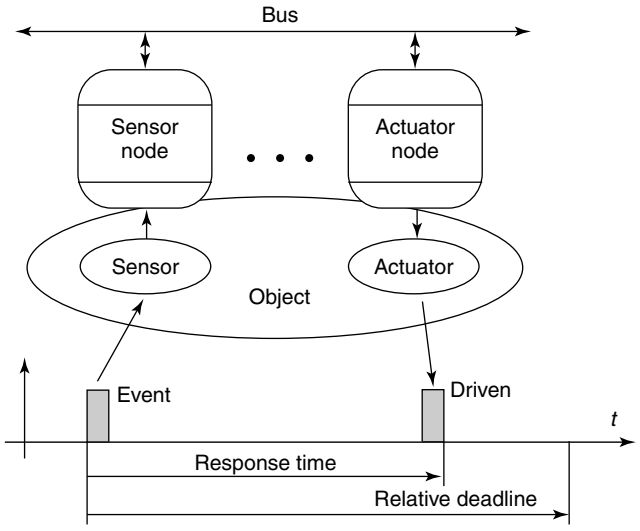


Figure 3. Two nodes NMCS.

5 TASKS MODELING

Consider the simple NMCS shown in Figure 3 with three basic components: communication bus and two nodes: source node (sensor node) and destination node (actuator node). The nodes cooperate throughout the network to provide the end-to-end functionality. In a real-time system, this end-to-end functionality, from event to response, must be provided within a specified deadline. The response time R_i on event i should be lower than deadline D_i. Timing analysis is applied to a resource being shared between multiple activities. According to the NMCS model shown in Figure 2, on an application level, the several sensor and actuator node functions are executed by a single microprocessor. On a communication stack level, the output and input messages wait in queue to be served and finally carry a number of messages sent by nodes on a network layer on one bus.

To simplify timing analysis of a sensor to actuator node communication, we partition the analysis into components that can be analyzed independently, as shown in Figure 4.

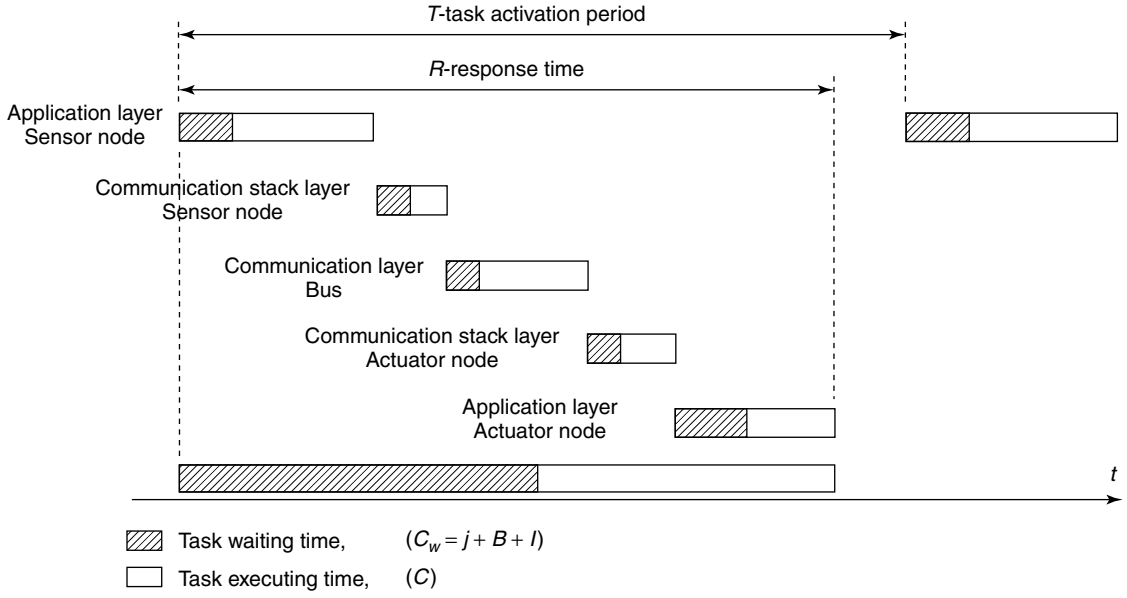


Figure 4. End-to-end communication timing analysis.

For task τ_i, release jitter time J_i, blocking time B_i, and interference time I_i are joined together and represented by the gray area (waiting time) to the left of the execution time windows shown in Figure 4.

End-to-end response time depends on the response time of each component involved in meeting the deadline.

Different offline scheduling policies could be investigated using this simple control loop as an example. If preemption is used, the blocking time corresponds to the longest duration of a low-priority code sequence that must run in mutual exclusion of the task. Without preemption, blocking is the duration of the longest lower-priority task. The interference time equals the total execution time of higher-priority tasks that may appear during the maximum response time.

In Figure 4 are shown execution times of tasks on the two nodes (sensor and actuator), and the communication task execution time on the bus.

An execution window represents each activity. The left end of the box represents task arrival and the right end represents the latest task completion. The length of the box represents the task response time on a given level. The activity may be finished at any time after the minimal execution time represented by the length of the white box. In the worst case, release jitter and blocking and interference from other tasks may delay task execution.

6 SCHEDULING TASKS ON NODE

In this section, the application of the scheduling theory to the design of a hypothetical, simple distributed real-time NMCS is reviewed.

Before going into the analysis, a few assumptions about tasks and node real-time operating system (RTOS) are needed:

- The set of tasks and their priority is fixed and known.
- Each task must have unique priority.
- The task deadline is no longer than its period.

During the designing stage, we can decouple scheduling of resources and analyze independently the scheduling of each node processor and each network. The same analysis technique for task scheduling on a processor node and message scheduling across the network will be used.

Consider that the utilization-based test and the worst-case response time evaluation, assuming a set of four independent, periodic tasks, with deadline lower than the period, is implemented in single processor node as shown in Figure 5.

Each task τ_i is characterized by a worst-case computation time C_i a period T_i and relative deadline D_i. The tasks deadline is lower than their period, so preemptive contexts

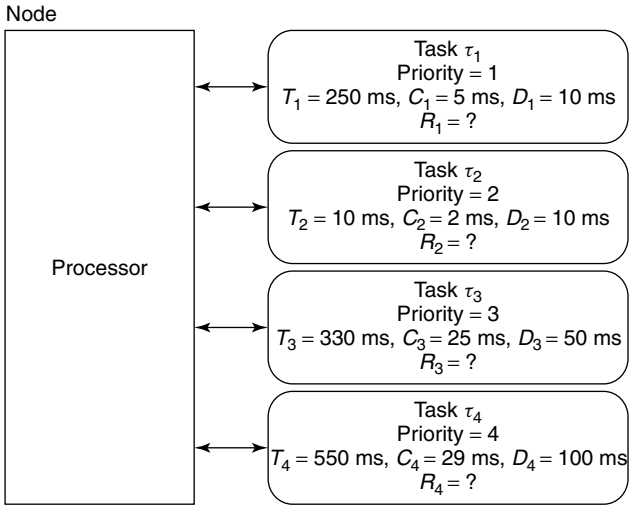


Figure 5. Single processor node. A set of node tasks with priority assignment according to DM priority order.

with both RM and DM scheduling policy can be used. Tasks shown in Figure 5 are in DM priority order.

To determine if a given set of independent tasks is schedulable, we use (1) and (3) introduced in Section 3. According to (1), the total processor utilization by four tasks $\tau_1-\tau_4$ is 0.29, which is less than 0.76, the bound for four tasks. Hence, these four tasks are schedulable.

In the next step, the worst-case response time for each task should be calculated. In Figure 6 is shown the results of the worst-case response time analysis, obtained from (3) for $\tau_1-\tau_4$ tasks with priority assignment according to the DM method and preemptive context.

The worst-case response time of a task for the preemptive context was introduced in Section 3. For each task, τ_i the

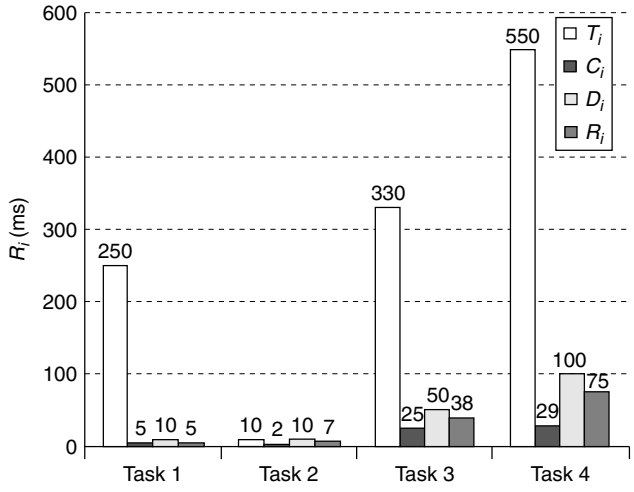


Figure 6. Results of a task scheduling with DM priority order for single processor node with task preemption.

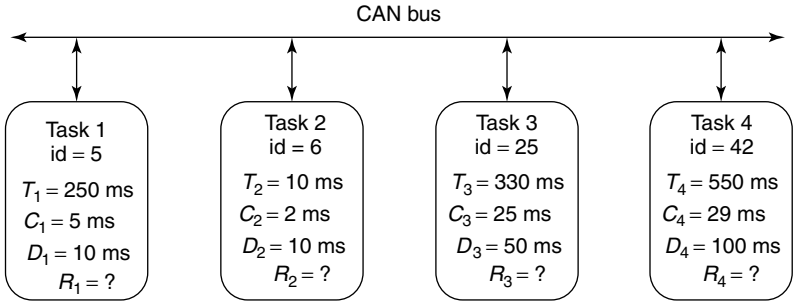


Figure 7. CAN network and a set of node messages with id assignments according to DM priority order.

worst-case response time R_i is lower than the task relative deadline D_i, so analyzed tasks set up in single processor node are schedulable. We can find that worst-case response time for task τ_2 is $R_2 = 7$ ms which is lower than the task deadline $D_2 = 10$ ms but, comparing with other tasks gives the smallest margin of safety $S_2 = 3$ ms.

7 SCHEDULING MESSAGES ACROSS A NETWORK

Suppose we have a four node network as shown in Figure 7, each with one independent, periodic task with given time parameters (T, C, and D). Here we have four periodic messages to be transmitted on a CAN network. CAN is a peer-to-peer communications bus for sending and receiving short real-time messages of up to 1 Mbps.

Each frame starts with an identifier that uniquely marks the frame – see Tindell, Hansson and Wellings (1994).

In our example, task τ_1 has id = 5, task τ_2 id = 6, task τ_3 id = 25 and the last one id = 42. This identifier is the transmitted most-significant bit first. It is important that no two nodes transmit frames with the same identifier as otherwise the arbitration will not work.

The effect of the arbitration scheme is to assign a priority to each frame, being the value of the identifier, where a smaller number is a higher priority. This priority-based scheduling of network is suited to the application of schedulability analysis.

Results of analysis and prediction of the worst-case response timing behavior of the NMCS based on a CAN bus, with non-preemptive DM approach for the communication level are shown in Figure 8.

Analysis was carried out by the use of (2) and (4) for the non-preemption approach. The analysis was done for each task and completed when the recurrence equation had converged to a consistent response time. Only the task τ_4 time requirements are met. Time requirements for the task τ_3 are almost met; for soft real-time systems it could

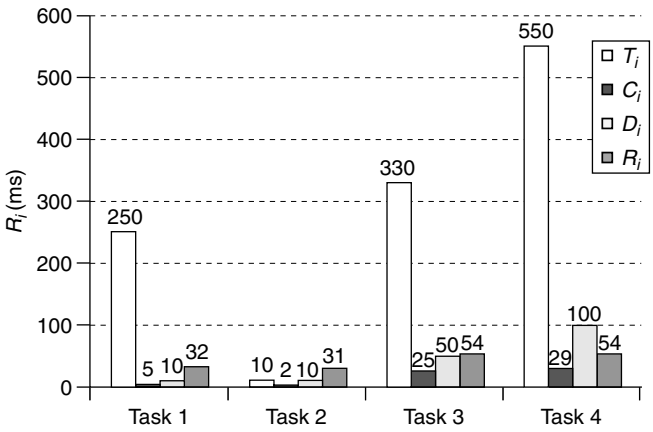


Figure 8. CAN network. Results of tasks scheduling with DM non-preemptive method.

be acceptable. Task τ_1 and task τ_2, for the DM offline scheduling approach, exceed its deadline significantly.

To reach schedulability on the CAN bus level we should increase the transmission rate to shorten the tasks execution time C.

As we can see, tasks with the same time parameters on the node and network level are scheduled only on the node level: this uses the preemption approach in access of its resources and thus can miss its deadlines on the CAN bus level for the three tasks.

8 TRENDS

The scheduling methods presented rely on deterministic techniques, which are based on worst-case execution times of tasks. This approach is sufficient for some types of timing constraints but is inadequate when relative constraints exist between tasks, and when execution times can range between lower and upper bounds. Developing novel techniques to solve this problem is needed.

Most of the real-time scheduling algorithms are based on the assumption that no error occurs in real-time systems. In order to make the real-time scheduling algorithm

more realistic, task precedence constraints need to be taken into account.

In mission critical applications, real-time tasks must tolerate faults. It would be interesting to combine proposed real-time analysis with real-time fault-tolerant scheduling algorithms for periodic real-time tasks in distributed systems.

With the popularity of the Internet and the Ethernet, a new design tool for Web embedded real-time systems is needed. Switched Ethernet can deliver IP packets deterministically at low cost. Ethernet and IP networking is becoming a good choice for real-time working.

RELATED ARTICLES

Article 78, Embedded Systems, Volume 2; **Article 158, Microcontroller Systems, Volume 3**; **Article 165, Introduction to Networked Instrumentation, Volume 3**; **Article 166, Instrument Interconnection, Volume 3**; **Article 172, Fieldbus Systems, Volume 3**.

REFERENCES

Audsley, N., Burns, A., Richardson, M., Tindell, K. and Wellings, A. (1993) Applying New Scheduling Theory to Static Priority Pre-emptive Scheduling. *Software Engineering Journal*, **8**(5), 285–292.

Burns, A. (1991) Scheduling Hard Real-Time Systems. *Software Engineering Journal*, Special issue on Real-Time Systems, **6**(4), 116–128.

Joseph, M. and Pandya, P. (1986) Finding Response Time in Real-Time System. *The Computer Journal*, **29**(5), 390–395.

Kopetz, H. (1997) *Real Time Systems: Design Principles for Distributed Embedded Applications*, Kluwer Academic Publishers, Boston, MA.

Lechoczky, J.P. (1990) Fixed Priority Scheduling of Periodic Tasks Sets with Arbitrary Deadlines, in *11th IEEE Real-Time System Symposium*, Lake Buena Vista, Florida, USA, pp. 201–209.

Leung, J. and Whitehead, J. (1982) On the Complexity of Fixed Priority Scheduling of Periodic Real-Time Tasks. *Performance Evaluation*, **22**(4), 237–250.

Liu, C. and Layland, J. (1973) Scheduling Algorithms for Multiprogramming in Hard-Real-Time Environment. *Journal of the ACM*, **20**(1), 46–61.

Michta, E. (2002) Scheduling Theory in Networked Measurement – Control System Design, in *Proceedings of the IMECO-TC7 Symposium*, Cracov (pp. 197–202).

Ramamritham, K. and Stankovic, J. (1994) Scheduling Algorithms and Operating Systems Support for Real-Time System. *Proceedings of the IEEE*, **82**(1), 55–67.

Sha, L. and Sathaye, S.S. (1995) *Distributed System Design Using Generalized Rate Monotonic Theory*, Technical Report CMU/SEI-95-TR-011, Carnegie Mellon University, Pittsburgh, PA.

Sha, L., Klein, M. and Goodenough, J. (1991) *Rate Monotonic Analysis for Real-Time Systems*, Technical Report CMU/SEI-91-TR-6, Carnegie Mellon University, Pittsburgh, PA.

Stankovic, J. (1988) *Real-Time Computing Systems: the Next Generation. Hard Real-Time Systems*. IEEE Computer Society Press, Los Alamitos, USA (pp. 14–38).

Tindell, K. (1992) *An Extendible Approach for Analysing Fixed Priority Hard Real-Time Tasks*, Technical Report YCS-189, Department of Computer Science, University of York, York, UK.

Tindell, K., Hansson, H. and Wellings, A. (1994) Analysis Real-Time Communications: Controller Area Network, in *IEEE Real-Time System Symposium*, San Juan, Puerto Rico (pp. 259–263).

174: Typical Signals Arising in Measurement

Eugen Georg Woschni
Technical University Chemnitz, Chemnitz, Germany

1 Relations Between Test Signals in the Time Domain 1167
2 Table of Important Signals 1167
3 Approximations and Estimations 1168
4 Noise Signals 1168
References 1170
Further Reading 1170

1 RELATIONS BETWEEN TEST SIGNALS IN THE TIME DOMAIN

In the time domain, especially for testing measuring systems, three forms of signals are used as often as an approximation. They also are used as typical model signals, see Section 3. These are

1. the unit pulse signal or the Dirac impulse $x(t) = \delta(t)$ (1a)
2. the unit step signal $x(t) = w(t)$ (1b)
3. the unit ramp signal $x(t) = ct; \quad c = 1$ (1c)

and last but not least, the sinusoidal test signal as treated in **Article 27, Signals in the Frequency Domain, Volume 1**.

The three signals are shown in Figure 1.

The three functions (1a, b, c) are related to each other: they are each the integral of the other

$$\int_0^t \delta(\tau)\,\mathrm{d}\tau = w(t) \tag{2a}$$

$$\int_0^t w(\tau)\,\mathrm{d}\tau = cx \text{ with } c = 1 \tag{2b}$$

Consequently, the inverse function yields

$$\frac{\mathrm{d}}{\mathrm{d}t} w(t) = \partial(t) \tag{3a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} cx = c = 1 \rightarrow w(t) \tag{3b}$$

It may be remarked that using classical analysis, (3a) is not strictly realizable because the unit step function is a discontinuous function and, therefore, cannot be differentiated at $t = 0$.

This difficulty is overcome using the modern distribution theory (Posthoff and Woschni, 1984; Smirnow, 1979). Here, the discontinuous function is replaced by a continuous one and then the limit $\Delta t \rightarrow 0$ is realized as Figure 2 shows.

These relations between the test functions are very important in system theory because in linear systems, the output signals resulting from these inputs must have the same relations with each other.

2 TABLE OF IMPORTANT SIGNALS

The essential properties of important signals are summarized in Table 1. Individual signals have already been presented as examples in the relevant articles. The harmonic oscillation, the unit impulse, and the unit step are typical test signals for system identification. White noise (wideband noise) has been included in the summary as it is the most important interfering signal. For this signal, there is no amplitude density $\hat{X}(\mathrm{j}\omega)$, as has already

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

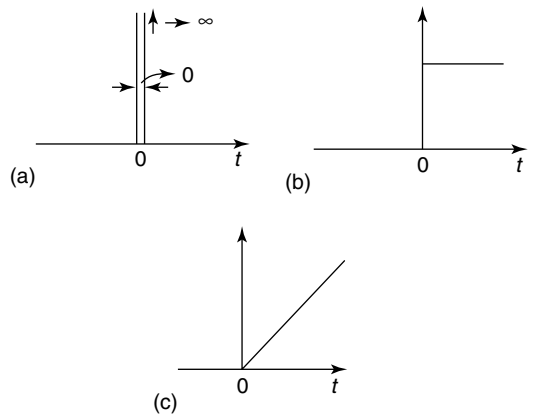


Figure 1. (a) Unit pulse signal; (b) unit step signal; (c) unit ramp signal.

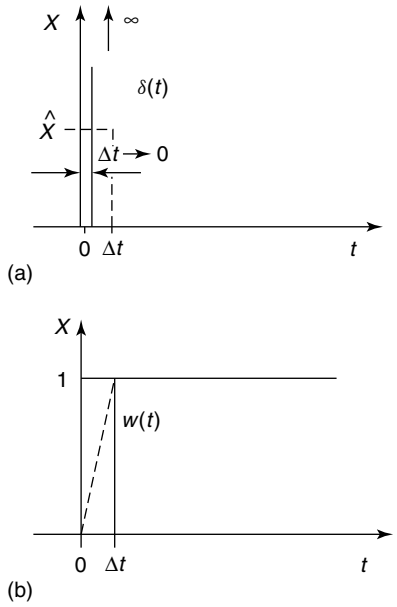


Figure 2. Interpretation of the relationship between unit pulse (a) and unit step; (b) function due to (3a).

been discussed in **Article 27, Signals in the Frequency Domain, Volume 1**. The relationships that have already been pointed out between time and frequency functions as Fourier transforms become evident: the narrower the time signals, the wider the band of the frequency signals present with the extreme case between the constant behavior in the frequency domain and the function behavior in the time domain, *see* **Article 29, Relationship Between Signals in the Time and Frequency Domain, Volume 1**.

3 APPROXIMATIONS AND ESTIMATIONS

Of importance in measurement technology, in particular, are estimates and approximate considerations. In contrast to other fields of information engineering, for instance, control engineering is where input signals are given (i.e. they are known) and it is the output signals that are to be calculated; in measuring technology, the opposite has to be done for the input signal is the unknown system variable.

For this reason, before choosing a suitable measuring device and processor, it is necessary to make an assessment of the signal behavior to be expected, on the basis of a priori information, and to then choose the device.

To this end, it is usual to approximate the quantity to be measured by either a pulse-shaped or a ramp-shaped signal. Therefore, these two signals have also been included in Table 1. The pulse duration and the duration of the ramp function respectively, can be taken from data on the technological process to be investigated, for instance, the speed in rotating machines. The ramp is a useful test signal for systems that cannot respond rapidly due to high inherent inertia.

According to the sampling theorem, a threshold frequency ω_g corresponds to signal duration $2T_1$

$$\omega_g = \frac{\pi}{2T_1} \tag{4}$$

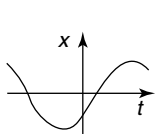
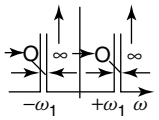
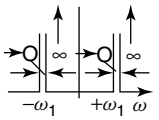
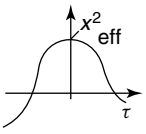
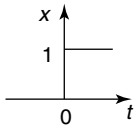
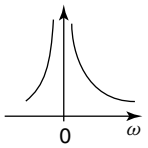
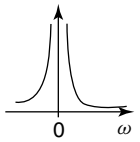
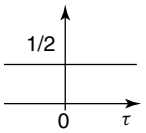
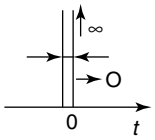
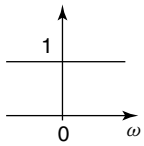
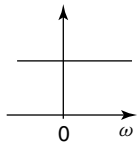
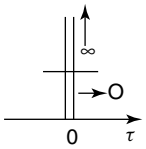
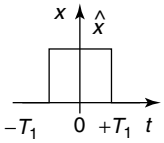
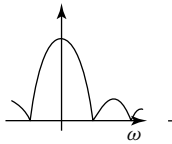
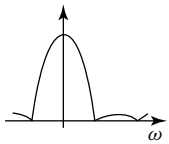
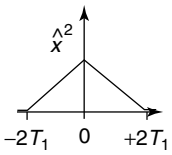
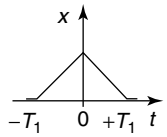
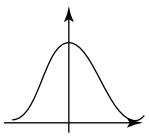
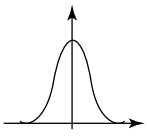
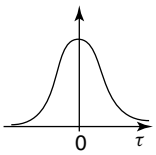
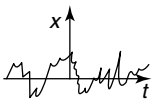
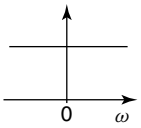
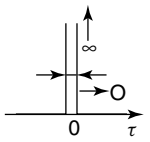
Above this frequency there are only spectral oscillations having a relatively low amplitude; the amplitude or power spectrum mainly lies below this threshold frequency (called *band-limited signals*). To avoid major measurement errors, it is therefore sufficient for the measuring device to cover this frequency domain, that is $0 \geq \omega > \omega_g$ (*see also* **Article 36, Systems in the Frequency Domain, Volume 1**). The dynamic response of typical linear systems is analyzed in detail in **Article 59, Introduction to the Dynamic Regime of Measurement Systems, Volume 1**; **Article 60, Zero-order System Dynamics, Volume 1**; **Article 61, First-order System Dynamics, Volume 1**; **Article 62, Second-order System Dynamics, Volume 1**.

4 NOISE SIGNALS

Noise signals are important in measurement technique firstly as test signals, and secondly because disturbances and electronic noise are inevitable in any measuring system. To describe these signals, Table 1 shows the power density function $S(\omega)$ and the autocorrelation function $\psi(\tau)$ of them. Between both exist relations that are treated in detail in **Article 29, Relationship Between Signals in the Time and Frequency Domain, Volume 1**, see especially Table 2 – the so-called Wiener–Chinchine theorem.

$$\Psi(\tau) = \int_{-\infty}^{+\infty} \delta(\omega) e^{j\omega\tau}\, d\omega \tag{5a}$$

Table 1. Properties of important signals.

Characteristic	Time function $x(t)$	Amplitude density $\lvert\hat{X}(\mathrm{j}\omega)\rvert$	Power density $S_{xx}(\omega)$	Autocorrelation function $\psi_{xx}(\tau)$	Remarks
Harmonic vibration					Test signal
Unit step $w(t)$					Test signal
Unit pulse $\delta(t)$					Test signal
Pulse					Important for approximations
Ramp function					Important for approximations
White noise					Typical disturbance

$$S(\omega) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \Psi(\tau)\mathrm{e}^{-\mathrm{j}\omega\tau}\,\mathrm{d}\tau \tag{5b}$$

Furthermore, there are the mean square value $\overline{x^2(t)}$ and the effective value

$$X_{\mathrm{eff}} = \sqrt{\overline{x^2(t)}} \tag{5c}$$

with the relation to the autocorrelation function of

$$\overline{x^2(t)} = \Psi(0) \tag{5d}$$

as dealt with in **Article 30, Statistical Signal Representations, Volume 1**.

Often the signals can be assumed to be of Gaussian type. Wideband or white noise, having a constant power density S_{o} – see Table 1, is important because it arises in all electronic components and devices. To measure the value of S_{o}, a color filter – also called *shaping filter* – as shown in Figure 3(a) is used, leading to the generation of colored noise

$$S(\omega) = \frac{S_{\mathrm{o}}}{1+\omega^2R^2C^2} \tag{6a}$$

Figure 3(b) demonstrates its nature. Often met names of colored noise types are pink and gray noise.

From (6a) it follows the whole power is at the output if the limiting frequency of the measuring device at the output is ω_{c}

$$P_{\omega_{\mathrm{C}}} = \int_{-\omega_{\mathrm{c}}}^{+\omega_{\mathrm{c}}} \frac{S_{\mathrm{o}}}{1+\omega^2R^2C^2}\,\mathrm{d}\omega = \frac{2S_{\mathrm{o}}}{RC}\arctan\omega_{\mathrm{c}}RC \tag{6b}$$

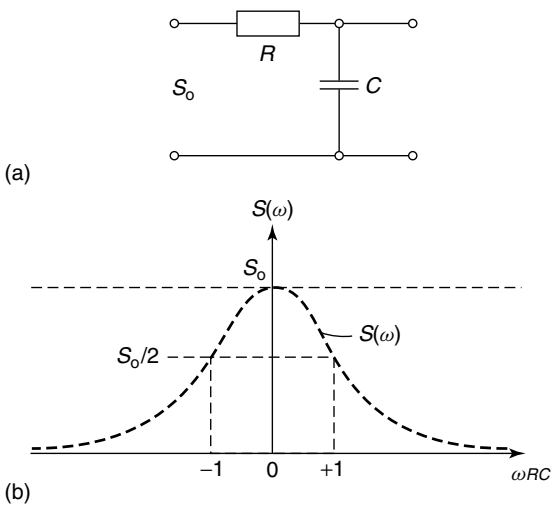


Figure 3. (a) Shaping filter (low-pass filter) and (b) colored noise.

and thus leading to the value of S_o

$$S_o = \frac{P_{\omega_c} RC}{2 \arctan \omega_c RC} \tag{6c}$$

The limiting case $\rightarrow$ yields (Woschni, 1988)

$$P = \frac{\pi S_o}{RC} \tag{6d}$$

Using this method, for instance, noise sources, such as noise diodes, may be calibrated.

REFERENCES

Posthoff, C. and Woschni, E.-G. (1984) *Funktional Transformationen der Informations Tecknik*, Akademie Verlag, Berlin.

Smirnow, W.I. (1979) *Lehrgang der hoheren Mathematik*, Deutscher Verlag d. Wiss., Berlin.

Woschni, E.-G. (1988) *Informations Technik*, Huthig, Heidelberg.

FURTHER READING

Levine, W.S. (1996) *The Control Handbook*, CRC Press, New York.

Woschni, E.-G. (2000) Approximations in Measurement – Today Still Necessary? in *Proceedings XVI IMEKO World Congress*, Vienna.

175: Comparison of Analog and Digital Signal Handling

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Comparison of Analog and Digital Signal Handling 1171
2 Analog and Digital Signal Processing 1171
3 Restoring Noisy Digital Signals 1173
Related Articles 1173

1 COMPARISON OF ANALOG AND DIGITAL SIGNAL HANDLING

The information carrying variable in measurement is a signal. It is beneficial to consider the straightforward case of signal propagation through a channel using either analog or digital signal formats. Figure 1 gives block diagrams, which illustrates a way to compare these two possibilities. First, propose the existence of a channel for analog signal propagation/transmission/storage as shown in Figure 1(a). When the analog signal is applied to this channel, it is contaminated by noise during its passage through the channel. Apart from including a conditioning filter at the output of the channel, there is comparatively little that the channel designer can do to compensate for the intrusion of the noise. Data acquisition systems (DAS), and the various parts as such, are covered in detail in **Article 132, Data Acquisition Systems (DAS) in General, Volume 3**; **Article 133, Amplifiers and Filters for DAS, Volume 3**; **Article 134, Analog Multiplexers, Volume 3**; **Article 139, Analog-to-Digital (A/D) Converters, Volume 3**; and **Article 141, Digital-to-Analog (D/A) Converters, Volume 3**.

When the signal to be transmitted or stored is a digital signal, it is relevant to examine the situation closely. The possibility of noise interference is illustrated in Figure 1(b). Notice the first important difference, which is the increase in the complexity of the process. For digital operation, there needs to be sampling, quantization, conversion using an *analog-to-digital converter* (called an ADC for short), *see* **Article 139, Analog-to-Digital (A/D) Converters, Volume 3**, and encoding at the channel input. The process of encoding results in *pulse code modulation*, or PCM – *see* **Article 34, Modulation Theory, Volume 1**. This complexity is also an essential feature at the output of the channel, where reconstruction is followed by decoding, conversion to analog using a *digital-to-analog converter* (called a DAC for short) and probably filtering – *see* **Article 141, Digital-to-Analog (D/A) Converters, Volume 3**. However, because of the insensitivity of the digital signal to noise and drift, the channel input signal can be reproduced with higher fidelity after reconstruction at the channel output than in the analog system of Figure 1(a).

2 ANALOG AND DIGITAL SIGNAL PROCESSING

Signal processing, which is now an established area of electronic and electrical engineering, is concerned with the extension of the human abilities to extract and interpret the information content of a signal. Different and diverse classes of signals may be processed using either *analog signal processing* (ASP) methods or *digital signal processing* (DSP) methods – *see* **Article 144, DSP Chip Sets, Volume 3** and **Article 145, DSP Tools, Volume 3**.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

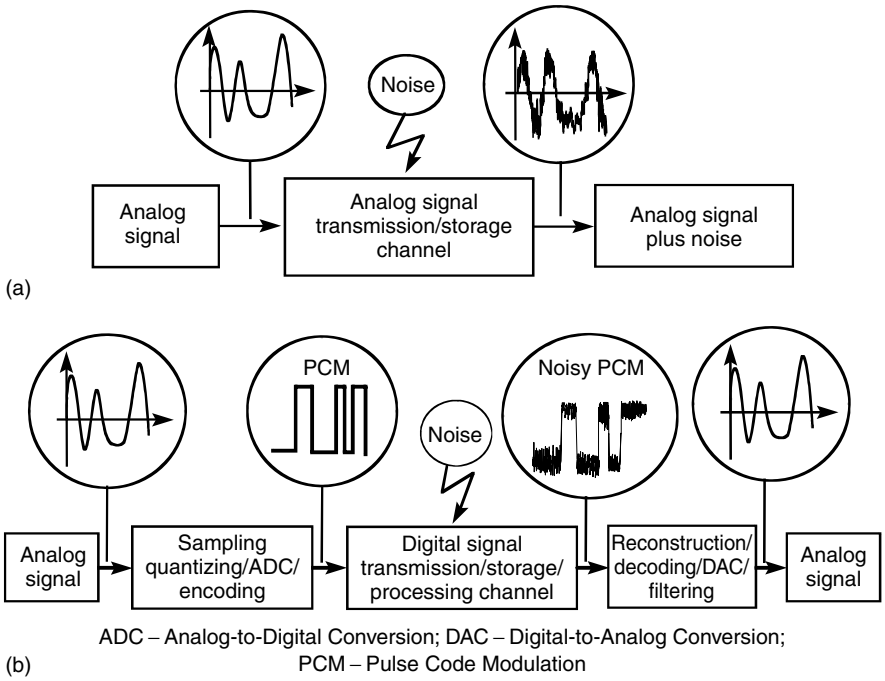


Figure 1. Transmission of (a) an analog signal through a noisy analog channel and (b) a digital signal through a noisy digital channel.

ASP, which involves the operations of amplification and filtering in the electrical energy domain, may also use optoelectronic methods, surface acoustic wave technology, and charge-coupled device technology. Alternatively, in the electrical energy domain, they may be converted into the digital form using an ADC, processed using digital techniques, and converted back into analog form using a DAC. This successive conversion between the two types of signal uses sophisticated electronic elements.

Digital signals are processed using digital computing methods, which involves the use of micro, mini or mainframe computers or nowadays more commonly with *personal computers* (PCs). The requirement for fast processing of digital signals has also contributed to the development of computers with a very fast processing capability called *array processors and supercomputers*. Special DSP devices such as the Texas Instruments TMS320 or the Motorola DSP56000 are also becoming widely used for certain applications.

Examples of ASP and DSP can be found in many different areas of application such as radar and sonar techniques, seismography, communications, optoelectronics, biomedicine, and all forms of measurement. Consumer applications, as occurs in sound and video recording and reproduction, are two additional areas in which signal processing is rapidly advancing. Radar and sonar systems, which may contain many analog filters in the receiver and demodulation stages, also contain increasing amounts of DSP technology to enhance the extraction of useful information from the signals. Because of this, they may be referred to as *HYBRID*. In seismography, the survey is processed and interpreted to detect the potential existence of oil and gas reserves. Data is recorded in digital format on magnetic tape. It is then processed '*off line*' in a data processing center. Many of the DSP techniques used in other areas of application originated in seismic data processing. In data measurement (DM), the DSP usually has to be performed in real time.

There are many situations in which it is necessary to maintain galvanic isolation between low voltage instrumentation circuits and the higher voltage, higher current levels associated with final control actuators. This can be arranged easily using optocouplers when the control signals are digital. In the case of analog control signals, the isolation barrier normally needs to be provided by a transformer. Rapid developments have occurred in digital signal processing hardware and software. Accompanying these developments are diverse methods that allow digital instruments to incorporate what may be loosely referred to as intelligence. In fact, this 'intelligence' is really an expansion in 'functional diversity'. In conclusion, it may be observed that digital DM systems are generally superior in most respects to analog instrumentation – *see* **Article 161, Knowledge-based Systems, Volume 3**; and **Article 128, Nature and Scope of AI Techniques, Volume 2**.

When analog signals are being processed, there is often an advantage in converting the signal to digital form so that the processing can be done digitally. For example,

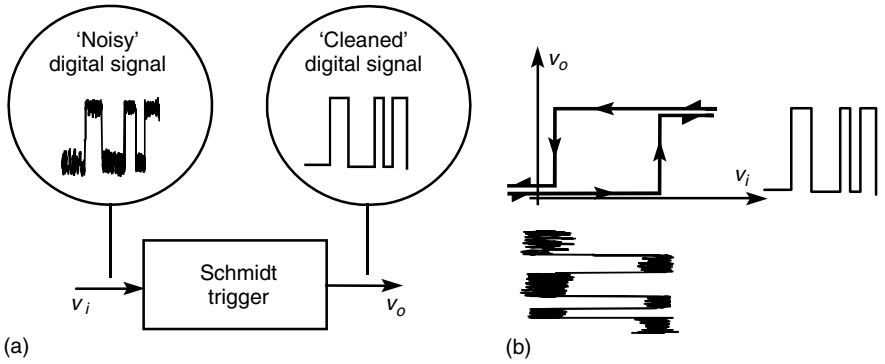


Figure 2. Illustration of the use of a Schmidt trigger in the restoration of digital signals (a) block schematic and (b) voltage transfer function.

the processing may consist simply of measuring an analog direct voltage. Using an analog voltmeter, the measurement is carried out by determining the position of a pointer on a scale. This measurement is usually subject to certain errors such as parallax. However, using a digital voltmeter, a more convenient and accurate numerical indication is given on a digital display panel.

The second example, which is illustrated in Figure 1(b), is based upon the ability to recover noisy digital signals using the method given in Section 3. The digital method enables an uncorrupted replica of a binary sequence to be recovered from a signal consisting of the originally uncorrupted sequence and a contaminating additive noise.

3 RESTORING NOISY DIGITAL SIGNALS

Information is the basis of all operations in machines for measurement, calculation, communication, and control. Propagation, transmission, storage, processing, and retrieval of information using digital signals have many advantages. One of these is the ability to detect the presence of a binary digit, or BIT, of information in the form of a noisy pulse even when the noise causes considerable distortion of the pulse. Using threshold detection circuits transforms a noisy digital signal into a 'cleaned' version of the same signal by arranging that the signal has only one of two possible levels. This operation, which is shown in Figure 1(b), is very similar to the tolerance in transistor–transistor logic, or TTL, logic circuits. For example, assume that the signals are active LOW, with 0 V as the '1'-level. When the signals are within the two prescribed ranges, they are regarded as either a logic HIGH, or '0', or as a logic LOW, or '1'. It is not permissible for the voltage to be anywhere between the lower threshold of the HIGH, '0'-level, and the upper threshold of the LOW level. When digital signals undergo attenuation due to loading effects, they can be restored or reconstructed to their correct TTL levels using a TTL circuit known as a *Schmidt trigger*. A Schmidt trigger is a special circuit that exhibits hysteresis in its voltage transfer function.

Consider Figure 2(a), where a noisy digital signal is applied to the input of the Schmidt trigger. The trigger circuit's operation latches the output to one particular level until the value of the input signal exceeds some designed threshold level. When this value is reached, the Schmidt circuit latches to the opposite polarity, where it will remain until the input signal reaches the reverse level switching threshold. The changes in signal level are shown in Figure 2(b). This illustrates the insensitivity of a digital signal to noise. It is also important to note that this insensitivity to noise is accompanied by insensitivity to drift. This ability to restore the levels of a digital signal allows the transmission, storage, and retrieval of signals much more easily. The fidelity of the digital data is preserved, whereas the fidelity of analog signals always degrades, sometimes quite substantially, with noise and drift.

RELATED ARTICLES

Article 25, Introduction to Signals in Physical Systems, Volume 1; **Article 26, Signal Classification, Volume 1**; **Article 121, Signals in the Presence of Noise, Volume 2**; **Article 174, Typical Signals Arising in Measurement, Volume 3**; **Article 176, Signals and Signal-to-noise Ratio, Volume 3**.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

176: Signals and Signal-to-noise Ratio

Richard Burdett
SIGNAL RECOVERY, Wokingham, UK

1 Signal Forms 1174
2 Signal-to-noise Ratio 1174
Related Articles 1174

1 SIGNAL FORMS

Suppose we were to look at a complex waveform on an oscilloscope. What is the signal? Is it the complete waveform? The peak (or rms or average) amplitude? The depth of modulation? The implied frequency spectrum? The difference in time or amplitude between two features of the waveform?

The answer, of course, is that the information-bearing signal could be any or none of the above. In this article, discussion is restricted to some commonly encountered types of signal for which enhancement is often required. Together with the enhancement technique normally used and described later in this section – **Part 12, Elements: F – Signals and Noise, Volume 3** – these are

1. base-band DC signals: low-pass filtering or autocorrelation;
2. amplitude modulated signals: band-pass filtering or phase-sensitive detection;
3. repetitive (not necessarily periodic) swept signals: signal averagers;
4. photon, electron, or ion beam signals: photon-counting systems.

The word signal is often used rather ambiguously to mean either the total signal being measured or a noise-free, information-bearing component of it.

2 SIGNAL-TO-NOISE RATIO

The following definitions should avoid such confusion. We will normally talk in terms of a total signal consisting of an rms signal component (S) accompanied by an rms noise component (N). Thus,

$$\text{Signal-to-noise ratio, SNR} = \frac{S}{N} \tag{1}$$

Note that

$$\text{measurement uncertainty, or inaccuracy} = \frac{1}{\text{SNR}} \tag{2}$$

$$\text{Signal-to-noise improvement ratio (SNIR)} = \frac{\text{SNR}_{\text{out}}}{\text{SNR}_{\text{in}}} = \frac{S_{\text{o}}/N_{\text{o}}}{S_{\text{i}}/N_{\text{i}}} \tag{3}$$

For unity gain (i.e. $S_{\text{o}} = S_{\text{i}}$), band-limited white input noise of bandwidth B_{ni} and output noise bandwidth B_{no},

$$\text{SNIR} = \frac{N_{\text{o}}}{N_{\text{i}}} = \sqrt{\left(\frac{B_{\text{ni}}}{B_{\text{no}}}\right)} \tag{4}$$

RELATED ARTICLES

Article 121, Signals in the Presence of Noise, Volume 2; **Article 178, Noise Matching and Preamplifier Selection, Volume 3**; **Article 179, Input Connections; Grounding and Shielding, Volume 3**; **Article 181, Amplitude Modulated Signals: The Lock-in Amplifier, Volume 3**.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

This article is based in part on a chapter by D.M. Munro originally published in Handbook of Measurement Science Vol 1, P.H. Sydenham, Wiley, 1982.

177: Grounding and Shielding

Kim R. Fowler
Technical Consultant, Cool Stream, Windsor Mill, MD, USA

1 Purpose of Grounding and Shielding 1175
2 General Principles of Grounding 1175
3 General Principles of Shielding 1177
4 Line Conditioning 1178
5 Noise Reduction in Cabling and Other Circuit Elements 1179
6 Enclosures for Noise Reduction 1180
7 Protecting Against Electrostatic Discharge (ESD) 1181
8 Limitations of Grounding and Shielding 1181
References 1181
Further Reading 1181

1 PURPOSE OF GROUNDING AND SHIELDING

Grounding and shielding increase safety and reduce noise within the operation of electronic circuits. Grounding and shielding also reduces the possibility of damage to components and circuit elements caused by electrostatic discharge (ESD).

General and specific accounts that review grounding and shielding, provide practical solutions for grounding and shielding, and provide key references are Fowler (1996), Johnson and Graham (1993), and Ott (1988).

Grounding has a two-fold purpose; it provides both reference for signals and safety for operators and personnel. The purpose for shielding is to prevent or reduce the transmission of disruptive electrical charge and signals.

2 GENERAL PRINCIPLES OF GROUNDING

The most important goal for grounding is to minimize the voltage differential between a circuit (or device, or instrument) and a reference point (Figure 1). Reducing the voltage differential provides both safety and signal reference.

An important distinction exists between a ground and a return. Ground connections and lines conduct no current during normal, undisrupted operation. Grounding provides a path for stray or harmful currents during times of atypical, disrupted operation. A ground is not a return path for either power or signal. A return path routinely conducts current; it completes the circuit for purposeful and designed current flow.

2.1 Safety ground

A safety ground must be a permanent, continuous, low-impedance conductor with adequate capacity that runs from the power source to the load (your instrument). Your aim is to reduce the voltage differential between external conducting surfaces (Figure 2).

Consider the following when developing a wiring and the grounding scheme for powering an instrument:

- Consider the instrument and the power mains as an integrated system.
- Draw the ground scheme with all possible circuit paths.
- Do not blindly rely on building steel to be the ground conductor; it can conduct significantly large currents, but through faulty installation and design elsewhere, it can be noisy or dangerous.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

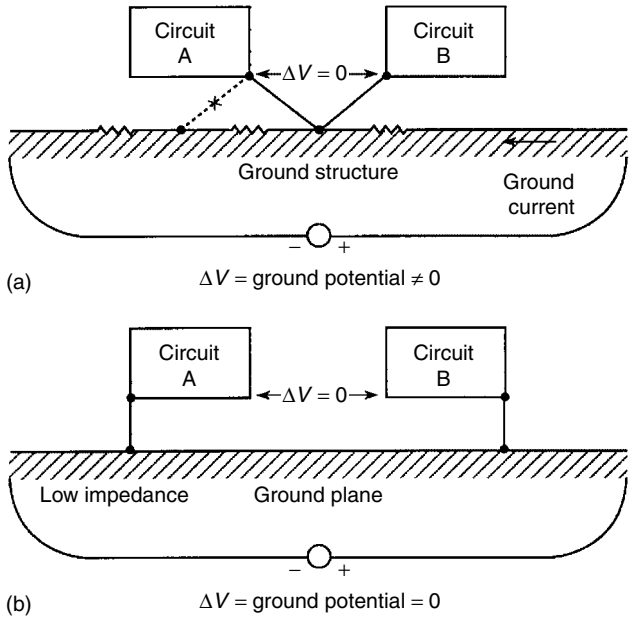


Figure 1. Grounding minimizes the voltage differential between device and reference point. (Reproduced from *Electronic Instrument Design: Architecting for the Life Cycle*, Fowler, K.R. Copyright (1996) © Oxford University Press.)

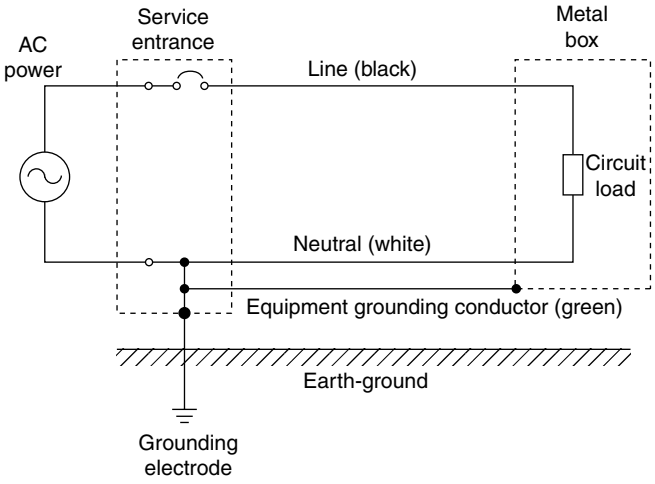


Figure 2. General scheme for safety ground in buildings. (Reproduced from *Electronic Instrument Design: Architecting for the Life Cycle*, Fowler, K.R. Copyright (1996) © Oxford University Press.)

2.2 Ground configurations

Grounding may have different configurations depending on the application and the operation. The main configuration used in small circuits and instruments is the single-point ground. For larger installations and high frequencies, a multipoint configuration maybe more appropriate.

2.2.1 Single-point grounding

Single-point grounding connects one conductor from the ground structure, possibly the return conductor, in each circuit or subsystem to a single point for the entire instrument or system. It looks like a star pattern with a point and radiating conductors to the various circuits or subsystems. The point connects through a single conductor to chassis and safety ground. Figure 3 illustrates an example of a single-point ground.

The main purpose for single-point grounding is signal reference. Single-point grounding is useful in low current and low-frequency (typically less than 1 MHz) applications.

2.2.2 Multipoint grounding

A multipoint ground plane or grid is better for high-frequency (generally above 100 KHz) or high-current

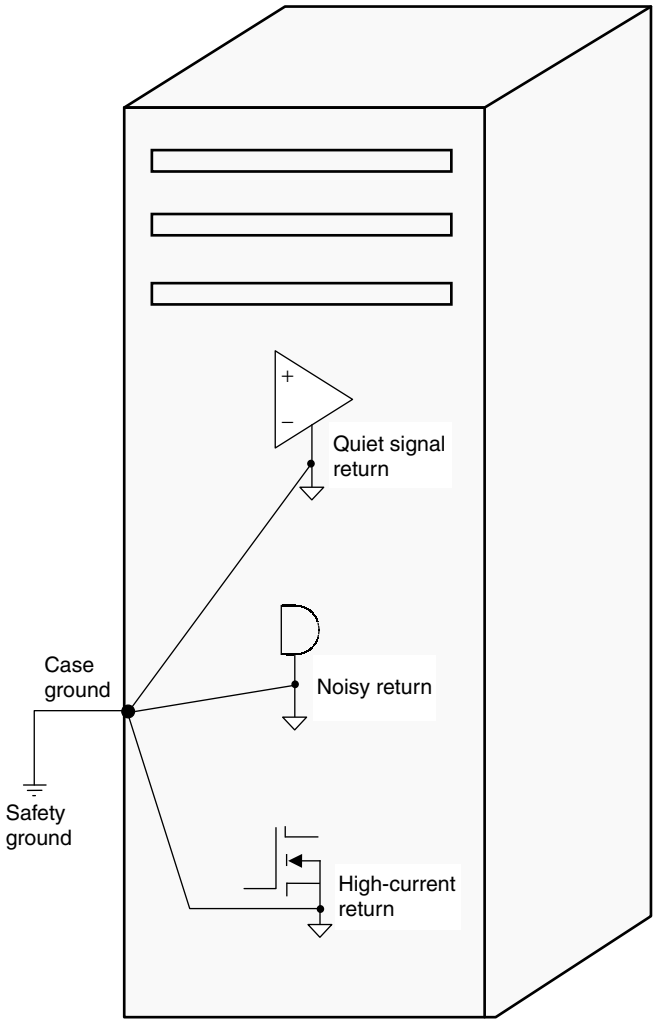


Figure 3. An example of a single-point ground. (Reproduced from *Electronic Instrument Design: Architecting for the Life Cycle*, Fowler, K.R. Copyright (1996) © Oxford University Press.)

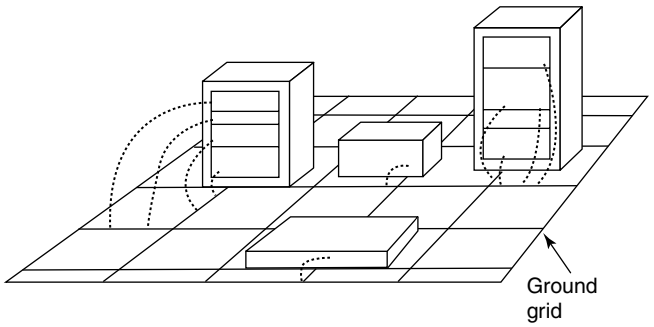


Figure 4. An example of a multipoint ground. (Reproduced from *Electronic Instrument Design: Architecting for the Life Cycle*, Fowler, K.R. Copyright (1996) © Oxford University Press.)

fault applications. Multipoint grounding provides lower impedance than a single cable in single-point grounding. Figure 4 illustrates an example of multipoint grounding.

2.3 Ground loops

A ground loop is a complete circuit comprising both signal path and part of the ground structure. External currents in the ground can generate potential differences between the ground connections and introduce noise into the signal circuit, see Figure 5.

Generally, the problem arises at low frequencies (below 10 MHz); high frequencies tend to follow the path of minimum impedance that might avoid higher-impedance ground loops.

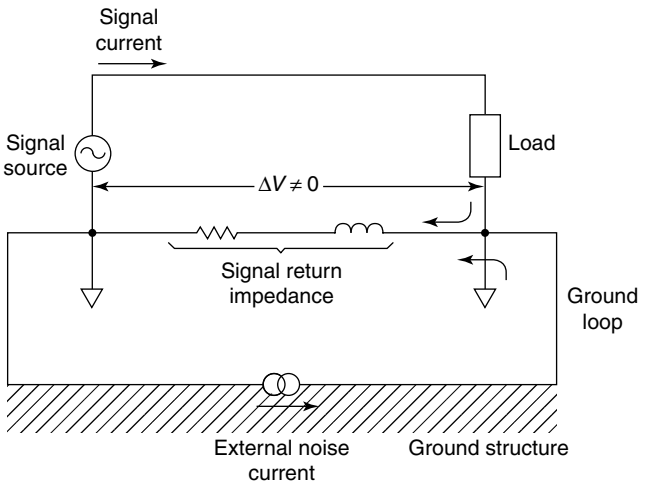


Figure 5. A ground loop occurs when a circuit makes several connections to the ground structure. (Reproduced from *Electronic Instrument Design: Architecting for the Life Cycle*, Fowler, K.R. Copyright (1996) © Oxford University Press.)

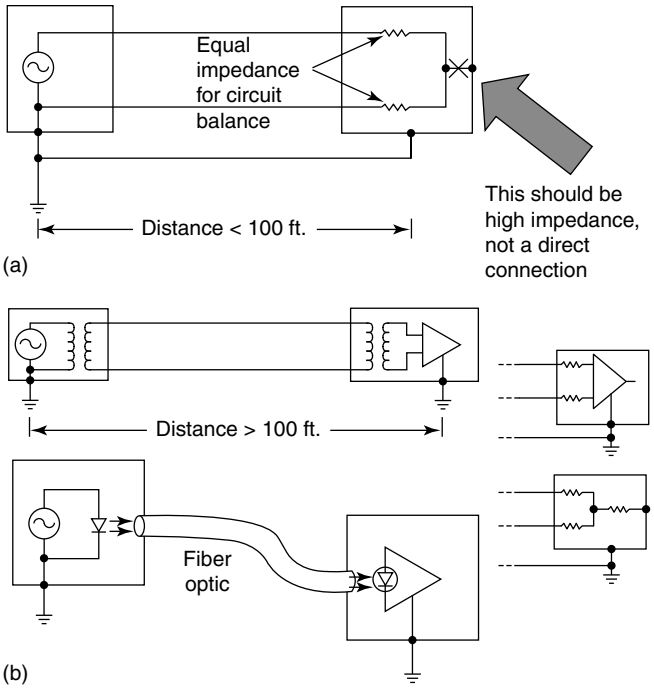


Figure 6. Examples of circuit balance and isolation that eliminate ground noise. (Reproduced from *Electronic Instrument Design: Architecting for the Life Cycle*, Fowler, K.R. Copyright (1996) © Oxford University Press.)

2.4 Remove ground problems in signal transmission

Long length signal transmission circuits suffer from ground loop problems. Circuit balance or isolation can eliminate the problems that arise from ground noise. For short distances of less than 30 m, a balanced transmission line and single-point grounding can reduce noise and safety concerns.

For distances over 30-m long, use of isolated signal transmission, through fiber optics or transformer isolation, allows multiple safety grounds while eliminating ground loops. Figure 6 illustrates some examples of circuit balance and isolation.

3 GENERAL PRINCIPLES OF SHIELDING

3.1 Analysis

First analyze the situation, *see* **Article 17, EMC and EMI, Volume 1**, for a practical basis on which to analyze the potential EMI. Four types of EMI problems exist:

- conductive
- inductive
- capacitive
- electromagnetic.

Once the problem situation is known, the appropriate shielding can be selected. Merely adding shields without that understanding can, in fact, lead to worsened situations.

Conductive problems can often be addressed by using more appropriate circuit connections (e.g. to eliminate ground loops) or filtering (e.g. to remove spikes). The various types of EMI problems and their shielding are addressed below.

3.2 Inductive shielding

Inductive shielding eliminates the generation, coupling, and reception of changing magnetic flux. It often is referred to as *self-shielding*. Do the following to achieve inductive shielding:

- Minimize the loop area between signal lines and their return paths.
- Separation can help reduce the problem somewhat.
- Understand that shielding enclosures help very little in preventing inductive coupling. Use of magnetic material shielding is an art and a science needing sophisticated materials, special treatment, and careful life management to maintain its magnetic shielding properties.

Twisting a signal wire with its return, for example, will eliminate the loop area. Running the twisted pair close to the ground will reduce the common-mode current, I_c in Figure 7, by reducing the loop area for inductive coupling.

3.3 Capacitive shielding

Capacitive shielding reduces and reroutes stray electrical charge. It prevents capacitive coupling, which provides a path for the injection of noise changes. Remember that a capacitance to ground reduces in its impedance as the frequency rises so at high frequencies and large capacitance sizes a low-impedance connection results in a situation apparently open circuited.

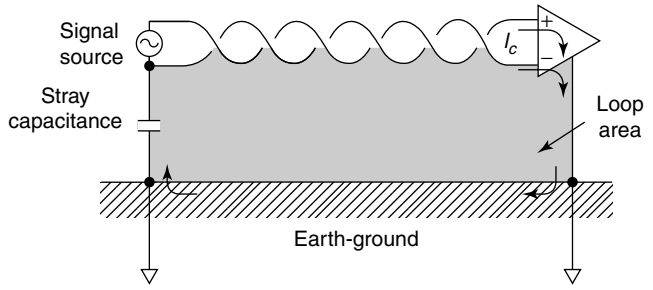


Figure 7. Twisting the signal with its return and running the pair close to ground will reduce the loop area. (Reproduced from *Electronic Instrument Design: Architecting for the Life Cycle*, Fowler, K.R. Copyright (1996) © Oxford University Press.)

An appropriately placed shield prevents the coupling of electrical charge between circuits by shunting the charge to ground (Figure 8).

3.4 Electromagnetic shielding

Electromagnetic shielding works to prevent the generation, coupling, and reception of high-frequency (above 20 MHz) signals. Remedies for EMI areas follows:

- Reduce EMI generated emissions and detected receptions within the circuits.
- Reduce the bandwidth of the circuits.
- Apply good signal routing within circuit boards.
- Use shielded enclosures with premeditation and care.

4 LINE CONDITIONING

One particular concern is the transmission of noise and disruption through local power lines. Power itself may suffer from a number of problems: dropout, low voltage, surge, overvoltage, and high voltage spikes.

Filtering the AC power source at the power converter will eliminate differential-mode and common-mode noise from the input conductors. An electrostatic shield in the transformer, appropriate chassis connections, and balancing impedances in the leads will help the filter eliminate common-mode noise from the output lines. These are conductive coupling problems – *see* **Article 17, EMC and EMI, Volume 1**.

Filtering can reduce high-voltage spikes and high frequency ripple. Only a battery-backed uninterruptible power supply can eliminate the other problems of dropout, low voltage, surge, and overvoltage.

Carefully selected, the power conductors within an instrument can help reduce EMI and inductive coupling – *see* **Article 17, EMC and EMI, Volume 1**. Reducing the separation between the power conductors or twisting the power cables will reduce the inductive loop and any magnetic coupling. Power and return planes do this effectively within circuit boards.

A problem often caused by instrumentation is the injection of noise harmonics into the AC power. This can happen because the input filter capacitors charge only near the voltage peaks. Such charging forces the current waveform into a nonsinusoidal form, thus generating harmonics. These harmonics do not contribute real power but increase the apparent power required; the ratio of real power to apparent power is the power factor. Power harmonics reduce the power factor of an instrument. Circuitry for correcting the power factor will ensure that the instrument does not generate power line harmonics.

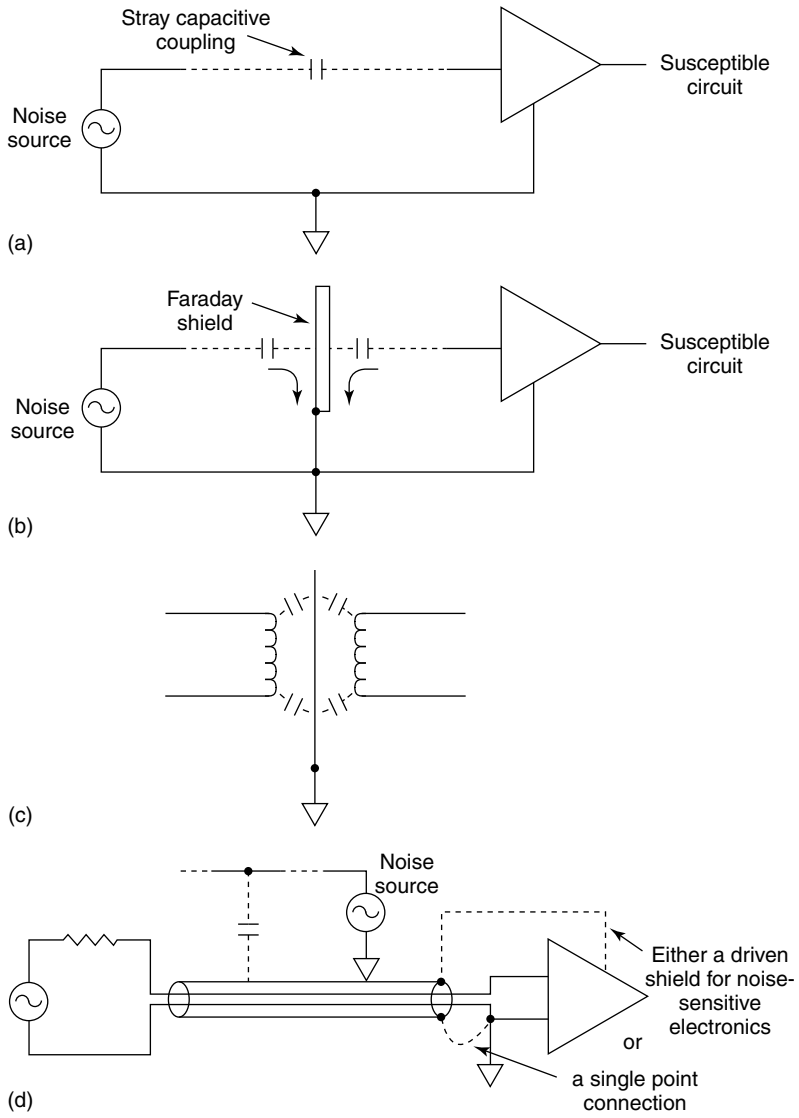


Figure 8. Examples of capacitive coupling and shielding. (a) outline of capacitive coupling, (b) proper placement of a shield to shunt noise charge, (c) a shielded transformer prevents capacitive coupling between windings, and (d) a cable shield usually should be connected at only one place to prevent coupling or shunt charge. (Reproduced from *Electronic Instrument Design: Architecting for the Life Cycle*, Fowler, K.R. Copyright (1996) © Oxford University Press.)

5 NOISE REDUCTION IN CABLING AND OTHER CIRCUIT ELEMENTS

Designing cables and laying out circuits is a matter of care and understanding. Fundamental principles, in more detail, are given in **Article 17, EMC and EMI, Volume 1**.

5.1 Cables

Cable configurations that best attenuate noise depend on the frequency of operation (Figure 9). The twisted-pair cable is usually effective for low frequencies up to 1 MHz, but it becomes lossy at higher frequencies. Its main advantage over coaxial (coax) cable is that it is cheaper and mechanically more flexible. Coax cable has low loss and less variance in its characteristic impedance from DC to very high frequencies (above 200 MHz).

Ribbon cable is suitable for low-frequency operation (below 1 MHz) when the signals are not a pair with return lines. For low-level signals or higher frequencies, pair each signal conductor with a return conductor or use a return plane (Figure 10).

Coaxial terminations must be carefully made to reduce leakage and ensure lowest signal reflection as the frequency rises (Figure 11). The pigtail connection is fine

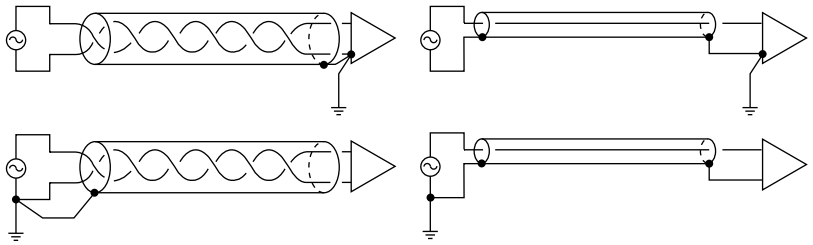


Figure 9. Cable configurations for shielding against noise. (Reproduced from *Electronic Instrument Design: Architecting for the Life Cycle*, Fowler, K.R. Copyright (1996) © Oxford University Press.)

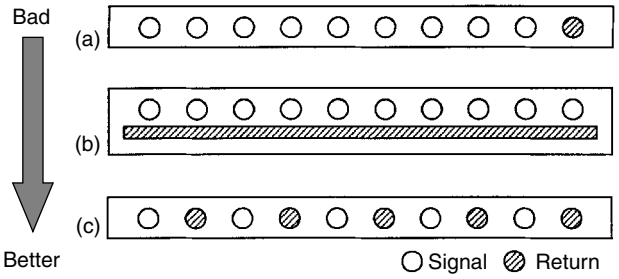


Figure 10. Ribbon cable configurations. (Reproduced from *Electronic Instrument Design: Architecting for the Life Cycle*, Fowler, K.R. Copyright (1996) © Oxford University Press.)

for low-frequency applications below 1 MHz. A complete 360° conducting seal is necessary for high frequencies above 1 MHz.

5.2 Component placement

Careful layout of components within circuits can provide a measure of self-shielding. Follow these guidelines for self-shielding:

- Place high-current circuits near the connectors to isolate stray currents.
- Place high-frequency circuits near the connectors to reduce signal path length, cross talk, and noise.
- Group the low-power and low-frequency circuits away from the high-frequency and high-current circuits to avoid disruption.
- Group analog circuits separately from digital circuits.

5.3 Signal lines and layout

A corollary to good component placement is to distribute the signal and return lines carefully to maximize self-shielding. Usually, this means pairing each signal line with a return line or plane. This will reduce large loop inductances and consequently the inductive impedance and coupled noise. This means one should not separate return lines long distances away from the signal line within connectors.

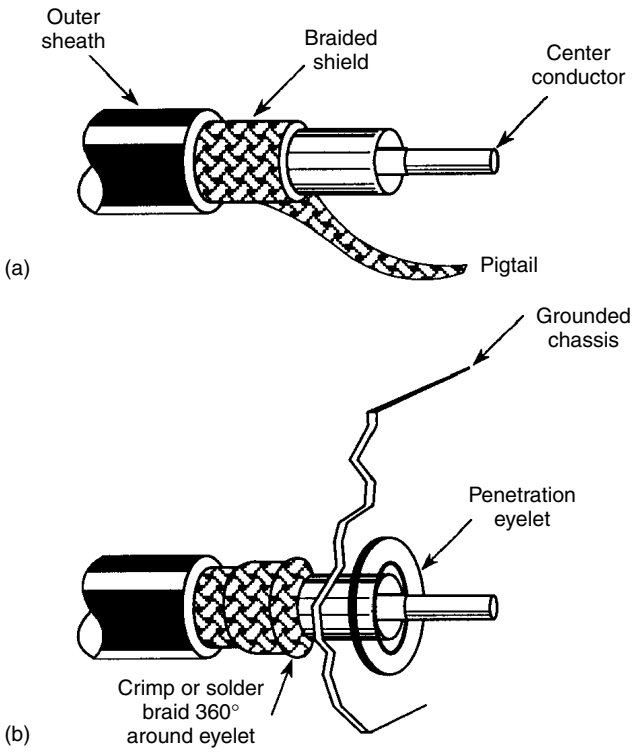


Figure 11. Coaxial cable terminations, the pigtail is only good for low frequencies below 1 MHz. (Reproduced from *Electronic Instrument Design: Architecting for the Life Cycle*, Fowler, K.R. Copyright (1996) © Oxford University Press.)

6 ENCLOSURES FOR NOISE REDUCTION

A shielded enclosure should ideally be a completely closed conducting surface. An effective enclosure is one that has watertight metallic seams and openings. Openings leak electromagnetic radiation all too easily (Figure 12). This will also help with ESD reduction. The design of enclosures that do not leak needs sophisticated design based on high-frequency transmission line design principles.

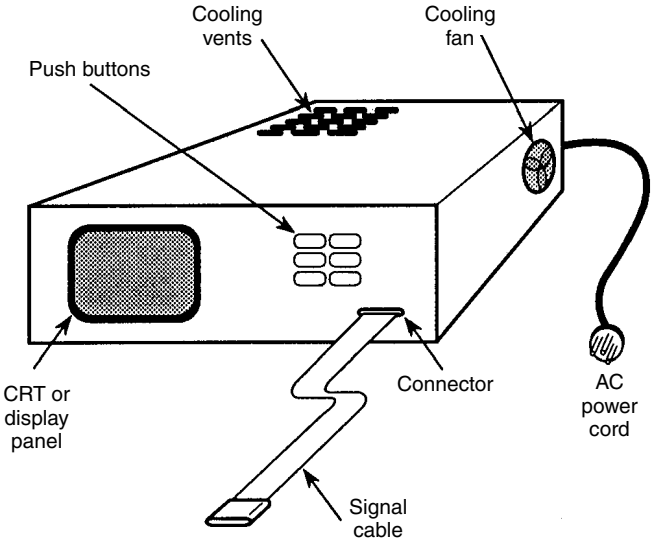


Figure 12. Examples of entry and exit points for electromagnetic energy. (Reproduced from *Electronic Instrument Design: Architecting for the Life Cycle*, Fowler, K.R. Copyright (1996) © Oxford University Press.)

7 PROTECTING AGAINST ELECTROSTATIC DISCHARGE (ESD)

Electrostatic discharge (ESD) is the transfer of electrical charge at very high voltage (100 s to 1000 s V) and very low current. ESD can easily damage sensitive electronic circuits. Any instrument containing integrated circuits is susceptible to ESD if not protected. Please refer to **Article 17, EMC and EMI, Volume 1** for more detail.

Control methods for ESD eliminate activities and materials that create high static charge and provide transient protection to sensitive circuits. Eliminating high static charge requires appropriate grounding, using protective handling and materials, and maintaining humidity.

Table 1. Checklist to reduce ESD.

Check activities
Wear a wrist ground-strap and use a static-free workstation.
Discharge static before handling devices.
Keep parts in their original containers.
Minimize handling of components.
Never slide a component over any surface.
Use conductive containers for storage and transport of components..
Clear all vinyl, plastic, and Styrofoam from the work area.

Source: Reproduced from *Electronic Instrument Design: Architecting for the Life Cycle*, Fowler, K.R. Copyright (1996) © Oxford University Press.

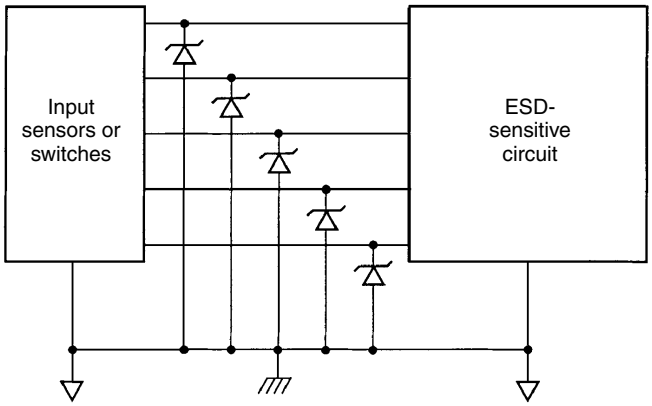


Figure 13. Preventing damage by shunting high-voltage transients away from circuits. (Reproduced from *Electronic Instrument Design: Architecting for the Life Cycle*, Fowler, K.R. Copyright (1996) © Oxford University Press.)

The checklist in Table 1 lists some of the activities that will reduce problems with ESD. Figure 13 illustrates one way to protective sensitive circuits by shunting ESD to chassis ground.

8 LIMITATIONS OF GROUNDING AND SHIELDING

No one method is a cure-all for EMC and EMI. It is necessary to carefully analyze the particular situation, understanding the modes and frequencies of operation, and the possible coupling mechanisms. Only then can a design begin to be an appropriate grounding and shielding system.

REFERENCES

Fowler, K.R. (1996) *Electronic Instrument Design: Architecting For The Life Cycle*, Oxford University Press, New York.

Johnson, H.W. and Graham, M. (1993) *High-Speed Digital Design: A Handbook of Black Magic*, PTR Prentice Hall, Englewood Cliffs, NJ.

Ott, H.W. (1988) *Noise Reduction Techniques in Electronic Systems*, 2nd edn, John Wiley & Sons, New York.

FURTHER READING

Van Doren, T. (1991) *Grounding and Shielding Electronic Systems*, Notes from video seminar on the NTU Satellite network.

178: Noise Matching and Preamplifier Selection

Richard Burdett
SIGNAL RECOVERY, Wokingham, UK

1 Preamplifier Selection 1182
2 Noise Figure 1182
3 Noise Matching with a Transformer 1183
4 Preamplifier Choices 1183
5 Cable Capacitance 1184
Related Articles 1185
Reference 1185

1 PREAMPLIFIER SELECTION

All preamplifiers add some additional noise. Whether this additional noise is significant will depend, of course, upon the noise level from the signal source. Since uncorrelated noise adds vectorially (in an rms fashion), the preamplifier noise can be neglected if it is less than about one-third of the source noise

$$\sqrt{[(1.0^2) + (0.3^2)]} \approx 1.0$$

We can think of a practical preamplifier as consisting of an ideal, noise-free amplifier with a (frequency-dependent) noise-voltage generator of voltage spectral density v_n (V/$\sqrt{}$Hz), and a noise-current generator of current spectral density i_n (A/$\sqrt{}$Hz), connected to its input as shown in Figure 1.

Figures 2 and 3, respectively, show separately the gain seen by the amplifier internal noise-voltage and -current generators. For voltage noise $v_{ov} = v_n(1 + Z_f/Z_s)$, where $Z_s = R_s/(1 + j\omega R_s C_s)$. Hence, any input shunt capacitance (see Figure 2) will decrease the input impedance and cause output noise that increases with frequency if Z_f is resistive.

The preamplifier noise may also be defined (Faulkner, 1966) in terms of an equivalent series noise resistance R_e, and an equivalent parallel noise resistance, R_i, where (from (2) in **Article 121, Signals in the Presence of Noise, Volume 2**.)

$$R_e = \frac{v_n^2}{4kT\Delta f} \text{ ohms}$$

and (from (3) in **Article 121, Signals in the Presence of Noise, Volume 2**.)

$$R_i = \frac{4kT\Delta f}{i_n^2} \text{ ohms}$$

2 NOISE FIGURE

Also defined is the noise figure (NF) of the preamplifier to be (when in dB units)

$$\text{NF} = 10\log_{10}\left(1 + \frac{R_e}{R_s} + \frac{R_s}{R_i}\right) \quad (1)$$

A perfect or noiseless preamplifier would have a 0 dB noise figure. Figure 4 shows the noise-figure contours that result when the noise figure of a particular practical preamplifier (the ***SIGNAL RECOVERY*** model 5113) is plotted as a function of source resistance and frequency. Notice from (1), that with high source resistance, $R_e/R_s \to 0$, and then

$$\text{NF} \approx 10\log_{10}\left(1 + \frac{R_s}{R_i}\right) \quad (2)$$

and the amplifier noise current, i_n, predominates.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

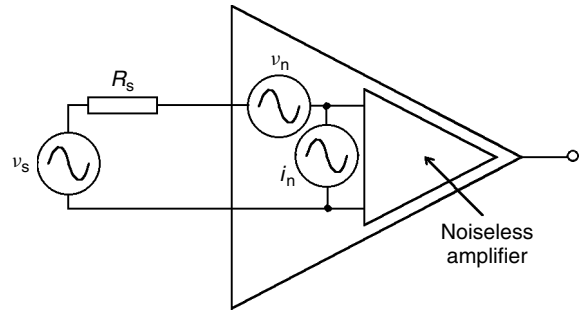


Figure 1. Amplifier noise – equivalent circuit. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

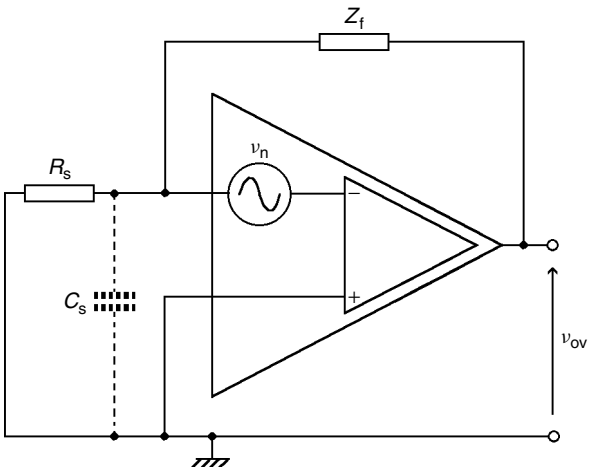


Figure 2. Amplifier noise – voltage noise. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

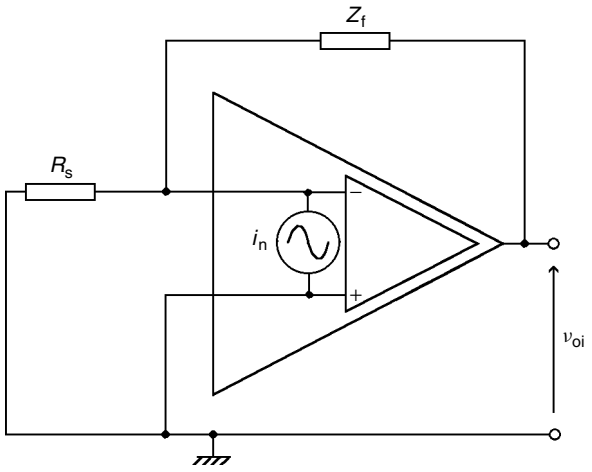


Figure 3. Amplifier noise – current noise. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

With low source resistances, the amplifier noise voltage, v_n, becomes the major noise source. Wherever possible, preamplifiers should be chosen so that their 3 dB noise-figure contour encloses the expected range of source resistance and frequency.

For a given preamplifier, the optimum source resistance, R_s, is given by

$$R_s(\text{opt}) = \frac{v_n}{i_n} = \sqrt{(R_e R_i)}\ \text{ohms} \tag{3}$$

Note that adding a series or parallel resistance between the signal source and the preamplifier always reduces signal and adds noise, and so cannot be used to obtain a better match.

Preamplifiers can be classified in many ways, one basic division being, for example, between differential input and single-ended input. All other things being equal, a differential preamplifier generates 3 dB (41.4%) more noise than a single-ended version. However, this disadvantage is significant only in situations in which preamplifier noise predominates and, in many cases, is outweighed by the flexibility of a differential input and its ability to remove ground-loop problems (*see* **Article 179, Input Connections; Grounding and Shielding, Volume 3**).

3 NOISE MATCHING WITH A TRANSFORMER

Transformers are often used to match very low source impedances (0.1 Ω–1 kΩ). Figure 5 shows an amplifier with an optimum, $\sqrt{(R_e R_i)}$, source resistance value of 1 MΩ being matched to a 100 Ω thermopile by means of a 100:1 voltage step-up transformer (10 000:1 impedance transformation). Note that, in general, such noise matching does not result in the same circuit values as would power matching; that is, the amplifier input resistance is not normally equal to $\sqrt{(R_e R_i)}$.

Transformers are a classic solution where high performance amplifiers are not available. Their use, however, should be avoided if possible since they reduce frequency response, may pick up magnetically induced interference, and may be microphonic (that is, they may generate noise when mechanical shock or vibration causes disturbance of their magnetic circuit).

4 PREAMPLIFIER CHOICES

For sources of approximately 100 Ω–10 kΩ, preamplifiers are available that use an operational amplifier with an input

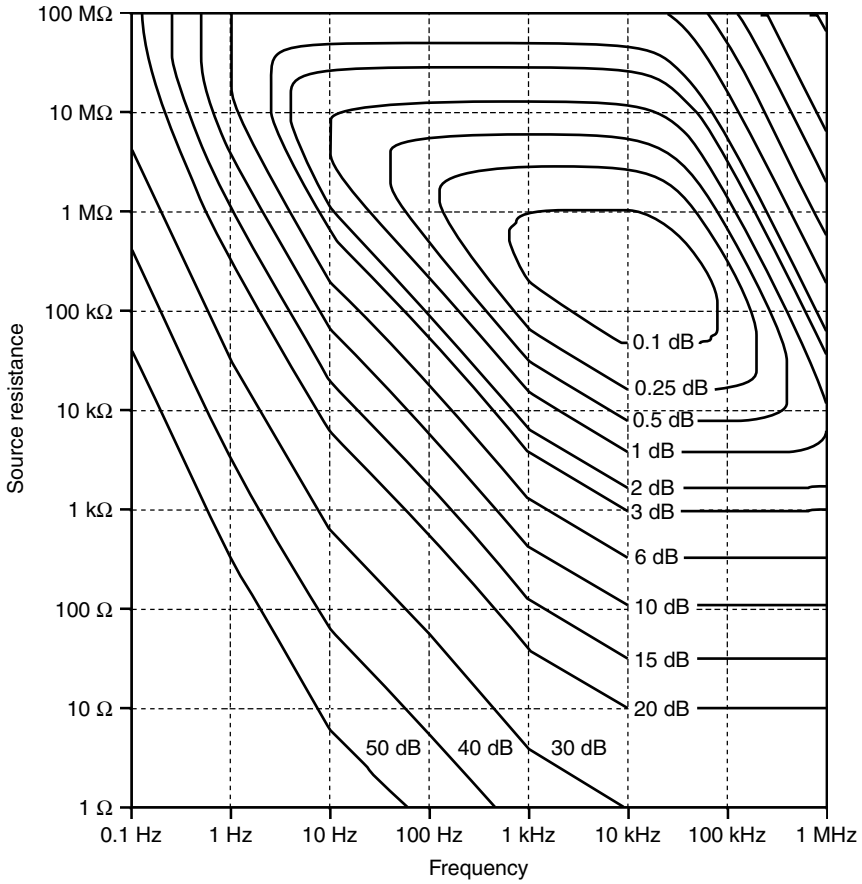


Figure 4. Noise-figure contours for a ***SIGNAL RECOVERY*** model 5113 voltage preamplifier. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

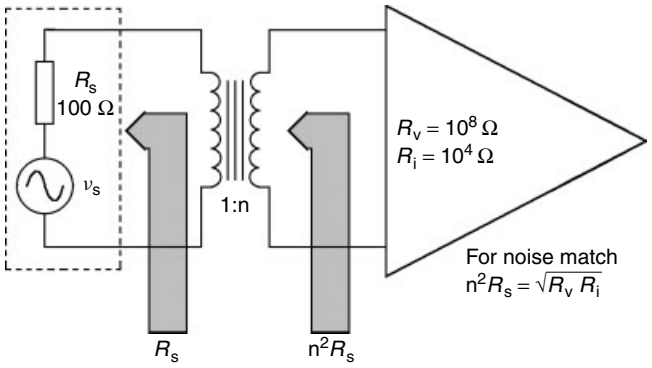


Figure 5. Transformer noise matching. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

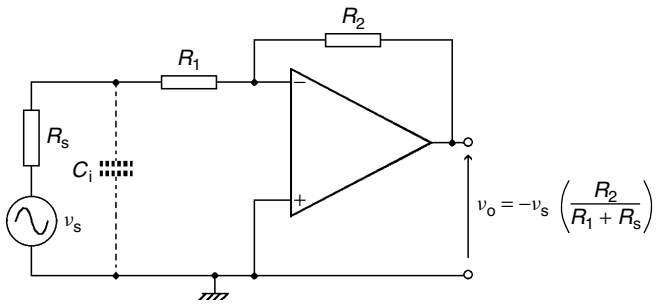


Figure 6. Inverting voltage amplifier. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

stage consisting of multiple bipolar transistors connected in parallel to provide a lower value of $\sqrt{(R_e R_i)}$.

Such preamplifiers avoid the bandwidth constraints imposed by input transformers. For higher impedance sources (1 kΩ–100 MΩ), preamplifiers usually employ junction-FETs as input devices and are available as voltage preamplifiers, charge amplifiers (for use with capacitive transducers), or current-input (transimpedance) amplifiers, see Figures 6, 7, 8, and 9 for these forms.

5 CABLE CAPACITANCE

In Figures 6 and 7 any cable capacitance or stray capacitance, C_i, will form a low-pass filter with the source

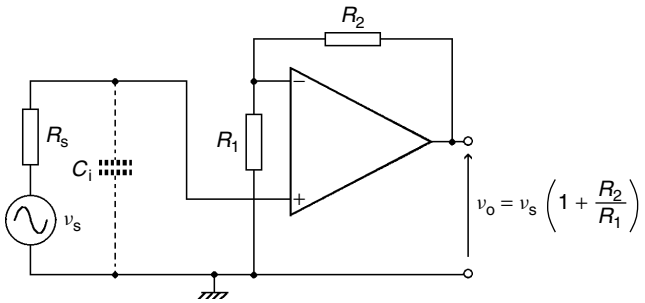


Figure 7. Noninverting voltage amplifier. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

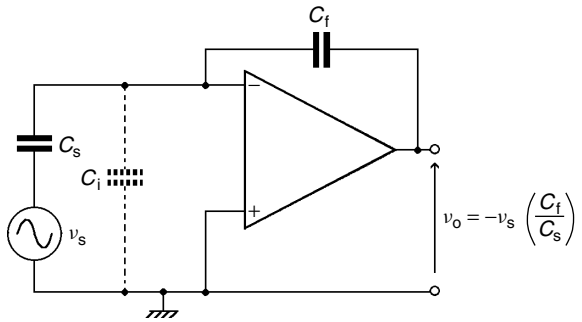


Figure 8. Charge amplifier. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

resistance, R_s, having a – 3 dB frequency given by

$$f_c = \frac{1}{2\pi R_i R_s C_i/(R_i + R_s)} \tag{4}$$

(refer to Figure 6)
or

$$f_c = \frac{1}{2\pi R_s C_i} \tag{5}$$

(refer to Figure 7)

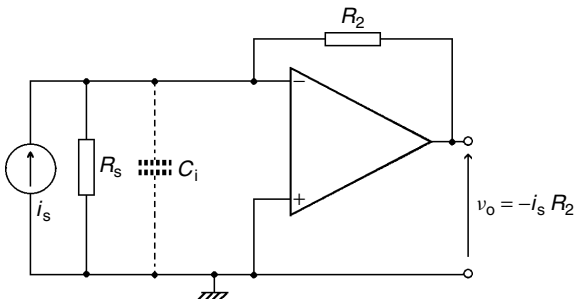


Figure 9. Transimpedance amplifier. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

In Figures 8 and 9, such shunt capacitance appears at first sight to have no effect since it is effectively short-circuited by the virtual-ground input. However, as shown in Figure 2, shunt capacitance will cause deterioration in the output signal-to-noise ratio, and also (by introducing an additional pole into the loop gain) may cause ringing in the amplifier response, or even oscillation.

By careful design (which usually includes adding a capacitor across the feedback resistor), these effects can be minimized and with high source impedance, commercial current and charge amplifiers usually provide significantly greater bandwidth than that given by the corresponding voltage amplifiers.

RELATED ARTICLES

Article 121, Signals in the Presence of Noise, Volume 2; **Article 179, Input Connections; Grounding and Shielding, Volume 3**; **Article 180, Bandwidth Reduction of Baseband DC Signals, Volume 3**; **Article 181, Amplitude Modulated Signals: The Lock-in Amplifier, Volume 3**.

REFERENCE

Faulkner, E.A. (1966) Optimum Design of Low-noise Amplifiers. *Electronics Letters*, **2**, 426–427.

This article is based in part on a chapter by D.M. Munro originally published in Handbook of Measurement Science Vol 1, P.H. Sydenham, Wiley, 1982.

179: Input Connections; Grounding and Shielding

Richard Burdett
SIGNAL RECOVERY, Wokingham, UK

1 Common-mode Sensors 1186
2 Ground Loops 1186
3 Reduction of Ground-loop Inputs 1186
4 Improving Ground-loop Situations 1188
5 Other Reduction Mechanisms 1189
Related Articles 1189
Reference 1189

1 COMMON-MODE SENSORS

Ideally, all grounds should have a zero-impedance connection to each other and to wet earth; in practice, they do not.

Owing to voltage drops existing across their finite impedance to earth, capacitively or inductively coupled interference, and other reasons, each ground point connection tends to be at a different potential from other nearby grounds. If two (or more) such adjacent grounds are connected together to form a ground loop (Figure 1), then the potential difference between the two grounds will cause a circulating current in the input of the detector stage.

The potential difference between grounds (v_{CM}) is called the *common-mode source* since it is common to both the signal (via loop 2) and ground (via loop 1) inputs of the preamplifier.

2 GROUND LOOPS

Figure 2 rearranges the circuit of Figure 1 and assumes the signal source v_s to be zero.

Note that the low resistance of the connecting coaxial cable shield (the braid) R_{CG} is in parallel with the series combination of the source resistance R_S, the coaxial cable center conductor resistance R_{CS}, and the preamplifier input impedance Z_{IN}.

Under normal circumstances, $(R_S + R_{CS} + Z_{IN}) \gg R_{CG}$ and $Z_{IN} \gg (R_S + R_{CS})z$ so that, as shown in Figure 3, the common-mode voltage dropped across R_{CG} is also applied across the preamplifier input terminals. More generally, with $v_s = 0$, the preamplifier input voltage, v_i, is given by

$$v_i = v_{CM} \frac{R_{CG}}{R_{CG} + R_{SG} + R_{PG}} \tag{1}$$

3 REDUCTION OF GROUND-LOOP INPUTS

From (1), the common-mode input to the preamplifier can be removed (i.e. $v_i = 0$) by making

1. $v_{CM} = 0$. This can be attempted by grounding the source and preamplifier to the same ground point (all connected at one point) and by using shielding to remove capacitively or inductively coupled interference. However, this procedure is rarely completely successful.
2. $R_{CG} = 0$. The usual approach here is to bolt both source and preamplifier chassis to a large metal plate. Unfortunately, it is fairly easy to develop large potential differences between points as small as a centimeter or two apart on a large metal plate, such as a mounting rack.
3. $R_{SG} = \infty$. Floating or disconnecting the source from ground is a good approach where practicable, but this

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

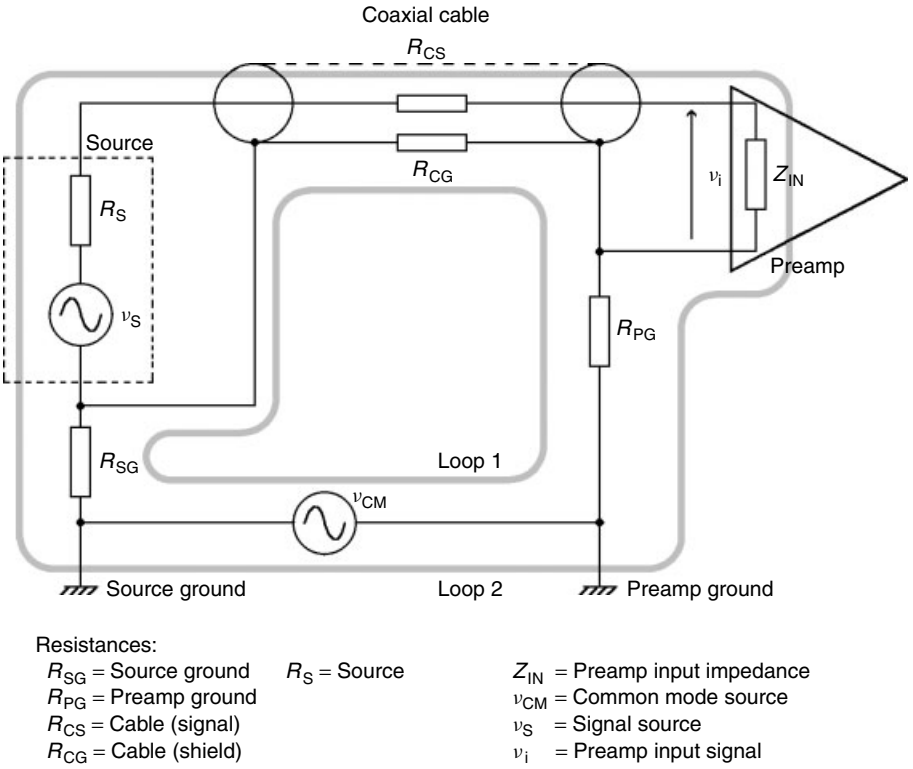


Figure 1. Ground loops – physical occurrence. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

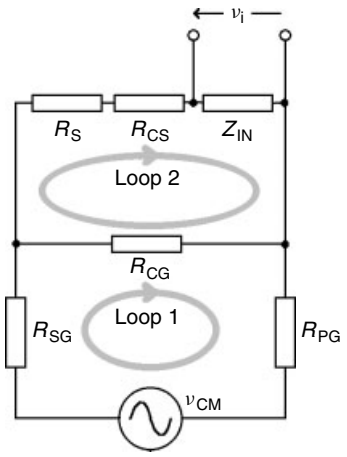


Figure 2. Ground loops – schematic equivalent circuit. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

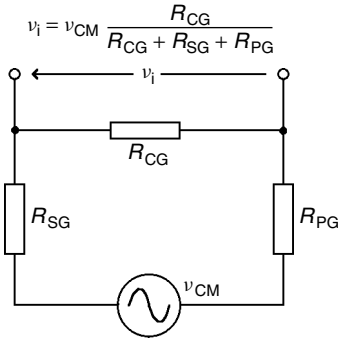


Figure 3. Ground loops – reduced equivalent circuit. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

needs careful design and maintenance to ensure it is truly floating.

4. $R_{PG} = \infty$. The preamplifier may also be floated – particularly, if it is battery powered. Note that disconnecting the power-line ground from an instrument can be extremely dangerous, and in some countries it could be illegal. In many instruments, R_{PG} consists of an internal 10 Ω to 1 kΩ resistor that can be switched into the circuit to effectively float the amplifier input terminals.

None of the above methods is usually able to produce perfect results. Ground loops are an important and challenging design issue that need specific measures to be applied for each application.

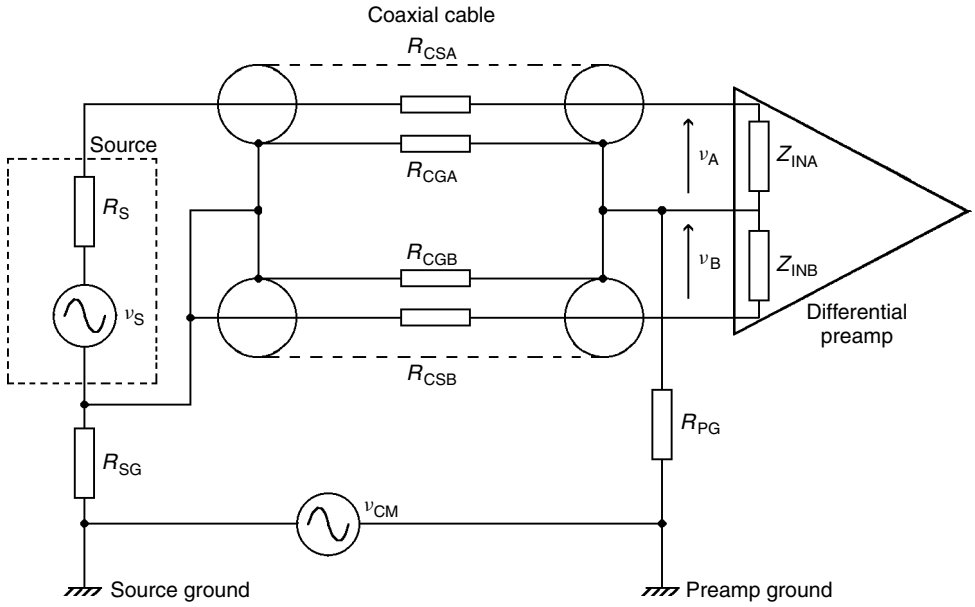


Figure 4. Differential preamplifier used with single-ended source – physical occurrence. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

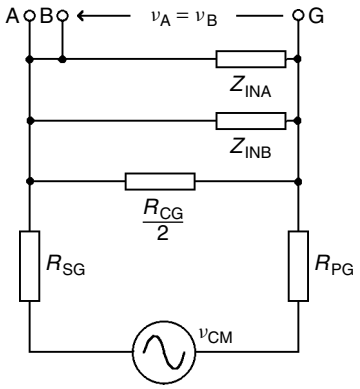


Figure 5. Differential preamplifier used with single-ended source – schematic equivalent circuit. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

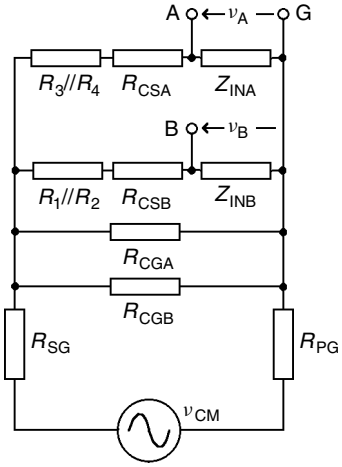


Figure 6. Differential preamplifier used with single-ended source – reduced equivalent circuit. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

4 IMPROVING GROUND-LOOP SITUATIONS

Figure 4 illustrates the use of a differential amplifier with an unbalanced (single-ended) source to eliminate or reduce ground-loop problems. As in Figures 1 to 3, the circuit simplifications shown in Figures 5 and 6 assume that the input impedance of each side of the differential amplifier (Z_{INA} and Z_{INB}) is much larger than source or cable resistances.

At low frequencies, this differential connection results in equal common-mode voltages at the amplifier's input terminals (A and B), and the amplifier's ability to discriminate against common-mode signals (i.e. its common-mode rejection ratio, CMRR or CMR) will determine the effectiveness of this configuration in suppressing ground-loop interference.

At higher frequencies, the cable capacitances will act with the unequal resistances in the A and B input circuits to form unequal low-pass filters so that v_A will no longer be equal to v_B and there will be a spurious differential (A – B) input to the preamplifier.

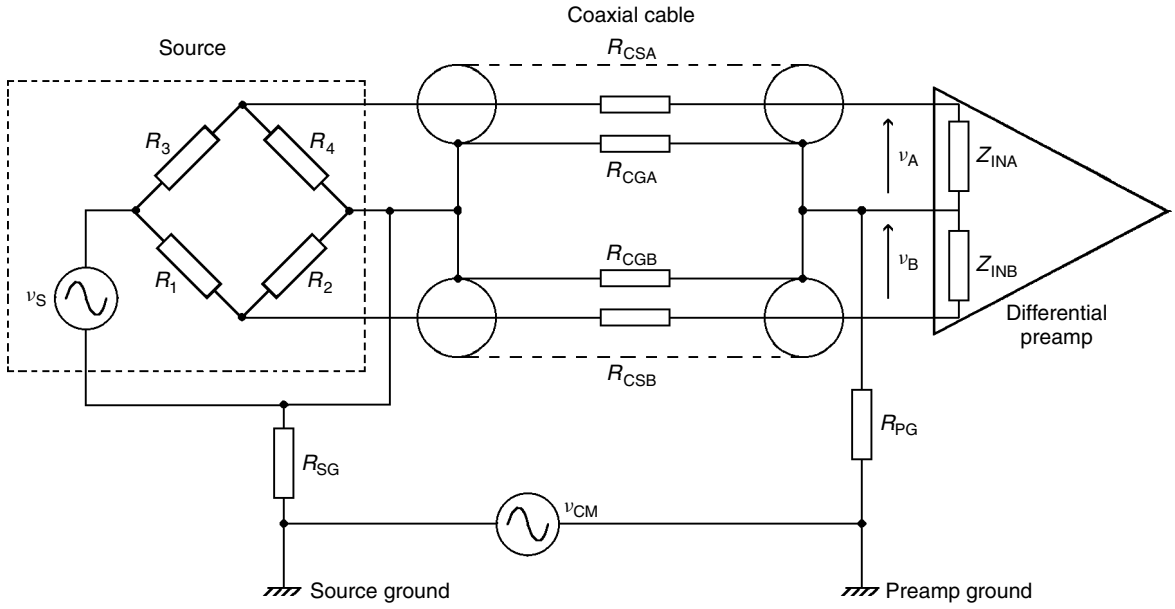


Figure 7. Differential preamplifier used with balanced source – physical occurrence. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

Although cable resistances and capacitances are shown for convenience as lumped parameters, it should be remembered that they are in fact distributed. As shown in Figure 7, high-frequency unbalance problems can be avoided by using a balanced source. Specific comment on connecting stages together is to be found in Morrison (1977), *see also* **Article 177, Grounding and Shielding, Volume 3**.

5 OTHER REDUCTION MECHANISMS

To end this article, here are a number of miscellaneous recommendations regarding good wiring and grounding practices.

1. Keep cable lengths short; for differential connections, keep them equal and following the same route.
2. Interference can be coupled into the ground (shield) or the outer conductor of a coaxial cable. Consider coiling the cable to form an RF choke to suppress high-frequency interference of this kind, use a transformer, or use a balun (which allows DC continuity).
3. Remember that a circuit loop of wire acts as an antenna; reduce the area of such loops as much as possible.
4. Separate low-level signals/cables from noisy ones. Where such cables must cross, cross them at right angles and with maximum separation.
5. For noncoaxial connections, use shielded twisted wire pairs.
6. Consider placing low-noise instruments in a shielded (screened) room when they are used with high-energy RF sources, such as pulsed lasers.
7. Keep analog and digital grounds separate.

RELATED ARTICLES

Article 121, Signals in the Presence of Noise, Volume 2; **Article 176, Signals and Signal-to-noise Ratio, Volume 3**.

REFERENCE

Morrison, R. (1977) *Grounding and Shielding Techniques in Instrumentation*, Wiley, Chichester.

This article is based in part on a chapter by D.M. Munro originally published in Handbook of Measurement Science Vol 1, P.H. Sydenham, Wiley, 1982.

180: Bandwidth Reduction of Baseband DC Signals

Richard Burdett
SIGNAL RECOVERY, Wokingham, UK

1 Bandwidth Reduction to Improve Signal-to-noise Ratio, SNR 1193
2 SNR Improvement from Photo Multiplier Tube (PMT) Detector 1193
3 Noise Sources in PMT Detectors 1194
4 Strategies for SNR Improvement in PMT Detectors 1195
5 Roll-off Rate of PMT Detectors 1195
Related Articles 1196

1 BANDWIDTH REDUCTION TO IMPROVE SIGNAL-TO-NOISE RATIO, SNR

The term DC signal is often used (and is in this article) to mean a signal that has a frequency spectrum that includes zero frequency DC.

Technically, of course, a truly DC voltage or current is unvarying and, therefore, cannot carry information (other than that it exists). Here, we are concerned with signals that are close to DC; an example is the temperature variation seen by a thermal detector as a person moves across its sensor in a room.

Such signals are also termed *baseband signals*, particularly when they are to be used to modulate a carrier frequency. They may also be referred to as a *pseudo-DC signal*.

Where the unwanted noise has a wider bandwidth than the signal, the simplest way to improve the signal-to-noise ratio SNR is to use a low-pass filter to reduce the noise bandwidth to the point where any further reduction would also change the wanted signal to an unacceptable extent. The total noise energy is thereby reduced but that of the signal is hardly altered. Thus, the SNR improves.

2 SNR IMPROVEMENT FROM PHOTO MULTIPLIER TUBE (PMT) DETECTOR

Figure 1 shows a typical source and preamplifier system for such a pseudo-DC signal. We will use this circuit – for a PMT detector – to show how the output SNR may by estimated and also how the SNR may be improved by reducing the noise bandwidth.

In this example, it is assumed that the photomultiplier tube PMT anode current consists of both a 5-Hz signal component with source current of ($i_S = 1\,\text{nA rms}$) and a DC current component of ($I_{DC} = 5\,\text{nA}$); typically, such DC currents in the PMT are due to stray light and dark/leakage currents.

The adjustable direct-current generator (I_{ZS}) is used to null, that is set the zero offset to zero, the DC component of the PMT current such that I_{ZS} is made equal and opposite to I_{DC}. This kind of zero suppression is often called *background subtraction*. Note in this simple arrangement that I_{ZS} must be readjusted manually each time the background changes.

The coaxial cable shown that connects the PMT and preamplifier has its own capacitance. Note that the virtual-ground input of this preamplifier configuration offers the following advantages in addition to those discussed in **Article 178, Noise Matching and Preamplifier Selection, Volume 3**. In this case then,

1. with zero volts existing across it, the cable capacitance cannot be charged and the cable will, therefore, be less microphonic than otherwise;

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

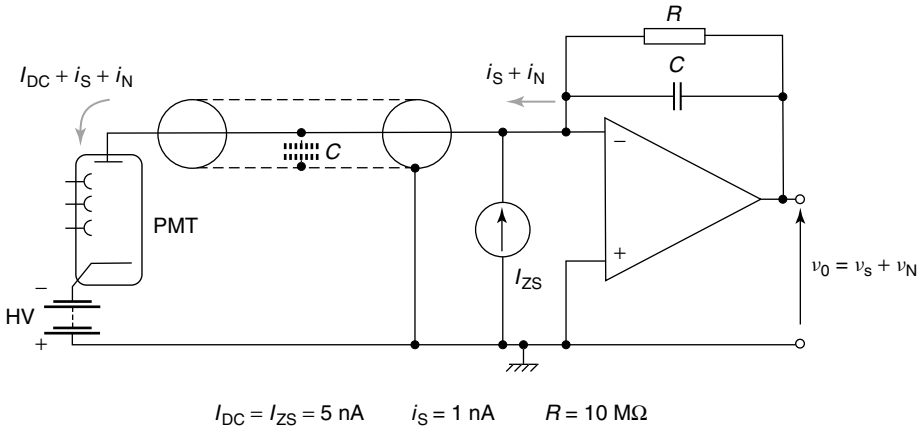


Figure 1. DC measurement example using photomultiplier (PMT) detector. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

2. since the PMT anode voltage is clamped at zero volts, the anode-to-last-dynode voltage is also held constant regardless of i_S (assuming that the dynode voltage remains fixed). Signal currents will, therefore, not change the PMT gain.

3 NOISE SOURCES IN PMT DETECTORS

There are five uncorrelated sources of noise in this circuit. These are as follows:

1. The DC component of the PMT current (I_{DC}) is produced by integrating anode pulses, each of charge $Q = Ae$ where A is the mean PMT gain. The interval in time between successive pulses arriving at the anode is random and governed by Poisson statistics (*see* **Article 184, Photon Counting, Volume 3**, Section 4). Assuming no additional dynode noise exists in the PMT, the rms value of the PMT shot noise current spectral density i_{N1} is given by

$$i_{N1} = \sqrt{(2AeI_{DC})} = \sqrt{A} \times \sqrt{(2eI_{DC})} \qquad (1)$$

For $A = 10^6$ (say), $R = 10^7\Omega$, $e \simeq 1.6 \times 10^{-9}$C, and $I_{DC} = 5$ nA, the resulting output noise voltage density, v_{N1} is given by

$$\begin{aligned} v_{N1} &= Ri_{N1} = 10^7 \times 10^3 \times \sqrt{2eI_{DC}} \\ &\simeq 10^{10} \times 4 \times 10^{-14} \\ &= 4 \times 10^{-4} = \frac{400\,\mu\text{V}}{\sqrt{\text{Hz}}} \end{aligned} \qquad (2)$$

2. For purposes of this example, we can assume that the zero-suppress current, I_{ZS}, is obtained from a transistor current source circuit so that it has a shot noise current spectral density, i_{N2}, given by

$$\begin{aligned} i_{N2} &= \sqrt{2eI_{ZS}} = \sqrt{(2eI_{DC})} \left(= \frac{i_{N1}}{10^3}\right) \\ &= \frac{4 \times 10^{-14}\text{A}}{\sqrt{\text{Hz}}} \end{aligned} \qquad (3)$$

Note that although $I_{DC} = I_{ZS}$, the shot noise component from the PMT is much larger than that from the transistor current source. It is thus detector limited. The resulting output noise voltage spectral density, v_{N2}, is given by

$$\begin{aligned} v_{N2} &= Ri_{N2} \simeq 10^7 \times 4 \times 10^{-14} \\ &= 4 \times 10^{-7} = \frac{400\,\text{nV}}{\sqrt{\text{Hz}}} \end{aligned} \qquad (4)$$

3. The feedback resistor, R, will generate (at $T = 290$ K) a Johnson noise output voltage density, v_{N3} given by

$$v_{N3} = \sqrt{(4kTR)} \approx 4 \times 10^{-7} = \frac{400\,\text{nV}}{\sqrt{\text{Hz}}} \qquad (5)$$

4. At 5 Hz, a typical value for the spot noise voltage density of the amplifier's internal noise voltage generator is $30\,\text{nV}/\sqrt{\text{Hz}}$. This amplifier voltage noise will experience unity gain (*see* **Article 178, Noise Matching and Preamplifier Selection, Volume 3**, Figure 2) since Z_{IN} (the PMT current source) is very high. The output noise voltage density v_{N4} due to this noise source is therefore given by

$$v_{N4} = \frac{30\,\text{nV}}{\sqrt{\text{Hz}}} \qquad (6)$$

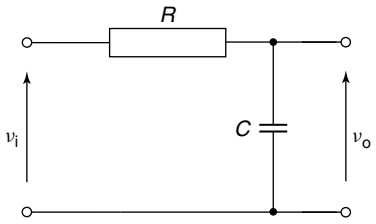


Figure 2. Low-pass filter, single pole. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

5. At 5 Hz, a typical value for the spot noise current density, i_{N5} of the amplifier internal noise current generator is $5fA/\sqrt{\text{Hz}}$. The resulting contribution v_{N5} to the amplifier output noise is given by

$$v_{N5} = Ri_{N5} = 10^7 \times 5 \times 10^{-15} = 5 \times 10^{-8} = \frac{50\,\text{nV}}{\sqrt{\text{Hz}}} \tag{7}$$

The total output noise voltage spectral density, v_N is given by

$$v_N = \sqrt{(v_{N1}^2 + v_{N2}^2 + v_{N3}^2 + v_{N4}^2 + v_{N5}^2)} \tag{8}$$

Since $v_{N1}^2 \gg v_{N2}^2$, v_{N3}^2, v_{N4}^2, and v_{N5}^2, then

$$v_N \approx v_{N1} \tag{9}$$

and the system is said to be detector limited or shot noise limited. An electrometer, an instrument characterized by extremely low leakage currents, is often used as a low-noise amplifier in DC measurements of this kind.

While discussing the limit of detection of radiant energy sensors, it is appropriate to cover the situation wherein the detector has geometry that also decides the sensitivity. Each design type and individual unit will then have different apparent sensitivities, but by addition of suitable optical and electrical set ups their differences can be normalized. Thus, it is not so meaningful to talk of their individual sensitivities without some means to normalize them all to a common method of comparison. For sensitive optical and other wavelength radiation detectors, this is often done using their detectivity or D* rating.

4 STRATEGIES FOR SNR IMPROVEMENT IN PMT DETECTORS

In Figure 1, the parallel resistor and capacitor in the feedback loop cause the circuit to act as low-pass filter of time constant RC seconds, so that the -3 dB cutoff frequency is given by $1/2\pi RC$ and $B_N = 1/4RC$ (*see* **Article 121, Signals in the Presence of Noise, Volume 2**, Section 2).

If no discrete capacitor is connected across R, the typical stray capacitance will (say) be about $C = 2.5$ pF, so that $RC = 10^7 \times 2.5 \times 10^{-12} = 25\,\mu\text{s}$ and $B_N = 10^4$ Hz. The output noise voltage (V_N) will, therefore, be

$$V_N = v_N\sqrt{B_N} = 4 \times 10^{-4} \times \sqrt{10^4} = 40\,\text{mV} \tag{10}$$

The output signal

$$v_S = i_S R = 10^{-9} \times 10^{-7}\,\text{V} = 10\,\text{mV} \tag{11}$$

Therefore,

$$\text{SNR} = \frac{S}{N} = \frac{10}{40} = \frac{1}{4} \tag{12}$$

The capacitance can be increased to 2.5 nF by adding discrete capacitors so that the noise bandwidth becomes $B_N = 10$ Hz. The -3 dB corner frequency will now be at 6.4 Hz (i.e. $10 \times 2/\pi$) so that the signal (for which the frequency is 5 Hz) is not significantly attenuated. The output noise voltage (V_N) is now reduced to

$$V_N = v_N\sqrt{B_N} = 4 \times 10^{-4} \times \sqrt{(10)} \approx 1.26\,\text{mV} \tag{13}$$

and

$$\text{SNR} = \frac{S}{N} = \frac{10}{1.26} \approx \frac{8}{1} \tag{14}$$

so that (*see* **Article 176, Signals and Signal-to-noise Ratio, Volume 3**; (3) and (4))

$$\text{SNIR} = \frac{S_o/N_o}{S_i/N_i} = \sqrt{\frac{B_{NI}}{B_{NO}}} = \sqrt{\frac{10000}{10}} \approx \frac{32}{1} \tag{15}$$

5 ROLL-OFF RATE OF PMT DETECTORS

A typical single pole low-pass filter section is shown in Figure 2, and exhibits a 6 dB/octave roll-off with frequency above the cutoff frequency (curve 1 in Figure 3).

The roll-off rate of this filter may be increased by adding more RC sections, and there are several ways of doing this. The simplest is shown in Figure 4, where two single pole stages are connected together via a unity gain buffer amplifier. The component values give a filter with the same overall equivalent noise bandwidth (ENBW) as that shown in Figure 2. This filter has the frequency response characteristic shown in curve 2 of Figure 3.

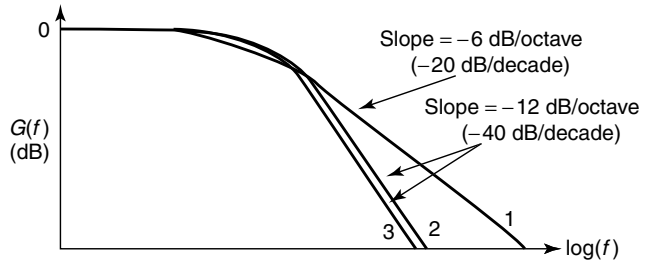


Figure 3. Low-pass filter frequency responses. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

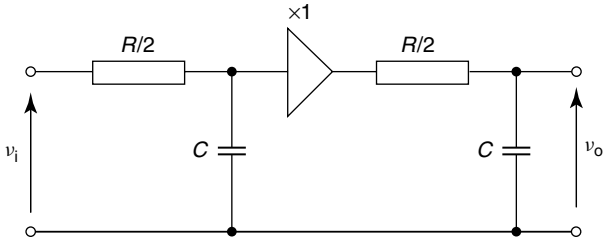


Figure 4. Low-pass filter, two equal buffered poles. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

More sophisticated multipole designs can be used, such as a Butterworth filter as shown in Figure 5, again which has component values giving the same ENBW as for the filter shown in Figure 2. This has the frequency response characteristic shown in curve 3 of Figure 3.

However, it must be remembered that the magnitude response of the filter with frequency is only one aspect of its performance. Its response in the time domain may also be important, particularly when measuring values that are changing as a function of time. Special care should be taken in using some filter configurations (Chebyshev or Butterworth, for example) since many such filters have undesirable overshoot characteristics (see Figure 6, where curves 1, 2, and 3 correspond to the filters shown in Figures 2, 4, and 5 respectively).

Note that the term time constant (τ) is meaningful only in connection with a single *RC* filter section and, even then, does not adequately convey a sense of the response time of the filter. With a voltage-step input, for example, such a single *RC* section requires about five time-constant

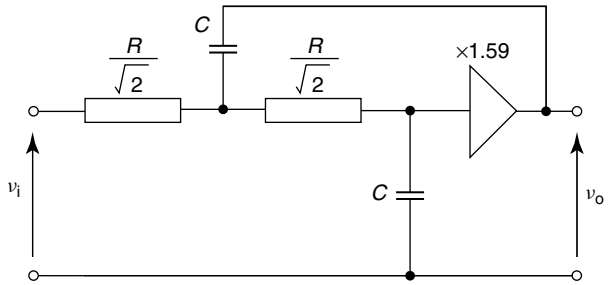


Figure 5. Low-pass filter, two pole Butterworth. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

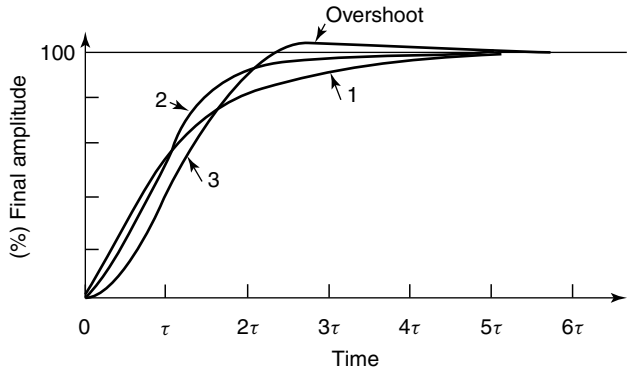


Figure 6. Low-pass filters, time response. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

intervals for its output to rise to within 1% of its final value.

Thus, it has been shown that addition of relatively small circuit changes can significantly improve the SNR in the baseband detection situation. When using other improvement methods, it is still necessary to carry out these kinds of improvements to gain the full effectiveness in signal recovery practice.

RELATED ARTICLES

Article 121, Signals in the Presence of Noise, Volume 2; **Article 178, Noise Matching and Preamplifier Selection, Volume 3**; **Article 179, Input Connections; Grounding and Shielding, Volume 3**.

This article is based in part on a chapter by D.M. Munro originally published in Handbook of Measurement Science Vol 1, P.H. Sydenham, Wiley, 1982.

181: Amplitude Modulated Signals: The Lock-in Amplifier

Richard Burdett
SIGNAL RECOVERY, Wokingham, UK

1 Flicker Noise 1197
2 Band-pass Filter in Modulated Systems 1197
3 Effect of Q Factor on Noise Bandwidth 1199
4 The Lock-in Amplifier 1199
5 Description of the Various Elements of a Lock-in Amplifier 1200
6 Phase-sensitive Detector Designs – Analog and Digital Techniques 1202
7 Single and Dual-phase Lock-in Amplifiers 1203
8 Limits of Detection for Lock-in Amplifiers 1204
9 Examples of Lock-in Amplifier Applications 1205
10 Advanced Techniques 1206
Reference 1207
Further Reading 1208

1 FLICKER NOISE

Most measurement systems are troubled by $1/f$ or flicker noise. By amplitude modulating the measurand (the quantity to be measured) at some reference or carrier frequency, f_r, the output noise can often be reduced and DC drift problems avoided (see Figure 1).

In optical systems, for example, rotating or vibrating mechanical chopper blades are often used to periodically block a light beam and thereby square-wave modulate the signal amplitude – even though, in most cases, such chopping means losing half of the light (signal). Measuring instruments that respond only to the modulation provide automatic background subtraction; as with DC systems, however, the noise component of the background remains. Such modulation also allows the use of transformers to noise-match preamplifiers to low-resistance sources.

2 BAND-PASS FILTER IN MODULATED SYSTEMS

As with base band signals and low-pass filtering, the SNR of a noisy amplitude-modulated signal can be improved by bandwidth reduction – in this case a band-pass filter is commonly used. In most applications, carrier frequencies are chosen from the 100- to 10-kHz range, where preamplifier and environmental noise is lowest; care should also be taken to avoid frequencies occupied by the fundamental or harmonics of the power-line frequency.

A second-order band-pass filter (see Figures 2 and 3) is specified by its resonant or center frequency f_r and its selectivity Q quality factor. For a given value of f_r, the higher the Q, the narrower will be the filter width.
The -3 dB frequencies are at $f_r \pm f_c$ and the signal bandwidth (B_s) is defined by

$$B_s = 2f_c = \frac{f_r}{Q} \tag{1}$$

For a second-order band pass, the signal bandwidth and the equivalent noise bandwidth (B_n are related by

$$B_n = \frac{1}{2}\pi B_s \tag{2}$$

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

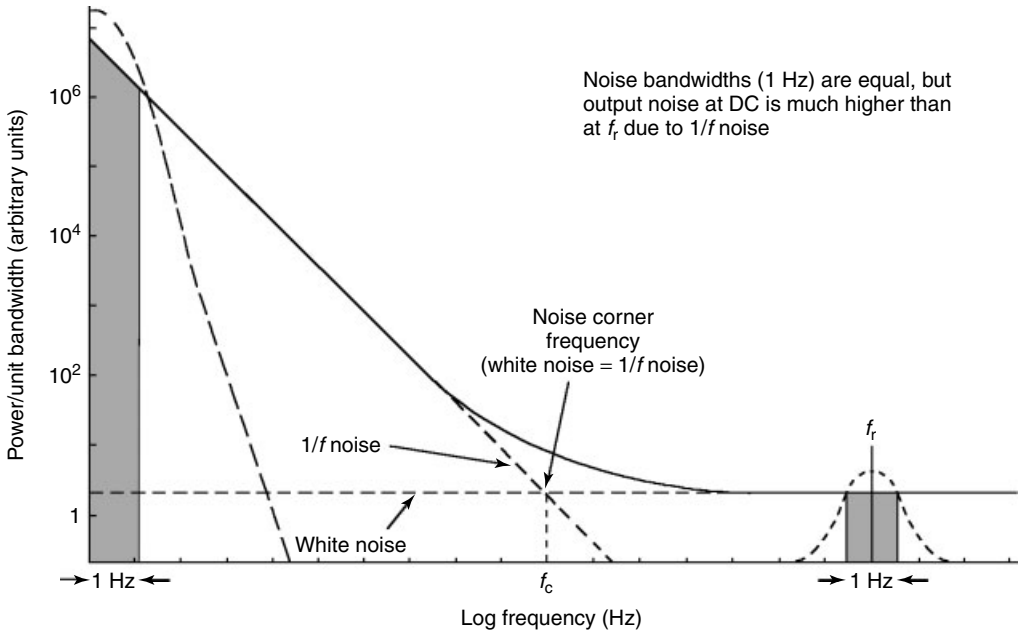


Figure 1. Amplitude modulation to avoid $1/f$ noise.

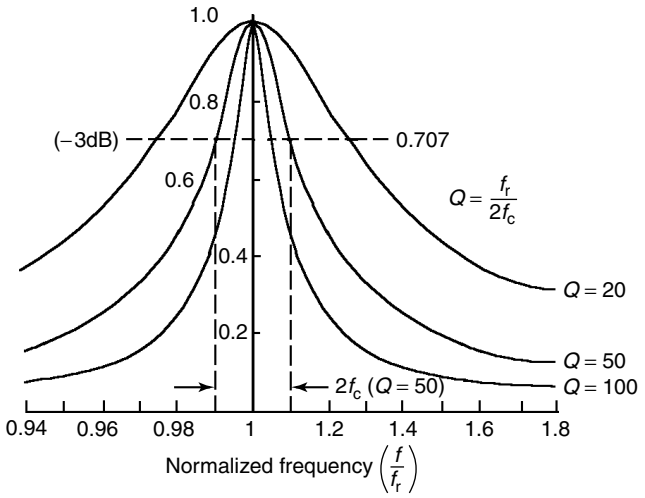


Figure 2. Amplitude transfer characteristics of a second-order band-pass filter.

so that,

$$B_n = \frac{\pi f_r}{2Q} \tag{3}$$

The band-pass filter has associated with it an effective time constant τ, where (as with the low-pass filter case discussed in **Article 121, Signals in the Presence of Noise, Volume 2**)

$$f_c = \frac{1}{2\pi\tau} \tag{4}$$

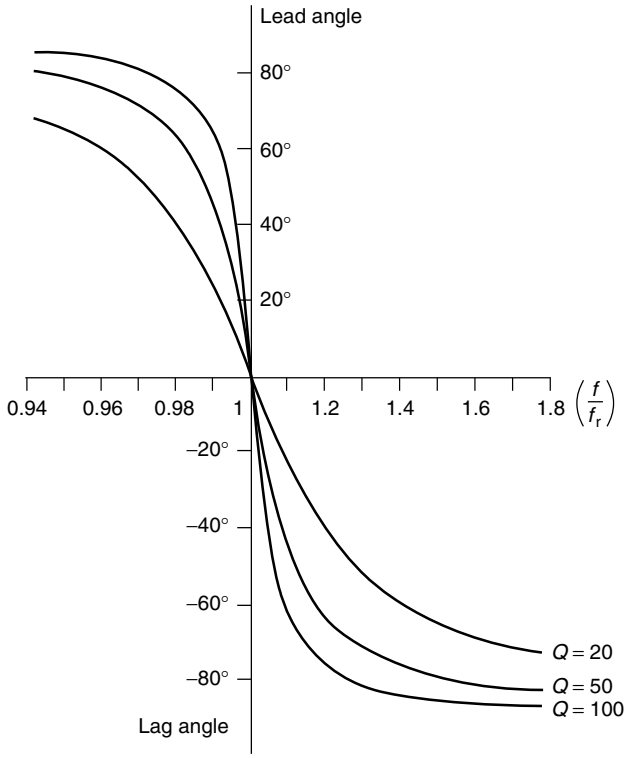


Figure 3. Phase transfer characteristics of a second-order band-pass filter.

so that,

$$B_n = \frac{1}{2}\pi B_s = \frac{1}{2}\pi 2 f_c = \frac{1}{2}\pi\left(\frac{2}{2\pi\tau}\right) = \frac{1}{2\tau} \tag{5}$$

3 EFFECT OF Q FACTOR ON NOISE BANDWIDTH

Also, from (1) and (4)

$$\tau = \frac{1}{2\pi f_c} = \frac{Q}{\pi f_r} \tag{6}$$

From (2), we can see that the higher the Q, the smaller the noise bandwidth and, therefore, for white noise or other broadband noise interference, the smaller the noise, the better the SNR. With a band-pass filter implemented by active *RC* (or *LC)* circuitry, frequency-stability problems limit the maximum practicable value of Q to about 100.

4 THE LOCK-IN AMPLIFIER

The *lock-in amplifier* (Burdett 2002a) can be considered as a tracking band-pass filter and amplifier that overcomes the Q-factor limitations of conventional circuits, offering noise bandwidths of less than 0.001 Hz and Q-factor values of 10^8. The lock-in amplifier can also provide amplification of more than 10^9 (180 dB) over a frequency range from 0.001 Hz to 200 MHz.

In common with most AC indicating instruments, it provides a DC output proportional to the AC signal being measured. In modern units, the DC output may be presented as a reading on a digital panel meter or as a digital value communicated over a computer interface, rather than a voltage at an output connector, but the principle remains the same.

The special rectifier, called a *phase-sensitive detector* (PSD), which performs this AC to DC conversion forms the heart of the instrument. It is special in that it rectifies only the signal of interest while suppressing the effect of noise or interfering components that may accompany that signal.

The traditional rectifier, which is found in a typical AC voltmeter, makes no distinction between signal and noise and produces errors due to rectified noise components. The noise at the input to a lock-in amplifier, however, is not rectified but appears at the output as an AC fluctuation. This means that the desired signal response, now a DC level, can be separated from the noise accompanying it in the output by means of a simple low-pass filter. Hence, in a lock-in amplifier, the final output is not affected by the presence of noise in the applied signal.

In order to function correctly, the detector must be 'programmed' to recognize the signal of interest. This is achieved by supplying it with a *reference* voltage of the same frequency (f_r) and with a fixed phase relationship to that of the signal. This is most commonly done by ensuring that they are derived from the same source. For example, in the case of an optical experiment using a rotating blade light chopper, the signal is generated from a detector focused on the light beam that has passed through the chopper blade, while the reference, at the same frequency, is taken from the chopper's reference output. The use of such a reference signal ensures that the instrument will 'track' any changes in the frequency of the signal of interest, since the reference circuit is 'locked' to it. It is from this characteristic that the lock-in amplifier derives its name.

This inherent tracking ability, allows extremely small bandwidths to be defined for the purpose of signal-to-noise ratio improvement since there is no frequency 'drift', as is the case with analog 'tuned filter/rectifier' systems. Because of the automatic tracking, lock-in amplifiers can give effective Q values (a measure of filter selectivity) in excess of 100 000.

The heart of the lock-in amplifier is the phase-sensitive detector, which is also known as a *demodulator* or *mixer*. The detector operates by multiplying two signals together, and the following analysis indicates how this gives the required outputs.

Figure 4 shows the situation where the lock-in amplifier is detecting a noise-free sinusoid, identified in the diagram as 'Signal in'. The instrument is also fed with a reference signal, from which it generates an internal sinusoidal reference, which is also shown in the diagram.

The demodulator operates by multiplying these two signals together to yield the signal identified in the diagram as 'Demodulator output'. Since there is no relative phase shift between the signal and reference phases, the demodulator output takes the form of a sinusoid at twice the reference frequency, but with a mean, or average, level that is positive.

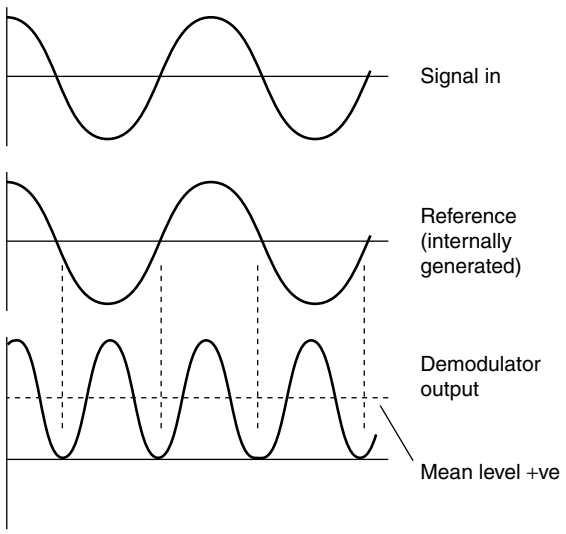


Figure 4. Operation of linear multiplier – signal and reference in-phase.

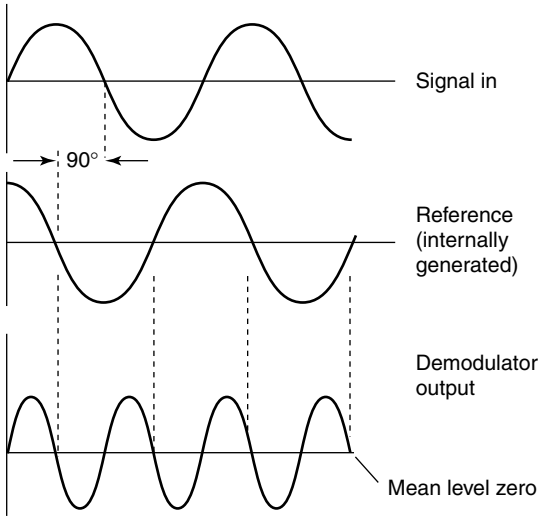


Figure 5. Operation of linear multiplier – signal and reference in quadrature.

Figure 5 shows the same situation, except that the signal phase is now delayed by 90° with respect to the reference. It can be seen that although the output still contains a signal at twice the reference frequency, the mean level is now zero.

From this, it can be seen that the mean level is

- proportional to the product of the signal and reference frequency amplitudes;
- related to the phase angle between the signal and reference.

It will be appreciated that if the reference signal amplitude is maintained at a fixed value, and the reference phase is adjusted to ensure a relative phase shift of zero degrees, then by measuring the mean level the input signal amplitude can be determined.

The mean level is, of course, the DC component of the demodulator output, so it is a relatively simple task to isolate it by using a low-pass filter. The filtered output is then measured using conventional DC voltmeter techniques.

Mathematically, we can derive the same conclusion as follows. Consider the case where a noise-free sinusoidal signal voltage V_{in} is being detected, where

$$V_{in} = A\cos(\omega t) \tag{7}$$

ω is the angular frequency of the signal, which is related to the frequency, f, in hertz by the equality

$$\omega = 2\,\pi f \tag{8}$$

The lock-in amplifier is supplied with a reference signal at frequency f derived from the same source as the signal and uses this to generate an internal reference signal of

$$V_{ref} = B\cos(\omega t + \theta) \tag{9}$$

where θ is a user-adjustable phase shift introduced within the lock-in amplifier.

The detection process consists of multiplying these two components together so that the PSD output voltage is given by

$$\begin{aligned} V_{psd} &= A\cos(\omega t) \times \mathrm{B}\cos(\omega t + \theta) \\ &= AB\cos\omega t(\cos\omega t\cos\theta - \sin\omega t\sin\theta) \\ &= AB(\cos 2\omega t\cos\theta - \cos\omega t\sin\omega t\sin\theta) \\ &= AB((\tfrac{1}{2} + \tfrac{1}{2}\cos 2\omega t)\cos\theta - \tfrac{1}{2}\sin 2\omega t\sin\theta) \\ &= \tfrac{1}{2}AB((1 + \cos 2\omega t)\cos\theta - \sin 2\omega t\sin\theta) \\ &= \tfrac{1}{2}AB(\cos\theta + \cos 2\omega t\cos\theta - \sin 2\omega t\sin\theta) \\ &= \tfrac{1}{2}AB\cos\theta + \tfrac{1}{2}AB(\cos 2\omega t\cos\theta - \sin 2\omega t\sin\theta) \\ &= \tfrac{1}{2}AB\cos\theta + \tfrac{1}{2}AB\cos(2\omega t + \theta) \end{aligned} \tag{10}$$

If the magnitude, B, of the reference frequency is kept constant, then the output from the phase-sensitive detector is a DC signal which is

- proportional to the magnitude of the input signal A;
- proportional to the cosine of the angle, θ, between it and the reference signal;
- modulated at $2\omega t$, that is, it contains components at twice the reference frequency.

The output from the PSD then passes to a low-pass filter, which removes the $2\omega t$ component, leaving the output of the lock-in amplifier as the required DC signal.

5 DESCRIPTION OF THE VARIOUS ELEMENTS OF A LOCK-IN AMPLIFIER

The block diagram of a typical lock-in amplifier is shown in Figure 6. Readers should be aware that the following discussion makes no assumptions as to the technology used to implement each of the circuit elements and that analog, mixed technology, and digital methods may be used.

5.1 Signal channel

In the signal channel, the input signal, including noise, is amplified by an adjustable-gain, AC-coupled amplifier, in order to match it more closely to the optimum input signal range of the PSD. Commercial instruments are usually fitted

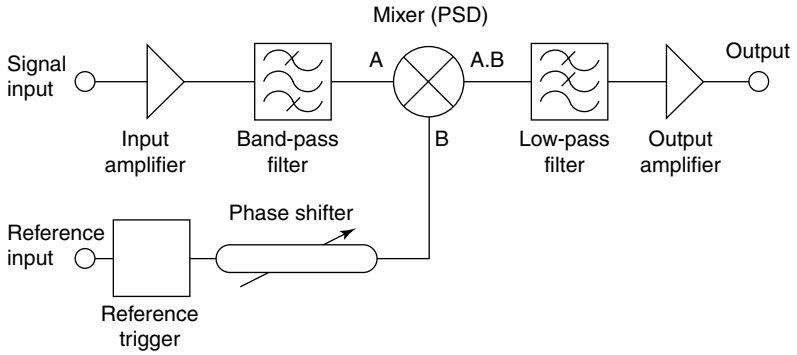


Figure 6. Basic lock-in amplifier (simplified). (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

with high impedance inputs for voltage measurements. Many also incorporate low impedance inputs for better noise matching to current sources, although in some cases the best results are obtained through the use of a separate external preamplifier.

The performance of the PSD is usually improved if the bandwidth of the noise voltages reaching it is reduced from that of the full frequency range of the instrument. To achieve this, the signal is passed through some form of filter, which may simply be a band rejection filter centered at the power-line frequency and/or its second harmonic to reject line frequency pick up, or alternatively a more sophisticated tracking band-pass filter centered at the reference frequency.

5.2 Reference channel

It has been shown that proper operation of the PSD requires the generation of a precision reference signal within the instrument. When a high-level, stable, and noise-free reference input is provided, this is a relatively simple task. However, there are many instances where the available reference is far from perfect or symmetrical, and, in these cases, a well-designed reference channel circuit is very important. Such circuits can be expensive and often account for a significant proportion of the total cost of the instrument.

The reference input waveform to the lock-in may be of almost any wave shape. Its zero crossings are used to define zero phase ($\phi_r = 0$) by the reference trigger circuit, which usually incorporates a phase-locked loop to lock onto the applied reference signal. The output of this phase-locked loop circuit is a precise waveform, locked in phase to the reference input, and at a frequency f_2. Normally, $f_2 = f_r$ (the reference frequency); most lock-ins also provide second-harmonic mode where $f_2 = 2f_r$. This mode is often used for derivative (signal rate of change) measurements. Modern instruments also often allow detection of even higher harmonics, allowing harmonic analysis of the input signal to be performed.

The internally generated reference is passed through a phase shifter, which is used to compensate for phase differences that may have been introduced between the signal and reference inputs by the experiment, before being applied to the PSD.

5.3 Phase-sensitive detector

As already mentioned, the purpose of the phase-sensitive detector is to multiply the signal by the reference waveform. At present, there are currently three common methods of implementing the PSD, these being the use of an analog multiplier, a digital switch, or a digital multiplier. These are discussed in greater depth in the following sections.

5.4 Low-pass filter and output amplifier

As discussed earlier, the purpose of the output filter is to remove the AC components from the desired DC output. Practical instruments employ a wide range of output filter types, implemented either as analog circuits or in digital signal processors. Most usually, however, these are equivalent to one or more stages of simple single-pole 'RC' type filters, which exhibit the classic 6 dB per octave roll-off with increasing frequency.

There is usually also some form of output amplifier, which may be either a DC-coupled analog circuit or a digital multiplier. The use of this amplifier, in conjunction with the input amplifier, allows the unit to handle a range of signal inputs. When there is little accompanying noise, the input amplifier can be operated at high gain without overloading the PSD, in which case little, if any, gain is needed at the output. In the case of signals buried in very large noise voltages, the reverse is the case.

The term dynamic reserve is used to specify noise overload performance and is defined as the ratio of the overload level (peak value of an asynchronous signal that will just cause significant nonlinearity), to the peak value of a full-scale synchronous signal.

5.5 Output

The output from a lock-in amplifier was traditionally a DC voltage that was traditionally displayed on an analog panel meter. Nowadays, especially when the instruments are used under computer control, the output is more commonly a digital number, although the analog DC voltage signal is usually provided as well. Instruments using an analog form of phase-sensitive detector use an analog-to-digital converter (ADC) to generate their digital output, whereas digital signal processor (DSP) lock-in amplifiers use a digital-to-analog converter (DAC) to generate the analog output.

5.6 Internal oscillator

All lock-in amplifiers use some form of oscillator within their reference circuits. Many units, however, also have a separate internal oscillator that can be used to generate an electrical stimulus for the experiment, usually with user-adjustable frequency and amplitude.

5.7 Computer control

All modern instruments include a microprocessor. This can simplify and automate manual measurements as well as support remote control of the instrument over common computer interfaces, such as the GPIB (IEEE-488) and RS232 links. The ability of the microprocessor to perform mathematical manipulations adds such useful functions as vector phase and noise measurements to the basic signal recovery capabilities of the lock-in amplifier.

6 PHASE-SENSITIVE DETECTOR DESIGNS – ANALOG AND DIGITAL TECHNIQUES

The phase-sensitive detector at the heart of the lock-in amplifier can be implemented in various ways, with the three most common methods being detailed in this section.

6.1 Analog multiplier

In an instrument with an analog multiplier, the PSD comprises an electronic circuit, which multiplies the applied signal with a sinewave at the same frequency as the applied reference signal. Although the technique is very simple in principle, in practice it is difficult to manufacture an analog multiplier that is capable of operating linearly in the presence of large noise, or other interfering, signals. Nonlinear operation results in poor noise rejection and thereby limits the signal recovery capability of the instrument. Consequently, instruments using this technique tend to provide only the most basic signal recovery performance.

6.2 Analog switching multiplier

The switching multiplier (Figure 7) uses a simple form of demodulator consisting of an analog polarity-reversing switch driven at the applied reference frequency. The great advantage of this approach is that it is very much easier to make such a demodulator operate linearly over a very wide range of input signals.

However, the switching multiplier not only detects signals at the applied reference frequency but also at its odd harmonics, where the response at each harmonic relative to the fundamental is defined by the Fourier analysis of a square wave. The frequency domain response of this type of multiplier is shown in Figure 8 and is characterized by a series of transmission 'windows' centered around the fundamental and odd harmonics of the reference frequency. Such harmonic responses may well be of use if the signal being detected is also a square wave but can give problems if, for example, the unit is being used at 1 kHz and there happens to be strong interfering signals at 7 kHz.

As discussed earlier, the use of a tuned low-pass or band-pass filter in the signal channel prior to the multiplier modifies the response of the unit so that it primarily detects signals at the reference frequency. However, in order to fully reject the $3f$ response, while still offering good performance at the reference frequency, very complex and

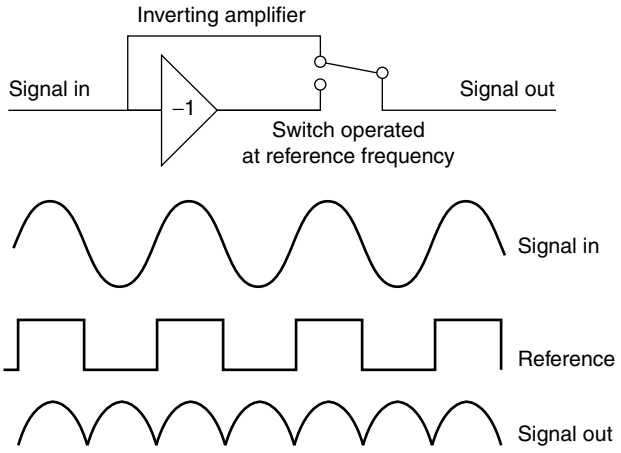


Figure 7. Analog switching multiplier.

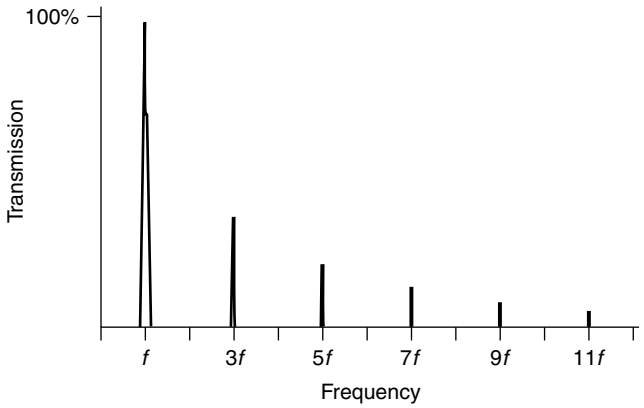


Figure 8. Analog switching multiplier – frequency domain response.

expensive filters would be required, since the filter needs to be tunable over the whole operating frequency range of the instrument. These are impractical for commercial instruments, so units fitted with filters tend to show some response to signals and noise at the third and fifth harmonics of the reference frequency and relatively poor amplitude and phase stability as a function of operating frequency.

Some analog lock-in amplifiers (e.g. the *SIGNAL RECOVERY* models 5209, 5210, 5110A, 5302) use an alternative and more sophisticated type of switching demodulator, which replaces the single analog switch with an assembly of several switches driven by a Walsh function. This may be thought of as a stepped approximation to a sinewave. Careful selection of components allows such a demodulator to offer all of the advantages of the switching demodulator with one additional benefit, which is the complete rejection of the responses at the third and fifth harmonics. The response to higher order odd harmonics is the same as for the normal switching modulator.

Such a demodulator, when used with a relatively slow roll-off, fourth-order, low-pass filter in the signal channel, produces an overall response very near to the ideal, which is an instrument that responds only to signals at the reference frequency. In this case, the demodulator rejects the third- and fifth-harmonic responses, while the signal channel filter removes the higher orders.

6.3 Digital multiplier using digital signal processing (DSP)

In an instrument employing this type of multiplier, the input signal is amplified and then immediately digitized. This digital representation is then multiplied by a digital representation of a sinewave at the reference frequency. A DSP is used for this task, and the output is therefore no longer an analog voltage but rather a series of digital values.

The technique offers the advantages of a perfect multiplication with no inherent errors and minimizes the DC-coupled electronics that are needed with other techniques, thereby reducing output drift. It has been used for a number of years in such applications as swept-frequency spectrum analyzers.

There are, however, a number of major problems with this technique when applied to recovering signals buried in noise. The most important of these is *dynamic range*. Consider the case of an input signal in the presence of 100 dB (100 000 times larger) of noise. If the signal is to be digitized to an accuracy of 'n' bits, then the input converter must handle a dynamic range of $2^n \times 100\,000$ to fully accommodate the signal and noise amplitudes. With a typical value for n of 15, this equates to a range of 3.2×10^9:1, corresponding to 32 bits. An ADC can be built with such accuracy, but would be extremely expensive and quite incapable of the sampling rates needed in a lock-in amplifier operating to 100 kHz.

Practical digital lock-in amplifiers use 14-, 16-, or 18-bit ADCs. Consequently, in the presence of strong interfering signals, the required signal may only be changing the least significant bit (LSB) of the converter, and indeed may actually be so small that there is no change at all in the ADC output. Hence, the measurement resolution of an individual output sample is very coarse. Resolution is however improved by averaging many such samples. For example, 256 samples of 1-bit resolution can average to 1 sample of 8-bit resolution, but this is at the expense of reduced response time. This averaging only operates predictably if the spectral power distribution of the interfering noise is known. If it is not, then noise has to be added by the instrument from its own internal noise source to ensure that it dominates, and to ensure that the signal does cause at least the LSB to change. The addition of this noise is only needed in demanding signal recovery situations, and so is not usually a problem.

7 SINGLE AND DUAL-PHASE LOCK-IN AMPLIFIERS

Descriptions of lock-in amplifiers usually start with the single-phase instrument, but, in recent years, the dual-phase (also known as the *vector lock-in amplifier*) type has become the most popular format. This is not, as initially may be thought, a dual-channel unit. Rather, it incorporates a second phase-sensitive detector, which is fed with the same signal input as the first but which is driven by a reference signal that is phase-shifted by 90°. This second detector is followed by a second output filter and amplifier and is usually referred to as the 'Y' output channel, the

original output being referred to as the 'X' channel. The X and Y outputs are also known as the in *phase* and *quadrature* outputs respectively.

The X and Y outputs can be processed mathematically to derive the vector magnitude R and signal phase θ where

$$R = \sqrt{(X^2 + Y^2)} = \sqrt{[(A\cos\theta)^2 + (A\sin\theta)^2]} \qquad (11)$$

and

$$\theta = \tan^{-1}\left(\frac{Y}{X}\right) = \tan^{-1}\left(\frac{A\sin(\theta_s + \theta_r)}{A\cos(\theta_s + \theta_r)}\right) = \theta_s + \theta_r \qquad (12)$$

where θ_s is the signal phase shift relative to the reference signal phase, and θ_r is the phase offset set by the phase-shift controls. Dual-phase lock-in amplifiers can, therefore, display their output signal in rectangular or polar form, with the phase controls (θ_r) allowing continuous vector rotation.

The principal advantage of the dual-phase unit is that if the signal channel phase (but not its amplitude) changes then the magnitude display will remain constant, even though the X and Y outputs change. Hence, if the lock-in amplifier is set to display vector magnitude R, changes in the signal phase will not affect the reading, and the instrument does not require the adjustment of the reference phase-shifter circuit. This capability is the main reason for the dual-phase instrument's widespread use.

Note that asynchronous signals ($f_s \neq f_r$) with beat frequencies within the response of the output low-pass filter will provide DC outputs and the instrument, therefore, acts as a *wave analyzer.* Commercial wave analyzers are essentially dual-phase lock-in amplifiers that are optimized for convenience in measuring frequency components of a signal rather than recovering a signal from noise.

8 LIMITS OF DETECTION FOR LOCK-IN AMPLIFIERS

8.1 Selection of output filter

The lock-in amplifier's output filter serves two main purposes. First, the action of the phase-sensitive detector generates a signal at twice the reference frequency, which must be rejected. Second, fluctuations caused by noise and interfering signals that are close to the reference frequency are translated down by the PSD to appear as frequencies close to DC. These signals, which do not have any fixed frequency or phase relationship to the reference, are also rejected by the filter and do not result in any change to the mean output DC level.

Consider the case shown in Figure 9, where the asynchronous (noise) of frequency $f_1 = f_2 + \Delta f$ is applied

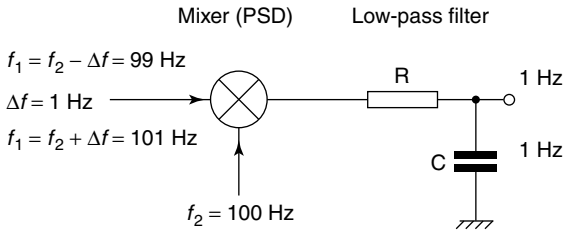


Figure 9. PSD operation with asynchronous (noisy) signal.

to the PSD input. The resulting mixer sum and difference signals will, therefore, be $2f_2 + \Delta f$ and Δf respectively. The output low-pass filter is set to reject all but the Δf component, which therefore appears at the output as noise. Suppose we change the frequency of this input noise to $f_1 = f_2 - \Delta f$. The resulting sum and difference frequencies will respectively be $2f_2 - \Delta f$ and $-\Delta f (= \Delta f)$. Again, only the Δf component can appear as output noise and the low-pass filter 'cannot tell' whether its Δf input resulted from an $f_2 + \Delta f$ input to the mixer or an $f_2 - \Delta f$ input.

In addition to its rectifying and phase-sensitive properties, therefore, the PSD filters noise as though it consisted of a band-pass response centered on f_2, where the effective *band-pass* response consists of the output low-pass filter response and its mirror image. Given that the PSD drive frequency is the reference frequency, the center frequency therefore automatically tracks the reference frequency, f_2.

8.1.1 Signal-to-noise improvement ratio (SNIR)

As with DC measurements using preamplifiers or electrometers, the signal-to-noise improvement ratio (SNIR) to be expected from a lock-in depends upon the input noise bandwidth (B_{ni}) to the lock-in, the noise bandwidth (B_{no}) of the lock-in, and the noise spectral characteristics. For random white noise and unity gain,

$$\text{SNIR} = \sqrt{\left(\frac{B_{ni}}{B_{no}}\right)} \qquad (13)$$

What is meant by the term *bandwidth*? In the simple low-pass filter circuit shown in Figure 10, for example, the transfer function is $G(f) = v_o/v_i$.

The *signal bandwidth* for this circuit (Figure 11) is usually, and somewhat arbitrarily, defined to be the *cut-off frequency,* f_c, where $v_o/v_i = 70.7\%$ (-3 dB) or $v_o^2/v_i^2 = 50\%$ (the half-power point).

In the case of the circuit above, $f_c = 1/2\pi RC$

Notice that frequencies above f_c will obviously pass (though attenuated) through the filter and, therefore, are not really cut off. For noise, it is convenient to think in terms of an equivalent noise bandwidth, B_n, defined by the

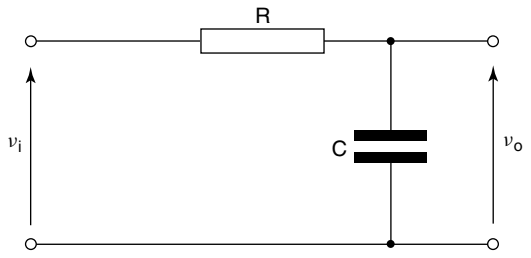


Figure 10. Low-pass filter circuit.

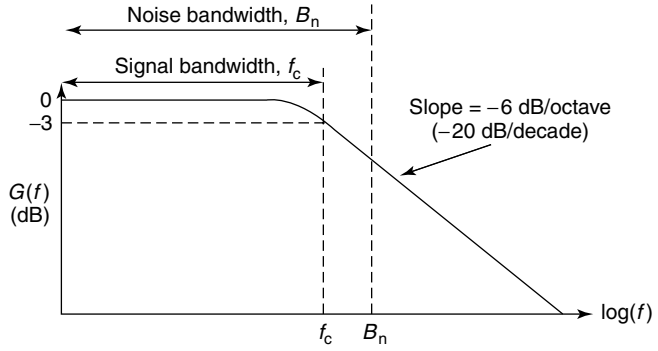


Figure 11. Low-pass filter transfer characteristic.

relationship

$$B_n = \frac{1}{G^2}\int_0^\infty |H(j\omega)|^2 \, df \tag{14}$$

where $H(j\omega)$ is the frequency response function of the system and G is a gain parameter suitably chosen to be a measure of the response of the system to some parameter of the signal. In the case of a low-pass filter such as that shown in Figure 10, G is usually taken to be the zero-frequency (DC) gain, whereas for band-pass responses, G is usually made equal to the maximum gain.

Using the above definition and taking G to be the zero-frequency gain (i.e. unity), it can readily be shown that for the simple RC filter shown in Figure 10 which has a −6 dB/octave roll-off rate

$$B_{no} = \frac{1}{4RC} \text{ Hz} \tag{15}$$

while for a two-stage buffered filter, which has a −12 dB/octave roll-off rate:

$$B_{no} = \frac{1}{8RC} \text{ Hz} \tag{16}$$

The lock-in amplifier's output filter is adjusted using the time constant control, where $TC = RC$. Hence, for example, if the time constant is set to 100 ms and the −12 dB/octave slope is selected, then $B_{no} = 0.1/8$ Hz, or 12.5 mHz. With an input noise bandwidth of say 100 kHz, the resulting SNIR is nearly 70 dB, demonstrating the power of the lock-in technique.

9 EXAMPLES OF LOCK-IN AMPLIFIER APPLICATIONS

9.1 Signal recovery

Figure 12 illustrates a basic optical detection setup using a mechanical light chopper and a lock-in amplifier. The light chopper consists of a motor, a speed control mechanism, and a rotating blade or chopper wheel. In some cases, all three of these components are in one assembly. In other choppers, the control unit may be in a separate housing.

The chopper wheel is a rotating metal disk, which contains one or more sets of equally spaced apertures, which

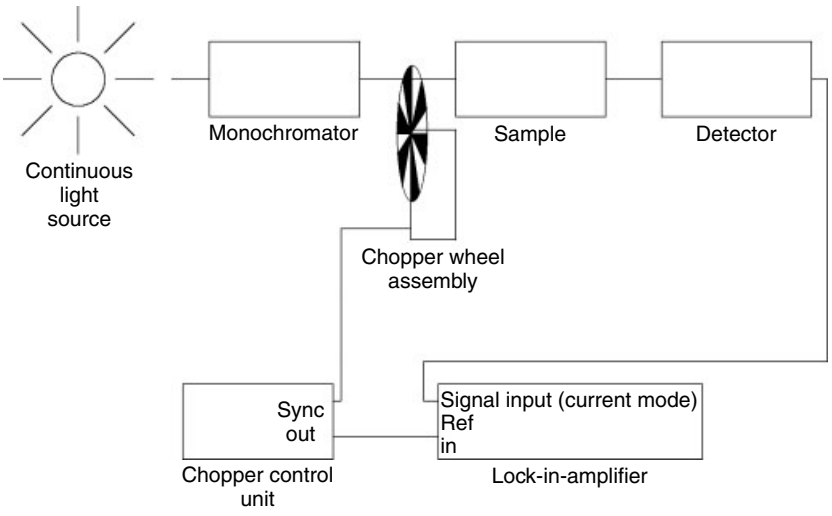


Figure 12. AC measurement system using a lock-in amplifier and mechanical chopper.

allow the light source to pass through or be blocked altogether. The number of apertures and the wheel rotation speed determine the chopping frequency. Since the rotation of the blade causes the optical signal path to be interrupted, the light source that stimulates the experiment is in the form of an AC excitation. One could visualize this excitation as an optical equivalent of a square wave, although this is only true if the aperture size is large compared to the beam diameter. The signal appearing at the detector output may or may not be a good representation of the optical stimulation since factors such as detector response time and cable capacitance must be considered.

In addition to modulating the light source, the chopper also provides a synchronous reference signal capable of driving the reference channel of a lock-in amplifier. This reference output voltage is a square wave, usually in the order of a few volts peak to peak.

The optical signal stimulating the experiment and thus falling on the optical detector generates an electrical current that can be measured by the lock-in. Any discrete frequencies or noise voltages not equal to the reference frequency will be rejected by the lock-in amplifier. The end result is a much lower limit on signals that can be measured. In fact, it is possible that the signal of interest may be completely obscured by noise if one were to view the detector output with an oscilloscope. Again, stray light falling on the detector is usually not a problem as long as the magnitude is insufficient to saturate the detector. However, the user still needs to ensure that stray light does not enter into the experiment via the chopped light path.

9.2 Vector voltmeter measurements

Figure 13 shows a typical application for a dual-phase lock-in where the signals being measured are relatively clean, but advantage is taken of the dual-phase capability to directly determine the wanted values.

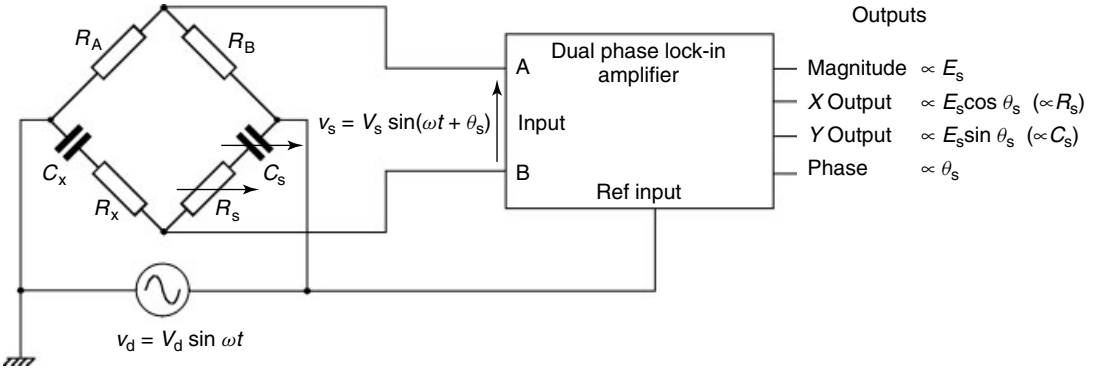


Figure 13. Bridge balancing using a dual-phase lock-in amplifier.

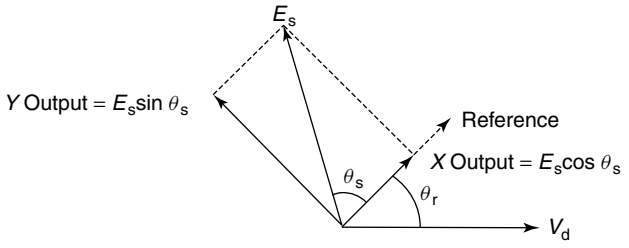


Figure 14. Dual-phase lock-in amplifier – vector relationships.

In this AC bridge application, the phase shift (θ_r) can be set to zero, so that the in-phase (X) signal responds only to the bridge resistance and the quadrature output (Y) to the bridge capacitance. The bridge can then be balanced very simply by separately nulling R_s and C_s. The phase relationships of the various waveforms are shown in Figure 14.

10 ADVANCED TECHNIQUES

The advent of DSP based lock-ins has allowed instruments (such as the *SIGNAL RECOVERY* models 7265 and 7280) to be designed that can simultaneously measure signals at two different frequencies. These operating modes are known as *dual reference* and *dual harmonic* modes.

One application for the dual reference mode is to build a very cost-effective dual-beam measurement system. This technique can eliminate variations in source intensity over several orders of magnitude, which is especially useful in two common situations:

- If the source output is unstable over time, such as with some discharge lamps;
- If the 'source' is the output of a spectrometer with a tungsten–halogen or other lamp as its input, and the spectrometer center wavelength is being scanned as part of the experiment.

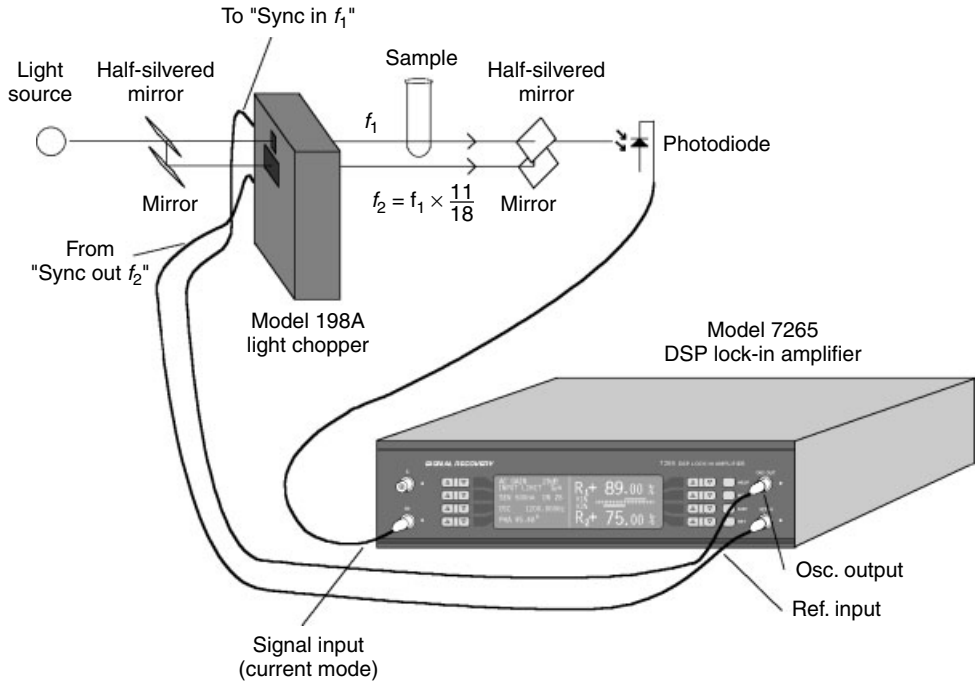


Figure 15. Dual-beam ratiometric system using model 198A light chopper and model 7265 DSP lock-in amplifier. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

A typical experimental arrangement is shown in Figure 15 above. A focused light source is split into two beams using either a half-silvered mirror or beam splitter, with the two beams being aligned so that they pass through the inner and outer apertures of the model 198A. One beam passes through the sample, while the other passes via a reference path that may include ND (neutral density) filters (not shown). Both beams are recombined back onto a single photodiode using a second mirror/beam-splitter combination. Hence, there will be two signals at the detector, one at frequency f_1, which is due to the light that passes through the sample and the second at f_2, due to light traveling via the 'reference' path.

A DSP lock-in amplifier with dual reference mode is set to this mode and with the internal oscillator set to 360 Hz. The oscillator amplitude is set to 2 V rms and the Osc. out connector coupled to the Sync in f_1 input on the 198A. The chopper locks to this input causing the f_1 frequency also to be 360 Hz, and thereby satisfying an instrument restriction in the dual reference mode that one of the signals being detected by the lock-in must be at the internal reference frequency.

The chopper also generates a reference signal at the f_2 frequency, which is coupled back to the lock-in's Ref. in input. Hence, its external reference channel operates at this frequency, which is in this case $360 \times 11/18 = 220$ Hz.

Once the lock-in amplifier's sensitivity and phase controls have been properly adjusted then it indicates in the R2 channel (i.e. the internal reference frequency) the magnitude of the signal due to light through the sample, while the R1 channel (the external frequency) show the magnitude of the signal due to light via the reference path.

The ratio of these two signals is independent over several orders of magnitude to changes in the source intensity, so if the ratio is calculated and displayed using the instrument's user equations menu such variations can be eliminated.

10.1 Advantages

The system described is considerably more cost effective than the traditional approach, which required two lock-in amplifiers. Since the same detector and analog signal channel is used for both signals, differential drift between channels is also eliminated. Note, however, that since the signals at f_1 and f_2 share a common path as far as the main ADC, best results are obtained when the experiment is arranged so that their absolute levels differ by no more than a factor of 100.

REFERENCE

Burdett, R.M. (2002a) *What is a Lock-in Amplifier? (Technical Note TN1000)*, available as PDF document from **SIGNAL RECOVERY** website at www.signalrecovery.com

FURTHER READING

Blair, D.P. and Sydenham, P.H. (1975) Phase sensitive detection as a means to recover signals buried in noise. *Journal of Physics E: Scientific Instruments*, **8**, 621–627.

Burdett, R.M. (2002b) *Specifying Lock-in Amplifiers (Technical Note TN1001)*, available as PDF document from **SIGNAL RECOVERY** website at www.signalrecovery.com

Meade, M.L. (1983) *Lock-in amplifiers : principles and applications*, *P. Peregrinus (London) on behalf of the Institution of Electrical Engineers*.

This article is based in part on a chapter by D.M. Munro originally published in Handbook of Measurement Science Vol 1, P.H. Sydenham, Wiley, 1982.

182: Boxcar and Signal Averagers

Richard Burdett
SIGNAL RECOVERY, Wokingham, UK

1 Boxcar Averager 1209
2 General Principles of Operation of the 'Boxcar' Integrator 1209
3 Static Gate Boxcar Averaging 1210
4 Waveform Recovery Boxcar Averager 1211
5 Signal Averagers – Overview 1211
6 Frequency Response of Signal Averagers 1213
7 Block Diagram of Signal Averager 1215
Related Articles 1216
Further Reading 1216

1 BOXCAR AVERAGER

The *boxcar averager* (***SIGNAL RECOVERY***, 1), also known as a *boxcar integrator* or *detector,* is a sampling instrument that integrates the applied input signal during a predefined *gate or aperture width* starting at a predefined *trigger, gate,* or *aperture delay* after an applied trigger. Each of these integrated samples of input signal can then be averaged, using either an analog averager or by digitizing each sample and then averaging the resulting digital values.

The boxcar therefore performs signal recovery by three methods. First, the input signal only affects the output during the period in which it is being sampled; at all other times its level is unimportant, other than the need to avoid causing an input overload for which the recovery time might affect a subsequent sample. This is temporal separation of the signal from the noise.

Second, the signal is integrated during the gate width, unlike common sample and hold circuits that simply take a 'snapshot' measurement of the signal level at one point in time. Hence, if there is noise or other interference present at the input at frequencies much higher than the reciprocal of the gate width, then these will be rejected.

Finally, the measured samples are themselves averaged, ensuring that low frequency fluctuation or noise that would cause sample-to-sample variation can also be removed.

In commercial instruments, the output was traditionally recorded using an analog chart recorder, with the Y input being connected to the output from the boxcar's output averager, and the X input being driven by a voltage generated by the boxcar and proportional to the trigger delay. Nowadays, the boxcar's output is usually digitized for recording by a computer and the trigger delay is set using a separate digital delay generator. This makes it possible to use a computer to run the whole experiment and record the resulting data.

Boxcar averagers can recover very fast waveforms and resolve features down to sub nanosecond level. However, they are time inefficient if used for waveform recovery and in such applications have been superseded by signal averagers and storage oscilloscopes. They do, however, remain cost-effective for static gate work.

2 GENERAL PRINCIPLES OF OPERATION OF THE 'BOXCAR' INTEGRATOR

The heart of any boxcar is the *gated integrator* circuit, shown in simplified form in Figure 1. This circuit is simply an *RC* low-pass filter gated by switch S_1, (the sampling gate). As shown, the gated integrator has unity DC signal gain.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

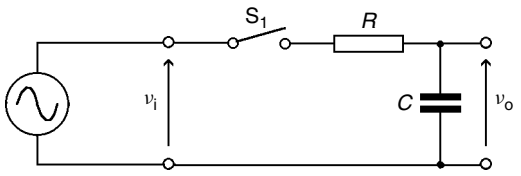


Figure 1. The gated integrator (simplified).

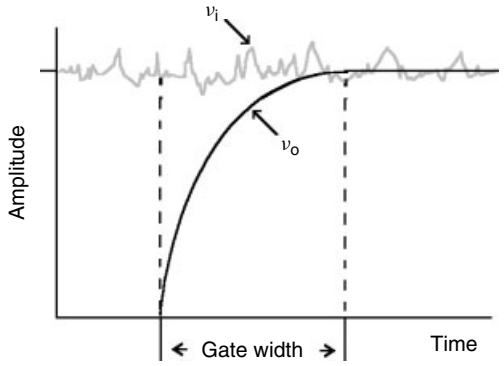


Figure 2. Gated integrator – gate time constant.

When the gate opens (switch S_1 closes) the output voltage v_o starts to rise exponentially toward v_i, as shown in Figure 2 where the gate width is shown as being small in relation to long-term changes in v_i. The gate time constant (set by the product RC) is adjusted so that the v_o is typically within a few percent of the input v_i at the end of the selected gate width. Consequently, high frequency components of the input signal are removed. The equivalent noise bandwidth is simply $B_{no} = 1/4RC$.

The integral of the gate sample is the voltage v_o at the end of the gate width. This can either be stored as an analog voltage or be digitized in order to be processed by the output averager, after which the sampling circuit is reset ready for the next trigger.

The rest of the boxcar consists of the output averager, discussed in another section, and trigger processing and delay circuits. Commercial instruments offer trigger inputs that are similar in design to those used on oscilloscopes, allowing triggering on a variety of waveforms. Following the trigger, a delay circuit allows the position in time start of the gate to be adjusted while a variable pulse generator creates the gate pulse.

The overall maximum trigger rate of the instrument is set by the gate delay, gate width, averager processing and reset times, and is typically up to 80 kHz.

3 STATIC GATE BOXCAR AVERAGING

In the *static gate* boxcar averager, the length of the trigger delay is fixed and the intention is usually to determine the amplitude of some 'spike' or feature of a waveform that is typically much narrower than the repetition period set by the overall trigger rate. Consider the situation shown in Figure 3.

The input signal consists of a repetitive waveform triggered at a 10 kHz rate so that each cycle lasts 100 µs. The trigger delay is set to open the gate just before the feature of interest and the gate width is set to 'bracket' this

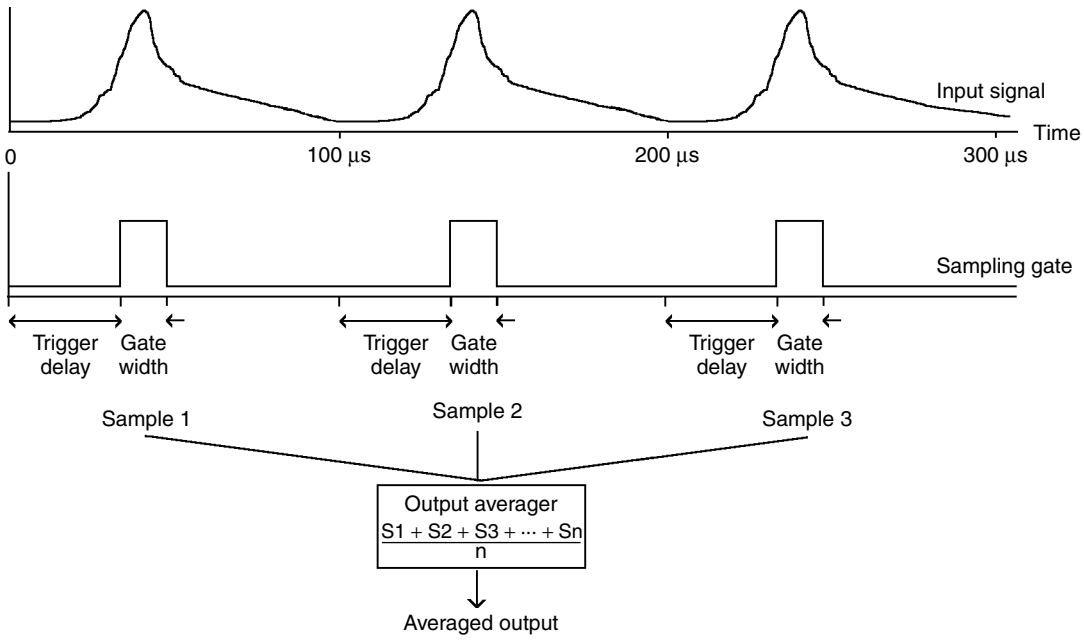


Figure 3. Boxcar averager with linear output averaging.

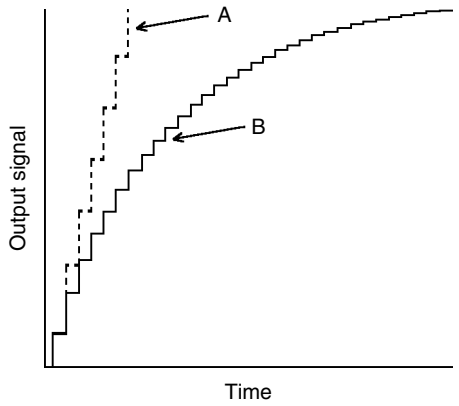


Figure 4. Boxcar output averager modes of operation: curve A – linear averaging, curve B: exponential averaging.

feature. Each sample results in an integral representing the area under the signal curve for the duration of the gate, and these samples are then averaged.

When using a linear output averager, all samples have equal weight and so the output will rise in a linear staircase fashion as shown in Figure 4, curve A. Each step in this curve represents a new sample becoming available, which in turn corresponds to a trigger cycle. In this *linear summation* mode, the desired number of signal samples (n) is selected; after n triggers have occurred, a switch or other method is used to reset the output averager. Since the signal component of the samples will add linearly, but random noise samples will add vectorially, after n samples of a constant-amplitude signal (S) plus white noise (N), and after maximizing the gate width to suit the signal wave shape, the output SNR (signal-to-noise ratio) is given by

$$\begin{aligned}\mathrm{SNR}_{\mathrm{out}} &= \frac{S_1 + S_2 + S_3 + \cdots + S_n}{\sqrt{({N_1}^2 + {N_2}^2 + {N_3}^2 + \cdots + {N_n}^2)}} \\ &= \frac{nS}{\sqrt{(nN^2)}} \frac{S}{N}\sqrt{n} = \mathrm{SNR}_{\mathrm{in}}\sqrt{n} \end{aligned} \quad (1)$$

so that

$$\mathrm{SNIR} = \frac{\mathrm{SNR}_{\mathrm{out}}}{\mathrm{SNR}_{\mathrm{in}}} \left(= \frac{\mathrm{SNR}\ (n\ \mathrm{samples})}{\mathrm{SNR}(1\ \mathrm{sample})} \right) = \sqrt{n} \quad (2)$$

Note that in this operating mode, it is easiest to think in terms of time averaging since the equivalent noise bandwidth of the gated integrator circuit is not constant but will decrease with increasing n.

The linear averager suffers from the disadvantage of needing to be reset after each set of triggers. Although this is not difficult if the output averager is implemented digitally, it is not as easy when using analog techniques. Historically, these were more common and so in such cases the output used *exponential averaging*, which essentially consists of nothing more than a further gated integrator stage, but this time with a time constant much longer than the gate width. The gate at this stage is operated for a short fixed period (typically a few microseconds) to apply the sample voltages out of the input averager to the output averager, with one such gate per trigger cycle.

If the samples from the input averager are similar in amplitude, then the output from the output averager will rise exponentially as shown in Figure 4, curve B. Again, each step corresponds to a new sample becoming available, and hence a trigger cycle.

4 WAVEFORM RECOVERY BOXCAR AVERAGER

Boxcar averagers can also be used for *waveform recovery,* where the intention is to record the waveform of the input signal. In this mode of operation, the trigger delay is not fixed but rather is incremented by a fixed amount on successive groups of n triggers so that it sweeps between *initial delay* and *final delay* values. In this mode, the boxcar output is a replica of the signal waveform and the boxcar can be regarded as a time-translation device that can slowdown and recover fast waveforms.

Figure 5 gives a simplified view of this mode. The gate width is now set to be much shorter than the signal period. The first set of n samples is taken after a trigger delay set to the initial delay setting. This group of samples is then averaged by the output averager and appears as point A on the output plot. The boxcar counts the applied triggers and after the first set of n has been detected, the trigger delay is incremented and the cycle repeats, resulting in point B. The process continues (generating point C and further points) until the trigger delay is equal to the preset final delay value.

The number of points m on the output waveform record is

$$m = \frac{\text{Final delay} - \text{initial delay}}{\text{Delay increment}} \quad (3)$$

It will therefore be seen that, as the delay increment is reduced, the time resolution of the recorded waveform will improve since there are more points. But of course the time to record the waveform will also increase since each point requires n triggers.

5 SIGNAL AVERAGERS – OVERVIEW

The boxcar is a *single-point* averager: it samples each signal occurrence (sweep) only once. A *signal averager* acts

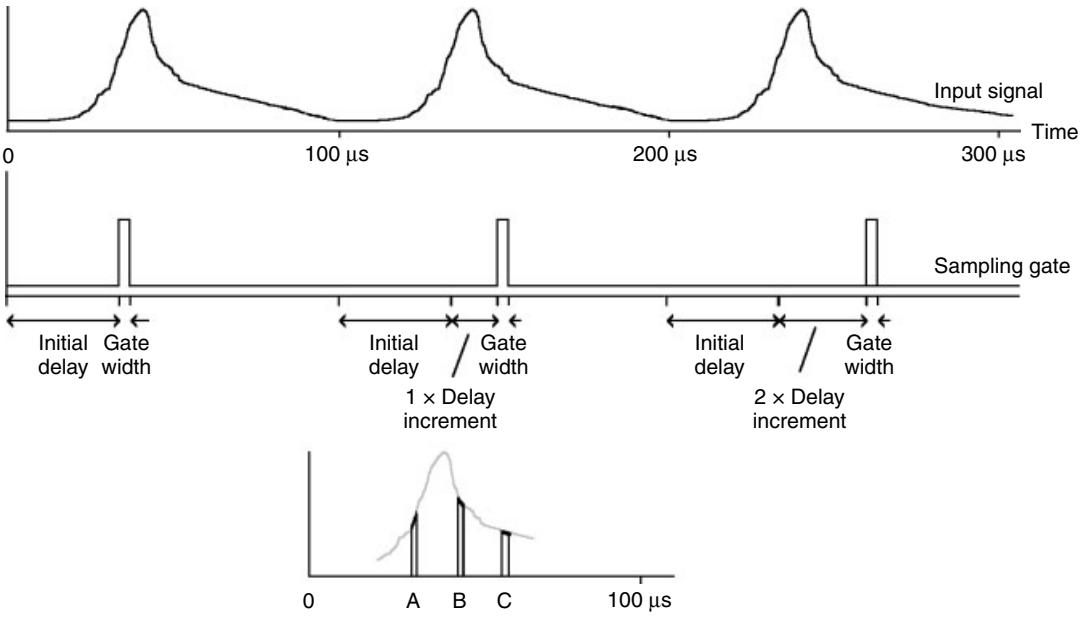


Figure 5. Boxcar averager – waveform recovery.

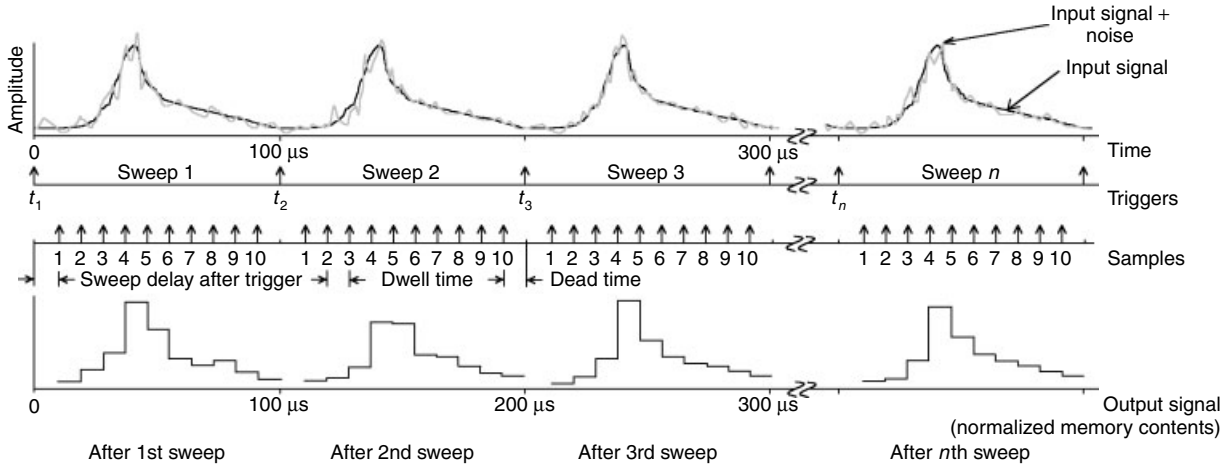


Figure 6. Multipoint signal averager.

much like a large number of boxcars connected in parallel, since it samples many points during each signal sweep. In such signal instruments, the analog storage capacitor of the boxcar is replaced by digital memory; each sample is digitized and the new data is added to the data from previous sweeps already in the memory location corresponding to that sampling point.

Figure 6 illustrates some typical waveforms and timing details for a multipoint signal averager; for simplicity $I = 10$ (i.e. only ten samples/sweep are shown). The total signal duration (T) is given by the product of the number of samples/sweep (I) and the dwell-time (gate width or sampling duration, t_g) of each sample. Note that T is less than the total sweep duration (τ) by the dead time (t_d), and that there is usually a fixed delay time (the *intrinsic delay*) between receipt of a trigger pulse and the beginning of the first sample. Although in most applications a signal averager is triggered at a constant rate ($f = 1/\tau$), it is not necessary that the trigger be periodic. Assume that the averager is set to continue averaging until n input sweeps have been sampled, at which point it will automatically stop.

Suppose the waveform of a noisy signal, $f(t)$, is to be recovered where

$$f(t) = s(t) + n(t) \tag{4}$$

For the ith sample of the kth sweep,

$$f(t) = f(t_k + it_g) = s(t_k + it_g) + n(t_k + it_g) \quad (5)$$

For any particular sample point (i), the input signal can be assumed to remain unchanged with each new value of k (i.e. with each new sweep) and the averaged signal will therefore be simply

$$S(i)_{out} = \sum_{k=1}^{n} s(t_k + it_g) = ns(it_g) \quad (6)$$

For random noise, samples (X_i) will add vectorially, so that the rms value (σ) of the averaged noise will be given by

$$\left(\sum (X_i)^2\right)^{1/2} = \sigma\sqrt{n} \quad (7)$$

The averager output can be described by

$$g(t_k + it_g) = ns(it)_g + \sigma\sqrt{n} \quad (8)$$

so that the output SNR is

$$SNR_{out} = \frac{s_{out}}{N_{out}} = \frac{ns}{\sigma\sqrt{n}} it_g = \frac{s}{\sigma} it_g \sqrt{n} \quad (9)$$

The input SNR is simply

$$SNR_{in} = \frac{s(it_g)}{\sigma} \quad (10)$$

so that

$$SNIR = \frac{SNR_{out}}{SNR_{in}} = \sqrt{n} \quad (11)$$

6 FREQUENCY RESPONSE OF SIGNAL AVERAGERS

In order to consider a signal averager from a frequency domain or filtering point of view, it is necessary to know its transfer function, $H(j\omega)$. This can be determined if the impulse response $h(t)$ is known, since $H(j\omega)$ and $h(t)$ are a Fourier-transform pair.

$h(t)$ can be determined heuristically by the following approach. In a multipoint averager, trigger pulses are used to synchronize the signal sweeps and allow the signal samples to be coherently added (CO-ADDed). Mathematically, this action can be thought of as convolving the input signal, $f(t)$, with a train of n unit impulses (triggers) spaced τ seconds apart. The averager's effective impulse response is, therefore, given by

$$h(t) = \sum_{k=1}^{n} \delta(t - k\tau) \quad (12)$$

By Fourier transforming this expression for $h(t)$, we find that the averager's transfer function is

$$|H(j\omega)| = \left|\frac{\sin(n\omega\tau/2)}{\sin(\omega\tau/2)}\right| \quad (13)$$

Notice (from L'Hopital's rule) that $H(j\omega) = n$ whenever ωt is an integral multiple of 2π. Figures (7) to (10) show the *comb filter* response of (13) for several values of n. Each band-pass response is centered at a harmonic (n/τ) of the sweep/trigger rate. (If the trigger rate is aperiodic, then this comb filter concept becomes meaningless.) Since the peak transmission of each band-pass response is n, the -3 dB points must occur at $n/\sqrt{2}$, so that

$$|H(j\omega)| = \left|\frac{\sin(n\omega\tau/2)}{\sin(\omega\tau/2)}\right| = \frac{n}{\sqrt{2}} \quad (14)$$

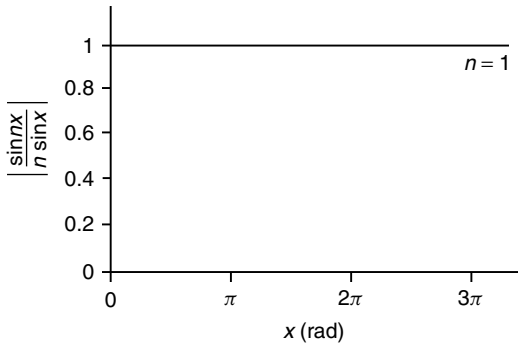


Figure 7. Comb filter response of multipoint signal averager, $n = 1$.

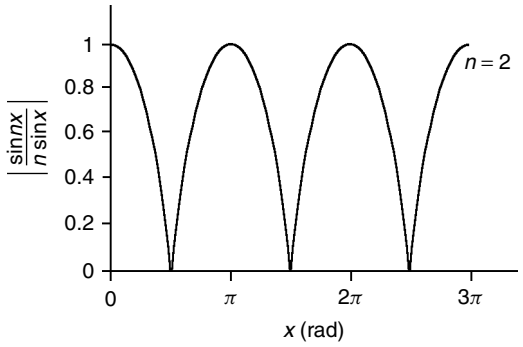


Figure 8. Comb filter response of multipoint signal averager, $n = 2$.

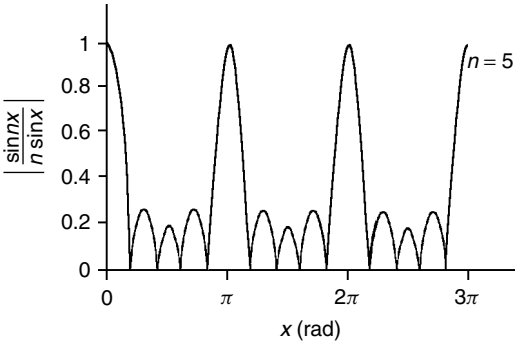


Figure 9. Comb filter response of multipoint signal averager, $n = 5$.

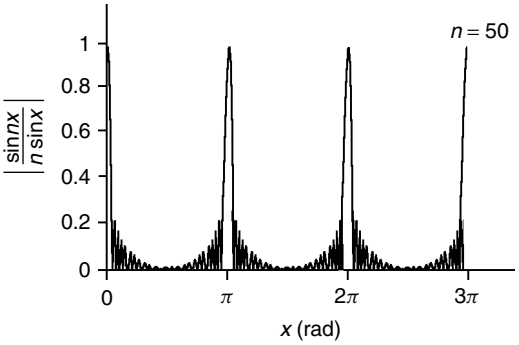


Figure 10. Comb filter response of multipoint signal averager, $n = 50$.

from which the −3-dB bandwidth B for large values of n, is found to be

$$B = \frac{0.886}{n\tau} \tag{15}$$

Large values of n are practicable, particularly at high sweep rates. With a trigger rate $1/\tau$ of 2 kHz and $n = 10^6$, for example, the total measurement time will be $n\tau = 10^6 \times 1/(2 \times 10^3) = 500\,\text{s} \simeq 8.3$ min, and $B = 1.772 \times 10^{-3}$ Hz.

Thus far in this discussion of signal averagers, a *linear summation* mode of averaging has been assumed that is, for the ith memory location, the average after n sweeps is given by

$$A_m = \sum_{k=1}^{n} f(t_k + it_g) = \sum_{k=1}^{n} I_k \tag{16}$$

where $I_k = f(t_k + it_g)$ is the value of the ith sample in the kth sweep.

This algorithm has the advantage of being simple to implement digitally. The output averaged signal, however, continually increases with each new sweep, so that rescaling is required to keep the displayed output at a useful size. A seemingly more convenient algorithm would be to normalize the data in memory after each sweep, that is, implement

$$A_k = \frac{1}{k}\sum_{k=1}^{n} I_k = A_{k-1} + \frac{I_k - A_{k-1}}{k} \tag{17}$$

During each sweep, the data (A_{k-1}) in each memory location are compared with the new sample value I_k and the computed value of $(I_k - A_{k-1})/k$ is added to memory to form the new average value A_k. Because of practical difficulties in implementing a division by k during or after each sweep, the algorithm shown in (17) is often approximated by

$$A_k = A_{k-1} + \frac{I_k - A_{k-1}}{2^J} \tag{18}$$

where J is a positive integer selected automatically such that 2^J is the closest approximation to k. Notice for $k = 6$ for example, that the closest 2^J values are $2^2 = 4$ or $2^3 = 8$. Though this *normalized averaging* mode is very slightly slower than the summation mode in enhancing the signal, we can assume that SNIR $= \sqrt{m}$ for all practical purposes. Note that the discrepancy between k and 2^J increases as larger values of J are used to deal with very noisy signals. In compensation, this averaging mode provides a stable, constant-amplitude display from which the noise appears to shrink with time.

If slowly varying noisy signals are to be recovered and monitored, the algorithm of (18) can also be used for *exponential* averaging if J is made a manually selectable constant. When $J = 0$, then $2^J = 1$ and $A_k = I_k$; with this setting, the input signal may be monitored in real time, since it is digitized and stored without averaging. In general, selecting a value of J will establish an effective time constant, τ_J, where

$$\tau_J = \frac{t_g}{-1n(1 - 2^{-J})} \tag{19}$$

or

$$2^{-J} = 1 - \exp\left(\frac{-t_g}{\tau_J}\right) \tag{20}$$

The larger the value of J selected, the greater the signal enhancement, and the more slowly the averager responds to changes in the input signal. For a large number of sweeps, the SNIR (signal-to-noise improvement ratio) is given by

$$\text{SNIR} \simeq \sqrt{2^{J+1}} \tag{21}$$

7 BLOCK DIAGRAM OF SIGNAL AVERAGER

Figure 11 shows the simplified block diagram of a commercial signal averager. The signal passes to an 8-bit flash ADC (analog digital converter) via a summing amplifier, the purpose of which is described later. The digitized samples are then summed into the averager memory, for which the sampling clock and control block provides suitable timing signals.

At the end of the preset number n of sweeps to averager, the sum data in the averager memory is transferred across to the output memory, allowing the averager memory to be cleared and the cycle to repeat. This allows the data to be transferred to the PC while the next set of sweeps is being acquired, thereby reducing the deadtime after each set of sweeps.

The maximum number of sweeps ($n_{\max}$) that can be digitized by an averager without data overflow, if the input signal is full scale and noise free, is given by 2^{h-c} where h is the memory size (bits) and c (bits) is the resolution of the A/D converter (ADC). For an 8-bit ADC and a memory of N words, each of 24 bits, then

$$n_{\max} = 2^{h-c} = \frac{2^{24}}{2^8} = 2^{16} = 32\,768 \tag{22}$$

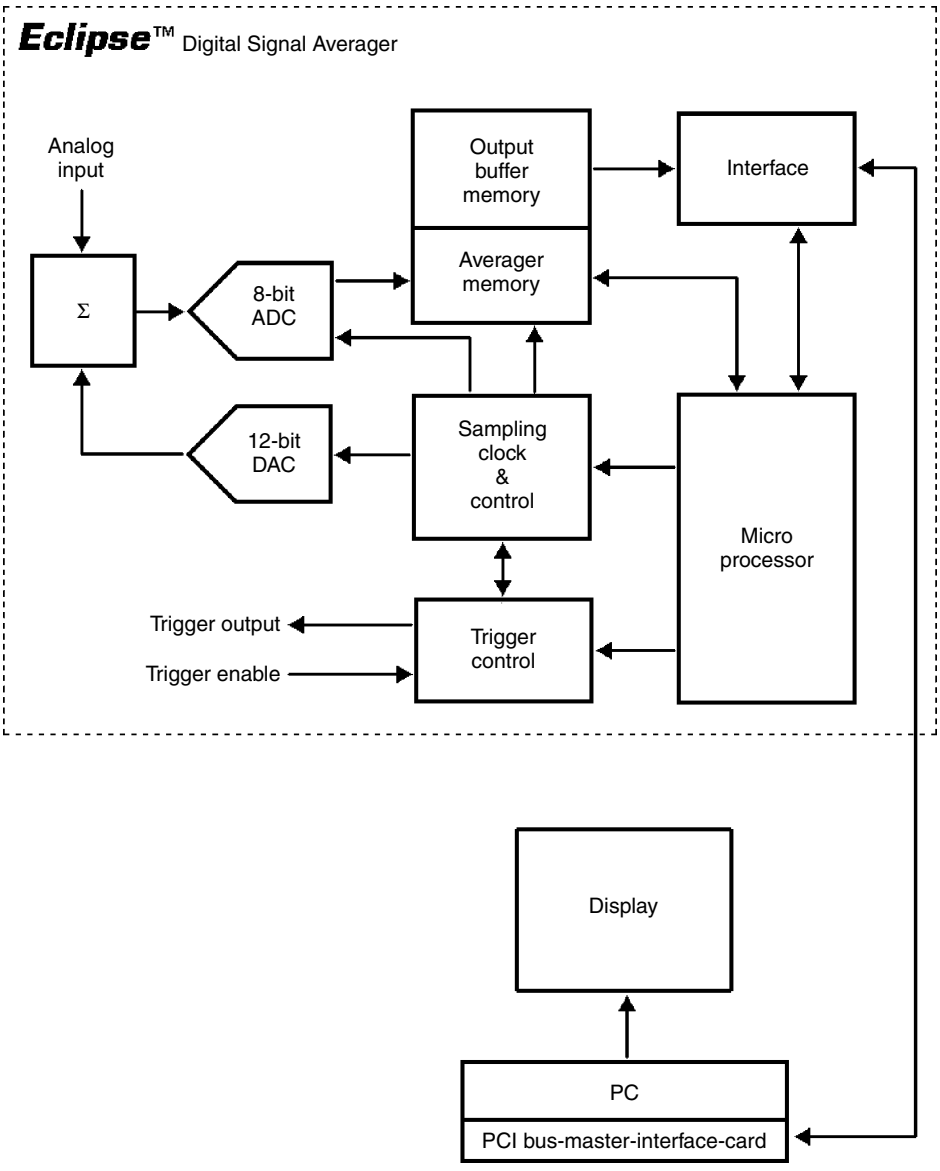


Figure 11. Typical multipoint signal averager. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

Suppose that the input SNR to an averager is 1:10; that is, the RMS noise (σ) is 10 times larger than the peak signal (S). The AC gain before the ADC, must be set such that the noise peaks do not exceed full scale. For Gaussian noise, it is 99.9% probable that the peak noise (N_p) amplitude is less than five times greater than the rms noise amplitude; that is, $N_p/\sigma \leq 5$, so that $N_p/s \leq 50$.

Assume that the input gain is set such that N_p is just equal (say) to the full-scale input level of an 8-bit ADC. Assume also that the resolution of the ADC is $2^8(=256)$, and the memory size $h = 2^{24}$, as before. Of these 8 bits, 6 bits ($=2^6 = 64$) will be required as dynamic reserve (i.e. to handle the input noise), and only 2 bits ($=2^2 = 4$) will be left to quantize the signal (S). In this example, then, the maximum number of sweeps before overflow would be

$$n_{\max} = \frac{2^{24}}{2^2} = 2^{22} \simeq 4.2 \times 10^6 \qquad (23)$$

Under the conditions of this example, the output (vertical) resolution will not be limited to 2 bits. Random noise accompanying the signal will *dither* the ADC; that is, the noise will modulate the quantization levels of the ADC so as to provide a resolution that increases as n increases. Note, however, that without noise and with the same full-scale setting, the averager output would indeed have a 2-bit amplitude resolution. In order to overcome this, the system shown in Figure 11 employs a Precision Enhancer, which is a separate 12-bit DAC, (digital analog converter) the output of which is added to the signal being digitized. The full-scale range of this signal is set to correspond to changes in only the bottom few bits of the main 8-bit ADC.

The 12-bit DAC is driven by a pseudorandom binary signal so that the analog output is effectively a noise source that dithers the true signal over a range of main ADC thresholds. Since the average of this extra noise is zero, it does not affect the averaged output, but it does improve the effective resolution of the main ADC.

It is useful to compare boxcar and signal averagers. Boxcar averagers can recover very fast waveforms and resolve features down to sub nanosecond level. However, they are time inefficient if used for waveform recovery and in such applications have been superseded by signal averagers and storage oscilloscopes. They do, however, remain cost-effective for static gate work.

RELATED ARTICLES

Article 121, Signals in the Presence of Noise, Volume 2; **Article 181, Amplitude Modulated Signals: The Lock-in Amplifier, Volume 3**; **Article 186, The Family of Signal Recovery Methods, Volume 3**.

FURTHER READING

SIGNAL RECOVERY (2002) *Model 4100 Boxcar Averager System Operating Manual*, available as PDF document from ***SIGNAL RECOVERY*** *website at www.signalrecovery.com*.

This article is based in part on a chapter by D.M. Munro originally published in Handbook of Measurement Science Vol 1, P.H. Sydenham, Wiley, 1982.

183: Correlators in Signal Extraction

Richard Burdett
SIGNAL RECOVERY, Wokingham, UK

1 Types of Correlation 1217
2 Loss of Phase Information in Correlators 1217
3 Simplified Autocorrelators 1218
4 Process of Autocorrelation 1218
5 Correlator Function for Sine Wave with Noise 1218
Further Reading 1219

1 TYPES OF CORRELATION

For the purpose of this article, correlation analysis is a method of detecting any similarity between two time-varying signals.

Autocorrelation consists of the point-by-point multiplication of a waveform by a delayed or time-shifted version of itself, this process being followed by an integration or summation process. Mathematically, the autocorrelation function $R_{xx}(\tau)$ of a time-varying function $f(t)$ is given by

$$R_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f(t) f(t+\tau)\, \mathrm{d}t \qquad (1)$$

where τ is the lag value or time shift between the two versions of $f(t)$.

Cross-correlation involves two waveforms and consists of the multiplication of one waveform $f(t)$ by a time-shifted version of a second waveform $g(t)$, followed by integration or summation. The cross-correlation function $R_{xy}(\tau)$ is given by

$$R_{xy}(\tau) = \lim_{x\to\infty} \frac{1}{2T} \int_{-T}^{T} f(t) g(t+\tau)\, \mathrm{d}t \qquad (2)$$

Note that cross-correlation requires two input signals, as is also true in the case of signal averaging (where a synchronizing input is required in addition to the signal input). Also, as with an averager, a cross-correlator preserves signal phase information. Unlike the averager, however, the cross-correlator output waveform, the correlogram, is affected by the waveform of the second input signal – an undesirable and unnecessary complication for signal-recovery applications since a multipoint averager may be used.

Cross-correlators are normally used in flow or velocity measurements where the time lag is detected to indicate flow rate; they are rarely used for simple signal-recovery purposes. (All that said, this is ignoring the lock-in amplifier and boxcar integrator, both of which can be regarded as a special type of cross-correlator.)

2 LOSS OF PHASE INFORMATION IN CORRELATORS

Phase information is lost in an autocorrelation function, as is also true for its Fourier transform, power spectral density. This lack of phase information means that in some cases, the input signal responsible for a given correlogram must be deduced by intelligent guesswork. For example, as shown in Figure 1, the autocorrelation function for band-limited white Gaussian noise is a spikelike peak at $\tau = 0$, with a width that would decrease as the noise bandwidth increases. A similar correlogram could have resulted from an input consisting of a single narrow pulse; the narrower the pulse width, the narrower the correlogram spike. Thus, the pulse and the band-limited white noise have similar power-density spectra. (The difference between them is

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

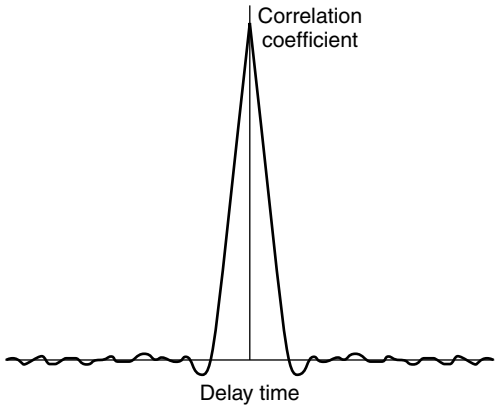


Figure 1. Correlation function of band-limited white noise. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

that frequency components of the noise have random phase relationships.)

3 SIMPLIFIED AUTOCORRELATORS

A simplified block diagram of an autocorrelator is shown in Figure 2. The ADC digitizes the input signal once every lag interval or dwell time, t_g, with each such A/D conversion requiring a conversion time, t_c, where $t_c \ll t_g$. The output digital word from the ADC, corresponding to the latest sample, provides one input to the digital multiplier and is also shifted, as word 0, into an n-word shift register. During this shift operation, the last word in the register, word $(n-1)$, is shifted out and discarded and the former word $(n-2)$ becomes the new word $(n-1)$. The control and timing circuits then cause the shift register to step n times, recirculating its contents one full cycle, and requiring a time interval t_r. During t_r, each shift-register word (i.e. $(n-1)$, $(n-2)$, $(n-3)$,. . ., 3, 2, 1, 0) is applied sequentially to the other input of the multiplier, which multiplies each of these words by the word at its other input, word 0. Each multiplier output is added to the contents of the corresponding bin of the N-bin main memory; for instance, word 0 $\times$ word $(n-16)$ bin$(n-16)$, word 0 $\times$ word 0 bin 0. After many such cycles, the contents of bin i (for example) of the main memory, will be the sum of products formed by multiplying each new signal sample by an it_g delayed version of itself. Bin 0, for example, corresponds to the signal multiplied by itself with zero delay.

The minimum time between successive samples is $(t_c + t_r)$. When $t_g \geq (t_c + t_r)$, the correlator is said to be working in a real-time mode. When $t_g \leq (t_c + t_r)$, samples can no longer be taken every t_g seconds and the correlator is said to be in a pseudo-real-time mode. For $t_g \ll (t_c + t_r)$, the correlator is in a batch mode.

4 PROCESS OF AUTOCORRELATION

The process of autocorrelation involves the concept of sliding a waveform past a replica of itself. For random noise, the two waveforms will match at only one point as they slide by each other, that is, at $\tau = 0$, when they are perfectly aligned. In contrast, a square wave sliding by another square wave will find a perfect match once in every period and will give rise to a triangular correlogram. More generally, signals that are periodic in time will produce a correlation function that is periodic in τ. Suppose, for example, that the input to an autocorrelator is $f(t) = A\cos(\omega t)$. Then,

$$R_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} A\cos(\omega t) A\cos(\omega t + \tau)\,\mathrm{d}t$$
$$= \frac{1}{2}A^2\cos(\omega\tau) \tag{3}$$

5 CORRELATOR FUNCTION FOR SINE WAVE WITH NOISE

Figure 3 shows the correlation function of a sine wave accompanied by band-limited white noise. Note that the

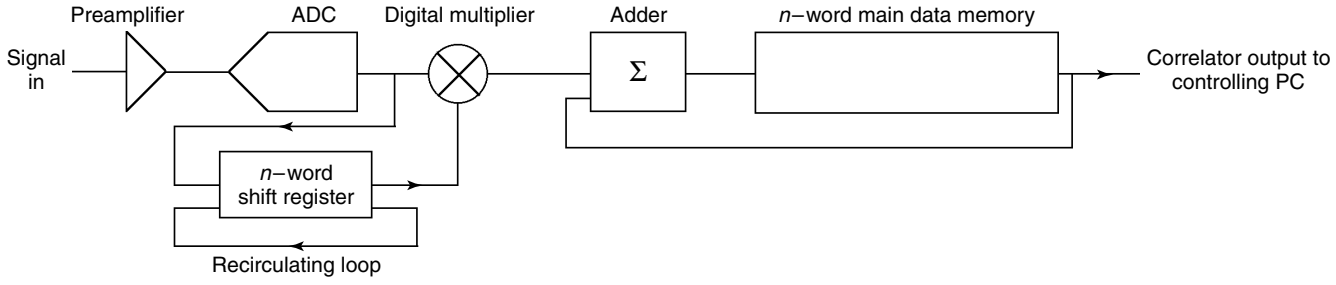


Figure 2. The autocorrelator (simplified). (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

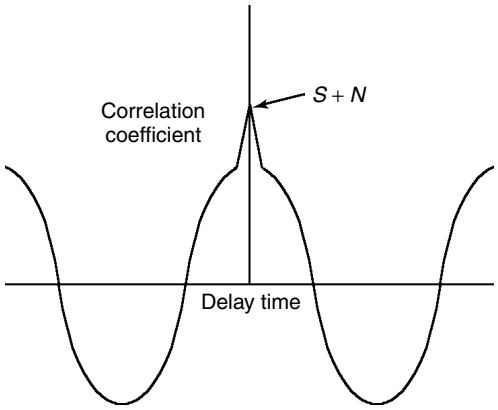


Figure 3. Correlogram of a noise sine wave. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

peak value at $\tau = 0$ in this correlogram is the mean-squared value of the signal plus noise (i.e. $S + N$). The mean-squared value of the sinusoidal signal component is given by the peak value of the sine wave at approximate values where the white noise spike has damped to zero. The SNR of the correlator output can, therefore, be determined since

$$\frac{S}{(S+N)-S} = \frac{S}{N} = SNR \tag{4}$$

Most importantly, notice that the signal and noise components have been separated by virtue of their different positions on the τ-axis. It is this separating ability that makes correlation a powerful signal-recovery technique.

FURTHER READING

Box, G.E.P. and Jenkins, G. (1976) *Time Series Analysis: Forecasting and Control*, Holden-Day.

This article is based in part on a chapter by D.M. Munro originally published in Handbook of Measurement Science Vol 1, P.H. Sydenham, Wiley, 1982.

184: Photon Counting

Richard Burdett
SIGNAL RECOVERY, Wokingham, UK

1 The Photomultiplier Tube, PMT 1220
2 Quantum Efficiency of PMT 1220
3 Ratemeters 1220
4 Signal-to-noise Ratio (SNR) of PMT 1220
5 Thermionic Emission and Dynamic Response 1221
6 Dark Current in PMT 1222
7 Practical Assumptions in PMT use 1222
Further Reading 1222

1 THE PHOTOMULTIPLIER TUBE, PMT

Photomultiplier tubes (PMTs) are used to measure the intensity or flux of a beam of visible photons (see Figure 1). With its photocathode removed, the PMT becomes an electron multiplier tube (EMT) that is widely used to detect ions and electrons.

One of the most important advantages of such detectors is that their high gain and low noise allow them to give one output pulse for each detected input particle. Since visible-light measurements are perhaps most commonplace, we consider here as an example a PMT and pulse-counting system of the type shown in Figure 2.

2 QUANTUM EFFICIENCY OF PMT

The probability of each incident photon causing an output pulse from the PMT is essentially equal to the quantum efficiency; ζ: typically, ζ is between 5 to 25%. In addition to the photon-derived pulses (i.e. the signal pulses), there will be spurious noise pulses at the PMT output due to thermionic emission.

Noise pulses caused by thermionic emission from the dynodes will experience less gain and will be smaller in amplitude than pulses due to cathode emission. The PMT output pulses are amplified and presented to a pulse-height discriminator circuit, where the peak amplitude of each pulse is compared to an adjustable threshold or reference voltage.

Ideally, the discriminator will reject all dynode-derived noise pulses and accept all cathode-derived pulses; in practice, the PMT gain is statistical in nature and cathode and dynode-derived pulses have overlapping amplitude distributions. The discriminator will therefore accept most cathode-derived pulses and reject most dynode noise pulses. Each accepted input pulse will cause a standardized output pulse. Such pulse-height discrimination also reduces the effect of PMT gain variations with time and temperature.

3 RATEMETERS

Ratemeters are used to give a continuous analog output voltage that is proportional to the rate of digital pulses being generated by the discriminator. Today, digital counter circuits are more commonly used to accumulate output counts for a preselected measurement time. Such counters allow very long integration times and give a direct digital output suitable for computer logging purposes.

4 SIGNAL-TO-NOISE RATIO (SNR) OF PMT

Suppose the PMT is set up to detect photons emitted from a thermal light source such as a tungsten filament lamp. The

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

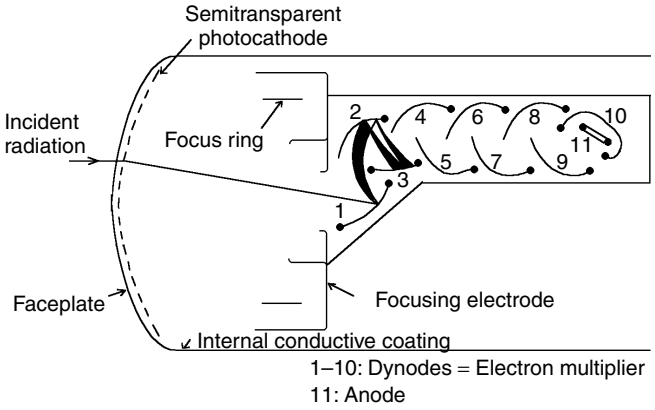


Figure 1. Cross-sectional view of typical end-window PMT. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

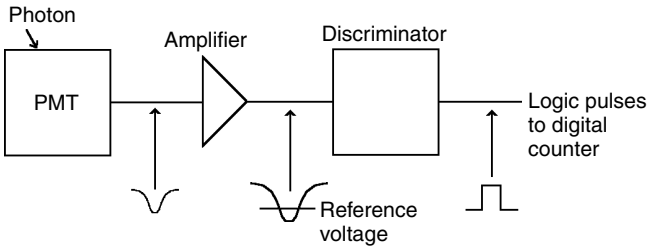


Figure 2. Typical photon counting system. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

time interval between successive photons impinging upon the PMT photocathode is random and governed by a Poisson distribution (see Figure 3, Curve A). The probability P of detecting n photons in time t following the last photon is given by

$$P(n,t) = \frac{(\zeta Rt)^n \mathrm{e}^{-\zeta Rt}}{n!} = \frac{N^n \mathrm{e}^{-N}}{n!} \tag{1}$$

where R is the mean photon rate (photons per second) and $N = \zeta Rt$ is the signal (i.e. the mean number of photon electrons emitted by the PMT photocathode during the time interval t). The noise, or uncertainty, in N is given by the standard deviation σ, where

$$\sigma = \sqrt{(\zeta Rt)} = \sqrt{N} \tag{2}$$

so that

$$\mathrm{SNR} = \frac{N}{\sqrt{N}} = \sqrt{N} = \sqrt{(\zeta Rt)} \tag{3}$$

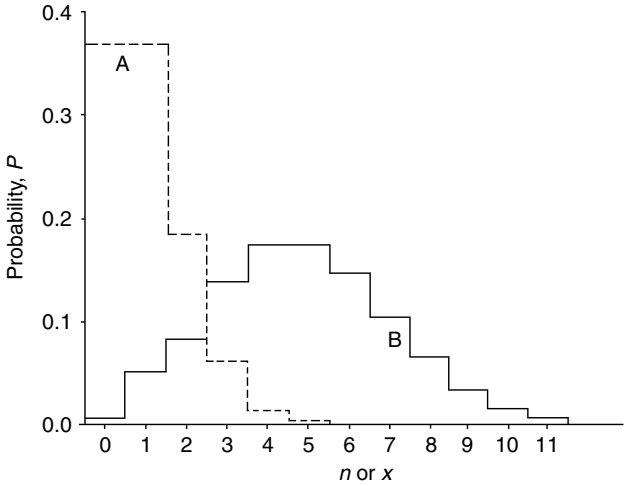


Figure 3. The Poisson distribution. Curve A: Probability of detecting n photons in time $t = (Rt)^n \exp(-Rt/n)$, $R = 10^8$ photons per second, $t = 10\,\mathrm{ns}$ and so $Rt = 1$, $\sigma = \sqrt{Rt}$. Curve B: Probability of gain magnitude $P(x) = M^x \mathrm{e}^{-M}/(x!)$ where $x =$ dynode gain, $M =$ mean dynode gain $= 5$, and $\sigma = \sqrt{M}$. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

5 THERMIONIC EMISSION AND DYNAMIC RESPONSE

Note that, as in all of the techniques examined in this section (with white noise), the SNR is again proportional to the square root of the measurement time (t). If we assume that there is no thermionic (dark) emission of electrons from the photocathode, then the photocathode (signal) current (in amperes) is given by

$$I_{\mathrm{pe}} = \zeta Re \tag{4}$$

where e is the charge of an electron ($\approx 1.6 \times 10^{-19}\,\mathrm{C}$).

The signal-to-noise ratio ($\mathrm{SNR_k}$) of the photocathode current (I_{pe}) is given by

$$\mathrm{SNR_k} = \sqrt{(\zeta Rt)} = \sqrt{\left(\frac{\zeta Rte}{e}\right)} = \sqrt{\left(\frac{I_{\mathrm{pe}}t}{e}\right)} \tag{5}$$

Now the measurement time t has associated with it a frequency range Δf, where

$$t = \frac{1}{2\Delta f},$$

so that

$$\mathrm{SNR_k} = \sqrt{\left(\frac{I_{\mathrm{pe}}}{2e\Delta f}\right)} = \sqrt{\left(\frac{I_{\mathrm{pe}}^2}{2eI_{\mathrm{pe}}\Delta f}\right)} = \frac{I_{\mathrm{pe}}}{\sqrt{(2eI_{\mathrm{pe}}\Delta f)}} \tag{6}$$

If we multiply both the numerator and denominator of (6) by the mean PMT gain A, then

$$\mathrm{SNR_k} = \frac{A I_{\mathrm{pe}}}{A\sqrt{(2eI_{\mathrm{pe}}\Delta f)}} = \frac{I_{\mathrm{a}}}{\sqrt{(2AeI_{\mathrm{a}}\Delta f)}} = \mathrm{SNR_a} \quad (7)$$

where I_{a} is the DC anode current and $\mathrm{SNR_a}$ is the signal-to-noise ratio of the anode current, if thermionic emission and other dynode noise contributions are ignored.

Note that the general expression for the shot noise of a DC current I is given by

$$\text{rms shot noise current} = \sqrt{(2AeI\Delta f)} \quad (8)$$

where A is the gain following the shot noise process. When $A = 1$, the expression simplifies to $\sqrt{(2eI\Delta f)}$. Notice also that shot noise is present in the light beam itself and that the PMT quantum efficiency (ζ) degrades the SNR by a factor of $\sqrt{\zeta}$.

6 DARK CURRENT IN PMT

In practice, with no input photons, the photocathode will emit electrons due to temperature effects. The dynodes will also emit thermionic electrons. The rate of such thermionic emission is reduced by cooling the PMT. Thermionic emission from the photocathode, that is, dark counts, can be further reduced by minimizing the cathode area and by selecting a photocathode material with no more red (long wavelength) spectral response than is necessary.

If the photocathode emits electrons randomly at a dark count rate R_{d}, then the noise components of the cathode current will increase to $\sqrt{(\zeta Rt + R_{\mathrm{d}}t)}$ and the signal-to-noise ratio of the cathode current will degrade to

$$\mathrm{SNR} = \frac{\zeta Rt}{\sqrt{(\zeta Rt + R_{\mathrm{d}}t)}} = \frac{\zeta R\sqrt{t}}{\sqrt{(\zeta R + R_{\mathrm{d}})}} \quad (9)$$

7 PRACTICAL ASSUMPTIONS IN PMT USE

The above expression, (9), will also be the PMT output's SNR if the dynode noise is assumed to be removed completely by pulse-height discrimination. For PMTs equipped with a high-gain first dynode with Poissonian statistics (see Figure 3, Curve B), this is not an unreasonable approximation.

Note that when a PMT is used in a noncounting or DC mode, as is discussed in **Article 180, Bandwidth Reduction of Baseband DC Signals, Volume 3**, all of the output electrons resulting from spurious cathode emission (i.e. dark counts) and dynode emission are integrated by the anode or preamplifier time constant into a DC dark current; the opportunity to remove dynode noise by pulse-height discrimination is lost.

FURTHER READING

Garrard, C.C. (2003) *Practical Photon Counting (Technical Reprint R/P096)*, Electron Tubes Ltd, London. Available from www.electron-tubes.co.uk Web site as PDF document.

This article is based in part on a chapter by D.M. Munro originally published in Handbook of Measurement Science Vol 1, P.H. Sydenham, Wiley, 1982.

185: Pulse Height Discrimination, Ratemeters and Pileup

Richard Burdett
SIGNAL RECOVERY, Wokingham, UK

1 Amplification in the PMT 1223
2 Adjustment of Discrimination in the PMT 1223
3 Setting the PMT High Voltage 1224
4 Ratemeters and Digital Counters 1224
5 Synchronous Detection 1224
6 Pulse Pileup 1225
7 Paralyzable Detectors 1226
8 Dead Time in PMTs 1226
Further Reading 1226

1 AMPLIFICATION IN THE PMT

Each electron emitted by a PMT photocathode will be amplified by the instantaneous value of the PMT gain. For a mean gain of 10^6, for example, each cathode electron will cause an average output charge q of $10^6 e$ C. This charge q will accumulate at the anode during a time t, given by the transit-time spread of the PMT. Typically, t will be about 10 ns so that the resulting anode current pulse ($i_a = dq/dt$) will have a full width (t_w) between half-maximum amplitude points (FWHM) of about 10 ns also. The peak value, I_{pk}, of the anode current pulse may be approximated by assuming the pulse to be rectangular so that in this example,

$$I_{pk} \approx \frac{q}{t_w} = \frac{10^6 e}{10 \times 10^{-9}} = \frac{10^6 \times 1.6 \times 10^{-19}}{10 \times 10^{-9}} = 16\,\mu\text{A} \quad (1)$$

In a photon-counting system, the anode-load resistor (R_a) of the PMT is kept small, usually 50 to 100 Ω. Therefore, the time constant (τ_a) formed by the anode stray capacitance (C_a) will be small compared to t_w and thus will not stretch the anode voltage pulse. Typically, $R_a = 50\,\Omega$ and $C_a = 20$ pF so that $\tau_a = 1\text{ ns} \ll t_w$. The anode voltage pulse will then have the same shape as the anode current pulse and a peak value of

$$E_{pk} = I_{pk} R_a = 16 \times 10^{-6} \times 50 = 0.8\,\text{mV} \quad (2)$$

It should be remembered that such pulse amplitudes depend upon the PMT gain, which in turn depends upon the dynode gains – which are statistical. In the above example, $E_{pk} = 0.8$ mV is the average pulse height to be expected. Actual pulse heights will be distributed above and below this value, and the better the PMT, the narrower this distribution will be. A suitable preamplifier is normally used to amplify the anode pulses to a suitable level for the pulse height discriminator.

2 ADJUSTMENT OF DISCRIMINATION IN THE PMT

Notice that the PMT gain, the preamplifier gain, and the discriminator threshold controls may all be used to adjust the effective discrimination level. Figure 1 shows a typical count-rate variation with PMT high voltage (i.e. PMT gain) at a fixed discriminator threshold level. This is one of the family of curves that could be plotted for different threshold levels; similar curves could be obtained by varying the

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

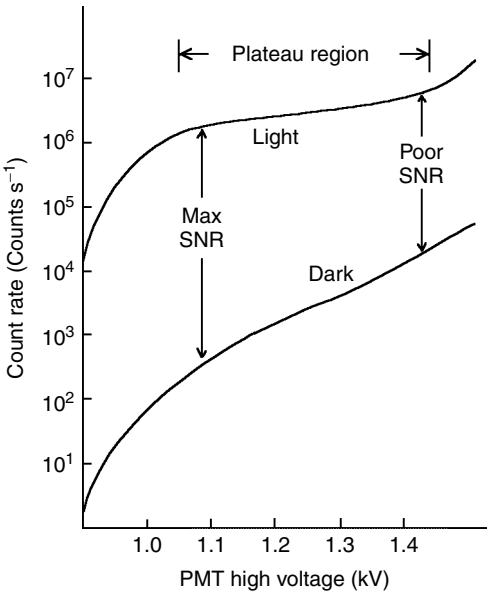


Figure 1. The PMT counting plateau. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

preamplifier gain rather than that of the PMT or by plotting count rate against discriminator threshold.

The upper curve in Figure 1 is plotted by allowing light to fall upon the PMT photocathode and by slowly varying the PMT voltage (which is nonlinearly related to the PMT gain). Notice that the steep slope at low PMT voltages begins to flatten and form a not-quite-horizontal plateau as proper focusing takes place in the PMT. The increasing slope at very high PMT bias voltages is due to increasing instability in the PMT.

The upper curve corresponds to $S + N$, since it is based on both signal and noise pulses. The lower curve was plotted with the PMT in darkness and therefore represents noise (N) pulses only. Note that, typically, the dark-count curve has no plateau; it has been suggested that the lack of a plateau is due to corona effects associated with microscopic protrusions from the dynode surfaces.

Since the count rate is plotted on a logarithmic scale, the vertical distance between the two curves corresponds to

$$\log(S+N) - \log(N) = \log\left(\frac{S+N}{N}\right) = \log\left(1\frac{S}{N}\right) \approx \log\left(\frac{S}{N}\right) \text{ for } \frac{S}{N} \gg 1 \qquad (3)$$

3 SETTING THE PMT HIGH VOLTAGE

A commonly used method for setting the PMT high voltage for a given preamplifier gain and discriminator threshold level is, therefore, to select a point at the beginning of the counting plateau, corresponding to maximum SNR.

4 RATEMETERS AND DIGITAL COUNTERS

Historically, the pulse rate generated by a PMT amplifier-discriminator was measured using an analog ratemeter. Essentially, this consists of a current pulse generator that produces a precise current pulse for each applied trigger as generated by the discriminator. These pulses are integrated using a low pass filter, whose output voltage is therefore proportional to the average value of the discriminator output count rate (R_{sig}). The voltage is displayed on an analog panel meter that is calibrated in counts per unit of time.

Nowadays, digital counters are almost always used as recording device in photon-counting experiments. These simply count the number of applied pulses and can give an output in terms of number of counts or counts per second. Dual counters allow more sophisticated work, such as ratio mode (also known as *source compensation mode*), where the signal count rate is proportional to both the measurand and (say) the intensity of a light source. By monitoring the light source with a separate PMT and amplifier-discriminator to produce a source-dependent count rate R_{sc}, R_{sig} can be normalized by R_{sc} to provide an output count that is independent of source fluctuations.

In the reciprocal mode, the counter measures the time (t, in μs) required for the cumulative signal counts to reach N; the smaller the signal count rate, the longer the elapsed time. If the dark count rate is negligible, then the measurement accuracy is $1/\text{SNR} = 1/\sqrt{(R_{sig}t)}$, and for a constant value of $R_{sig}t\ (= N)$, all measurements will have the same SNR and accuracy.

5 SYNCHRONOUS DETECTION

A synchronous photon-counting system that provides automatic background subtraction is shown in Figure 2. When the chopper blade blocks the input light, the output pulses from the amplifier-discriminator are, by definition, background noises; these pulses (N) are gated into counter B by the timing circuit – which is itself synchronized by the chopper reference signal. When the chopper blade allows light to reach the PMT, the discriminator output consists of signal-plus-background pulses ($S + N$), and these pulses are gated into counter A.

After each measurement interval, an arithmetic circuit provides two outputs:

$$A - B = (S+N) - N = S = \text{signal} \qquad (4)$$

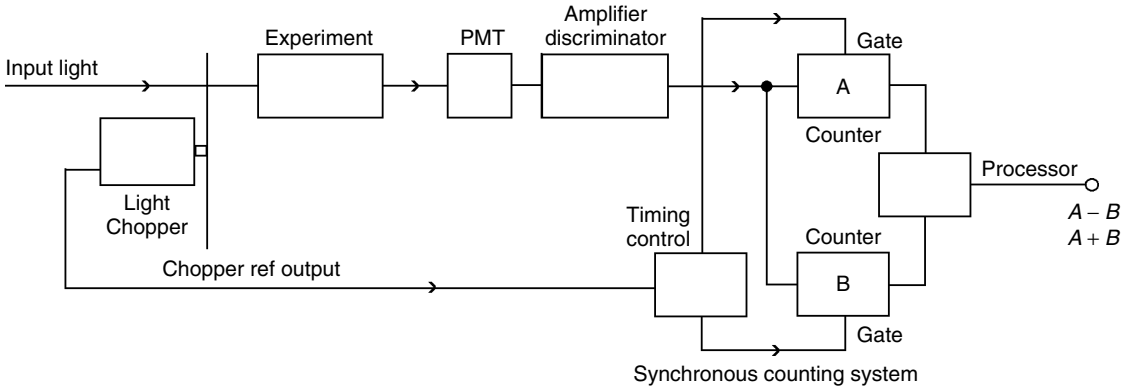


Figure 2. Photon counter with background subtraction. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

and

$$A + B = (S + N) + N = \text{total counts} \tag{5}$$

where A and B are the numbers of counts in counters A and B respectively. For Poissonian noise,

$$\text{SNR} = \frac{\text{signal}}{\sqrt{\text{total counts}}} = \frac{A - B}{\sqrt{A + B}} \tag{6}$$

Suppose, for example, that $A = 10^6$ counts and $B = 9.99 \times 10^5$ counts, then $S = A - B = 10^3$ counts and $\sqrt{(A + B)} = 1.41 \times 10^3$, so that

$$\text{SNR} = \frac{A - B}{\sqrt{A + B}} = \frac{10^3}{1.41 \times 10^3} = 0.71$$

and (in)accuracy

$$\frac{1}{\text{SNR}} = \frac{1}{0.71} = 141\%$$

or, expressed in words, the measurement is worthless! The $A + B$ output is important since it allows the measurement accuracy to be estimated in this way.

6 PULSE PILEUP

The dynamic range of photon-counting measurements is limited at low light levels by PMT dark count and at high light levels by pulse pileup in the PMT or electronics. As the mean rate (R) of photons arriving at a PMT photocathode increases, so does the probability of two or more photons arriving with too short an interval between them to be resolved by the PMT.

The time resolution of a PMT is effectively equal to its output pulse width t_w, and each output pulse from a PMT, therefore, occurs whenever an electron is emitted after a time greater than t_w following the previous electron. The probability of this happening is the same as that for zero photoelectron events in a time t_w (dark counts at high light levels can be neglected). As shown in Figure 3 of **Article 184, Photon Counting, Volume 3** and from (1) in that article,

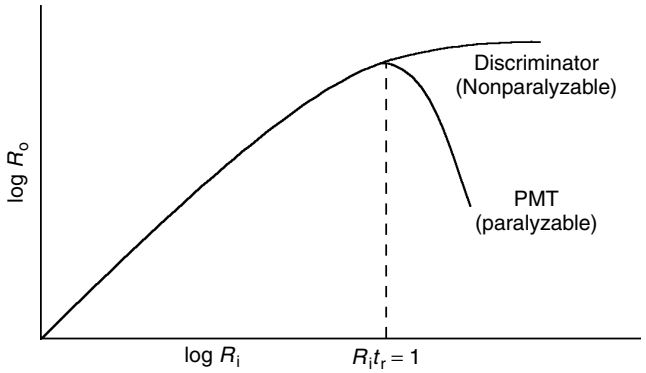


Figure 3. Counting error due to pulse pileup. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK Advanced Measurement Technology.)

$$P(0, t_w) = \exp(-\zeta R t_w) \tag{7}$$

and the output count rate, R_o, is given by

$$R_o = P(0, t_w) R_i = \zeta R \exp(-\zeta R t_w) = R_i \exp(-R_i t_w) \tag{8}$$

The resulting PMT pulse pileup error is given by

$$\varepsilon_{pmt} = \frac{R_i - R_o}{R_i} = \frac{R_i - R_i \exp(-R_i t_w)}{R_i} = 1 - \exp(-\zeta R t_w) \tag{9}$$

7 PARALYZABLE DETECTORS

The PMT is a paralyzable detector; that is, when the input count rate exceeds a certain value ($R_i = 1/t_w$), the output count rate will begin to decrease for an increasing input count rate and will become zero when the PMT is completely paralyzed (saturated) (See Figure 3).

Discriminators and counters, on the other hand, are usually nonparalyzable. Suppose a discriminator, for example, has a pulse-pair resolution or dead time t_d. That is, each time it accepts an input pulse, it cannot accept a new pulse until after a time t_d. Then, for a measurement time t, an input pulse rate of R_i, and an output rate R_o; the total number of output pulses, N_o, is given by

$$N_o = R_o t \tag{10}$$

and the

$$\text{total dead time} = N_o t_d = R_o t t_d \text{s} \tag{11}$$

so that

$$\text{total live time} = t - R_o t t_d \tag{12}$$

The total number of input pulses accepted is therefore given by

$$N_o = R_o t = R_i(t - R_o t t_d) \tag{13}$$

so that

$$R_o = \frac{R_i}{1 + R_i t_d} \tag{14}$$

8 DEAD TIME IN PMTs

A fast discriminator and counter have a dead time, t_d, of about 10 ns – similar to the time resolution, t_w, of a reasonably fast PMT. Note, however, that any PMT pileup will act as a prefilter to the discriminator; that is, such pileup will decrease the input count rate to the discriminator. PMT pileup usually provides the upper limit to the system dynamic range. Photon-counting systems cannot be used in pulsed light measurements where the peak photon rate (during the light pulse) will cause unacceptably high pulse pileup errors.

FURTHER READING

Garrard, C.C. (2003) *Practical Photon Counting (Technical Reprint R/P096)*, Electron Tubes Ltd., London. Available from www.electron-tubes.co.uk website as PDF document.

This article is based in part on a chapter by D.M. Munro originally published in Handbook of Measurement Science Vol 1, P.H. Sydenham, Wiley, 1982.

186: The Family of Signal Recovery Methods

Richard Burdett
SIGNAL RECOVERY, Wokingham, UK

1 Signal Recovery Family 1227
2 Multichannel Scalers and Multichannel Averagers 1227
3 Using Several Instruments Together 1227
4 Time Taken Versus Signal Recovery Sensitivity 1228
Related Articles 1228
Further Reading 1228

1 SIGNAL RECOVERY FAMILY

Many of the signal recovery considerations discussed in this section in **Part 12, Elements: F – Signals and Noise, Volume 3**, such as instrument selection, are summarized in Figure 1.

Note that this figure includes several instruments not mentioned in the preceding articles of this section, including the multichannel scaler (MCS) and multichannel analyzer (MCA).

The choice of signal recovery instrument is often limited to that which is available so that, for example, a combination of an MCS and voltage-to-frequency converter (VFC) can be used in place of a digital signal averager. Similarly, a boxcar averager in swept gate (waveform recovery) mode can be a substitute for a high performance digital signal averager.

2 MULTICHANNEL SCALERS AND MULTICHANNEL AVERAGERS

An MCS consists of a digital counter, whose output is recorded by a bank of digital memory, in much the same way that the output of the analog-to-digital converter (ADC) is processed in a digital signal averager.

During each sweep, the counter records the number of input pulses during each dwell time and adds that number to the cumulative count in the corresponding memory address. By using a VFC in front of an MCS, analog signals can be time-averaged, and the VFC/MCS combination is essentially an alternative approach to using a sampled ADC as in a digital signal averager.

An MCA performs pulse-height analysis by using the amplitude of each pulse to determine an address location. The memory at that address is then incremented by one count, and the cycle repeats for many pulses.

In other words, each input pulse with amplitude between 63.85 and 63.95% (say) of full-scale, will add one count to the memory address No. 639. In this way, a pulse-height distribution or spectrum is built up.

Another common use of an MCA is to precede it by a time-to-amplitude converter (TAC), so that each input pulse to be digitized corresponds to a time interval. Low-level measurements of short fluorescent lifetimes, for example, may be made in this manner.

3 USING SEVERAL INSTRUMENTS TOGETHER

In some cases, two or more signal recovery instruments may be used at the same time. For example, the output of a lock-in amplifier can be digitally averaged in order to reduce $1/f$ noise problems. Similarly, a boxcar averager followed by digital signal averager can offer the picosecond or nanosecond time resolution of the boxcar, but without the need to scan so slowly that the system being measured may change during a scan (sweep).

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

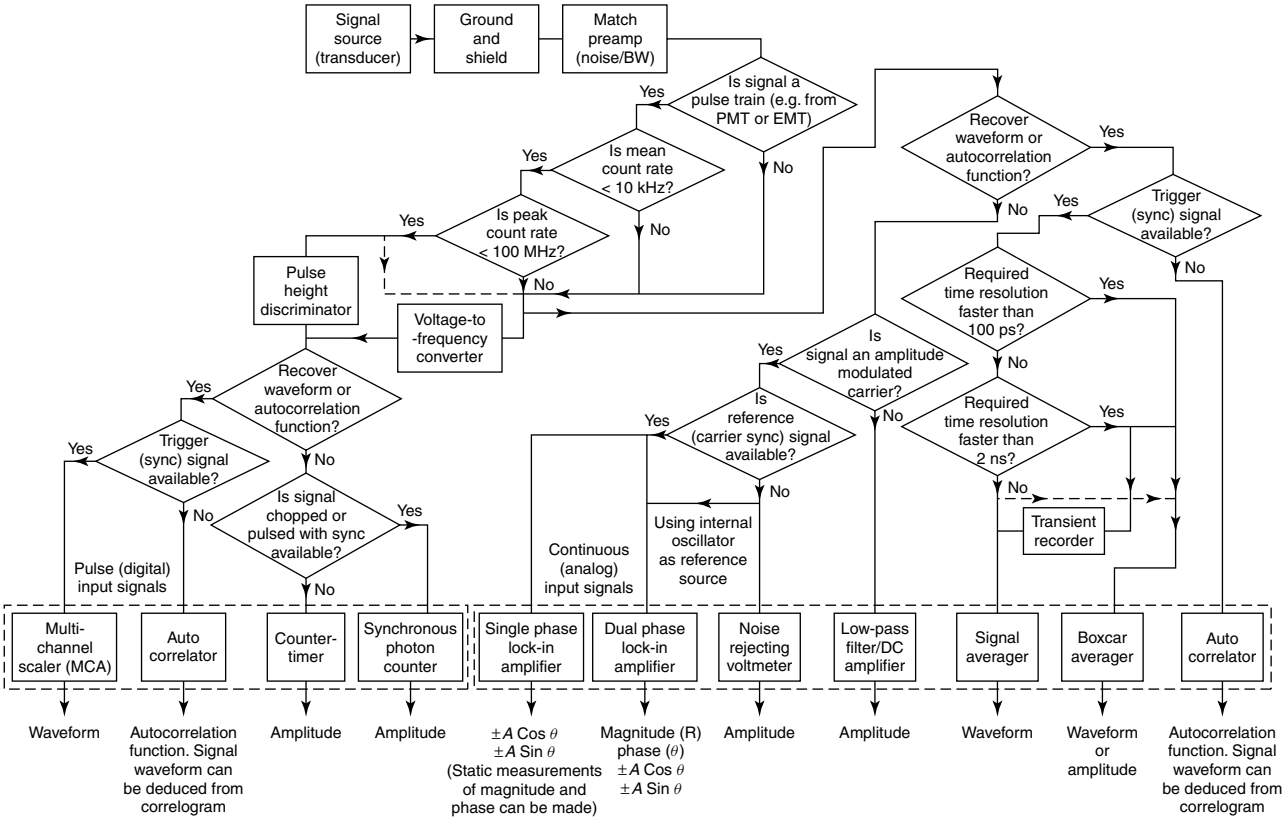


Figure 1. Signal recovery flowchart. (Reproduced by permission of SIGNAL RECOVERY, a part of AMETEK advanced measurement technology.)

Two lock-in amplifiers can be used for tandem demodulation experiments, where the signal being measured consists of an amplitude-modulated carrier of frequency f_1 with a modulation of frequency f_2. The first lock-in amplifier is run at a reference frequency f_1 and, assuming the output time-constant is short enough to allow it to pass, generates an analog output signal that is at f_2. This is then coupled to a second lock-in, running at f_2, which therefore measures the modulation amplitude.

4 TIME TAKEN VERSUS SIGNAL RECOVERY SENSITIVITY

The object in signal recovery is not to maximize the signal-to-noise improvement ratio (SNIR) alone but to minimize the measurement time required to reach a particular output signal-to-noise ratio (SNR). Similarly, in selecting a preamplifier, the real object is to minimize noise, not the noise figure. The noise and/or bandwidth of the signal source or transducer should, therefore, always be first minimized before using other signal recovery techniques to further improve the SNR. As a general rule, increasing the discrimination needed for a given SNR increases the time required to achieve it. Furthermore, care must be taken to ensure that the signal is stable during the extraction time.

RELATED ARTICLES

Article 121, Signals in the Presence of Noise, Volume 2; **Article 176, Signals and Signal-to-noise Ratio, Volume 3**.

FURTHER READING

ORTEC. (2002) TRUMP-pci Multichannel Analyzer on a PCI Card with Software (product data sheet), available as PDF document from ORTEC website at www.ortec-online.com.

ORTEC. (2003) MCS-pci Multichannel Scaler on a PCI Card with Software (product data sheet), available as PDF document from ORTEC website at www.ortec-online.com.

This article is based in part on a chapter by D.M. Munro originally published in Handbook of Measurement Science Vol 1, P.H. Sydenham, Wiley, 1982.

187: Flowmeter Selection and Application

Michael Reader-Harris
National Engineering Laboratory, Scotland, UK

1 Survey of Flowmeters and Benefits of their Use 1233
2 Flowmeter Types 1233
3 An Overview of Different Types of Flowmeters 1233
4 Installation Issues 1235
Related Articles 1236
References 1236

1 SURVEY OF FLOWMETERS AND BENEFITS OF THEIR USE

In the United Kingdom alone there are around 15 million flowmeters in use, employing as many as 120 different operating principles. These meters have to operate at temperatures between −260 and +800 °C and at pressures from vacuum to 800 bar. Flowmeter diameters range from less than 1 mm to several meters, and the fluids are as diverse as water, natural gas, hot nitric acid, cryogenic liquids, and complex mixtures of liquids and gases. Flowmeters are essential for buying or selling fluids, for taxation, for process control, for evaluation of the performance of pumps, turbines or fans, and for many other purposes. Some flowmeters need to operate at the lowest possible level of uncertainty (e.g. large gas meters); some only need to be repeatable (e.g. some meters used for process control).

The benefits of using flowmeters are varied: they may be used to obtain fairness in purchase, sale, or taxation, or to give the required quality in a manufacturing process, or to ensure in water distribution systems that reservoirs are appropriately sized and that there are no leaks. Picking the best meter for a particular application is therefore a difficult task and one that the inexperienced plant designer or engineer can find very confusing. This is especially true when the plant designer is not familiar with flow measurement, or when the instrument engineer familiar with the technology is not involved early in the design process.

2 FLOWMETER TYPES

To assist in the selection of a flowmeter, a classification guide was produced within BS 7405 (British Standards Institution 1991). It is of use in helping to clarify what types of meter are available and hence to assist in selection. BS 7405 divides flowmeters into 10 groups, as shown in Table 1. The groups are loosely based on each flowmeter's operating principles rather than on their applications. The basic principles of each of these types of flowmeters is described in **Article 189, Basic Principles of Flow Measurement, Volume 3**.

3 AN OVERVIEW OF DIFFERENT TYPES OF FLOWMETERS

3.1 Conventional differential-pressure flowmeters

The orifice plate is by far the most widely used type of flowmeter; consequently, more is known about its behavior than about the behavior of any other flowmeter. It has good standards, which have recently been revised (International Organization for Standardization 2003a,b). It is affected by flow profile and hence needs a sufficient straight installation length. The upstream straight length can be reduced by

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

Table 1. Classification of flowmeters.

Group	Meter type	Description
1	Conventional differential pressure	Sharp-edged concentric orifice plates, Venturi tubes, and flow nozzles.
2	Other differential pressure	All other differential-pressure devices, including those that can only loosely be termed differential pressure, including sonic nozzles and variable area meters.
3	Positive displacement	All PD meters, sliding vane, gear, piston types, etc.
4	Rotary turbine	Rotary turbine meters, of which the most common is the axial-bladed turbine meter.
5	Fluid oscillatory	Meters in which the operating principle is based on fluidic oscillations, normally generated by the meter itself, e.g. vortex meters, fluidic meters.
6	Electromagnetic	Meters in which an electromagnetic principle is used directly to sense the flow.
7	Ultrasonic	Meters in which ultrasonic sensors directly measure the flow, e.g. time of flight and Doppler meters. This excludes meters in which ultrasonic methods are used to detect another principle.
8	Direct and indirect mass	Mass flowmeters of both the direct and indirect type are included. This would probably include Coriolis meters used as volume meters.
9	Thermal	Meters using thermal heat loss or gain to measure the flow rate. Thermal types include hot-wire anemometers.
10	Miscellaneous	All those meters that cannot easily be classified into any of the other groups, and so this group contains such types as the laser Doppler anemometer, and tracer, cross-correlation and velocity-area techniques.

use of a flow conditioner. The orifice plate will work for liquids and gases and perform in the range 0.5 to 1%. Its main attraction is that this performance can be achieved by careful manufacture and measurement without calibration in a flowing fluid. Its discharge coefficient is a function of Reynolds number, and, if calibration is required to reduce the uncertainty, it is not necessary to use the same fluid for calibration and for use. If the upstream sharp edge is even marginally rounded, the measurement is significantly affected.

Nozzles and Venturi tubes perform in a similar manner, but when used uncalibrated, they generally give a larger uncertainty. Venturi tubes give much less pressure loss than orifice plates or nozzles. Where possible, single-phase flow differential–pressure devices should use multiple tappings or annular chambers, and not a pair of single pressure tappings.

Further information on this important class of flowmeter can be found in **Article 188, Differential Pressure (DP) Flowmeters, Volume 3**.

3.2 Other differential-pressure methods (including sonic nozzles)

The variable area meter is a cheap and simple meter for both gas and liquid applications at about 2% accuracy. It can be very good for process-control and indication applications.

Many other differential-pressure devices are available such as V-cones and laminar-flow elements. Each has particular advantages. Even instrumenting a valve or pipe restriction may provide the measurement required.

For gas, the sonic or critical-flow nozzle provides a good calibration reference and also a good measurement device for mass flow. It has a very high pressure drop and hence is not used for applications in which pressure loss is important.

3.3 Positive-displacement flowmeters

Positive-displacement (PD) meters provide highly accurate and repeatable measurement of most *clean* fluids. They can be heavy and bulky in larger sizes and hence expensive. They are ideal for batch measurement at high accuracy and are extensively used for fiscal and custody transfer work. PD meters are generally unaffected by changes in flow profile generated by pipe installation. They are often the only meter that is capable of being used with high-viscosity fluids, but they can block the flow if they seize up. A variant is the metering pump that pumps and measures at the same time.

3.4 Turbine flowmeters

Turbine meters are used for many high accuracy applications in both gas and liquid. They have a restricted range with viscosities greater than 30 cSt. They are very repeatable but sensitive to viscosity, flow profile, and pulsations. Wear to the bearings has a significant effect on the meter's performance. Fast response and good repeatability ensure a wide range of applications from fiscal measurement of

hydrocarbon liquids and gases through to process control. For fiscal oil-flow measurement they are often used with a dedicated meter prover.

3.5 Oscillatory flowmeters

Vortex meters are the most common type of oscillatory meter. They provide good process control and are used in both liquid and gas applications. They are affected by flow profile and pulsations in the flow line. Vortex meters can be characterized by Reynolds number: this feature makes the nature of the fluid less important and calibration possible in a different fluid from the one in the application. They have no moving parts. Pulsations at a frequency close to that of the shed vortices lead to large errors.

In very small sizes, fluidic meters find some specialist niche applications. Special designs for domestic gas and water meters have been produced and are in service.

3.6 Electromagnetic flowmeters

Restricted to conductive fluids, electromagnetic meters are used extensively for water-flow measurement. Their very wide rangeability (up to 1000 : 1) and apparent long life without significant drift makes them the first choice for water distribution systems. Electromagnetic flowmeters are relatively unaffected by flow disturbances.

3.7 Ultrasonic flowmeters

In its various forms, the ultrasonic meter is the meter that is being most rapidly improved at present. Currently, multipath meters provide fiscal, custody transfer or allocation measurements in both oil and gas. Single-path clamp-on meters often provide the only solution to nonintrusive measurement but cannot be expected to give better than 2 to 5% uncertainty. Owing to their principle of operation, ultrasonic flowmeters can be affected by acoustic noise generated by valves or pipe fittings.

3.8 Mass flowmeters

Coriolis meters provide direct mass measurement of most liquids. The inclusion of density measurement means that volume also is measured. They also measure gas but perform best when the pressure (density) is high. Achieving better than 0.5% uncertainty, this type of meter is starting to compete with the turbine meter and the positive-displacement meter for liquid applications in both 'easy' and difficult fluids. Long calibration intervals are claimed assuming verification of zero drift can be achieved in situ. No evidence has been shown that Coriolis meters are affected by flow-profile effects. When using Coriolis meters, it is vital to avoid the vibration of the meter by providing firm clamping and to avoid flow pulsations. Valves must be installed to allow zero setting.

3.9 Thermal flowmeters

Thermal meters can be divided into two main types: the thermal meters used to measure low gas flow rates and the insertion hot-wire anemometers usually used for research purposes.

Generally, capillary thermal mass flowmeters are not affected by skewed or swirling flows. In-line thermal mass flowmeters (ITMF), however, are sensitive to both skew and swirl and the manufacturer's recommendations for minimum distances upstream and downstream of sources of disturbance should be observed. ITMF meters are not sensitive to pipework vibrations, but, like all insertion meters, the insertion version should be firmly fixed to the pipe wall to prevent resonance.

3.10 Miscellaneous

It is possible using various techniques (pitot tubes, insertion electromagnetic meters, hot-wire anemometers, laser Doppler anemometers) to measure the velocity at different points across the pipe and then to integrate the measurements to give the flow rate; such techniques are called *velocity-area methods*.

Tracers can be added to the flowing fluid, and the measured velocity of the tracer between two points can be used to give the flow rate of the fluid.

Meter provers, syringes, and automatic weighing dispensers are some other devices that can be considered for specialist applications.

4 INSTALLATION ISSUES

Most flowmeters are designed to operate in ideal conditions. In practice, it is rarely possible to achieve these conditions and the way in which the meter is installed may generate a measurement error. Common sources of installation error include

- upstream and downstream disturbances to the flow caused by bends, valves or other fittings;
- flow pulsations;

- acoustic noise sources and vibration;
- particles or bubbles in the flow.

Flow disturbances are most important as they affect most types of flowmeters to a greater or a lesser degree. Other potential installation problems have been highlighted in Section 3 on a meter-by-meter basis. A high proportion of meters are installed incorrectly because they were an afterthought in the project design. It is good project design to involve the metering engineer early in the design process, as it can be very difficult and costly to modify the design later to allow for the correct installation of the meter.

Pipework fixtures disturb the flow by

- changing the distribution of flow in the pipe (distortion of the axial velocity profile);
- causing the flow in the pipe to swirl;
- changing the turbulence of the flow in the pipe.

Meters may be affected by one or more of these factors. In most cases, the percentage error generated by flow disturbances tends to be only slightly affected by Reynolds number or flow rate.

In general, disturbing elements that are in the same plane generate distorted axial velocity profiles without bulk rotation of the flow. If, for example, the fluid passes through a single bend, the flow will be skewed to one side of the pipe generating higher velocities near one wall. Contractions and expansions cause a flattened and a peaked velocity profile respectively. Rough pipework changes the axial velocity profile. Distortions of the axial profile usually become sufficiently small to give negligible effect on flow measurement within about 45 pipe diameters from the disturbance. In many cases, a much shorter length will be required.

Swirl, in the sense of bulk rotation of the flow, can be more persistent than distortion of the axial velocity profile. It can persist for 150 pipe diameters. When the pipe fittings are such that the flow changes direction twice in quick succession and in two different planes, swirl is often generated. The most typical situation in which swirl occurs is two 90° bends in planes at right angles to each other.

In nearly all cases of interest to engineers, the flow in pipes includes random turbulent fluctuations. Passing fluid through pipework fixtures effectively stirs up the flow, usually causing increased turbulence.

Work on the effect of upstream fittings has been carried out on all types of flowmeters. The most comprehensive program of work has been carried out on orifice plates. The fruit of this work is seen in ISO 5167-2 (International Organization for Standardization 2003a), which contains a table of straight lengths required downstream of particular fittings.

There are two problems with tables of straight length requirements such as the one in ISO 5167-2 (International Organization for Standardization 2003a): the user's fitting may not be included in the table, and the required length may be longer than that which is available. One solution to both problems is to use a flow conditioner. Instead of requiring the use of particular flow conditioners ISO 5167-1 (International Organization for Standardization 2003b) contains a compliance test so that, if a flow conditioner provides sufficiently small shifts in discharge coefficient downstream of certain flow disturbers (giving major distortion of the axial profile and strong swirl) and in good flow conditions, then it can be used with the same type of flowmeter downstream of any upstream fitting. The ranges of distances, expressed in terms of numbers of pipe diameters, between the flow conditioner and the flowmeter and between the upstream fitting and the flow conditioner, which are used in the tests, determine the acceptable ranges of distances when the flowmeter is used. The compliance test includes tests at two Reynolds numbers and in two pipe sizes. In the case of an orifice plate, the total upstream length from fitting to orifice plate can be reduced to 17 diameters by use of a Zanker Flow Conditioner Plate.

Standards and manufacturers' recommendations usually specify only a short length of straight pipe downstream of the meter as being necessary. Ten diameters will normally be sufficient and in some cases a much shorter length is necessary.

For information on testing and calibration of flowmeters, *see* **Article 190, Calibration and Standards in Flow Measurement, Volume 3**.

Further information on the selection and application of flowmeters can be found in Baker (2000), Miller (1996) and Spitzer (2001).

RELATED ARTICLES

Article 188, Differential Pressure (DP) Flowmeters, Volume 3; **Article 189, Basic Principles of Flow Measurement, Volume 3**; **Article 190, Calibration and Standards in Flow Measurement, Volume 3**.

REFERENCES

Baker, R.C. (2000) *Flow Measurement Handbook: Industrial Designs, Operating Principles, Performance and Applications*, Cambridge University Press, Cambridge.

British Standards Institution (1991) *Guide to Selection and Application of Flow Meters for the Measurement of Fluid Flow in*

Closed Conduits, British Standard BS 7405, British Standard Institution, London.

International Organization For Standardization (2003a) *Measurement of Fluid Flow by Means of Pressure Differential Devices Inserted in Circular Cross-Section Conduits Running Full – Part 2: Orifice Plates*, ISO 5167-2:2003, Geneva.

International Organization For Standardization (2003b) *Measurement of Fluid Flow by Means of Pressure Differential Devices Inserted in Circular Cross-Section Conduits Running Full – Part 1: General Principles and Requirements*, ISO 5167-1:2003, Geneva.

Miller, R.W. (1996) *Flow Measurement Engineering Handbook*, 3rd edn, McGraw-Hill, New York.

Spitzer, D.W. (ed.) (2001) *Flow Measurement: Practical Guides for Measurement and Control*, 2nd edn, ISA International, Research Triangle Park, Durham, NC.

188: Differential Pressure (DP) Flowmeters

Michael Reader-Harris
National Engineering Laboratory, Scotland, UK

1 Principles of Operation of DP Flowmeters 1238
2 Hardware and Software of DP Flowmeters 1239
3 Performance Overview of DP Flowmeters 1239
4 Design Issues of DP Flowmeters 1240
5 Selection of DP Flowmeters 1244
6 Installation Issues of DP Flowmeters 1245
7 Limitations of DP Flowmeters 1245
8 Errors in DP Flowmeters 1246
9 Standards Relevant to DP Flowmeters 1246
10 Testing of DP Flowmeters 1246
Related Articles 1246
References 1246

1 PRINCIPLES OF OPERATION OF DP FLOWMETERS

Differential pressure (DP) flowmeters partially obstruct the flow and hence operate by creating a difference in static pressure between the upstream and downstream side of the device. If a converging section of pipe, Figure 1, typical of the entry cone of a Venturi tube, is considered and Bernoulli's Equation for incompressible flow is applied between two planes, then

$$p_1 + \tfrac{1}{2}\rho U_1^2 = p_2 + \tfrac{1}{2}\rho U_2^2 \tag{1}$$

where p, ρ, and U are the pressure, density, and mean velocity respectively and the subscripted 1 and 2 represent the upstream and downstream (throat) planes.

From conservation of mass for incompressible flow

$$q_m = \rho A_1 U_1 = \rho A_2 U_2 \tag{2}$$

where q_m is the mass flow rate and A is the area of each plane. The area ratio, m, and the diameter ratio, β, are defined by

$$\frac{A_2}{A_1} = m = \left(\frac{d}{D}\right)^2 = \beta^2 \tag{3}$$

where d and D are the throat and upstream pipe diameters.

From these relations, it can be shown that

$$q_m = \frac{1}{\sqrt{1-\beta^4}} \frac{\pi d^2}{4} \sqrt{(2\rho(p_1 - p_2))} \tag{4}$$

The expression $[1/\sqrt{(1-\beta^4)}]$ is known as the velocity of approach factor and has usually been denoted by E. $p_1 - p_2$ is the differential pressure, Δp.

Hence,

$$q_m = \frac{1}{\sqrt{1-\beta^4}} \frac{\pi d^2}{4} \sqrt{2\rho\Delta p} \tag{5}$$

assuming no losses occur and the moving fluid completely fills the pipe.

In reality, some loss will occur and the equation is multiplied by the discharge coefficient C to take this into account.

Thus,

$$q_m = \frac{C}{\sqrt{1-\beta^4}} \frac{\pi d^2}{4} \sqrt{2\rho\Delta p} \tag{6}$$

The product $[C/\sqrt{(1-\beta^4)}]$ is sometimes referred to as the *flow coefficient* of the meter and represented by α.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

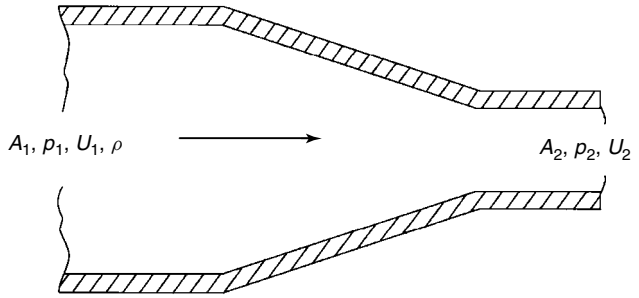


Figure 1. Converging flow.

In nozzles and Venturi tubes, the flow follows the boundary of the tube closely and the value of C is usually close to unity. However, in the case of an orifice plate, the flow continues to converge downstream of the plate forming a vena contracta. Bernoulli's Equation can be applied between an upstream plane and the vena contracta, but the area of the vena contracta cannot practically be measured and is thus not known accurately; therefore, in (6) the area of the orifice bore is used. This leads to a value of C of approximately 0.6, which, in effect, includes a coefficient of contraction.

If the fluid being metered is compressible, there will be a change in density when the pressure of the fluid falls from p_1 to p_2 on passing through the device. As the pressure changes quickly, it is assumed that no heat transfer occurs and because no work is done by or on the fluid, the expansion is isentropic. In nozzles and Venturi tubes, Bernoulli's Equation is applied, and an expansibility factor, ε, is calculated:

$$q_m = \frac{C}{\sqrt{1-\beta^4}}\varepsilon\frac{\pi d^2}{4}\sqrt{2\rho_1\Delta p} \tag{7}$$

where

$$\varepsilon = \left(\left(\frac{\kappa\tau^{2/\kappa}}{\kappa-1}\right)\left(\frac{1-\beta^4}{1-\beta^4\tau^{2/\kappa}}\right)\left(\frac{1-\tau^{(\kappa-1)/\kappa}}{1-\tau}\right)\right)^{1/2} \tag{8}$$

ρ_1 is the density at the upstream tapping, τ is the pressure ratio, p_2/p_1, and κ is the isentropic exponent. Because of the assumptions on which this equation is based, it has an uncertainty, which increases with $\Delta p/p_1$.

The theoretical values of ε do not apply to orifice plates because the effective area ratio is not readily obtained and the flowing jet can also expand laterally within the orifice. Hence, empirical values are used, and for all three recognized tapping arrangements (7) is used together with the equation

$$\varepsilon = 1 - (0.351 + 0.256\beta^4 + 0.93\beta^8)\left\{1 - \left(\frac{p_2}{p_1}\right)^{1/\kappa}\right\} \tag{9}$$

2 HARDWARE AND SOFTWARE OF DP FLOWMETERS

The chief types of differential pressure meters are

orifice plate	low-loss device (e.g. Dall tube)
nozzle and Venturi nozzle	inlet flowmeter
Venturi tube	

DP flowmeters may be installed either 'in-line', that is, with pipework both upstream and downstream of it, or at the inlet to a length of pipe. It is estimated that at least 40% of industrial flowmeters in use at present are differential pressure devices, with the orifice plate being the most popular. The orifice plate may be installed in an orifice fitting to enable the orifice plate to be changed easily, or it may be installed between flanges. Processing methods for data vary from the simple to the use of flow computers.

3 PERFORMANCE OVERVIEW OF DP FLOWMETERS

The main advantages of DP flowmeters are that

1. they are simple to make, containing no moving parts,
2. their performance is well understood,
3. they are cheap – especially in larger pipes when compared with other meters,
4. they can be used in any orientation,
5. they can be used for most gases and liquids,
6. they have generally good standards, based on years of research, and
7. some types do not generally require calibration.

A typical orifice plate designed and manufactured according to a recognized standard will measure flow with an uncertainty around 1% (or a little lower) in good conditions depending on its area ratio. Uncertainty can be reduced by calibration. Because of the square root relationship between differential pressure and flow rate, the relative uncertainty of the flow rate measurement increases at low flow rates, and the rangeability of a differential pressure meter is limited normally to 3 : 1 and to about 8 : 1 at most. The use of modern 'smart' pressure transmitters is increasing the range of the meter.

In some situations where brief interruption of the flow is permissible, a greatly extended flow range can be obtained by the use of multiple orifice plates. Interruption of the flow will not be necessary if the multiple meter runs downstream of a header are used or a suitable orifice fitting is employed.

4 DESIGN ISSUES OF DP FLOWMETERS

4.1 Orifice plate

There are many forms of orifice plates in the sense that various profiles of the actual orifice are used and alternative positions of pressure tappings are permitted by ISO 5167-2 and ISO TR 15377. The normal profile used in orifice plates is the square edge, Figure 2. The downstream edge of a square-edged plate is beveled, unless the plate is thin.

Three alternative positions of the pressure tappings are permitted corner, flange, and D and $D/2$, *see* Figure 3. Their choice is largely a matter of local custom and engineering convenience, although experience indicates that D and $D/2$ tappings are generally less sensitive to disturbed flow than flange or corner tappings. In industry, especially in the United States, flange tappings are commonly used.

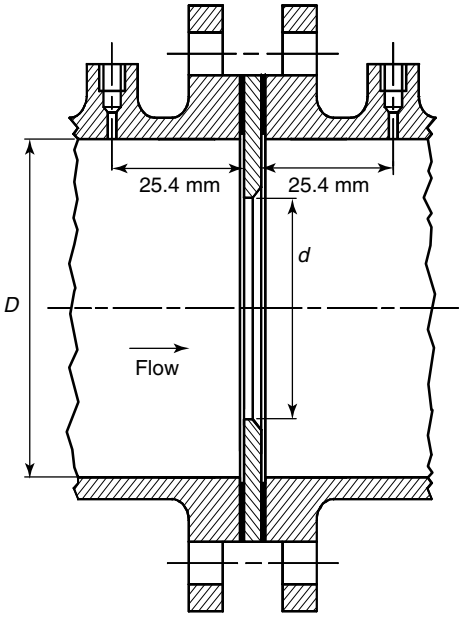


Figure 2. Orifice plate with flange tappings.

To measure pressure, it is generally better to use a piezometer ring or annular groove than a single tapping. A triple-tee arrangement (Figure 4) has been shown to be the most satisfactory, especially where flow disturbances exist (Blake, 1976).

The outstanding advantage of a square-edged orifice plate is that, if designed and manufactured according to ISO 5167-2, the coefficient can be predicted by an internationally accepted equation based on a wide range of good data, and this removes the need for calibration. No other type of flowmeter has quite this capability.

A very large international program of work led to the creation of a database of orifice plate discharge coefficients: the diameter ratios range from 0.1 to 0.75, orifice Reynolds numbers from 1700–7×10^7, and pipe diameters from 50 to 600 mm. As a result of this work, the Stolz Equation for the discharge coefficient in ISO 5167-1:1991 was replaced by the Reader-Harris/Gallagher (1998) Equation first in ISO 5167-1:1991/Amd 1:1998 and then in ISO 5167-2:2003. The Reader-Harris/Gallagher (1998) Equation is as follows:

$$C = 0.5961 + 0.0261\beta^2 - 0.216\beta^8 + 0.000521\left(\frac{10^6\beta}{Re_D}\right)^{0.7} + (0.0188 + 0.0063A)\beta^{3.5}\left(\frac{10^6}{Re_D}\right)^{0.3} + (0.043 + 0.080e^{-10L_1} - 0.123e^{-7L_1})(1 - 0.11A)\frac{\beta^4}{1-\beta^4} - 0.031(M_2' - 0.8M_2'^{1.1})\beta^{1.3} \quad (10a)$$

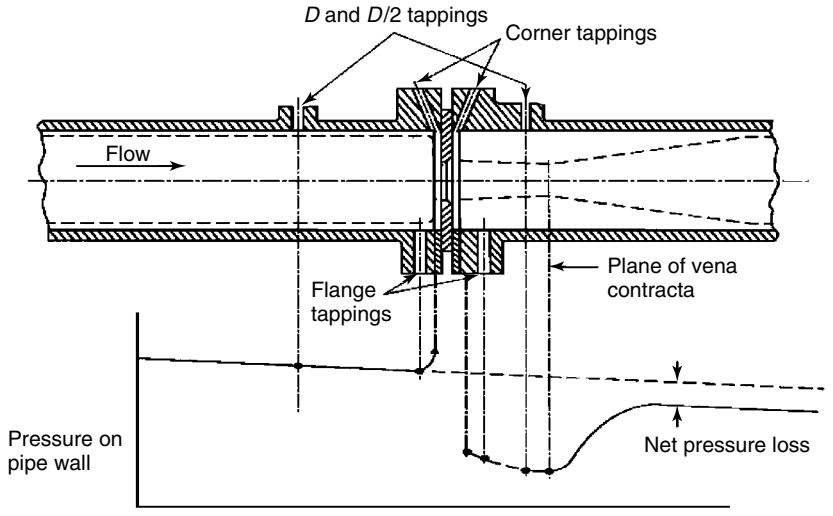


Figure 3. Tapping locations and pressure profile upstream and downstream of an orifice plate.

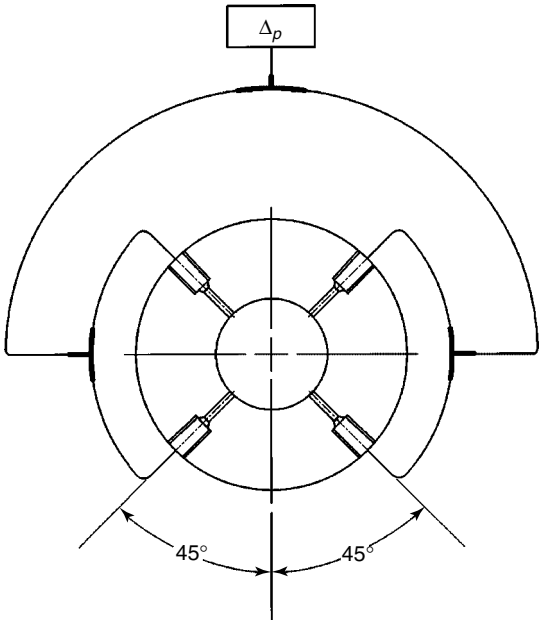


Figure 4. Triple-tee arrangement.

Where $D < 71.12$ mm (2.8 inch) the following term should be added to the above Equation:

$$+0.011(0.75 - \beta)\left(2.8 - \frac{D}{25.4}\right)(D : \text{mm}) \qquad (10b)$$

In this Equation,

Re_D is the Reynolds number related to D;
$L_1 = l_1/D$ is the quotient of the distance of the upstream tapping from the upstream face of the plate and the pipe diameter;
$L'_2 = l'_2/D$ is the quotient of the distance of the downstream tapping from the downstream face of the plate and the pipe diameter;

$$M'_2 = \frac{2L'_2}{1 - \beta}; \qquad (11)$$

$$A = \left(\frac{19000\beta}{Re_D}\right)^{0.8}$$

The limits of use in terms of Reynolds numbers and pipe roughness are given in ISO 5167-2. A similar but not identical equation is included in API 14.3 Part 1.

Square-edged orifice plates and nozzles may be used in pipes smaller than 50-mm diameter: this is covered by ISO TR 15377, which also includes conical-entrance, quarter-circle and eccentric orifice plates. Conical-entrance and quarter-circle plates are used especially for viscous flow. For special purposes, eccentric or even noncircular orifice plates are used; for example, in metering suspended solids, a chord-type orifice plate can be used.

Further guidance on the use of pressure differential devices is given in other standards (*see* Section 9).

4.2 Venturi tubes

The profile of the classical Venturi tube is shown in Figure 5. Although the included angle of the conical divergent section can be between 7 and 15°, it is recommended that an angle of 7 to 8° be chosen where possible. Various forms of construction are employed, depending upon size, but all are considerably more expensive than the orifice plate. When the outlet diameter of the divergent section is less than the upstream pipe diameter, a Venturi tube is called *truncated*. The divergent portion can be truncated by about one-third of its length without greatly affecting the pressure loss in the device.

The discharge coefficient of a Venturi tube with a machined convergent is typically 0.995. Because it has pressure tappings in the throat, a Venturi tube is much more affected by burrs on the tappings than an orifice plate.

There is an increasing desire to use Venturi tubes in gas flows and especially in wet gas. However, on undertaking the calibrations it has been found that on many occasions discharge coefficients significantly greater than 1 have been found (e.g. $C \approx 1.02$). This problem has been described by Jamieson *et al*. (1996). Part of the cause of this problem is static hole error. A full understanding of the problem is a matter of current research (*see* Reader-Harris *et al*., 2001).

4.3 Nozzles

The nozzle (see ISO 5167-3) has a curved entry and cylindrical throat, but no divergent outlet section (Figure 6). Therefore, the discharge coefficient is similar to that of a Venturi tube. However, the pressure loss is the same as that of an orifice plate for the same flow rate in the same size of pipe at the same differential pressure.

There are two types of standard nozzles, the ISA 1932 nozzle and the long-radius nozzle. The profile of the ISA

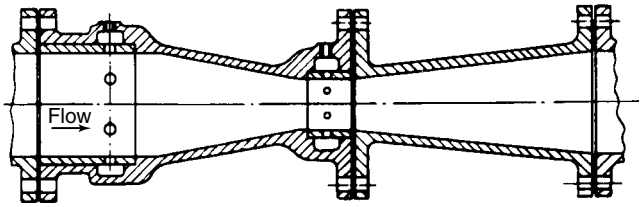


Figure 5. Venturi tube. (*Source*: Extracts from British Standards reproduced with the permission of BSI under licence number 2002SK/0147.)

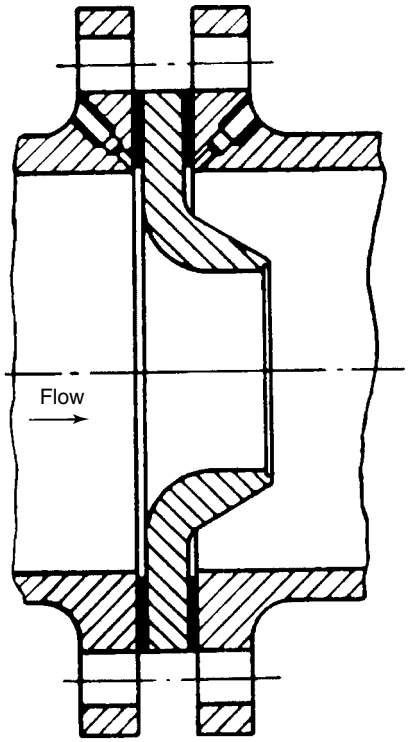


Figure 6. Nozzle mounted between pipe. (*Source*: Extracts from British Standards reproduced with the permission of BSI under licence number 2002SK/0147.)

1932 nozzle depends on whether the throat diameter d is less or greater than two-thirds the pipe diameter D. Corner pressure tappings are usually used with this nozzle, which may be mounted between pipe flanges or in a carrier.

Two standard designs of long-radius nozzles are used according to the value of the diameter ratio β. The upstream pressure tapping is about one pipe diameter from the inlet face. The downstream tapping is half a pipe diameter from the inlet face unless the nozzle is less than $D/2$ in length in which case it must be no further downstream than the nozzle outlet.

Long-radius nozzles with throat tappings have been standardized in ASME PTC-6. This standard is commonly used in the electrical power generation industry.

In order to reduce the pressure loss caused by a nozzle it can be fitted with a divergent section similar to that used for a Venturi, hence becoming a Venturi nozzle. The upstream face is identical with an ISA 1932 nozzle. Upstream corner and throat pressure tappings are used, Figure 7.

4.4 Low-loss meters

Given the importance of energy conservation there is every incentive to develop devices with a low pressure loss, and yet to retain the basic advantages of differential pressure meters.

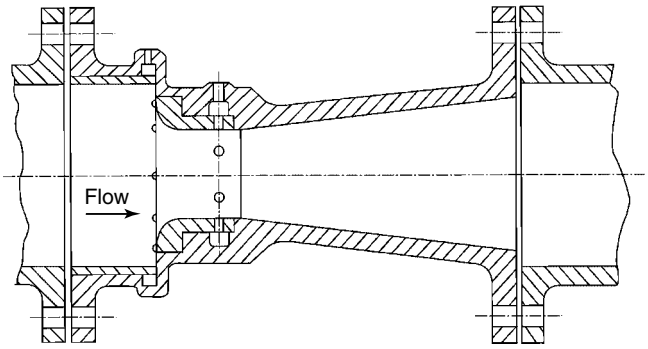


Figure 7. Venturi nozzle (with truncated divergent section). (*Source*: Extracts from British Standards reproduced with the permission of BSI under licence number 2002SK/0147.)

Many variations of low-loss meters have been invented, but only a few have been exploited commercially. The main problem perhaps is that to be competitive with conventional pressure differential devices, an immense amount of test data would need to be accumulated to match the installation experience on the latter.

One of the earliest such developments was probably the Dall tube (Dall, 1960), Figure 8. It comprises two cones, each with a substantial included angle, between which is a circumferential slot. The abrupt change of boundary contour results in flow curvature, which increases the differential pressure produced. The sudden reduction in section immediately downstream of the upstream pressure tapping gives a local pressure increase, which also augments the differential pressure produced by this device.

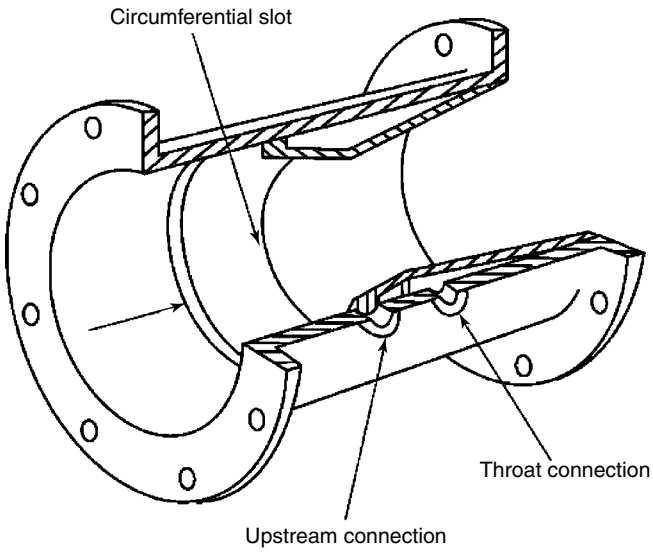


Figure 8. Dall tube.

Other low-loss devices are the Dall orifice, virtually a shortened version of the Dall tube intended for insertion between flanges, the Epiflo device (Lewis, 1975), which features a cylindrical ring protruding upstream and a conical diffuser downstream to obtain maximum pressure recovery, and the Gentile Flow Tube (Gentile, 1950). The Gentile Flow Tube is about $1.5D$ long and has two groups of pressure sensing holes, one group pointing upstream, the other downstream, and can measure flow in either direction since it is symmetrical.

4.5 Inlet flowmeters

Four devices are commonly used as a flowmeter at the inlet to a pipe: the nozzle, bellmouth entry, conical inlet, and Borda inlet. Such flowmeters are most commonly used for airflow measurements; although there is no reason in theory why they should not be used with liquid flows, there are, in practice, few situations in which they are appropriate.

In every case, the 'upstream' pressure is simply that in the reservoir (normally the atmosphere) from which the air is drawn into the pipe. Details of the dimensions and construction of the devices are given in Figure 9.

The use of *nozzles* at an inlet is described in Jaumotte and Nagel (1976). Alternatively, an ISA or a Venturi nozzle can be used in accordance with ISO TR 15377.

The design of *bellmouth entries* is not standardized, but any long-radius curved inlet will act as a good flowmeter provided that the walls are smooth, and will have a discharge coefficient of about 0.99.

The *conical inlet* has been standardized (BS 848) and has been shown (Whitaker, 1968) to have a discharge coefficient of about 0.96, although this value can vary by $\pm 2\%$ even with carefully made devices.

The *Borda inlet* has the advantages that it is easy and cheap to make, and generates about four times the differential pressure produced by a conical or a bellmouth entry with the same throat diameter (the diameter of the 'throat' being that of the downstream pipe). It does, however, have

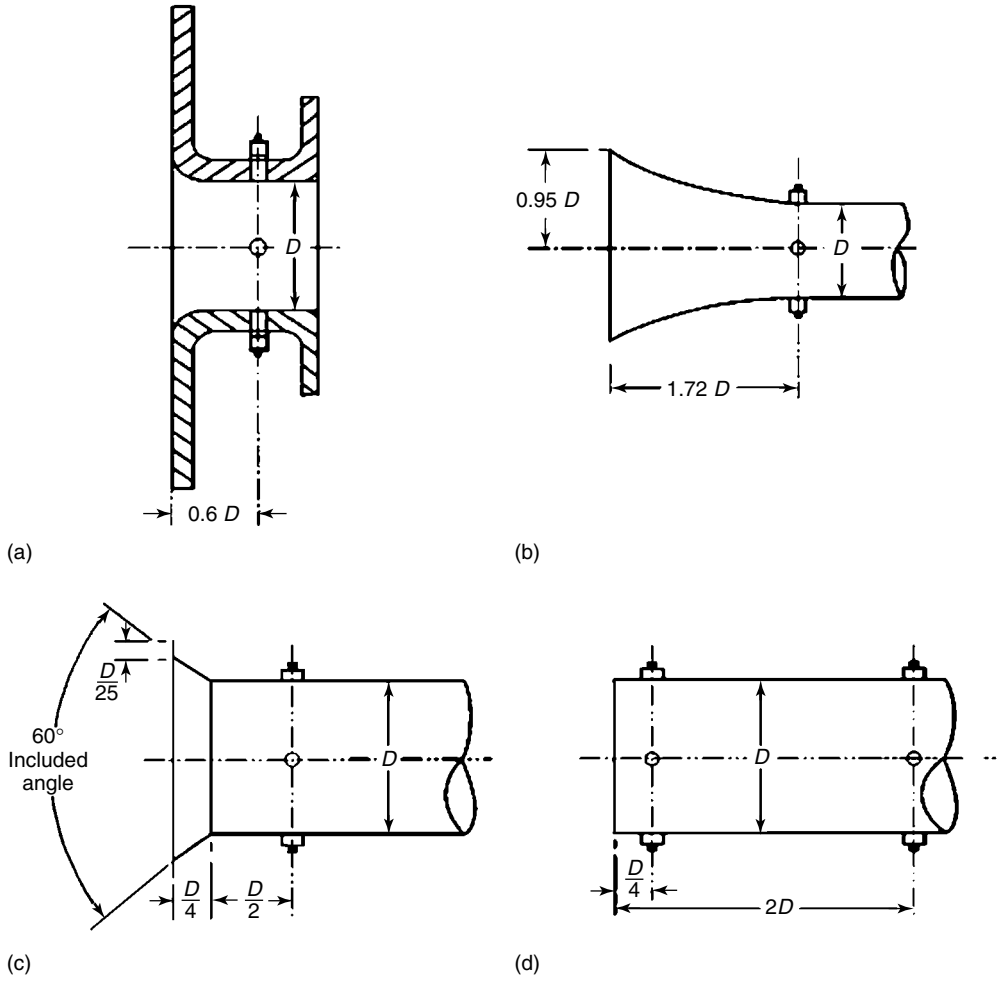


Figure 9. Inlet flowmeters: (a) inlet nozzle, (b) bellmouth entry, (c) conical inlet, and (d) Borda inlet.

a higher head loss – about 0.5 times the velocity head. The discharge coefficient is approximately 0.5, and details of the design and use of Borda inlets are given in Kinghorn and McHugh (1978).

When using inlet flowmeters, care has to be taken that the area from which the air is drawn is free from draughts or crosswinds, since otherwise significant errors can occur (Kinghorn, McHugh and Crawford, 1975).

4.6 V-Cone flowmeter

A modern patented differential flowmeter that has found significant use is the McCrometer V-Cone flowmeter (Figure 10). Its advantages include the fact that it requires a shorter straight length upstream than an orifice plate (Ifft, 1995).

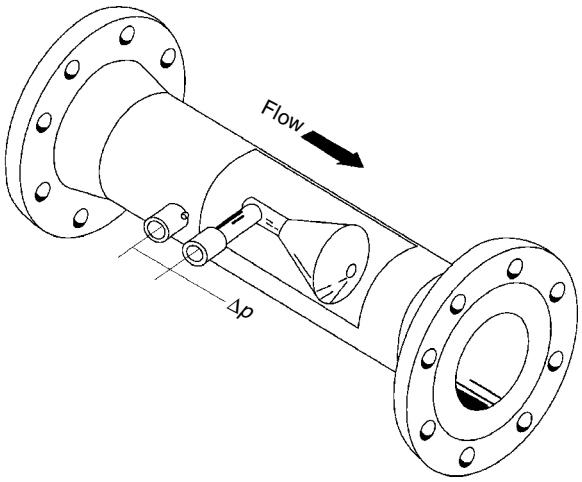


Figure 10. V-Cone flowmeter.

4.7 The spring-resisted, differential pressure, variable-area flowmeter

Because of the square law relationship between differential pressure and flow rate, one option to extend the range of differential pressure flowmeters is the spring-resisted, differential pressure, variable-area flowmeter, Figure 11. The cone takes up a position of equilibrium under the combined effects of the flow, the spring, and the area of the orifice. By suitable design of the profile to match the choice of spring, the differential pressure can be made a linear (or even some other) function of flow rate.

5 SELECTION OF DP FLOWMETERS

An orifice plate is probably the least expensive differential pressure flowmeter, and it has the great advantage that it is normally sufficient to manufacture it in accordance with ISO 5167-2 without calibrating it in a flowing fluid. If the flow rate changes after installation, then it may be possible to change the plate for one of a different diameter ratio very simply and inexpensively.

One advantage of a nozzle over an orifice plate is that there is no sharp edge to erode, but it is more expensive to manufacture and more difficult to install and to remove from the pipe.

If it is important to minimize the length required upstream of the meter, Venturi tubes will have an advantage over orifice plates.

The pressure loss in a system is shown in Figure 12. Therefore, if it is important to reduce the pressure loss a Venturi tube or, for even lower loss, a low-loss meter such as the Dall tube will be desirable.

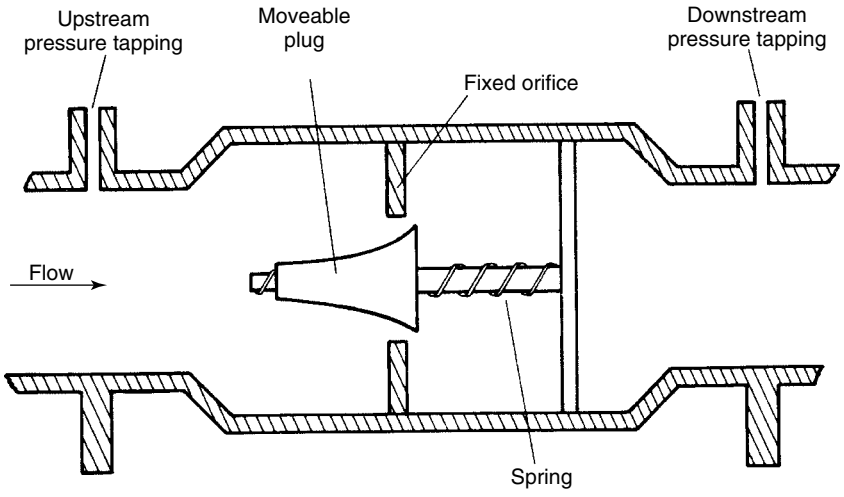


Figure 11. Variable-area differential pressure meter.

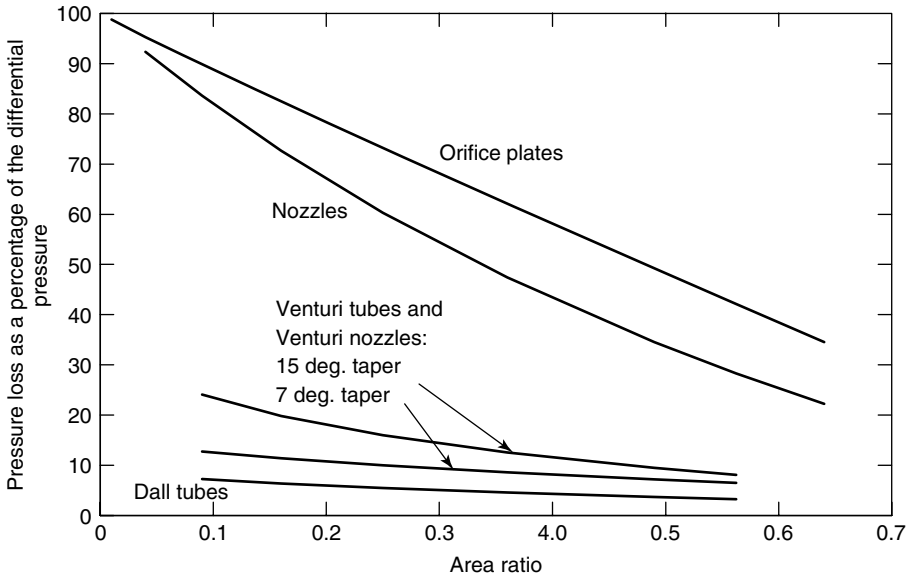


Figure 12. Pressure loss as a percentage of the differential pressure.

Throat-tapped devices may have an unexpected performance at high Reynolds number; so it is desirable to calibrate them or to use an orifice plate.

In wet gas, Venturi tubes may have advantages because of their shape.

6 INSTALLATION ISSUES OF DP FLOWMETERS

DP flowmeters are significantly affected by upstream installation. The standards state the lengths of straight pipe required between specified upstream fittings and the flowmeters. Generally, the lengths increase with diameter ratio. In recent years, in addition to the work on the orifice plate discharge coefficient, there has also been a significant amount of work on installation effects on orifice plates: data have been collected on the effect of single 90° bends, two 90° bends in the same plane (S-configuration) and in perpendicular planes, single 45° bends, two 45° bends in the same plane (S-configuration), a tee, reducers, expanders, a gate valve, and headers. A large part of the data is in Studzinski *et al*. (1997). These data have helped to determine the required upstream lengths in ISO 5167-2. The straight lengths required upstream of a Venturi tube have also been recently reevaluated (Reader-Harris, Brunton and Sattary, 1997). The lengths are much shorter than those required upstream of an orifice plate, but longer than those in ISO 5167-1:1991. Revised lengths have been included in ISO 5167-4:2003.

In many cases, it is possible to use a flow conditioner to reduce the required upstream straight length. ISO 5167-1:2003 contains a compliance test for flow conditioners: this provides that if in good flow conditions, in strong asymmetry and in strong swirl, a flow conditioner gives a sufficiently small shift in discharge coefficient from a baseline in a long straight pipe, then the flow conditioner can be used with any upstream fitting, given the same lengths upstream and downstream of the flow conditioner as in the compliance test.

Rough pipe and steps in the upstream pipe diameter also affect the discharge coefficient and again the standards contain limits which have been revised in the case of orifice plates in the light of recent data.

7 LIMITATIONS OF DP FLOWMETERS

The main disadvantages of DP flowmeters are

1. rangeability (turndown, dynamic range) – *see* **Article 57, Static Considerations of General Instrumentation, Volume 1** – is less than for most other types of flowmeter,
2. significant pressure loss may occur,
3. output signal is nonlinear with flow,
4. coefficient and accuracy may be affected by pipe layout or nature of flow, and
5. they may suffer from aging effects, that is, the buildup of deposits or erosion of sharp edges.

8 ERRORS IN DP FLOWMETERS

Errors in DP flowmeters can be due to poor manufacture and consequent failure to meet the standard and thus to obtain the predicted discharge coefficient, to poor installation, and to the effects of aging. The effects of some errors are described in ISO TR 12767; the effect of pulsation is described in ISO TR 3313.

As time goes by, gradual changes to DP flowmeters may occur, which eventually cause significant errors to be introduced unless remedial measures are taken. The sharp edge of an orifice plate may be eroded away causing the discharge coefficient to rise. The convergent section and throat of a nozzle or Venturi tube may roughen resulting in a lower discharge coefficient. Unfortunately, no general rules for the rate of deterioration are available; each installation must be assessed according to its own set of circumstances and renewals or recalibrations made as appropriate.

9 STANDARDS RELEVANT TO DP FLOWMETERS

9.1 ISO standards

ISO TR 3313:1998. Measurement of fluid flow in closed conduits – guidelines on the effects of flow pulsations on flow measurement instruments.

ISO TR 9464:1998. Guide to the use of ISO 5167-1:1991.

ISO TR 12767:1998. Measurement of fluid by means of pressure differential devices – guidelines to the effect of departure from the specifications and operating conditions given in ISO 5167-1.

ISO TR 15377:1998. Measurement of fluid flow by means of pressure differential devices – guidelines for specification of nozzles and orifice plates beyond the scope of ISO 5167-1.

ISO 5167-1:2003. Measurement of fluid flow by means of pressure differential devices inserted in circular cross-section conduits running full – Part 1: General principles and requirements.

ISO 5167-2:2003. Measurement of fluid flow by means of pressure differential devices inserted in circular cross-section conduits running full – Part 2: Orifice plates.

ISO 5167-3:2003. Measurement of fluid flow by means of pressure differential devices inserted in circular cross-section conduits running full – Part 3: Nozzles and Venturi nozzles.

ISO 5167-4:2003. Measurement of fluid flow by means of pressure differential devices inserted in circular cross-section conduits running full – Part 4: Venturi tubes.

9.2 British standards

BS 848: Part 1:1980. Fans for general purposes. Part 1: Methods of testing performance.

9.3 API standards

API 14.3 Part 1:1990. Manual of Petroleum Measurement Standards, Chapter 14 – Natural Gas Fluids Measurement, Section 3 – Concentric Square-Edged Orifice Meters, Part 1 – General Equations and Uncertainty Guidelines.

9.4 ASME standards

ASME PTC-6:1996. Performance Test Code 6 on Steam Turbines.

10 TESTING OF DP FLOWMETERS

Orifice plates are rarely individually calibrated. Nozzles and Venturi tubes may be calibrated. Other DP flowmeters probably require calibration. It is necessary to have the same Reynolds number for the calibration as for the use of the meter. It is not necessary to have the same fluid. It is possible to calibrate in water a device for use in low-pressure air. Calibrations for use in gas generally assume that the expansibility factor equation is correct and determine the discharge coefficient. For some DP flowmeters, it may be necessary to determine the expansibility factor equation first.

General texts on flow have material on DP flow metering (Liptak, 1993; Miller, 1996; Spitzer, 2001; Baker, 2000).

RELATED ARTICLES

Article 187, Flowmeter Selection and Application, Volume 3; **Article 189, Basic Principles of Flow Measurement, Volume 3**; **Article 190, Calibration and Standards in Flow Measurement, Volume 3**.

REFERENCES

Baker, R.C. (2000) *Flow Measurement Handbook: Industrial Designs, Operating Principles, Performance, and Applications*, Cambridge University Press.

Blake, K.A. (1976) The Design of Piezometer Rings. *Journal of Fluid Mechanics*, **78**(2), 415–428.

Dall, H.E. (1960) Flow Tubes and Non-Standard Devices for Flow Measurement with Some Coefficient Considerations, *Symposium on Flow Measurement in Closed Conduits*, Vol. 2, National Engineering Laboratory, East Kilbride (pp. 385–394), Published by HMSO in 1962.

Gentile, V. Jr. (1950) Theory and Application of the Flow Tube, in *Proceedings of Tenth Annual Appalachian Gas Measurement Short Course*, Tech. Bulletin 31, Engineering Experiment Station, West Virginia University, Morgantown, WV.

Ifft, S.A. (1995) Installation Effects on the V-Cone Flowmeter, *Proceedings of 3rd International Symposium on Fluid Flow Measurement*, Gas Research Institute, San Antonio, TX; Chicago, IL.

Jamieson, A.W., Johnson, P.A., Spearman, E.P. and Sattary, J.A. (1996) Unpredicted Behaviour of Venturi Flowmeter in Gas at High Reynolds Numbers, *Proceedings of 14th North Sea Flow Measurement Workshop*, Paper 5, National Engineering Laboratory, East Kilbride.

Jaumotte, A. and Nagel, P. (1976) Measurement of flowrate at the inlet to a pipe using standardised nozzles (in French). *Promoclim E. (Etudes Thermiques et Aeroliques)*, Tome 7E-No-4, 246–262.

Kinghorn, F.C. and McHugh, A. (1978) The Performance Characteristics of Borda Inlet Air Flowmeters. *Journal of Physics E: Scientific Instruments*, **11**, 87–93.

Kinghorn, F.C., McHugh, A. and Crawford, D.A. (1975) *Effect of Crosswind on Three Types of Differential Inlet Flowmeter*, NEL Report No 605, National Engineering Laboratory, East Kilbride.

Lewis, D.C.G. (1975) Further Development of a Low-Loss Flow Metering Device (Epiflo) Based on the Pressure Difference Principle, *Proceedings of Conference on Fluid Flow Measurement in the mid 1970s*, Paper J3, National Engineering Laboratory, East Kilbride (pp. 633–644).

Liptak, B.G. (1993) *Flow Measurement*, CRC Press.

Miller, R.W. (1996) *Flow Measurement Engineering Handbook*, McGraw-Hill Professional.

Reader-Harris, M.J., Brunton, W.C. and Sattary, J.A. (1997) Installation Effects on Venturi Tubes, *Proceedings of ASME Fluids Engineering Division Summer Meeting*, Vancouver, Canada, FEDSM97-3016, American Society of Mechanical Engineers, New York.

Reader-Harris, M.J., Brunton, W.C., Gibson, J.J., Hodges, D. and Nicholson, I.G. (2001) Discharge Coefficients of Venturi Tubes with Standard and Non-Standard Convergent Angles. *Flow Measurement and Instrumentation*, **12**, 135–145.

Spitzer, D.W. (2001) *Flow Measurement: Practical Guides for Measurement and Control*, ISA – The Instrumentation, Systems, and Automation Society.

Studzinski, W., Karnik, U., LaNasa, P., Morrow, T., Goodson, D., Husain, Z. and Gallagher, J. (1997) *White Paper on "Orifice Meter Installation Configurations with and without Flow Conditioners"*, American Petroleum Institute, Washington, DC.

Whitaker, J. (1968) *Calibrations of British Standard Conical Inlets*, NEL Report No 349, National Engineering Laboratory, East Kilbride.

189: Basic Principles of Flow Measurement

Richard Thorn
University of Derby, Derby, UK

1 Introduction 1248
2 Basic Terminology 1248
3 Characteristics of Fluid Flow 1249
4 Flowmeter Classification 1249
5 Basic Principles of Single-phase Flowmeters 1250
6 Multiphase Flow Measurement 1255
Related Articles 1255
References 1255
Further Reading 1255

1 INTRODUCTION

Flow measurement is something that nearly everyone has experience of. Everyday examples include the metering of household utilities such as water and gas. Similarly, flowmeters are used in nearly every sector of industry, from petroleum through to food manufacture and processing. It is therefore not surprising that today the total world flowmeter market is worth over $3000 million and expected to continue growing steadily in the future.

However, what is surprising, given the undoubted importance of flow measurement to the economy, is the accuracy and technology of the most commonly used flowmeters that are poor and relatively old fashioned in comparison to instruments used to measure other measurands such as pressure and temperature. For example, the orifice plate flowmeter, which is still one of the most frequently used flowmeters in the process industry, only has a typical accuracy of $\pm 2\%$, and was first used commercially in the late-1800s. The conservative nature of the flow measurement industry means that traditional techniques such as the orifice plate, Venturimeter, and variable area flowmeter still dominate, while ultrasonic flowmeters, which were first demonstrated in the 1950s, are still considered by many to be 'new' devices. This article will consider the principles of operation of the most commonly used commercially available methods of flow measurement. For recent research developments in flow measurement, see the journal Flow Measurement and Instrumentation.

2 BASIC TERMINOLOGY

The term flow measurement is a general term, and before selecting a flowmeter, it is important to be sure what type of flow measurement is actually required. For a fluid flowing through a pipe, flow measurement may mean any one of six different types of measurements.

1. Point velocity measurement – the fluid's velocity at a fixed point across the pipe's cross section ($\mathrm{m\,s^{-1}}$)
2. Mean flow velocity measurement – average fluid velocity across the cross section of the pipe ($\mathrm{m\,s^{-1}}$)
3. Volumetric flowrate measurement – the rate of change in the volume of fluid passing through the pipe with time ($\mathrm{m^3\,s^{-1}}$)
4. Total volume measurement – the total volume of fluid that has passed through the pipe ($\mathrm{m^3}$)
5. Mass flowrate measurement – the rate of change in the mass of the fluid passing through the pipe with time ($\mathrm{kg\,s^{-1}}$)
6. Total mass measurement – the total mass of fluid passing through the pipe ($\mathrm{kg\,s^{-1}}$).

Although the most common type of flow measurement is that of a fluid through a closed conduit or pipe, open channel flow measurements are also regularly needed in

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

applications such as sewage and water treatment. For further information on open channel flow measurement techniques, see Grant (2001).

3 CHARACTERISTICS OF FLUID FLOW

The fluid being metered is usually a liquid or gas and is known as *single-phase flow*. However, there is an increasing need for the flowrate of multiphase mixtures to be measured (See Section 6).

There are a number of important principles relating to the characteristic of flow in a pipe that should be understood before a flowmeter can be selected and used with confidence. These are the meaning of Reynolds number and the importance of the flow's velocity profile (Mulley, 2004).

The Reynolds number *Re* is the ratio of the inertia forces in the flow ($\rho \bar{v} D$) to the viscous forces in the flow (η), and it can be used to determine whether a fluid flow is laminar or turbulent in nature. Reynolds number can be calculated using

$$Re = \frac{\rho \bar{v} D}{\eta} \tag{1}$$

where ρ is the density of the fluid, $\bar{v}$ is the mean velocity of the fluid, D is the pipe diameter, and η is the dynamic viscosity of the fluid. If *Re* is less than 2000, viscous forces in the flow dominate and the flow will be laminar. If *Re* is greater than 4000, inertia forces in the flow dominate and the flow will be turbulent. If *Re* is between 2000 and 4000, the flow is transitional and either mode can be present. The Reynolds number is mainly calculated using properties of the fluid and does not take into account factors such as pipe roughness, bends, and valves, which also affect the flow characteristic. However, the Reynolds number is a good guide to the type of flow that might be expected in most situations.

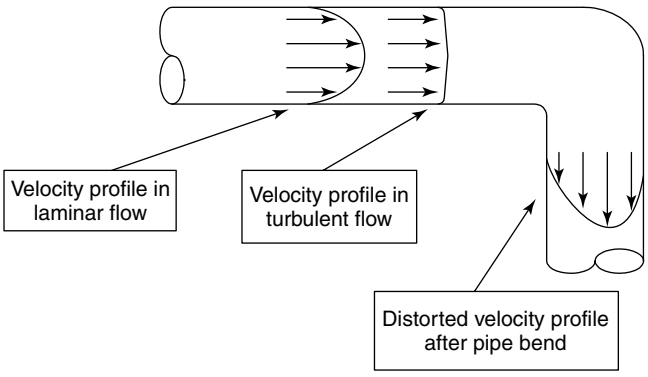


Figure 1. Flow velocity profiles in laminar and turbulent flow.

The fluid velocity across a pipe's cross section is not constant and depends on the type of flow present (Figure 1). In laminar flow, the velocity at the center of the pipe is twice the average velocity across the pipe cross section, and the flow profile is unaffected by the roughness of the pipe wall. In turbulent flow, pipe wall effects are less and the flow's velocity profile is flatter, with the velocity at the center being about 1.2 times the mean velocity. The exact flow profile in a turbulent flow depends on pipe wall roughness and Reynolds number. In industrial applications, laminar flows are rarely encountered unless very viscous fluids are being metered. The pipe Reynolds number should always be calculated since some flowmeters are not suitable for use in both laminar and turbulent flow conditions.

A flow's velocity profile will only be symmetrical at the end of a very long pipe. Bends and obstructions such as valves will cause the profile to become distorted or asymmetric. Since the calibration of many flowmeters is sensitive to the velocity profile of the flow passing through the meter, in order to have confidence in the performance of a flowmeter, the velocity profile of the flow passing through the flowmeter should be stable and known.

4 FLOWMETER CLASSIFICATION

Although there at least 80 different types of commercially available flowmeter, they may all be classified into 10 main groups (British Standards Institution, 1991) – Differential pressure, Variable Area, Positive Displacement, Turbine, Oscillatory, Electromagnetic, Ultrasonic, Mass, Thermal, and Miscellaneous (*see* **Article 187, Flowmeter Selection and Application, Volume 3**).

Traditional flow measurement technologies are represented by the differential pressure, variable area, positive displacement, and turbine categories. Newer techniques are represented by the electromagnetic, ultrasonic, oscillatory, and mass categories. Although differential pressure flowmeters are still the most commonly used method of flow measurement, especially in the process industrial sector, in general, traditional methods are being increasingly replaced by newer techniques. These techniques are now often preferred because, in most cases, they do not obstruct the flow and yet match many of the traditional flowmeters in terms of accuracy and reliability.

The following sections will consider the principles of operation, advantages, and limitations of the most popular types of flowmeters. For information on the selection and application of these flowmeters, *see* **Article 187, Flowmeter Selection and Application, Volume 3**. For further information on flowmeters, see one of the many textbooks

on flow measurement such as Baker (2000), Miller (1996), Spitzer (2001), and Webster (2000).

5 BASIC PRINCIPLES OF SINGLE-PHASE FLOWMETERS

5.1 The differential pressure flowmeter

Over 40% of all liquid, gas, and steam flow measurements made in industry are still accomplished using the common types of differential pressure flowmeter, that is, the orifice plate, Venturi tube, and nozzle. The operation of these flowmeters is based on the observation made by Bernoulli that if an annular restriction is placed in a pipeline, then the velocity of the fluid through the restriction is increased. The increase in velocity at the restriction causes the static pressure to decrease at this section and a pressure difference to be created across the element. The difference between the pressure upstream and pressure downstream of this obstruction is related to the rate of fluid flowing through the restriction and therefore, of course, through the pipe. A differential pressure flowmeter consists of two basic elements – an obstruction to cause a pressure drop in the flow (a differential producer) and a method of measuring the pressure drop across this obstruction (a differential pressure transducer).

One of the major advantages of the orifice plate, Venturi tube, or nozzle is that the measurement uncertainty can be predicted without the need for calibration if it is manufactured and installed in accordance with one of the international standards covering these devices. In addition, this type of differential pressure flowmeter is simple, has no moving parts, and is therefore reliable. The main disadvantages of these devices are their limited range (typically 3:1), the permanent pressure drop they produce in the pipeline (which can result in higher pumping costs), and their sensitivity to installation effects (which can be minimized using straight lengths of pipe before and after the flowmeter). The combined advantages of this type of flowmeter are still quite hard to beat, and although it has limitations, these have been well investigated and can be compensated for in most circumstances. Unless very high accuracy is required, or unless the application makes a nonintrusive device essential, the differential flowmeter should be considered. Despite the predictions of its demise, there is little doubt that the differential pressure flowmeter will remain a common method of flow measurement for many years to come. For further information on this important class of flowmeter, *see* **Article 188, Differential Pressure (DP) Flowmeters, Volume 3**.

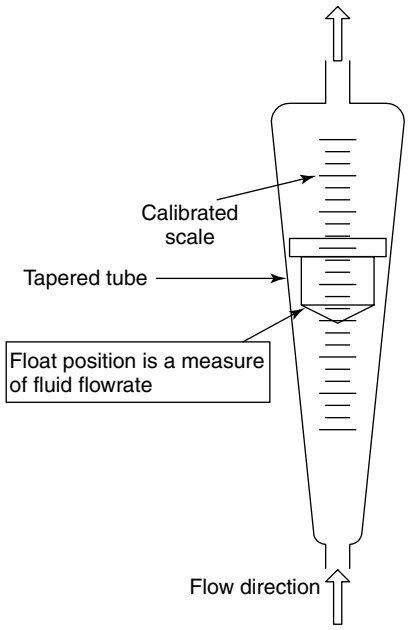


Figure 2. Tapered tube and float variable area flowmeter.

5.2 The variable area flowmeter

Variable area flowmeters are also based on using an obstruction in the flow to create a differential pressure, except, in this case, the differential pressure is constant and the area of the obstruction varies as the flowrate being measured changes. Probably the best-known type of variable area flowmeter is the taper tube and float flowmeter known almost universally as a *rotameter* (Figure 2). This type of flowmeter consists of a vertical tapered tube into which a float or bob is fitted. The fluid being metered enters the tube at the bottom and forces the float up the tube, which also increases the cross-sectional area available around the float for the fluid to pass through. Increasing the flowrate will move the float further up the tube, and so the position at which the float comes to rest is a direct function of flowrate.

Rotameters are extremely simple and reliable, and have an output, which changes linearly with flowrate (unlike differential pressure flowmeters) and a typical range of 10:1 (compared to 3:1 for differential pressure flowmeters). Accuracy is typically $\pm 2\%$ of full scale, but will depend on range and cost of the device. In addition, the flowmeter's calibration is insensitive to changes in the velocity profile of the flow. If the tube can be made of glass or clear plastic then a visual indication of flowrate is directly available and the flowmeter requires no external power supply in order to function. As a result, such flowmeters are commonly found in many process and laboratory applications where gases or liquids need to be metered. If high temperature,

high pressure or corrosive fluids need to be metered then the rotameter's tube can be made of metal. In such cases, a mechanism for detecting and displaying the position of the float is required.

A major limitation of the rotameter is that it can usually only be used vertically and so causes installation difficulties if the pipeline being metered is horizontal. Some manufacturers produce spring-loaded rotameters, which can be used in any position; however, in general, these have poorer accuracy than standard rotameters. Other limitations are that the calibration of the meter is dependent on the viscosity and density of the fluid being metered, and producing an electrical output signal suitable for transmission requires extra complexity. However, the use of optical or magnetic limit switches to enable the flowmeter to be used in high or low flow alarm applications is common.

5.3 The positive displacement flowmeter

Positive displacement flowmeters are based on a simple measurement principle. The flow being measured is 'displaced' or moved from the inlet side of the flowmeter to the outlet side using a series of compartments of known volume. The number of compartments of fluid that have been transferred are counted to determine the total volume that has passed through the flowmeter, and if time is also measured then volumetric flowrate can be measured. There are many designs of positive displacement flowmeter commercially available. For liquids, the most common designs are piston, sliding vane, oval gear, birotor, trirotor, and disc types of flowmeter, while for gases, roots, bellows (or diaphragm), or CVM flowmeters are popular. Despite this wide range of design, all are based on the same principle and all are predominantly mechanical devices.

The advantages of positive displacement flowmeters are that they are capable of high accuracy measurement (typically $\pm 0.5\%$ of reading for liquids and $\pm 1\%$ of reading for gases) over a wide range of flowrates. They can be used to meter fluids with a wide range of viscosity and density. In addition, unlike most other flowmeters, they are insensitive to changes in flow velocity profile and so do not require long lengths of straight pipe work before and after the flowmeter.

Figure 3 shows the principle of the oval gear flowmeter and illustrates the limitations of positive displacement flowmeters. They are relatively complex mechanical devices with moving parts, which of course wear with time. Their measurement accuracy depends both on the initial quality of manufacture and a regular maintenance schedule once in use. Fluids being metered should also be free of solid particles so as to reduce wear of the seals and

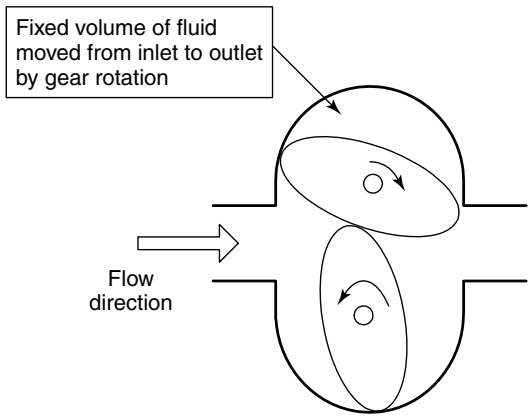


Figure 3. The oval gear positive displacement flowmeter.

the need for excessive maintenance. Positive displacement flowmeters can also be heavy and expensive for larger pipe sizes, and some designs can result in a complete blockage of the pipeline if the flowmeter seizes up.

5.4 The turbine flowmeter

Like the positive displacement flowmeter, turbine (or vane) flowmeters are mechanical devices capable of achieving high measurement accuracy. The principle of operation of this type of flowmeter is that a multibladed rotor is placed in the flow and rotates as fluid passes through it. The rotor's speed of rotation is detected using a sensor (rf, magnetic, and mechanical types being the most common) and is proportional to the velocity of the fluid flowing through the meter. These flowmeters measure the average velocity of fluid in a pipeline, and since the pipe diameter is known, volumetric flowrate can be determined.

Despite the fact that the turbine flowmeter is a mechanical device that may appear to be old fashioned when compared to many of the other technologies available, it is still one of the most accurate and repeatable flowmeters available today. Measurement accuracy of better than $\pm 0.1\%$ of reading for liquids, and better than $\pm 0.25\%$ of reading for gases, are possible using this type of flowmeter. For this reason, the turbine flowmeter is one of the most commonly used instruments in custody transfer applications. These flowmeters have a linear output and a range of at least 10:1, with 100:1 possible in some applications.

The main limitation of the turbine flowmeter is the fact that key mechanical components such as the rotor bearings will wear with use and, in doing so, degrade the instrument's repeatability and alter its calibration. Regular maintenance and recalibration is therefore necessary with this type of flowmeter. Care should also be taken to ensure that the fluid being metered is clean since solid particles in

the flow will cause more rapid bearing wear. The flowmeter's calibration is also sensitive to changes in fluid viscosity and upstream flow velocity profile.

Other types of flowmeters that use the turbine principle include the Pelton wheel and propeller meter, although they have poorer measurement accuracy than axial designs.

5.5 Oscillatory flowmeters

The vortex shedding flowmeter, now more commonly known as the *vortex flowmeter*, is the most common type of oscillatory flowmeter. This flowmeter relies on the phenomena of vortex shedding, which was first experimentally studied by Strouhal in 1878. Figure 4 shows the principle of the vortex flowmeter. A nonstreamlined obstruction known as a *shedder bar* or *bluff body* is placed across the middle of the flow stream. As the fluid stream in the pipe hits this obstacle, it must separate to pass around it, with fluid layers nearer the surface of the bar moving slower than those further away. As a result, when these fluid layers recombine after passing the bar, vortices are generated and shed alternately from either side of the shedder bar. The frequency of generated vortices is related to the upstream velocity of the fluid and the width of the shedder bar and is defined by the K factor of the flowmeter. For a given geometry of shedder bar, the K factor of a flowmeter is relatively constant over a wide range of pipe Reynolds number, and so, in these circumstances, the volumetric flowrate of the fluid is linearly related to the vortex shedding frequency.

The frequency of generated vortices is usually detected using sensors integrated into the sides of the shedder bar. Pressure, capacitance, thermal, and ultrasonic are the most common types of sensors used for this purpose.

The vortex flowmeter is capable of accurate measurement of liquid or gas (typically $\pm 1\%$ of reading) over a minimum flow range of 30:1. The flowmeter can also be used over a wide range of fluid temperatures and so is commonly used for metering process fluids at the extreme ends of the temperature range such as liquid nitrogen and steam. The instrument's calibration is also insensitive to changes in fluid density, and so a meter's calibration holds for any fluid as long as the flowmeter is used within the Reynolds number range specified by the manufacturer. The vortex flowmeter has a simple and reliable construction and so can be used with flows containing small amounts of particles, although more extreme multiphase flows such as slurries will cause rapid wear of the shedder bar and so a change in calibration. The relatively small obstruction that the shedder bar causes results in a permanent pressure loss of about half that of an orifice plate over the same range of flowrate.

The main limitation of the vortex flowmeter is that it can only be used in turbulent flow conditions. It is therefore not usually suitable for use in large pipe diameters, or in applications where the flow velocity is low or the fluid viscosity high. Most manufacturers set a minimum Reynolds number of typically 10 000 at which the specified flowmeter performance can be achieved. While many flowmeters will continue operating at Reynolds numbers less than this, the generated vortex stream becomes less stable and so accuracy is reduced. At a Reynolds of less than around 3000, vortices will not be generated at all and so the flowmeter will stop operating. The vortex flowmeter is also sensitive to changes in upstream flow velocity profile and other disturbances such as swirl and so should be installed with a sufficient straight length of pipe upstream and downstream of the measurement point. The flowmeter should not be used in applications where pipe vibration or local sources of electrical interference are high since this will corrupt the vortex signal being detected and possibly generate false readings under no-flow conditions.

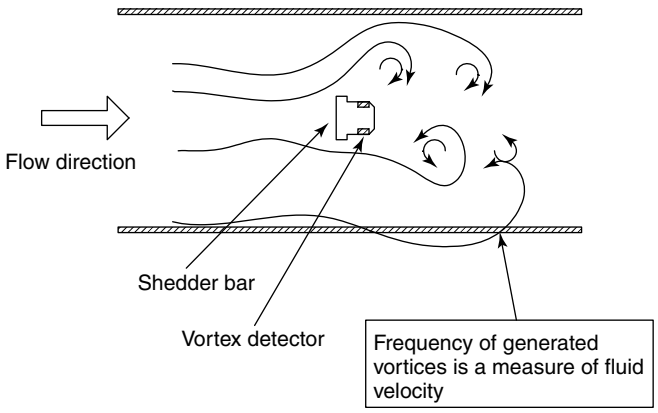


Figure 4. Principle of the vortex shedding flowmeter.

5.6 The electromagnetic flowmeter

The operation of the electromagnetic flowmeter is based on Faraday's law of induction that is when a conductor is moving perpendicular to a magnetic field, the voltage induced across the conduction is proportional to its velocity. In the case of the electromagnetic flowmeter, the conductor is the fluid being metered, while the induced voltage is measured using electrodes in the pipe wall. Since in most applications, the pipe wall of the flowmeter is made from a conductive material such as stainless steel, an inner nonconducting liner is required to insulate the electrodes and prevent the generated voltage signal from being dissipated into the pipe wall. Coils on the outside of the pipe are used to generate a magnetic field across the fluid, with simpler AC coil excitation methods, which suffer from zero drift problems, being

increasingly replaced by pulsed DC excitation techniques, which do not.

The electromagnetic flowmeter has a number of advantages over traditional flow measurement techniques and some characteristics of an ideal flowmeter. The flowmeter has no moving parts, does not obstruct the pipe at all, is available in a very wide range of pipe sizes, and may be used to measure bidirectional flows. A measurement accuracy of typically $\pm 0.5\%$ of reading over a range of at least 10:1 is possible. The flowmeter's accuracy is also unaffected by changes in fluid viscosity and density and may be used to meter difficult mixtures such as slurries and paper pulp.

The major limitation of the electromagnetic flowmeter is that it can only be used with fluids with a conductivity of typically greater than $5\,\mu S\,cm^{-1}$, although special designs are available for use with liquids with conductivities as low as $0.1\,\mu S\,cm^{-1}$. The flowmeter is therefore not suitable for use with gases, steam, or nonconducting liquids such as oil. The flowmeter's calibration is also sensitive to changes in flow velocity profile, although it requires a shorter straight length of pipe upstream of the meter than the orifice plate or turbine meter. Although electromagnetic flowmeters do not require significant maintenance, care must be taken during operation to ensure that the liner does not become damaged and that significant deposits do not build up on the electrodes since these can cause changes in the calibration or, in some cases, cause the flowmeter to stop functioning altogether. Even if these effects are minimized, electromagnetic flowmeters will require periodic recalibration using either traditional techniques or an electronic calibrator now available as an accessory from most manufacturers.

5.7 The ultrasonic flowmeter

The dream of producing a universal noninvasive flowmeter has been the catalyst for the many different ultrasonic flowmeter configurations, which have been investigated over the last 40 years (Lynnworth, 1989). However, most ultrasonic flowmeters commercially available today can be placed into one of two categories; Doppler and transit-time.

The ultrasonic Doppler flowmeter is based on the Doppler shift principle. Ultrasound at a frequency of typically 1 MHz is transmitted at an angle into the moving fluid being monitored. Some of this energy will be reflected back by acoustic discontinuities such as particles, bubbles, or turbulent eddies. The difference in frequency between the transmitted and received signals (the Doppler frequency shift) is directly proportional to the velocity of the flow.

The ultrasonic transducers, which are used to transmit and receive the ultrasound, are usually located in a single housing, which can be fixed onto the outside of the pipe, and so a simple clamp-on flowmeter, which is easy to install and is completely noninvasive, is possible.

The reflected ultrasound does not consist of a single frequency but a spread of frequencies resulting from reflections from a range of different-sized discontinuities, which may also be travelling at different velocities through the detection area. For liquids, the frequency of the transmitted ultrasound may lie in the range from 500 kHz up to a few MHz. At 500 kHz, discontinuities must have a diameter of approximately $50\,\mu m$ in order to reflect ultrasound back to the receiver. Increasing the operating frequency will allow the detection of smaller particles but at the cost of reducing the penetration of the transmitted signal into the fluid. The flowmeter is also sensitive to changes in flow velocity profile and the spatial distribution of discontinuities in the flow. As a result, the accuracy of Doppler flowmeters is poor, typically $\pm 5\%$ of full scale. However, this can be improved by calibrating the flowmeter online. Since there is a large acoustic mismatch between steel and air, clamp-on Doppler flowmeters cannot be used for metering gas flows or, of course, totally clean liquids where there are insufficient reflecting particles or bubbles to produce a reliable Doppler signal.

Figure 5 illustrates the basic principle of the ultrasonic transit-time flowmeter. Two ultrasonic transducers are mounted on either side of the pipe so that ultrasound can be transmitted across the fluid flowing in the pipe. The difference in the time taken for a pulse of ultrasound to travel between transducer 1 and 2 (with the flow) and transducer 2 and 1 (against the flow) is given by

$$\Delta T = \frac{2lv\cos\theta}{c^2 - v^2\cos^2\theta} \tag{2}$$

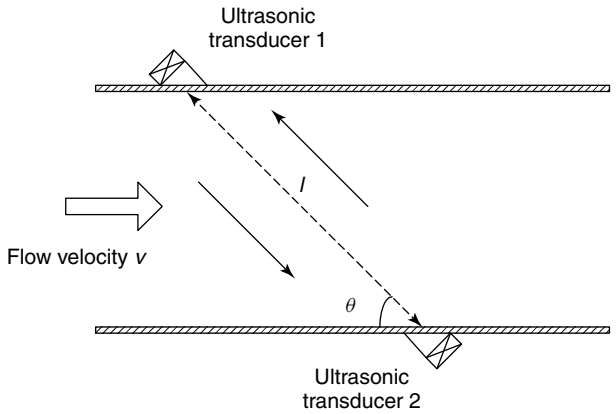


Figure 5. Principle of the transit-time ultrasonic flowmeter.

Since the velocity of sound in the fluid c is much greater than the velocity of the fluid v,

$$\Delta T \approx \frac{2lv\cos\theta}{c^2} \tag{3}$$

Therefore, if the velocity of sound in the fluid is constant, then there is a linear relationship between ΔT and v.

Although this method is elegant and straightforward in principle, in practice, there are difficulties since ΔT can be small and the change in ΔT that occurs with changing fluid velocity can be even smaller (typically fractions of μs per meter). In addition, as 3 shows, if the temperature of the fluid changes, c will change. Measurement of, and correction for, changes in the fluid temperature are usually needed. Transit-time flowmeters therefore require complex signal conditioning and processing.

The flow velocity v that is calculated using (3) is the average velocity along the transmission path between the two transducers, and so the flowmeter's calibration will be very dependent on the flow velocity profile. The problem of flow regime dependency can be significantly reduced by using a configuration of several parallel ultrasonic beams and averaging the measured mean velocity along each beam to calculate the overall fluid velocity. This is analogous to numerical integration, and a wide of range of multibeam configurations have been proposed, each with their own advantages.

Single path ultrasonic transit-time flowmeters have a typical accuracy of ±2% of reading over a range of at least 100:1. Although this type of flowmeter can be used with liquids or gases, clamp-on designs can only be used with liquids. Multibeam flowmeters have an improved accuracy but are more expensive. However, they are finding increased use in high value applications like the custody transfer of natural gas. Unlike most other flowmeters, the cost of transit-time flowmeters does not increase significantly with pipe diameter. Transit-time flowmeters are intended for use with clean fluids, although most can still operate if there are a small amount of impurities present in the flow.

See **Article 118, Ultrasonic Sensors, Volume 2**; **Article 119, Ultrasonic Instrumentation Principles, Volume 2**; and **Article 120, Ultrasonic Instrumentation Design, Volume 2** for information on ultrasonic sensors, and ultrasonic instrumentation principles and design.

5.8 The mass flowmeter

The most common type of mass flowmeter used today is the Coriolis flowmeter. As the name suggests, the principle of operation makes use of the Coriolis effect, which was discovered by Gustave Coriolis in 1835. The fluid being metered passes through a vibrating tube, and as a result of the Coriolis force acting on it, the fluid will accelerate as it moves towards the point of maximum vibration and decelerate as it moves away. This will result in flexure of the tube, the amount of flexure being directly proportional to the mass flowrate of the fluid.

The first commercial Coriolis flowmeter used a U-shaped tube but now many different configurations exist, including dual loops and straight through designs. Each design has its own advantages, with factors such as accuracy, repeatability, and pressure drop varying from design to design.

Whichever design is used, the Coriolis flowmeter is a high accuracy instrument, which may be used to meter gas or liquid with an accuracy of typically ±0.25% of reading. Measurement range varies with design but 100:1 is possible for U-tube designs and 50:1 for straight tube designs. Since the flowmeter measures mass directly, changes in density, viscosity, pressure, and temperature does not affect the calibration of the flowmeter. The flowmeter is also not affected by changes in flow velocity profile or other flow disturbances such as swirl. The flowmeter does not obstruct the flow at all and can be used to meter flow in both directions. However, the pressure drop across U-tube designs can be a limitation with viscous fluids.

The major disadvantage of the Coriolis flowmeter is its cost, which is high in comparison to most other flowmeters. This cost may be justified in applications where the product cost is high or where mass flowrate of the fluid is required. The cost and weight of the Coriolis flowmeter increases significantly with increasing pipe diameter and, as a result, are usually limited to pipe diameters of less than 100 mm.

Unlike most of the flowmeters discussed so far, the Coriolis flowmeter can meter some difficult two-phase flows. For example, reliable measurements of the mass flowrate of liquid/gas mixtures are possible if the gas component is small and well distributed, and therefore, the mixture is acting like a pseudohomogeneous fluid. The percentage of gas that can be tolerated by the flowmeter will depend on the viscosity of the liquid component. The less viscous the liquid, the more likely the gas is to separate out and cause problems. Liquid/solid flows (slurries) may also be metered, although the user has to compromise between avoiding particle dropout and avoiding excessive fluid velocities, which would result in accelerated wear of the flow tube.

Details of less commonly used types of mass flowmeters such as those based on thermal principles can be found in Baker (2000).

6 MULTIPHASE FLOW MEASUREMENT

There is a growing demand in areas such as the petroleum and food industries to be able to measure two-phase flows such as liquid with entrained gas or liquid with solids. Yet, the measurement of such flows nearly always presents difficulties to the process engineer.

For example, an ideal flowmeter would be able to directly measure the volumetric flowrate of a liquid, whether it is all liquid or contains a second phase of gas. Unfortunately, most of the flowmeters described above can usually only be used to meter two-phase flows when the second component is small. A review of the performance of conventional flowmeters in two-phase flows can be found in NEL (1997).

The alternative to direct flow measurement is to use an inferential method. An inferential method for liquid/gas flow would require the velocity of the gas and the liquid phases (v_g and v_l) and cross-sectional fraction of the gas phase (α) to be independently measured in order to calculate the volumetric flowrate of the mixture Q_m:

$$Q_m = v_l A(1-\alpha) + v_g A\alpha \tag{4}$$

The overall uncertainty of the flowrate measurement would depend on the accuracy with which the independent measurements can be made. The velocity of the liquid and gas phases cannot be assumed to be the same, and the way the gas is distributed in the liquid (the flow regime) will change depending on factors such as gas content, individual phase velocity, and pressure. Even in a simple case such as this, it is clear that multiphase flow measurement is by no means straightforward. For this reason, commercial multiphase flowmeters are generally expensive and target at very specific applications.

The most common two-phase flows are liquid/gas (e.g. gas in water), liquid/liquid (e.g. water in oil), gas/solid (pneumatically conveyed solids), and liquid/solids or slurries (e.g. coal in oil). Each presents its own measurement problems, and it is not feasible to discuss all possible metering combinations for these types of flow in a overview article such as this. For further details of two-phase flow measurement techniques, see Rajan, Ridley and Rafa (1993), while the important area of three-phase flow measurement is considered in a paper by Thorn, Johansen and Hammer (1997).

RELATED ARTICLES

Article 120, Ultrasonic Instrumentation Design, Volume 2; **Article 187, Flowmeter Selection and Application, Volume 3**; **Article 188, Differential Pressure (DP) Flowmeters, Volume 3**.

REFERENCES

Baker, R.C. (2000) *Flow Measurement Handbook: Industrial Designs, Operating Principles, Applications and Performance*, Cambridge University Press, Cambridge, MA.

British Standards Institution. (1991) *Guide to Selection and Application of Flow Meters for the Measurement of Fluid Flow in Closed Conduits*, British Standard BS 7405: 1991, British Standards Institution, London.

Grant, D.M. (2001) Open Channel Flow Measurement, in *Flow Measurement: Practical Guides for Measurement and Control*, 2nd edn (ed. D.W. Spitzer), ISA, Research Triangle Park, NC, (pp. 249–293).

Lynnworth, L.C. (1989) *Ultrasonic Measurements for Process Control: Theory, Techniques, Applications*, Academic Press, Boston, MA.

Mulley, R. (2004) *Flow of Industrial Fluids*, CRC Press, Boca Raton, FL.

Miller, R.W. (1996) *Flow Measurement Engineering Handbook*, 3rd edn, McGraw Hill, New York.

National Engineering Laboratory. (1997) *Effects of Two-phase Flow on Single-phase Flowmeters*, Flow Measurement Guidance Note No. 3, National Engineering Laboratory, UK.

Rajan, V.S.V., Ridley, R.K. and Rafa, K.G. (1993) Multiphase Flow Measurement Techniques – A Review. *Journal of Energy Resource Technology*, **115**, 151–161.

Spitzer, D.W. (ed.) (2001) *Flow Measurement: Practical Guides for Measurement and Control*, 2nd edn, ISA, Research Triangle Park, NC.

Thorn, R., Johansen, G.A. and Hammer, E.A. (1997) Recent Developments in Three-phase Flow Measurement. *Measurement Science and Technology*, **8**(7), 681–701.

Webster, J.G. (ed.) (2000) *Mechanical Variables Measurement: Solid, Fluid, and Thermal*, CRC Press, Boca Raton, FL.

FURTHER READING

Flow Measurement and Instrumentation, Elsevier Science, Oxford (1996).

190: Calibration and Standards in Flow Measurement

Richard Paton
National Engineering Laboratory, Scotland, UK

1 General Principles 1256
2 Gravimetric Calibration of Liquid Flowmeters 1257
3 Volumetric Calibration of Liquid Flowmeters 1258
4 Calibration of Gas Flowmeters 1258
5 Mixed or Multiphase Flowmeters 1259
6 In Situ or Field Calibration 1260
References 1260

1 GENERAL PRINCIPLES

Calibration is defined as follows: 'The set of operations that establish, under specified conditions, the relationship between values of quantities indicated by a measuring instrument or measuring system and the corresponding values realized by standards.'

It is important to recognize at the outset the definition of a 'calibration' and to note that the 'comparison' applies only to the conditions at the time of the calibration. The purpose of calibration is to increase the confidence in the reading obtained from the flowmeter in service.

Standards for flow measurement are based on a comparison of the quantity of fluid passed, or passing, through the flowmeter with the quantity measured by the standard. Standards can be based on the measurement of mass or volume. The required mass or volume quantity can be calculated from the measured quantity from a knowledge of the fluid density at the test flowmeter. Standards may be static or dynamic, and this choice is made on the basis of the output and end use of the flowmeter. Some flowmeters are used to measure quantity and have a fast response time; others are designed to measure flow rate and have a slow response time. Calibration of a flowmeter should cover a significant flow-rate range for the flowmeter and establish a performance across that range.

A 'standard' for flow calibration should replicate as closely as practicable the conditions under which the flowmeter will be used. Full replication is impossible, and the key to success is to be as close as practicable, and to recognize the nature and performance of the flowmeter type being tested. The factors that should be addressed are the fluid viscosity, the installation effects due to bends and fittings, temperature, and pressure. The standard should also have a defined traceability and uncertainty chosen to match the expectation of the final measurement.

The choice of standard must also recognize the dynamic performance of the flowmeter and the nature and resolution of the output. As a calibration is a comparison, the quantity measured by the standard must match the quantity measured by the flowmeter and this must take cognizance of the resolution of the flowmeter. A flowmeter used to measure very large quantities of fluid over a long period of time may not have a resolution suitable for measuring the smaller quantities measured by a standard.

The result of a flowmeter calibration will normally provide two related figures: one related to the flow rate and the other as a performance indicator. Flow rate will be expressed as mass per unit time, volume per unit time, Reynolds number, or some other flow rate–related measure. The performance indicator relates the expected performance of the meter to the measured performance. Examples of performance indicators are K-factor, error, and meter factor.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

2 GRAVIMETRIC CALIBRATION OF LIQUID FLOWMETERS

A flowmeter can be calibrated gravimetrically by weighing the quantity of liquid collected in a vessel. The vessel is weighed and the weight (in air) of the fluid collected is noted.

As the quantity of fluid has to be expressed as mass, the weight has to be corrected for the effect of air buoyancy. As a weighing machine is calibrated using weights with a conventional density of $8000\,\mathrm{kg\,m^{-3}}$, and the fluid collected will have a significantly different density, the upthrust of the air on the tank will have a significant effect. This amounts to around 0.1% for water. This correction is given in (1).

$$M = W \times \left\{1 + \rho_{\mathrm{air}} \times \left[\frac{1}{\rho_{\mathrm{f}}} - \frac{1}{\rho_{\mathrm{w}}}\right]\right\} \qquad (1)$$

where M is the mass, W is the measured weight, ρ_{air} is the density of air, ρ_{f} is the density of the fluid, and ρ_{w} is the density of the calibration weights (by convention, $8000\,\mathrm{kg\,m^{-3}}$).

To determine the volume, the mass is divided by the density determined at the flowmeter. Density can be measured using a densitometer but is more often calculated from a knowledge of the fluid properties and measurements of temperature and pressure at the test meter.

2.1 Standing-start-and-finish method

This method is generally preferred for flowmeters that are used for measuring the quantity of liquids, especially flowmeters for batch quantities (Figure 1).

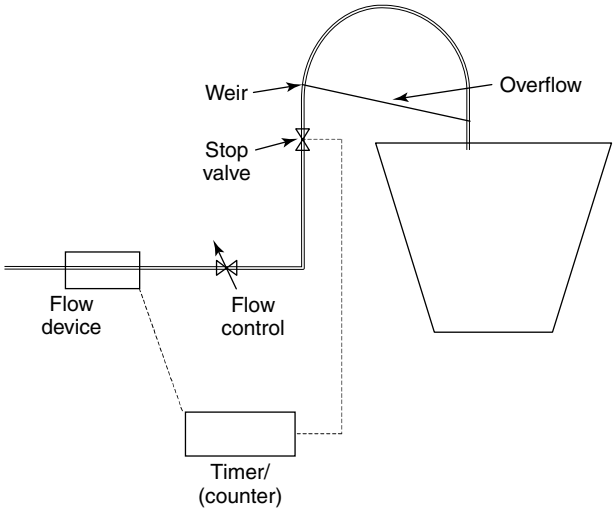


Figure 1. Standing-start-and-finish method for the gravimetric calibration of liquid flowmeters.

The required flow rate is established into the tank. The flow is then stopped using a fast-acting valve, the container drained, and the drain valve closed. The flow is restarted, the container filled, and the flow stopped. The weight of liquid is noted along with the time taken to fill the tank. The reading from the flowmeter is also noted. Temperature and pressure of the liquid at the flowmeter are also noted during the fill.

2.2 Flying-start-and-finish method

This is sometimes called the *diverter method*. In this method, the flow through the flowmeter is not stopped but the flow is diverted between a return to the supply and the collection tank (Figure 2). A switch on the diverter mechanism starts and stops a timer to time the filling of the collection measure and counter to totalize pulses from the test device.

In this method, the key to accurate measurement is a clean separation between fluid entering the tank and fluid returning to the supply.

Flying-start-and-finish methods are used primarily for flowmeters with slow response times and where flow rate is the primary measurement rather than quantity passed. Flowmeters with visual displays cannot be calibrated by this method.

The main source of uncertainty for both gravimetric calibration methods lies in the timing error. In the case of the standing-start method, this is caused by the response time of the flowmeter and the changing flow at each end of the test. For the diverter methods, this is introduced by not triggering at the hydraulic center of the liquid jet.

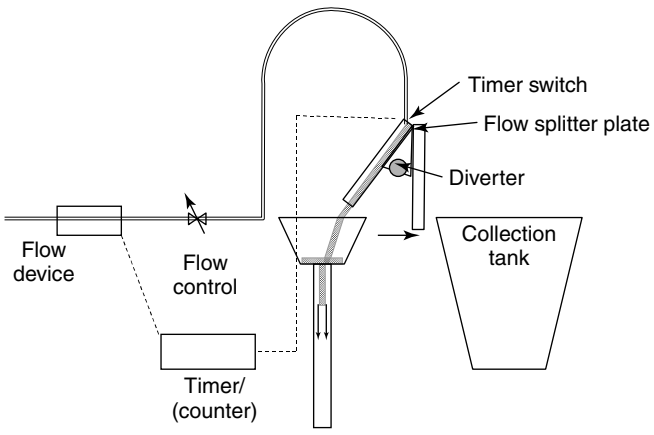


Figure 2. Flying-start-and-finish method for the gravimetric calibration of liquid flowmeters.

3 VOLUMETRIC CALIBRATION OF LIQUID FLOWMETERS

The measurement of the quantity of liquid collected may be carried out volumetrically, that is, by collecting a known volume of liquid in a container. In the volumetric method, the standard vessel takes the form of a container with calibrated volume. Normally, this would be a pipette with conical ends to facilitate drainage and to reduce the risk of air entrapment. The neck of the pipette is normally fitted with a sight glass and a scale marked in volumetric units. A typical volumetric tank is shown in Figure 3.

The tank volume must be determined by calibration against weighed water contained in the vessel, or by using smaller volumetric measures that are themselves traceable to national standards by weighing methods.

Volumetric systems are normally used with standing-start-and-finish methods owing to the difficulty of diverting flow into the tank and controlling the finish of the fill. Drainage time (after the tank is empty) is vitally important, as liquid clingage to the wall can be significant. Each tank has a calibrated drain time and this must be maintained.

Reference volume tanks and pipe provers have their volume defined at a stated reference temperature (and pressure). Normal reference temperatures are 15 °C or 20 °C. For all volumetric methods, corrections due to the expansion and contraction of both the standard and the device being calibrated have to be accounted for along with the expansion and contraction of the fluid between the standard and the flowmeter. This applies to both thermal and pressure expansion.

Pipe provers provide a dynamic volumetric calibration standard method for calibration. They provide a sealed system providing high-accuracy calibrations in situ as well as in laboratories. 'Proving' is used extensively in the oil industry and is generally synonymous with calibration.

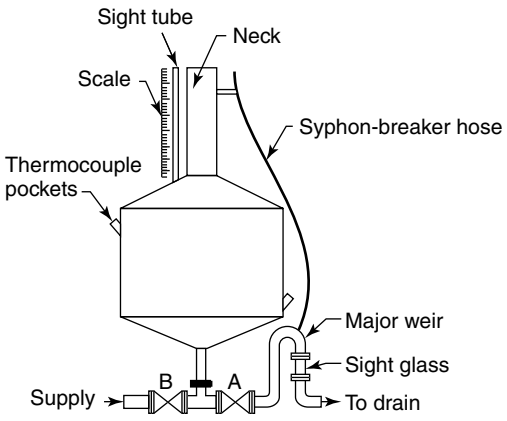


Figure 3. A volumetric tank.

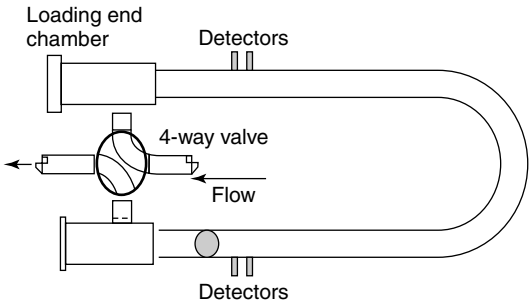


Figure 4. A bidirectional prover.

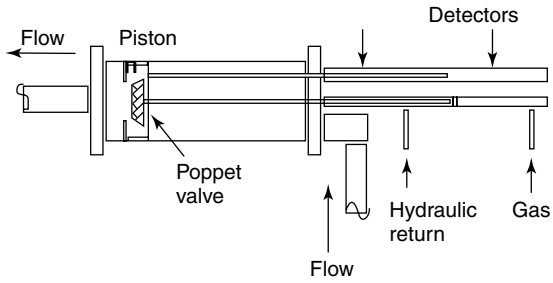


Figure 5. A small volume prover.

A length of pipe is fitted with switches such that the volume between the switches is known. If a displacer, or 'pig', is introduced to the flow, the time it takes to travel between the switches will give a measure of the flow rate. If the switches are used to gate a pulse counter, totalizing pulses from a flowmeter, a measure of the flowmeter factor (pulse per liter) can be found.

A variety of designs are available, and one of the most common, the bidirectional prover, is illustrated in Figure 4.

A four-way valve, of very high integrity, changes the flow path without breaking the flow. The sphere is held in special end chambers. These are designed to launch the sphere and absorb the shock of capture. One chamber also provides a means of removing the sphere. Note from Figure 4 that two switches are provided at each end. This provides better integrity of the measurement by giving redundancy and a means of checking results by developing four separate volumes.

A small volume prover is a pipe prover where the length, diameter, or volume is smaller than that required to obtain 10 000 pulses from the flowmeter. A proprietary device is illustrated in Figure 5, where a precision pipe and a piston along with optical detectors are used. Flowmeter resolution is effectively increased using pulse interpolation.

4 CALIBRATION OF GAS FLOWMETERS

In general, all the methods for the calibration of gas flowmeters have analogies with liquid methods. The main

difference between the calibration of a gas and a liquid flowmeter is the compressibility of the gas and the fact that the gas has to be contained in a closed container. As gas is compressible, the volume measured at the standard and the volume measured at the test device have to be corrected to a common or to a 'standard' condition.

Volume standards generally take the form of displacement devices much akin to flowmeter provers. Three common examples are discussed.

A common standard used extensively for lower pressure calibrations is the 'bell prover' (see Figure 6). In this device, air is displaced into a calibrated bell, which rises or lowers as the bell is filled. A bath of oil or water acts as the seal.

Piston provers for gas are based on a very light and low-friction piston displacing gas from a smooth cylinder. The piston may be driven by the gas or from a mechanical mechanism. The example of the mercury-seal piston prover is illustrated in Figure 7.

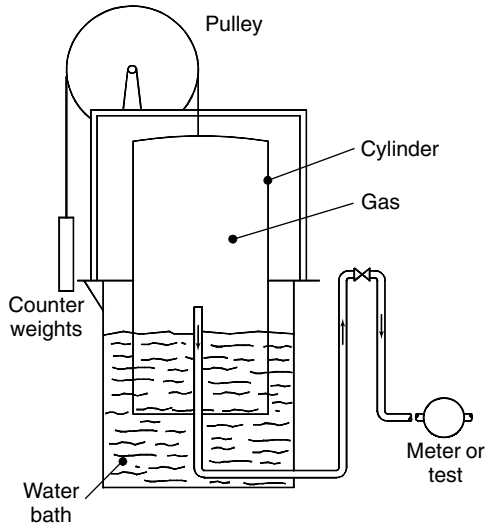


Figure 6. A bell prover.

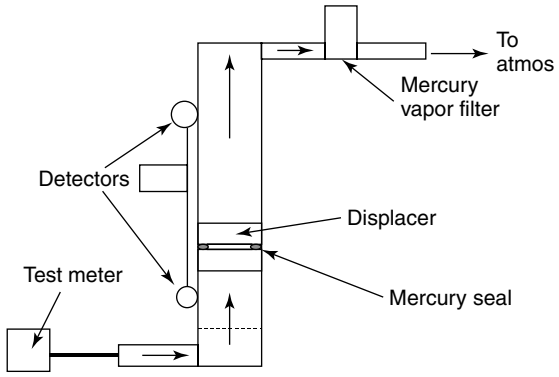


Figure 7. A mercury-seal piston prover.

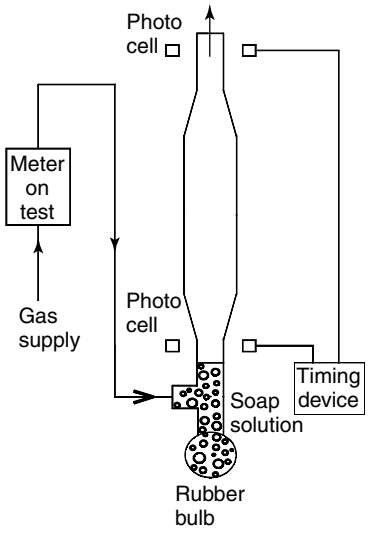


Figure 8. A soap-film burette.

Another device for the measurement of very low-pressure gas is the soap-film burette where the piston is replaced by a soap film (Figure 8).

Gas flowmeters can also be calibrated using mass as the reference quantity. This can be done gravimetrically by weighing high-pressure gas collected from, or delivered to, a test meter. Alternatively, the mass can be calculated using PVT (pressure/volume/temperature) calculations if a fixed volume is used.

Critical flow nozzles provide an extremely stable calibration device. In this device, when the velocity of gas reaches the speed of sound in the throat of the nozzle, the mass flow will be a function of the upstream pressure and the properties of the gas only. This allows measurement of mass flow to 0.3% using a standard nozzle and down to 0.1% if the nozzle has been calibrated. The equation is given by (2).

$$\frac{\mathrm{d}m}{\mathrm{d}t} = C_\mathrm{d} C^* A_\mathrm{t} P_\mathrm{o} \frac{1}{\sqrt{RT_\mathrm{o}}} \tag{2}$$

where C_d is the discharge coefficient, C^* is the critical flow factor, A_t is area of the nozzle throat. P_o and T_o are the upstream pressure and temperature.

5 MIXED OR MULTIPHASE FLOWMETERS

Multiphase flowmeter applications are commonly used for two main applications.

The first application is for mixtures of oil, gas, and water, particularly in the offshore production of hydrocarbons. Meters can be calibrated at a dedicated laboratory such

as NEL in the United Kingdom. Flows of oil, water, and gas are separately metered and combined upstream of the meter. They are then separated and recirculated. The multiphase flowmeter reading is compared with the single-phase measurements. In the field, flowmeters are 'calibrated' or assessed against a 'test separator'. The metered multiphase flow is introduced to a separator and the single-phase measurements from the separator are measured over long time periods.

The second application of multiphase meter is to meter 'wet' gas. This type of meter can also be calibrated in a specialist laboratory, such as NEL, by injecting measured quantities of liquid to a measured gas flow. In the field, dilution tracer techniques can be used to estimate the flow rate of the liquid phase, hence giving a measure of the correction to be applied to the primary gas meter previously calibrated in dry gas.

6 IN SITU OR FIELD CALIBRATION

In an ideal world, all flowmeters would be calibrated in situ, hence avoiding differences in fluid and installation changing the calibration of the flowmeter. In practice, this cannot be done with low uncertainty as the field installations and the possible standards are not suitable.

In the oil industry and for fiscal installations, it is common to have flowmeter provers permanently installed and valve arrangements in place to allow calibration of a bank of flowmeters to be carried out at any time. In other applications, mobile flowmeter provers or even volume tanks can be taken to the installation. This concept is used for high-accuracy systems and for lower-uncertainty systems such as petrol dispensers.

Where a portable volume standard is impractical, reference flowmeters can be installed in series. The various effects of installation and fluid conditions have to be recognized. This applies also to differences between the in situ use and the laboratory calibration of the meter. A different type of reference flowmeter that can be used is the clamp-on ultrasonic meter. In this type, the sensors are strapped to the outside of the pipe wall and the flowmeter measures the velocity of the fluid. If the pipe diameter is known, the flow rate is calculated. Although improved greatly in recent years, uncertainties of better than 1% cannot be expected and 5 to 10% is not uncommon.

Insertion meters can be used to 'verify' the calibration of in situ meters. An insertion meter can be a pitot tube, an electromagnetic sensor, a turbine type sensor, or a thermal sensor fitted to the end of a probe, which can be inserted into the pipe through a fitting on the wall. The point velocity of the fluid is measured, and hence the bulk flow rate calculated. 1 to 10% uncertainties can be expected.

A final technique is to use tracer fluids injected into the fluid to measure the flow. Two methods are used. In the first, a radioactive tracer with a very short half-life is injected as a pulse into the flow. The tracer is injected some distance upstream of a detector strapped to the pipe wall. A second detector is located a measured distance downstream. The time taken for the tracer to travel between the detectors gives a measure of the flow. In the second technique, a chemical tracer is continuously injected at a measured rate into the flow. A sample is withdrawn from the flow some distance downstream and the concentration of the chemical is analyzed. The concentration gives a measure of the flow.

Further information on the calibration of flowmeters can be found in Baker (2000) and Spitzer (2001).

REFERENCES

Baker, R.C. (2000) *Flow Measurement Handbook: Industrial Designs, Operating Principles, Performance and Applications*, Cambridge University Press, Cambridge.

Spitzer, D.W. (ed.) (2001) *Flow Measurement: Practical Guides for Measurement and Control*, 2nd edn, ISA International, Research Triangle Park, NC.

191: Displacement and Angle Sensors Performance and Selection

Halit Eren
Curtin University of Technology, Perth, Western Australia, Australia

1 Displacement and Angle Sensor Classification 1263
2 Key Parameters in Selection of Method 1264
3 Selection Displacement Measuring Methods 1265
4 Selection Strategy for Angle Measuring Methods 1266
5 Excitation Methods and Their Choice 1266
6 Relative Merits of Resistive, Magnetic, Capacitive, and Optical Sensing 1267
Further Reading 1267

1 DISPLACEMENT AND ANGLE SENSOR CLASSIFICATION

There is a large range of displacement sensors that operate on different physical principles. Some of the basic principles may be capacitance and capacitive effects, magnetism and magnetic effects, optics and its properties, piezoelectric materials and their properties, conductivity of electric conductors, radio and microwave techniques, mechanical means, and so on. Also, within each range many sensors may be available in various sizes and shapes to suit different application requirements. In general, displacement sensors can be classified as

1. *Capacitive displacement sensors*

   Variable plate-distance capacitive sensors
   Variable-dielectric capacitive displacement sensors
   Variable plate-area capacitive displacement sensors
   Differential capacitive sensors.

2. *Magnetic displacement sensors*

   Linear variable-reluctance displacement sensors
   Single-coil linear variable-reluctance sensors
   Variable-differential reluctance sensors
   Linear variable-inductive sensors
   Variable-coupling sensors
   Eddy current displacement sensors
   Linear variable-differential transformers.

3. *Piezoelectric and piezoresistive displacement sensors*

   Resistive displacement sensors
   Strain gauge sensors
   Piezoresistive displacement sensors.

4. *Optical displacement sensors*

   Optical fibers
   Interferometric sensors
   Time-of-flight and frequency shift sensors
   Encoders and digital displacement sensors.

5. *Radio Frequency (RF) and microwave displacement sensors*

   RF position sensing
   Microwave sensors
   Global Positioning System (GPS) systems.

6. *Acoustic and ultrasonic displacement sensors*
7. *Angular displacement sensors*

   Tachogenerators
   Microsyn
   Synchros

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

Induction potentiometers
Rotary variable-differential transformers.

2 KEY PARAMETERS IN SELECTION OF METHOD

The key parameter associated with these measurements is the magnitude of displacement. The nature of the process that results in the displacement is also significant in determining the type of instrument to be selected. From the basic principles, the displacement is closely associated with acceleration and velocity. Double integration of acceleration and single integration of velocity yield the displacement of an object. Therefore, some devices that resemble acceleration and velocity sensors can easily be modified for displacement measurements.

There will usually be available a variety of suitable sensors that could be used for a particular application, as shown in Figure 1. For a particular application, if a choice is available, the selection may be based on factors such as cost, speed in readings, accuracy and precision, availability, and so on. For example, capacitive displacement sensors may satisfy the requirements of applications in which high linearity and wide ranges from a few centimeters to a couple of nanometers in a cost effective manner whereas similar characteristics with greater accuracy may be obtained by optical methods but at a greater expense.

Capacitive and magnetic displacement sensors enjoy a wide range of industrial applications. In capacitive sensors, precision signal processing allows sensing of capacitance changes down to a few femtofarads. There are a number of different capacitive sensors – *see* **Article 193, Specialty Displacement and Angle Sensors, Volume 3**. For example, the variable plate-distance capacitive sensors are often used for measuring small incremental displacements without making contact with the object. Variable-dielectric capacitive displacement sensors are mostly used in the form of two concentric cylinders for measuring the level of nonconducting fluids in tanks. These types of sensors are used as rotating capacitors for angular displacement measurements.

Similar to capacitive displacement sensors, the magnetic displacement sensors come in various shapes and sizes. They have a diverse range of applications. Linear variable-reluctance displacement sensors are used in acceleration measurements or they can be configured for sensing displacements and velocities. Whereas, eddy current displacement sensors are mainly the probe type for sensing the presence of electrically conductive targets, variable-reluctance magnetic sensors are used in specialized applications such as pressure measurements. Variable-differential reluctance sensors respond to static and dynamic displacements.

In the linear variable-differential transformer (LVDT), there is no physical contact between the core and the coil

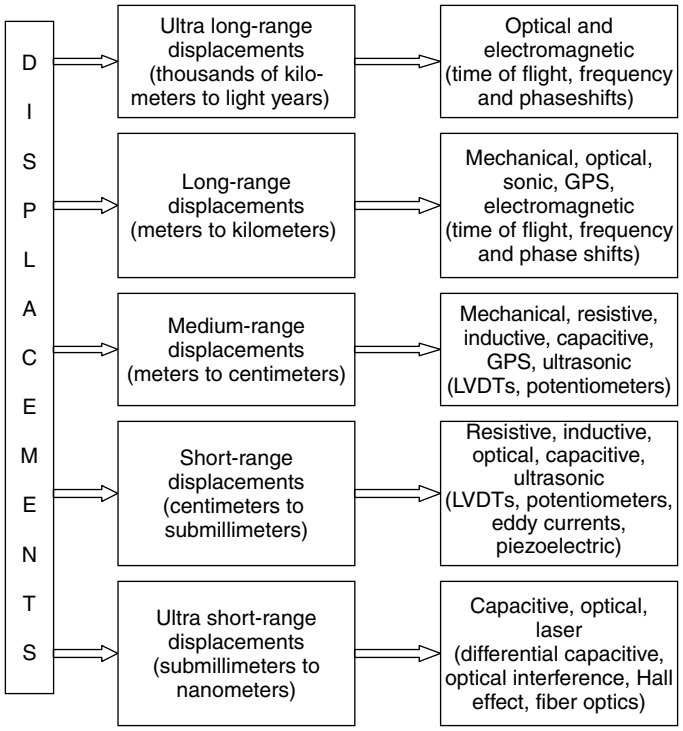


Figure 1. Most commonly applied displacement measurement techniques.

form; thus, there is no friction or wear. Hence, they find a variety of applications in which continuous displacement measurements need to be made, such as in jet engines or in measuring roll positions in the thickness of materials in hot-slab steel mills. After some mechanical modifications of LVDTs, force and pressure measurements can also be made.

A typical commercial LVDT sensor is illustrated in Figure 2. This particular sensor is suitable for high-temperature applications.

The piezoelectric displacement sensors are useful for general-purpose motion and for vibration measurements from low- to medium-frequency displacements. They are often manufactured in the form of piezoelectric ceramic cartridges giving inexpensive but highly reliable displacement measurements.

Piezoresistive displacement sensors are used in vibration measurements at low frequencies (e.g. below 1 Hz).

Optical methods are commonly used in displacement sensing, particularly where high precision and small displacements are involved.

An example of commercial fiber-optic distance measurement devices is illustrated in Figure 3. The resolution of this device is 2.5 μm.

The laser interferometric sensors operate on precise wavelength monochromatic light and are the most accurate in this family of sensors. An accuracy better than 1.0×10^{-9} m can be obtained. Encoders are used mainly for linear and rotary displacement sensing. Some of the encoders are manufactured to give resolution better than 1 μm.

For angle measurements, Microsyns are extensively used in applications involving gyroscopes. By using Microsyns, very small motions can be detected, giving output sensitivity as low as 0.01° of change in angles. The output of the device can be made as high as 5 V/1° of rotation. Synchros are primarily used in angle measurements and are commonly applied in control engineering as parts of servomechanisms, machine tools, antennas, and the like. In most rotary linear variable-differential transformers, the rotor mass is very small, usually less than 5 g. The motion in the radial direction produces a small output signal that can affect the overall sensitivity.



Figure 2. A typical LVDT displacement sensor. (Courtesy of RDP Electrosense, Inc., http://rdpelectrosense.com.)



Figure 3. A typical displacement sensor based on fiber optics. (Courtesy of MTI Instruments Inc., http://www.mtiinstruments.com/.)

3 SELECTION DISPLACEMENT MEASURING METHODS

Displacement measurement techniques employ various transduction media, including sound waves, magnetic fields, electric fields, light, RF and microwave, and mechanical means. Selection of an appropriate method depends on the application characteristics, range, and other requirements. As mentioned earlier, the magnitude of the displacement is one of the important parameters; the selection of an appropriate displacement sensor to suit this requirement is essential. Some examples are as follows:

A simple way of sensing distance is the use of a linear potentiometer. The moving object is connected to the slider of the potentiometer so that each position along the axis corresponds to movement of the object. These position sensors can give a precision of 0.1%. Alternative methods are capacitive, induction, or other magnetic sensors. For very small distances, strain gauges can be used. Fiber-optic methods and laser interferometers are applicable when very precise changes within micrometer range must be sensed. They find applications in high-precision measurements such as numerically controlled machine tools. They basically compare the changes in optical path length to the wavelength of light. One particular advantage of the laser interferometer is that its output can readily be converted to digital form, since it is based on counting of the wave peaks.

In general, capacitive, magnetic, resistive, and piezoelectric sensors are suitable for subcentimeter measurements. Capacitive sensors are particularly suitable in applications such as automotive, in robotic systems as miniature control elements, in medical applications for tissue deformation, in implant micromotion, in materials science, in civil engineering, in structural deflections, and so on.

In robotic grasp applications, electro-optical techniques are often used. These sensors are relatively small in size, and are known to have a large range of operation. In this particular application, relatively new ultrasonic and capacitive sensors are used.



Figure 4. A typical optical distance sensor. (Courtesy of MTI Instruments Inc., http://www.mtiinstruments.com/.)

Optical methods, RF, GPS, ultrasonic, and microwave techniques are selected for sensing large displacements. In optical, RF, ultrasonic, and microwave techniques, time of flight, phase shift, and frequency change methods are employed by means of comparing the transmitted signals with the returned signals.

A typical commercial optical sensor is illustrated in Figure 4. This device measures distance, displacement, vibration, or thickness without contact for distances of up to 152.4 mm.

A linear digital encoder gives an output, which is a binary number that is proportional to the distance of the encoder relative to a fixed point. Digital encoders provide a set of binary digital signals directly, they require slides with precision printing, good light sources and detectors.

4 SELECTION STRATEGY FOR ANGLE MEASURING METHODS

In the angle-measuring methods, the sensing and measurement of rotational movement is absolutely necessary in many industrial applications since almost all industrial machines include rotations. The simplest form of sensing angular velocity is by using the AC or DC generators commonly known as *the tachogenerators*.

In some applications, signals are generated for each revolution of the wheel or the shaft optically or by means of piezoelectric or magnetic pulsing arrangements. These pulses enable the determination of the angular velocities as well as positions of shafts or wheels.

Angular methods for displacement measurements can be grouped to be small or large angular measurements. Small displacements are measured by the use of capacitive, strain gauge, or piezo sensors. For larger displacements, potentiometric, or inductive techniques are more appropriate. The rotary digital encoders also find wide applications.

Resistive angle and displacement sensors (also known as potentiometers or pots) are commonly used. In potentiometers, an electrically conductive wiper slides against a fixed resistive element according to the angle of displacement of an external shaft. The pot operates on the voltage division principle; therefore, the analog voltage output can directly be used as the indication of angle or it can be digitized.

An alternative resistive angle measurement is the magnetoresistive potentiometer available for linear and rotary applications. In this case, a nonconducting, permanent magnet wiper rides above a pair of magnetoresistive elements.

Microsyn transducers are used extensively in applications involving gyroscopes. By this method, very small angles can be detected, giving measurements as low as 0.01° of change in angles. The sensitivity of this device can be made as high as 5 V per degree of rotation.

5 EXCITATION METHODS AND THEIR CHOICE

Excitation requirements vary from sensor to sensor. Resistive angle and displacement sensors are known to operate on the simplest excitation method by means of supplying a stable DC voltage source across the terminals. Nevertheless, the outputs of these sensors can be highly nonlinear owing to the loading effects that may require some form of linearization. Some of these linearization techniques include terminal-based linearity, absolute linearity, and zero-based linearity – *see* **Article 58, Description of Accuracy, Linearity, and Drift, Volume 1**. Potentiometers are available with linearities ranging from 0.1 to 1%.

A stable AC excitation is essential for magnetic sensors. For example, an inductive linear-variable sensor uses a bridge that is excited typically with 5 to 25 V AC source with a range of frequency of 50 Hz to 5 kHz. They usually deliver a DC output after internal processing of the phase sensitive system.

Another example is the linear variable-differential transformers or the LVDT that requires a stable AC supply with frequency of 50 Hz to 20 kHz. The typical recommended excitation voltage for an LVDT is 4 to 6 V AC with a nominal frequency of 5 kHz.

In the case of capacitive sensors, the frequency responses can be several kHz (typically 7 kHz). These sensors are used as part of a bridge network to process the output signals. In many situations, the signals obtained from the sensors are amplified by using suitable charge amplifiers. Modern IC (integrated circuit) sensors can operate with 2.5 V power supplies.

Excitation of optic, RF, and microwave sensors can be complex. They need highly stable voltage and current sources. For example, a balanced constant-current excitation uses a pair of matched constant-current sources to excite the sensor. This technique provides immunity from noise pickup. Manufacturers usually supply extensive information on excitation requirements of these sensors.

6 RELATIVE MERITS OF RESISTIVE, MAGNETIC, CAPACITIVE, AND OPTICAL SENSING

The relative merits of angle and displacement sensors can be explained in terms of their uses and characteristics. The main merits are as follows:

Inductive and capacitive sensors operate on magnetic and electric field principles respectively and they are suitable for small displacement and angle measurements such as the proximity determinations.

The LVDT works by moving a slug of high permeability material inside part of a transformer. An AC signal is applied to one side of the transformer, the voltage on the other side (across a resistor) is proportional to the displacement. They are suitable for small displacement measurements where friction is not allowed and precision is necessary.

Potentiometers are simply resistors that give a voltage output in response to displacements. They are suitable for small angle and displacement measurements.

Laser, microwave, or acoustic methods use different techniques such as the time of flight or frequency variations. They are generally used for large displacement measurements.

Optical and laser techniques use interferometric and optical gratings methods. They are a highly accurate measurement system but relatively more expensive.

FURTHER READING

Titles in instrumentation rarely cover specific sensor types. The following books will variously cover the sensors discussed here.

Brignell, J. and White, N. (1996) *Intelligent Sensor Systems*, Rev. ed., Institute of Physics, Bristol, CT; Philadelphia, PA.

Buchla, D., McLachlan, W. (1992) *Applied Electronic Instrumentation and Measurement*, McMillan Publishing Company, New York.

Coombs, C.F. Jr (2000) *Electronic Instrument Handbook*, 3rd edn, McGraw-Hill, New York (pp. 47.1–47.26).

Doebelin, E.O. (1990) *Measuring Systems. Application and Design*, 4th edn, McGraw-Hill, New York (p. 992).

Dyer, S.A. (ed.) (2001) *Survey of Instrumentation and Measurement*, John Wiley & Sons, New York.

Eren, H. (2003) *Electronic Portable Instruments – Design and Applications*, CRC Press, Boca Raton, FL.

Nachtigal, C.L. (1990) *Instrumentation & Control, Fundamentals and Applications*, *Wiley Series in Mechanical Engineering Practice*, Wiley-Interscience, Washington, DC (p. 890).

Soloman, S. (1999) *Sensors Handbook*, McGraw-Hill, New York.

Webster, J.G. (ed.) (2000) *Mechanical Variables Measurement – Solid, Fluid, and Thermal*, CRC Press, Boca Raton, FL.

Webster, J.G. (ed.) (1999) *Wiley Encyclopedia of Electrical and Electronic Engineering*, John Wiley & Sons, New York.

Webster, J.G. (ed.) (1999) *The Measurement, Instrumentation, and Sensors Handbook*, CRC Press, Boca Raton, FL.

192: Strain Sensors

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Basics of the Strain Variable 1268
2 Properties of Materials 1269
3 Installation and Calibration of Strain Gauges 1270
4 Electrical Resistance Strain Gauges 1273
5 Semiconductor Strain Gauges 1274
6 Compensation for Thermal Error 1274
7 Interrogation Circuitry 1275
8 Practical Points 1276
9 Signal Processing Systems 1278
10 Other Strain Measurement Methods 1279
Related Articles 1279
References 1280

1 BASICS OF THE STRAIN VARIABLE

An object is said to be strained when any of its length dimensions is altered by a small fraction from its nominal length. It can occur in tension or compression.

Strain ε is defined as a change in length ΔL, occurring in the length interval L when a force F is applied, the sense being positive for tensile strain and negative for compressive strain, Figure 1(a).

Being a ratio of two lengths, strain is a dimensionless number. To convey a more meaningful value, it is sometimes preferable to also state the units used as a ratio, such as μm/m. It can also be stated as a percentage but that can lead to confusion where the value has tolerance associated with the stated value. The term *microstrain* is often used, being a strain of 10^{-6}.

Strain is a three-dimensional property and hence is a vector quantity requiring three measurements to uniquely define its magnitude and direction. To ease measurement difficulty, they can often be devised to use a measurement in a known line or in a plane; it is rarely practical to make all three measurements at exactly the same point.

Several sources of strain can arise at the measuring point on an object under test. These are as follows:

- *Real or true strain* is the strain that is actually arising in the component. It is not necessarily produced only by mechanical force but can also occur because of temperature-induced forces.
- *Indicated strain* is the result at the output of the complete strain-gauge measuring system. Again, it is not necessarily the strain measurement that is needed because of instrumentation error.
- *Apparent strain* is the algebraic difference between real and indicated strain.
- *Mechanical strain* is the strain induced by mechanical forces, the measurement that is normally needed. Usually, it is the difference between real and thermal strain.
- *Thermal or temperature-induced strain* is caused solely by temperature, causing length change to the object according to its thermal coefficient of length, see Section 6.

Various delineated stresses exist, Figure 1(b):

Normal stress is a single component measurand occurring in pure tension or compression.

Bending stress arises when a rod is bent, causing the longitudinal fibers to elongate or compress as some function from a defined datum axis. The actual nature of the stress and the related strain is identical with normal stress.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

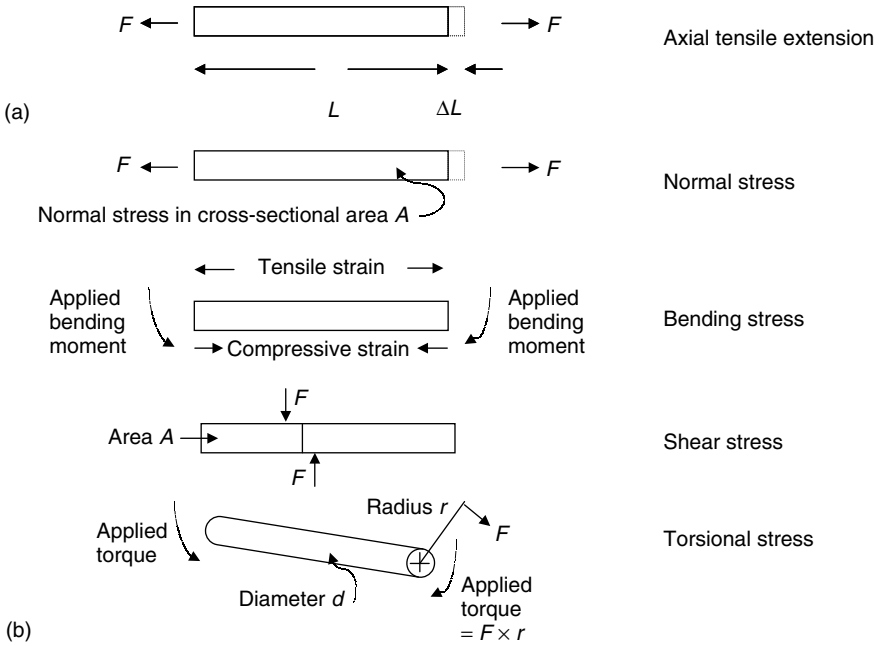


Figure 1. Types of stress and strain (a) definition of strain; (b) types of strain.

Shear stress is stress produced by strain invoked by forces acting to shear a bar along its axis of unbalanced forces.

Torsional stress is the stress arising when the subject is subjected to a rotational, twisting action.

Changes to shape because of the various stress situations all acting together produce complex stresses. Appreciation of these kinds of complex situations requires an advanced understanding of the strength of materials, which is not given here.

Finally, on this primer on the strain situation, correct application of strain gauges requires knowledge of material properties and calculation of the strength of materials data (Dally and Riley, 1991).

The following knowledge on strength of materials should suffice to appreciate how to apply strain gauges. Useful texts are Measurement Group Inc. (1988), Window (1992), and Khan and Wang (2000).

2 PROPERTIES OF MATERIALS

The elasticity of a material is characterized by its elastic or Young's modulus E where

$$E = \frac{\text{normal stress}}{\text{strain}} = \frac{\sigma}{\varepsilon} = P \times \frac{L}{A} \times \Delta L \qquad (1)$$

where P is the applied force, A the area of cross section of interest, and L the gauge interval over which the length change ΔL occurs.

This relationship only holds in the elastic, or Hooke's law region, of a material – *see* **Article 85, Elastic Regime of Design – Design Principles, Volume 2**. Stress levels can, therefore, be calculated from strain measurements. Direct sensing of stress is actually not easy to achieve.

Where the strain measured represents shear stress, the appropriate modulus to use for conversion into stress levels is the shear modulus G:

$$G = \frac{F}{\gamma} \qquad (2)$$

where F is the shear force applied and γ the angular shear distortion.

It can be shown that

$$G = \frac{E}{2(1+\mu)} \qquad (3)$$

where E is Young's modulus and μ is another material property constant known as *Poisson's ratio*.

Thus, appropriately placed strain gauges can measure the angular shear distortion γ, allowing computation of the existing force F at the point of measurement. From this, the stress level at that point can be deduced. Poisson's ratio μ is a measure of the amount to which a material will expand or contract in its longitudinal direction when compressed or tensioned transversely in that direction.

These considerations enable calculation of stresses from strain measurements in the ideal situation. Errors can arise

Table 1. Temperature coefficient of length for common construction materials.

Material	Temperature coefficient of length, α, $10^{-6}\,K^{-1}$
Ferritic steel	11
Austenitic steel	16
Aluminum	23
Titanium	9
Plastic	65
Molybdenum	5.4
Quartz	0

from many influence sources – *see* **Article 16, Reduction of Influence Factors, Volume 1** – the most prevalent being the effect of temperature.

All materials change dimension according to their coefficient of thermal expansion (TEMPCO), Table 1. While there exists some materials with very low values (such as fused silica, invar, carbon fiber), the materials used for construction (such as steels, concrete, glass, plastics, brick, wood, aluminum) have high values such that normal operating conditions produced significant strains in materials that will produce error on strain measurement if not catered for.

As an indication, steel expands at typically 16 parts per million per degree Centigrade. An ambient temperature change of, say, 30 °C will alter the size by 360 ppm, which, in a typical strain-gauge length of 5 mm, will produce a spurious absolute strain of 2.5 μm. Thus, temperature alone can generate strain without any external force being applied. Precision gauging needs to allow for this, especially where the object under test is subjected to a large temperature excursion, such as, for example, the temperature rise in the case of turbine blades in a jet engine.

Closely allied to this effect is the little known, but important, thermoelastic coefficient. This shows how Young's modulus varies with temperature. Steels, for example, become more elastic as their temperature increases, meaning that less stress is needed for the object to obtain a given strain excursion.

At times, another effect can arise if the material under test has significant magnetostriction. This is mentioned because steels with a nickel component (to make it stainless) might be considered for forming the test section (called a *load cell*). In such cases, it will experience a change in length if placed in a large DC magnetic field. For example, a load cell used to weigh the aluminum billets forming on an electrode in an electrolytic refinery process will exhibit this error significantly as the large through electrical current will generate large magnetic fields.

The third error source in strain-gauge measurements to watch for is the thermoelectric effect – *see* **Article 196, Thermocouple Temperature Sensors, Volume 3**. This arises when two dissimilar materials are subjected to a temperature differential along their length; it generates a small electric current due to the thermocouple (t/c) effect. Even a different batch of the same metal can sometimes exhibit a sufficient t/c effect such that small DC currents are generated in strain-gauge circuits that appear to be a strain. Wiring up sensitive strain-gauge circuits needs to allow for any t/c effect to ensure that it is small enough to be ignored.

With this background, it is now possible to appreciate the practical use of strain gauges.

3 INSTALLATION AND CALIBRATION OF STRAIN GAUGES

Strain gauges are produced in many forms of which the majority is a thin planar device glued, or (less used) clamped, to the surface of the test object, Figure 2. Their characteristics are covered for the two main types in Sections 4 and 5.

The major use of strain gauges is its application to a suitable surface of the object under test. This object could be a bridge support, the surface of a wing, or an automobile chassis member. They are also applied to rotating shafts, using slip rings or wireless communication to read their values. Sometimes, the object is specially formed into a load cell on which the gauges are applied; this permits the measurement of loads on crane hooks, in tensile test machines, weighing scales, portable tensile load measurement, and more.

To measure more than one of the possible strain components, it is necessary to apply a strain gauge for each component of strain. If it is known that pure tensile strain can be measured, then one gauge element is sufficient. In many cases, at least two components are required so manufacturers offer a range of multiple sets of gauges forming one unit – see Figure 3 for a small selection.

The sequence of tasks for a strain-gauge application that need early consideration is given in the following sequence:

1. Determine what kind of strain measurement is required. Determine how many axes are needed. Approximate calculations of expected stress and strain levels are performed to size up the gauge type that is needed. As a rule, the smaller the length of the gauge is, the lower is their sensitivity and the larger the effect of influences. Even this early, the designer needs to consider whether the task will be able to use a simple DC bridge interrogator – see Section 7. Use a gauge maker's catalog to see what is available.

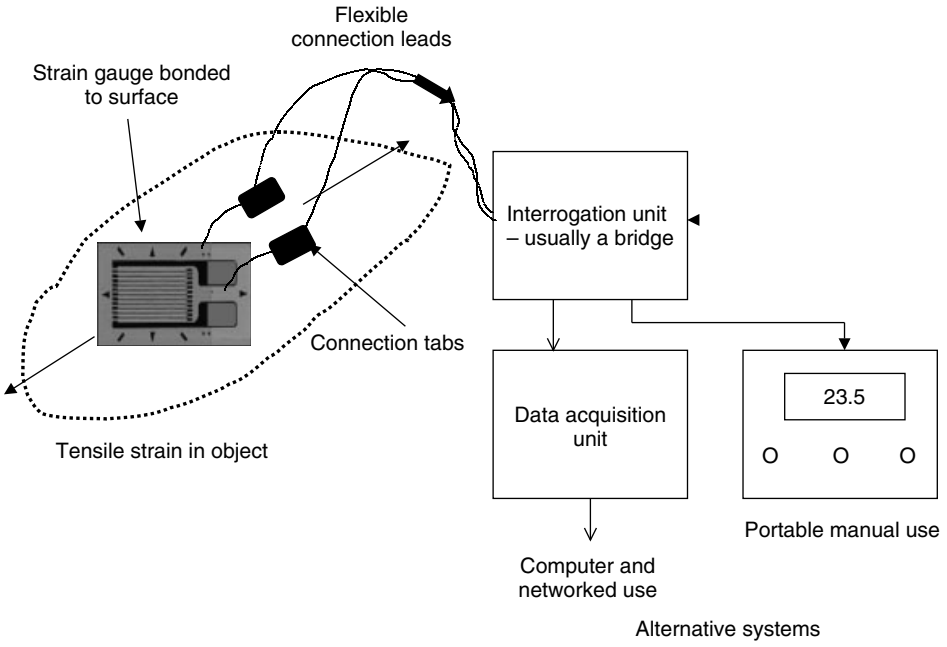
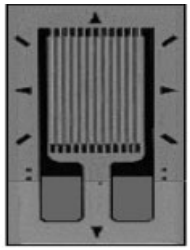


Figure 2. Basic units of a strain-gauge system.

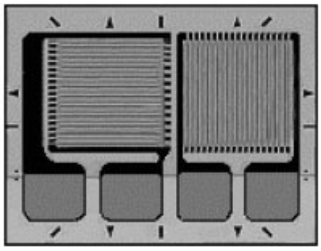
2. Decide where to place the gauge(s).
   This is dictated by there being sufficient strain to detect (a minimum of around 10-μ strain is needed as the discrimination) and there being good access for applying the gauge(s) where they can also be wired in and protected. The type of gauge initially selected is important as multiaxis types integrate several gauges into one unit that takes less space than when using several single-axis sensors. Orientation of the axes is likely to be more accurate in a self-contained unit than it would be when laying down a set of singles.
3. Learn how to apply them for a long reliable and repeatable lifetime.
   While the gauge type to be used is not yet decided, it is useful to look ahead at the process of application. Gauge makers often provide full pictorial instructions for cementing gauges to the various kinds of surfaces. The process needs some degree of hand skill, with lots of care and attention to detail. Every step they suggest is needed to obtain a long-lasting installation. The steps are as follows:

   (a) Roughen the application area to make it clean and ready.
   (b) Clean the area with the proprietary cleaner provided in the kit supplied.
   (c) Position the gauge with tape to hold it in place.
   (d) Apply the provided proprietary adhesive suited to the application.
   (e) Spread the adhesive with an applicator.
   (f) Press the gauge to the surface to displace spare adhesive. The thinnest possible layer of the adhesive is essential to keep self-heating effects to the minimum by the layer being less of a thermal insulator than the surface it is mounted upon.
   (g) Where possible, cure the adhesive with heat at the recommended time and temperature for the cycle time recommended. Where the gauge is not mounted on a portable object, heat lamps may be used. Follow the maker's recommendations.
   (h) Lay down the connection tabs and solder on the strain-gauge ribbons.
   (i) Solder in the special light gauge connection wires to the fixing cable tabs and secure the ends of the cables well.
   (j) Coat the gauge with the right kind of coating compound. This protects the gauge from a modest amount of mechanical impact and from chemical attack. Use the proprietary coatings supplied. Definitely, do not use the commonly available acetic acid–based silicon compounds as they will eat the gauge foil away in days!

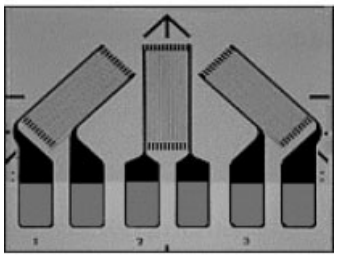
   If all goes well, the gauge readings, when first set up, will give a reasonably well-balanced bridge condition, at least within ±2%. If not, then something has gone wrong and application of a new gauge is needed from the beginning of the process given here. Take care not to test arm resistances with too high a voltage test set as that can open-circuit the gauge. Foils can be very small



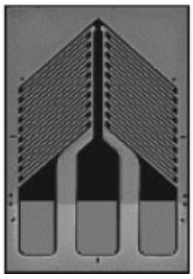
Uniaxial pattern



Biaxial 'tee' pattern



Three element rosette pattern



Shear strain pattern

Figure 3. Range of strain-gauge foils. (Images Copyright of Vishay Technology, Inc. Used by permission.)

in cross-sectional area and are thus unable to tolerate more than milliamps of current.

One's first application of a gauge is not likely to be totally successful. It is suggested that a trial run is made with a gauge mounted on a surrogate object to develop some experience. The extra cost and time usually pays off.

4. Establish what are the influence effects that are likely to introduce error.

   Here, one is looking to establish the temperature excursions that may arise from consideration of the minimum- and maximum-operating and storage/transport temperatures. The latter are often much larger than the operating range. These decide the type of adhesive needed. It also points up the kind of compensation required; it is always needed. A range of options are available. Extra gauges are used to overcome thermal error, see Section 6.

   Ensure that any t/c, magnetostrictive, and other unwanted effects are not large enough to corrupt the measurement values. Simple calculations will usually suffice to find the order of magnitude needed to make this design decision. Where they are significant, use of AC modulation interrogation methods may be the way to go – *see* **Article 184, Photon Counting, Volume 3**. They can usually overcome these. Good design, however, can often avoid the need for an AC system.

   Also, investigate what level of gauge mechanical protection is needed.

5. Decide the appropriate kind of interrogation circuit needed.

   Before any gauges are bought, the full design of the gauges and influence effects is needed as this will dictate the interrogation method and the number and type of gauges to be purchased. Choices are to purchase a suitable off-the-shelf interrogation box or to construct one from basic components to fit a certain space. Time, and available skills and experience will come into play here. Makers offer special IC units for building the block into an installation. Portable general-purpose instruments are offered for testing many gauges.

6. Decide how the influence errors will be controlled.

   Compensation for temperature will be needed and there are several means available – see Section 6. Decide which scheme is appropriate.

7. Consider how the installation will be calibrated initially and, thereon, over its useful life.

   Before committing to any purchases it is a good idea to look into how the units will be calibrated, for this could well influence some of the design decisions.

Any sensor, once installed, should be subjected to several tests to ensure it is operating correctly. The tests need also to ascertain that the sensor is only responding to the measurand of interest. These tests should establish the discrimination, stability, and repeatability before any calibration is carried out.

In theory, each strain gauge of a multiple gauge installation can be directly calibrated by applying a force that generates a known length change in the gauge. This method is fine for load cells and for other self-contained mechanical objects that can be set up to experience known forces.

Often, however, this is quite impractical because the strain gauge is mounted on an object that cannot be

deformed under the controlled conditions needed to perform a calibration. In such cases, an indirect method of calibration has to be employed. This is done by viewing the full system as being formed of two parts, the gauge set, and the rest of the electrical parts forming the interrogation system.

The strain gauge has its resistance changed by the input strain in a reasonably well-known manner. This means, for purposes of calibrating and setting the interrogation system, the gauge arm(s) can be replaced by a calibrated stepwise variable resistance(s) for which each of its increments of resistance takes the interrogator unit through known steps of simulated strain input. This setup will allow zeros and offsets to be set in the bridge and amplifiers and the overall gain to be established. This calibrates the electrical system leaving the only unknown, the gauge set. Gauge calibrators, constructed using precision resistors and low thermoelectric switch contacts are offered by manufacturers for setting up a system.

With the interrogation stage calibrated, it leaves the gauge set to be calibrated. The performance of a gauge when cemented in place can vary from the unbonded gauge value. Well-executed installations can be expected to reproduce the unbonded performance within a band from 0.04 to 5%. Thus, the weak link in the overall calibration of a strain-gauge sensor system, a fixed type of installation, is the gauge calibration. This limits accuracy of use in such cases. If the uncertainty in the calibration must be low, then the structure carrying the gauge must be built so as to allow the strain-gauge element of the structure cemented to its host object to be removed for traceable calibration in a quality test facility.

When all of the above issues are understood and roughly sized out, the complete design and installation details should start to become clear. This will take some time, for iteration will surely be needed.

4 ELECTRICAL RESISTANCE STRAIN GAUGES

The information given above is generic to all types of strain gauges. Now it is time to discuss the two mostly used types of electrical strain gauge – the metal resistance foil and the semiconductor element. Here, there are more design decisions to make.

In brief, the metal gauge is the most stable type, with more certainly known performance and linearity in operation. The semiconductor gauge has the main advantage of being much more sensitive but, otherwise it does not compete as well as metal gauges in precision open-loop uses.

The metal gauge is formed from fine (15–30 μm diameter) wire or thin metal foils (3–5 μm thick) etched to create a long path in a grid form. Several forms, Vishay (2004), are shown in Figure 3. The grid is made to give maximum sensitivity on one axis when strained with a much lower strain effect in the transfer direction. Active lengths of gauges range from 0.5 to 150 mm. Values of resistance offered range from 60 to 1000 Ω, the most commonly used being 120 and 350 Ω. Different gauges variously cater for strains from 0.1 to 5%.

They are offered with a range of pure and specially alloyed metals to allow them to be matched in TEMPCO to the material of the object under test. The carrier base of the gauge, being placed between the grid and the test object surface, is designed to ensure that the gauge surface moves with the object's surface. Once bedded down to the surface, special connector pads are added to the host for the connection wires.

The theory of the metal strain gauge is well understood. For a metal gauge, the resistance R of a thin element of conducting metal material is given by

$$R = \rho \frac{L}{A} \tag{4}$$

where ρ is the specific resistivity of the material, L the gauge interval, and A the cross-sectional area of the gauge filament or foil element.

When tensioned longitudinally, for a strain less than 2%, all parameters of the above expression will vary linearly giving

$$\frac{\Delta R}{R} = \frac{\Delta \rho}{\rho} + \frac{\Delta L}{L} - \frac{\Delta A}{A} \tag{5}$$

Using Poisson's ratio μ for the metal used gives

$$\frac{\Delta R}{R} = \frac{\Delta \rho}{\rho} + (1 + 2\mu) \times \frac{\Delta L}{L} \tag{6}$$

This shows that the resistance change is made up of a strain plus a component, depending on the metal used. The value μ for most metals is much the same, being around 0.3.

Rather than working with material constants that change a little with each batch of metal formed, gauge users prefer to use the gauge factor (symbols used vary greatly – K, k, F, G, GF, s, etc.). This is defined in direct engineering terms as

$$K = \frac{\Delta R/R}{\Delta L/L} = \frac{\text{Proportionate change in resistance}}{\text{Proportionate change in strain}} \tag{7}$$

From this, indicated strain ε from the gauge stage is

$$= K \times \Delta R$$

which is easy to apply once K is known. Note that $K \sim 2.0$ for metals.

A more defined value of K is obtained from production run samples and is stated with each gauge. Strain sensors formed with several single units should use matched sets of K for the best results. Provided the thermal effects are reduced by common-mode operation, the temperature effect in the gauge factor in metal gauges is small and linear.

5 SEMICONDUCTOR STRAIN GAUGES

Semiconductors also exhibit resistance change with strain. Here, two strain effects arise – mechanical and structural. A filament of silicon or germanium exhibits strain change as follows

$$\frac{\Delta\rho}{\rho} = \Pi_L \times \sigma \tag{8}$$

where ρ is the electrical resistivity of the material, Π_L is the longitudinal piezoelectric coefficient, and σ is the applied stress. These allow more design freedom, resulting in gauges with a higher sensitivity than the metal ones; there is, however, a price paid for, the higher the sensitivity, the greater are the effects of temperature on the gauge readings.

K for semiconductor gauges can be as high as 45 to 200. However, K is not a constant when it exceeds 45 or so. Compensation is needed as both temperature and the strain magnitude alter the K value for a gauge. The input–output relationship is also not linear with strain magnitude as K varies.

An advantage of the semiconductor gauge is that its temperature dependency can be tailored to become negative as well as positive. This lends them for use on bridges where cancellation compensation can be used, see Section 6.

6 COMPENSATION FOR THERMAL ERROR

A strain gauge mounted on a surface that experiences no mechanical strain can generate three dominant thermally induced strain error effects:

1. changes in the resistance of the gauge according to its temperature coefficient of resistance (TCR);
2. changes in the physical and mechanical size of the element according to its temperature coefficient of length α;
3. changes to the gauge factor K according to the thermal coefficient of the gauge factor (TCGF).

The sum effect of these error strain signals is called the *apparent strain*. Clearly, such signals need to be controlled, by design, to adequately small magnitudes.

Other errors that can arise because of temperature effects are lead resistance and interrogation bridge changes.

Temperature-error sources in the gauge itself can be largely controlled by gauge design. Manufacturers offer a range of gauges matched to widely used materials (Table 1). These values have an uncertainty of $\pm 1 \times 10^{-6}$ and are valid for use in the range 10 to 120 °C. However, the actual spot value will be affected by the actual temperature, so they should not be taken as constants.

Semiconductor gauges use two methods of compensation for thermal effects.

1. Matched gauges with complementary coefficient polarities, that is, the same resistance value for each, one having a negative temperature effect and the other being positive. This compensates for temperature and also doubles the bridge sensitivity.
2. Using two filaments that have opposite TCGFs.

In both cases, the performance is tailored to make the common-mode action closely matched as the actual temperature varies.

Temperature-induced mechanical changes can be compensated in three ways, as shown in Figure 4:

- a dummy gauge mounted on an unstressed component;
- a dummy gauge placed orthogonally to the active gauge;
- a second active gauge placed in the undergoing compression that matches the magnitude of the tension of the other. This is suited for bending situations. Load cells often use a bending moment, not simple tension, load sensing member for this reason.

Each has its merits, depending on the situation. For best results, the gauge pair used has to have matched properties. Makers offer sets of gauges in many different configurations for this reason.

It is also possible to add a temperature sensor to the material under stress, some temperature sensors being formed integrally with the strain-gauge unit. Measuring the surrounding air temperature is not a suitable means of obtaining the compensation temperature as it does not follow the test object's temperature in the dynamic state nor is it close enough to the actual temperature of the test object.

Valuable application notes are offered by gauge makers. These are well worth obtaining as they contain very detailed explanations of all manner of issues.

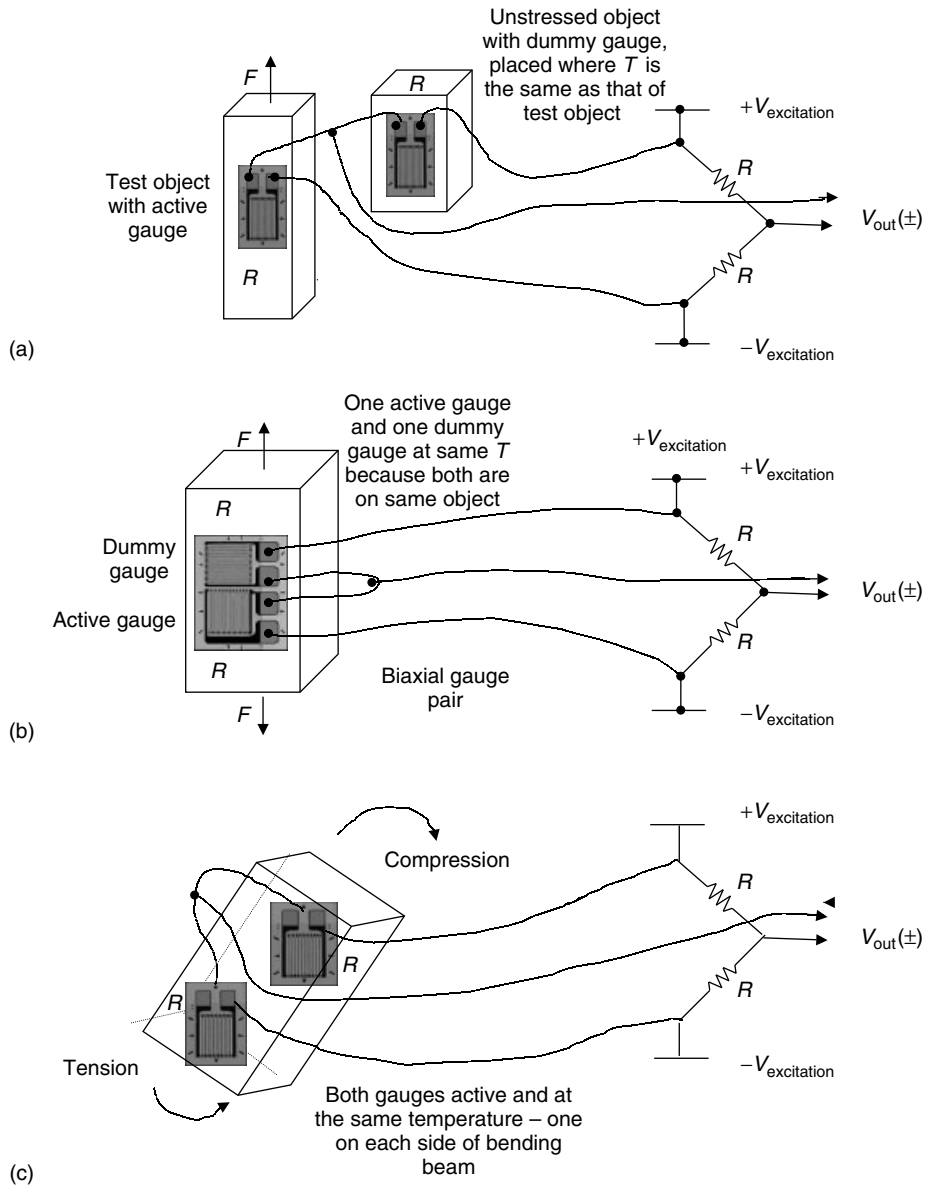


Figure 4. Temperature compensation for strain gauges (a) use of dummy gauge on separate object; (b) second gauge mounted to not experience strain; (c) use of two active gauges.

7 INTERROGATION CIRCUITRY

Resistance change of a metal gauge is proportional to strain, so interrogation needs a means to measure relatively small resistance changes with the required discrimination and uncertainty. Discerning applications can seek to determine resistance changes to as little as 0.01%, this being just 12 mΩ for a standard 120-Ω gauge. A good understanding of the circuitry needs is essential to reach such precision in practical use.

The most practical means for reading strain gauges is the Wheatstone bridge, Figure 5(a).

As discussed in **Article 126, Electrical Bridge Circuits – Basic Information, Volume 2**, the basic equation describing the four-arm bridge of Figure 5(a) is

$$\frac{V_1}{V_2} = \frac{V_3}{V_4} \tag{9}$$

This can be expressed in terms of the resistances

$$\frac{R_1}{R_2} = \frac{R_3}{R_4} \tag{10}$$

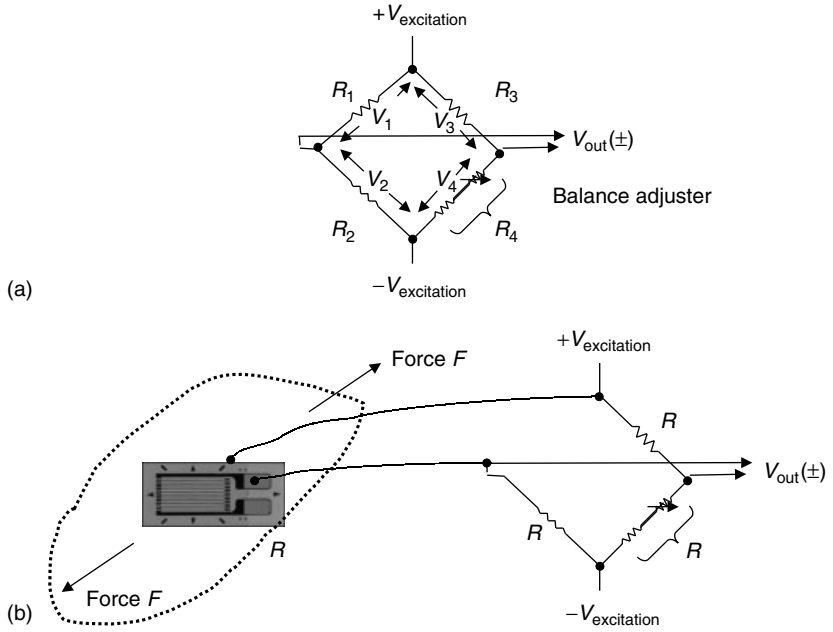


Figure 5. Circuit of Wheatstone bridge with parameters shown (a) circuit; (b) one arm active bridge.

The excitation voltage $V_{\text{excitation}}$ can be either in DC or AC form, the former being the one most commonly used, but it is not able to obtain the best performance.

When the resistance ratios are equal, the bridge output is zero and it is then in the balanced or null state. Assuming that the fixed resistances do not change, a variation in the value of a gauge changes the bridge, indicating a strain proportional to the unbalanced bridge output voltage V_{out}.

8 PRACTICAL POINTS

8.1 Bridge sensitivity and self-heating

The output measuring sensitivity of a strain-gauge bridge is directly proportional to the magnitude of the excitation voltage. This implies that large-value excitation could be used to obtain greater bridge sensitivity. There is, however, a limit to this.

What is not always appreciated is that even in the balanced condition, current flows through each arm of a bridge, the magnitude being decided by the total resistance of the arm and the magnitude of the excitation voltage. Thus, the gauge will experience a steady through current that heats it up according to the power being dissipated in each component.

It can be shown that the power W generated in the gauge is given by

$$W = V_{\text{excitation}}^2 \times \frac{R_1}{(R_1 + R_2)^2} \tag{11}$$

where, in Figure 5(a), R_1 is the gauge resistance, R_2 the other resistance in the same arm, and $V_{\text{excitation}}$ the excitation voltage.

This self-heating effect causes the gauge temperature to rise to a steady state temperature, also heating the host to a small extent, producing apparent strain. The amount of strain occurring will depend on the thermal conducting properties of the gauge, its mount, the host material, and the geometry. The actual effect is complex to model. Also, obtaining the thermal dissipation coefficient for gauges in given mounting situations is not always possible. Manufacturers of gauges sometimes offer useful advice here, HBM (2004) and Vishay (2004).

Excitation levels can vary from 1 to 40 V, or 5 to 25 mA. The larger gauges are able to dissipate more heat and, thus, can be driven harder.

A good way to test for excessive self-heating in a specific gauge installation is to monitor the strain-gauge sensor system output during its initial warm-up period. Any noticeable drift that took place when the gauge was first turned on after a long period of being off, which exceeded the allowable discrimination, indicated that the excitation was most likely too large.

In AC bridge operation, the equivalent voltage magnitude to be used is the rms value as that value represents the same power level as a DC signal of the same magnitude.

8.2 Lead error

If all gauges and resistors could be made with an exact resistance size, the bridge would balance. This cannot be achieved, so a variable balancing resistor is needed. Several means are employed, such as the use of

- a very small R in series in an arm (as shown in Figure 6a);
- a small R, center tapped, to allow resistance to be apportioned to each side of the bridge (as shown in Figure 6d);
- a high resistance R placed in parallel with an arm.

Several bridge configurations are commonly used. Many others have been reported for special purposes. The options are shown in Figure 6.

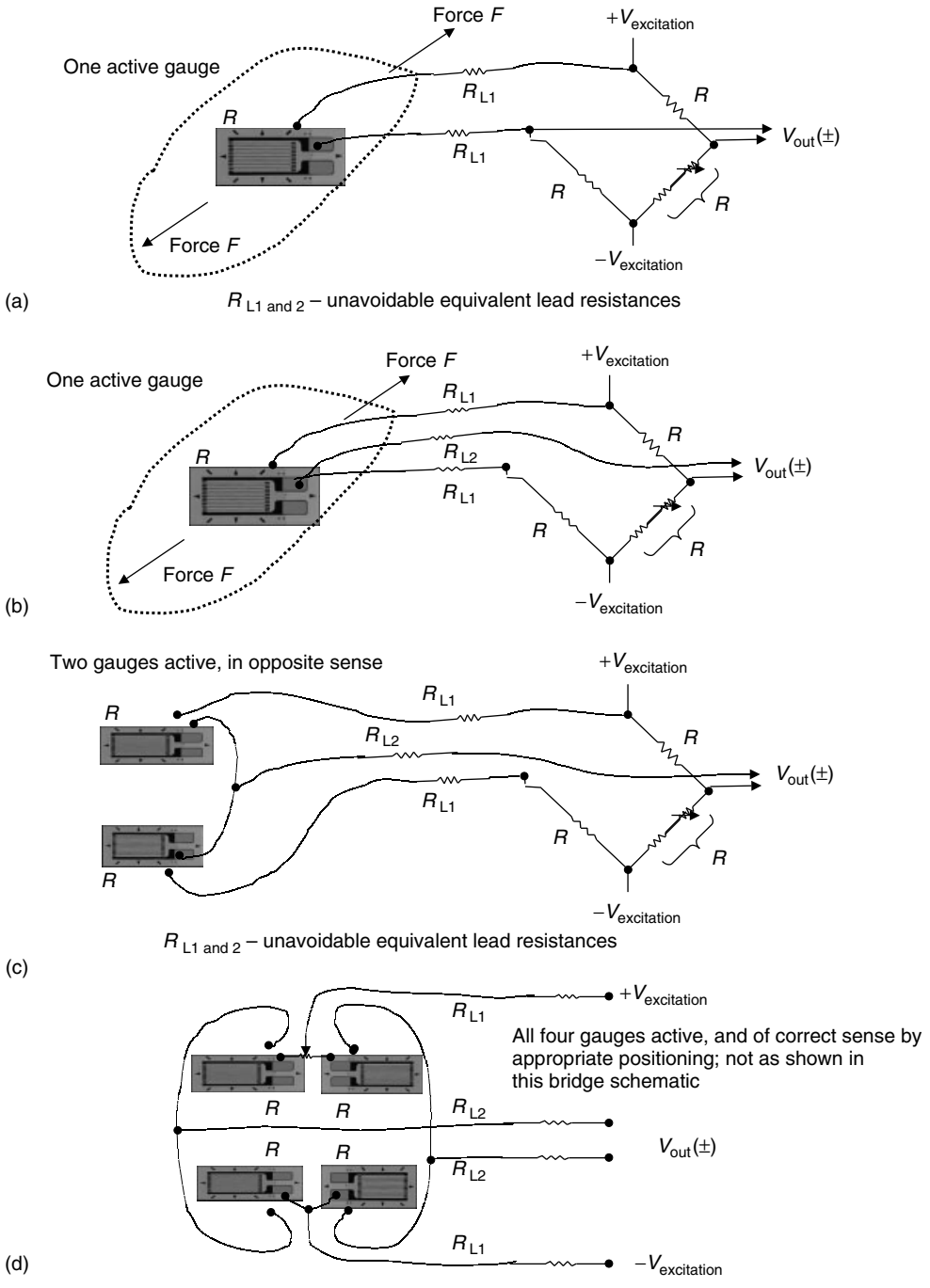


Figure 6. Reduction of lead error (a) no correction; (b) correction using 3 leads; (c) half bridge; (d) all arms active in full bridge.

The quarter bridge, Figure 6(a), is rarely used as it does not provide any temperature compensation. Also, note how the lead resistances are doubled and, thus, more temperature effect is added, possibly making the bridge a temperature sensor.

The single gauge system can be made more acceptable by adding three connection leads to the gauge, Figure 6(b). The leads R_{L1} for the gauge connection into the bridge must be identical in length and type. This provides common-mode rejection for the lead temperature effect. The third lead R_{L2} is in the output detector stage where its resistance is of little consequence.

A still better method is to use two active gauges with opposite strain senses, Figure 6(c). This is the most used method. The gauges are shown there as a schematic circuit wired to show how they form a bridge and not how they would necessarily be mounted.

An even more superior method is to make all gauges active, as shown on the schematic given in Figure 6(d). In this case, no lead resistance is significant, provided gauge connections are short.

In both cases, Figures 6(c) and (d), the gauges need to be placed to get the appropriate tension and compression. It is rarely possible to lay down four gauges that will all be active when they are all on the same surface. One exception is the use of a special foil that contains both radial and circumferential gauge grids that are used on a pinned edge circular diaphragm of a pressure gauge. Strain theory of such a membrane shows that the disk experiences both tension and compression on a single side in different places.

In any application, it is wise to make simple calculations to determine the effect of lead resistances, especially where they are long.

8.3 Transverse sensitivity

A gauge is often subjected to transverse strain. This effect on a gauge is kept minimal by their design but is not totally avoidable. Application of the gauge that is not exactly aligned with the longitudinal axis will create transverse action on the gauge.

With gauges mounted on an immovable host, its extent is hard to assess other than by carefully constructed calibration. Transverse effects can be expected to remain reasonably constant over time.

8.4 Dynamic response

In theory, a gauge becomes part of its parent material, and thus would exactly mimic the surface strains of the object being measured. In practice, however, the gauge does not move exactly with its host during dynamic operation because the bond allows some degree of compliance. Keeping the adhesive layer thin and using the recommended adhesive and curing process helps keep this effect to the minimum.

As a guide, upper limits for vibration measurement using strain gauges is 10 KHz for continuous oscillations and is higher up to 500 KHz for impulsive shocks. Application notes are available that detail how to get the best upper frequency response.

Periodic resetting is usually needed if the sensor system is pushed too hard. In long-term vibration measurement, the gauge must not suffer significant fatigue over the lifetime of its use. The constant reversing vibration will, however, result in some drift in the balance of the bridge. The main cause of this is relaxation of stresses in the host material, but the gauges also suffer from this effect. Gauge makers offer data on this, such as the recommended number of reversals for a given amplitude and frequency of reversals. Where possible, the design is derated as far as is reasonable to reduce this effect.

Some materials exhibit a noticeable amount of elastic after-effect wherein the residual strain in the host gradually relaxes back to an earlier state.

Fatigue behavior is also influenced by the adhesive used, its application, and the surface properties of the host material.

With AC interrogation systems, the modulating frequency should be around 10 times the highest frequency occurring in the host.

9 SIGNAL PROCESSING SYSTEMS

Strain signals can produce vast amounts of data, not all of which need to be captured. For example, vibration contains a large amount of data if detailed information of every cycle is recorded. For example, a 1-KHz sine wave signal logged at every 1.0% of amplitude for 100 min will produce close to 10^9 test data values. Processing all of this, while now being possible, is not always a sound way to proceed. It is better to decide just what needs to be known from the test and set up a more economical data extraction system.

Some special data sets needed in strain testing include peak values, strain pairs, frequency content, special event logging, and the period of vibration frequency.

Peak hold is a commonly used facility that captures the peak value of strain during a test, such as in destructive testing, where the breaking value is needed. This is a sampling process that tracks the maximum value. It can be achieved relatively simply in analog systems if short

hold time is satisfactory. It is a software-controlled process in digital signal processors (DSP).

In airframe testing, hundreds of hours of fatigue cycling are needed, so it is expedient to measure strain reversal pairs, for these are known to represent fatigue strength. A strain cycle counter is used.

A vibration is sometimes not at a steady single level but varies widely over time. Examples are in weighing moving railway cars where track noise adds to the situation as it passes over the automatic weighbridge. There, signal integration is needed to extract the wanted signal from the noisy situation.

Another data economizer design feature can be employed where many gauges are installed for relatively low frequency use, such as over the hull of a ship. A switching system is used to time-multiplex each strain gauge into a commonly used central interrogation unit. Strain gauges, being typically 120 Ω, need wiring, and they need switches that have much less resistance than this in the circuit when measuring.

Resetting offsets and zeros can be demanding, especially in multigauge installations. Application of smart sensor features – *see* **Article 160, Smart Sensor System Features, Volume 3** – can automate these operations. Process supervision is essential to monitor the degree of resetting that is taking place lest it be caused by a shifting parameter resulting from a significant error source, such as the gauge gradually becoming detached.

Today, the inclusion of sophisticated digital data processing allows the power of Fourier analysis and the like to be employed in smart ways to make decisions on the run, for example, as needed in machine health monitoring.

While the DSP parts of a strain sensor system have high proven reliability, the analog strain components and their practical use can be more problematic. Addition of smart features will assist in the creation of 'fit and forget' measuring systems but only if the gauge and interrogation circuits are soundly engineered and installed.

10 OTHER STRAIN MEASUREMENT METHODS

In the main, electrical strain gauges are the best way to determine stress where precision, or an electronic output, is needed. There are, however, many applications where electrical methods are not suited for cost, time to install, and installation reasons (Kuske and Robertson, 1974); (Asundi, 2002). Several of these methods might be useful:

Etched grid. Some surfaces are suitable for the application of a grid pattern, either by direct etching, printing, or gluing on a grid. Stress causes the grid to deform, the shape change being detected with a suitably graduated optical magnifier.

Brittle coating. These are applied coatings that crack if stressed beyond limits. Lacquer made from methylene chloride is one example. It is odorless, nonflammable, and is of low toxicity. It is sprayed on and left to air-dry. About 10 coats result in a film of 120 μm. The threshold strain level in room-temperature use is 500 $\mu m\, m^{-1}$.

As the sensitivity varies a lot with time and temperature, such methods need to follow manufacturer's guidelines.

A calibrator is offered for these methods that strains a sample piece in a small handheld tool.

10.1 Photoelasticity

Certain optically transparent dielectric materials exhibit optical property variations with stress level. This is the photoelastic effect.

When unstressed, the material shows no coloration when white light is used to illuminate it from behind. When stressed, the refractive index of the material changes, causing different colors to appear on the viewing side. The strain-optical coefficient K is available for the material. A special polariscope is used to view the object.

Photoelasticity can be used in several different ways to evaluate strains in a modeled member:

- modeling it in 2D
- modeling it in 3D
- coating the object.

An advancement on this method is to use scattered light photoelasticity. A sensitive optical radiation detector scans the object to produce a computer-based image that is interpreted with various schemes.

For these methods to work well enough, any residual stresses in the photo-optical material must first be removed. Stressing it in installation in the test rig will add error.

Overall, the method is useful for low-precision measurements where the object can be modeled and viewed in the special viewer.

RELATED ARTICLES

Article 117, Materials in Measuring Systems, Volume 2; **Article 191, Displacement and Angle Sensors Performance and Selection, Volume 3**; **Article 193, Specialty Displacement and Angle Sensors, Volume 3**; **Article 195, Characteristics of Temperature Measurement,**

Volume 3; **Article 206, Practice of Vibration Measurement, Volume 3**.

REFERENCES

Asundi, A. (2002) *MATLAB(R) for Photomechanics – A Primer*, Elsevier Science.

Dally, J.W. and Riley, W.F. (1991) *Experimental Stress Analysis*, McGraw-Hill Higher Education.

HBM. (2004) http://www.hbmhome.com.

Khan, A.S. and Wang, X. (2000) *Strain Measurements and Stress Analysis*, Prentice Hall, Upper Saddle River, NJ.

Kuske, A. and Robertson, G. (1974) *Photoelastic Stress Analysis*, Wiley, New York.

Measurement Group Inc. (1988) *Strain Gage Based Transducers: Their Design and Construction*, Measurements Group Raleigh, NC.

Vishay. (2004) http://www.vishay.com/brands/measurements_group/strain_gages.

Window, A.L. (1992) *Strain Gauge Technology*, Kluwer Academic Publishers.

193: Specialty Displacement and Angle Sensors

Halit Eren
Curtin University of Technology, Perth, Western Australia, Australia

1 Special Needs for Displacement and Angle Sensors 1281
2 Principles and Specialty Displacement, and Angle Sensors 1282
3 Specialty Displacement and Angle Sensor Types and their Applications 1282
4 Sensitivity and Interference Issues 1285
5 Proprietary Specialty Displacement and Angle Sensors 1285
Further Reading 1286

1 SPECIAL NEEDS FOR DISPLACEMENT AND ANGLE SENSORS

This article continues discussion from **Article 191, Displacement and Angle Sensors Performance and Selection, Volume 3** in which the performance and selection of the most commonly used methods of measurement of displacement and angle are discussed.

As technology improves, many different special needs for displacement sensors appear, mainly depending on the application requirements. For example, in the automotive industry, capacitive noncontact displacement and thin gap sensors are being used often to monitor vehicle condition for safety. As in the case of fuel injection systems, due to technological improvements, the dimensions of components get smaller hence the sensors also get smaller to fit these confined spaces. Sensors often are required to fit into locations with diameters <1 mm, where they measure gaps in submillimeter region. Other special needs for the displacement sensors include high operating temperatures up to 1000 °C, small overall diameter within 100 μm, high and consistent accuracy down to 0.2 μm, immunity to magnetic fields, and so on.

In many industrial applications, displacement, angle, and gap sensors are required to have the following typical characteristics: noncontact measurements, ability to withstand high temperatures and pressures, minimum calibration requirements, operating in unusual and hostile environments such as oils or other fluids, very small size for installation in tight spots, good immunity to magnetic and electromagnetic interference, and so on. Magnetic type displacement sensors, in particular, need careful shielding against external interferences.

Optical and laser displacement sensors require careful designing and construction. They need highly stabilized light sources and the reflective index of light medium must be factored in carefully in the design process.

By suitable constructions and application of appropriate signal-processing techniques, most of the displacement sensors can be used in acceleration and vibration measurements, and velocity, inclinometer, and rotation sensors.

Optical encoder type displacement sensors are often used: they can be configured and manufactured in an amazing number of geometric forms. In those using fringing effects, careful design and precision building is necessary during constructions.

Large-scale displacement sensors that are based on optics, electromagnetic, sonic, and microwaves, generally need very bulky and expensive electronic support units for signal generation and signal processing.

The trend in small-scale displacement sensors is the miniaturized sensors based on microelectronics. Many modern microdisplacement sensors can be classified as smart

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

sensors. These sensors need the support of nonconventional components plus the software to drive them. They can be networked using techniques such as the IEEE 1451 standards for connecting transducers to digital networks. They can be connected to a variety of networks including Ethernet interfaces that will support TCP/IP, FTP, and HTTP protocols with a fully functional Web server, *see also* **Article 162, Principles of MEMS, Volume 3**; and **Article 163, Uses and Benefits of MEMS, Volume 3**.

2 PRINCIPLES AND SPECIALTY DISPLACEMENT, AND ANGLE SENSORS

Specialty displacement and angle sensors are often based on capacitive, magnetic, optical piezoelectric, and radio frequency (RF) and microwave propagation principles. They are briefly visited here. The principles and theory of some of the lesser-used displacement angle sensors are explained below.

2.1 Capacitive displacement sensors

The basic sensing element of a typical capacitive displacement sensor consists of two simple electrodes with capacitance C between them. Capacitance is a function of the distance d (meters) between the electrodes of a structure, the surface area A (square meters) of the electrodes, and the permittivity ε ($8.85 \times 10^{-12}\,\mathrm{F\,m^{-1}}$ for air) of the dielectric between the electrodes; therefore,

$$C = f(d, A, \varepsilon) \tag{1}$$

There are three basic methods for realizing a capacitive displacement sensor, namely, by varying d, A, or ε. The capacitive displacement sensors can be classified as variable plate-distance, variable-dielectric, variable plate-area, and differential capacitive sensors. More detail in capacitance measurement is found in **Article 204, Capacitance and Inductance Measurement, Volume 3**.

2.2 Magnetic displacement sensors

The majority of the magnetic displacement sensors are inductive types and they are based on Faraday's law of induction. They can be classified as reluctance, transformative, tachogenerator, induction potentiometer, eddy current, and magnetometer types. They are designed to measure linear as well as rotary displacements.

Magnetic displacement sensors can be classified as linear variable-reluctance, single-coil linear variable-reluctance, variable-differential reluctance, linear variable-inductive, variable-coupling, eddy current displacement sensors, and linear variable-differential transformers.

2.3 Resistive, piezoelectric, and piezoresistive displacement sensors

The resistive displacement sensors consist of a sliding or rotational arm of a potentiometer that is connected to the moving object while the main body of the potentiometer is kept in a fixed position. In strain gauge, displacement sensors are based on resistance properties of electrical conductors. Piezoelectric sensors are based on the change in electrical properties of materials under stress. In the case of piezoresistive displacement sensors, when a conductor is stretched or compressed, its resistance alters because of two-dimensional changes and changes in the fundamental electrical properties of the material. Resistance measurement is covered in **Article 203, Resistance Measurement, Volume 3**.

2.4 Optical displacement sensors

There are many different types of optical displacement sensors available, such as those that make use of properties of optical fibers, laser methods, encoders, and interferometric techniques.

2.5 Angular displacement sensors

Angular displacement sensors generally involve rotational capacitive or magnetic devices. The most commonly used devices involve tachogenerators, microsyns, synchros, and induction potentiometers.

3 SPECIALTY DISPLACEMENT AND ANGLE SENSOR TYPES AND THEIR APPLICATIONS

3.1 Variable-dielectric capacitive displacement sensors

The displacement may also be sensed by the relative movement of the dielectric material, attached to the object under

investigation, between the plates. The corresponding equations will be

$$C = \varepsilon_0 w\{\varepsilon_2 l - [\varepsilon_2 - \varepsilon_1]x\} \quad (2)$$

where ε_1 is the relative permittivity of the dielectric material and ε_2 is the permittivity of the displacing material (e.g. liquid).

In this case, the output of the transducer is linear. These types of sensors are often used in level measurement of fluids.

3.2 Variable-coupling magnetic displacement sensors

Variable-coupling sensors consist of a frame holding a center-tapped coil and a ferromagnetic plunger. The plunger and the two coils have the same length l. As the plunger moves, the inductances of the coils change. The two inductances are usually placed so that they form two arms of a bridge circuit with two equal balancing resistors. The bridge is excited with 5 to 25 V AC at a frequency of 50 Hz to 5 kHz. At the selected excitation frequency, the total transducer impedance at null is set to 100 to 1000 Ω ranges. The resistors are chosen to have about the same value as the transducer impedances. The load for the bridge output must be at least 10 times the resistance R-value. When the plunger is in the reference position, each coil will have equal inductances of value L. As the plunger moves by δl, changes in inductances $+\delta L$ and $-\delta L$ create a voltage output from the bridge. By constructing the bridge carefully, the output voltage may be made a linear function of the displacement of the moving plunger, within a rated range. Bridges are covered in **Article 126, Electrical Bridge Circuits – Basic Information, Volume 2**; and **Article 127, Unbalanced DC Bridges, Volume 2**.

3.3 Eddy current displacement sensors

Inductive transducers based on eddy currents are mainly probe types, containing two coils. One of the coils, known as the *active coil*, is influenced by the presence of the conducting target. The second coil, known as the *balance coil*, completes the bridge circuit and provides temperature compensation. The magnetic flux from the active coil passes into the conductive target by means of a probe. When the probe is brought closer to the target, the flux from the probe links with the target, producing eddy currents within the target.

The eddy current density is greatest at the target surface and becomes negligibly small about three skin depths below the surface. The skin depth depends on the type of material used and the excitation frequency. Even though thinner targets can be used, a minimum of three skin depths may often be necessary to minimize temperature effects. As the target comes closer to the probe, the eddy currents become stronger, causing the impedance of the active coil to change, altering the balance of the bridge in relation to target position. This unbalanced voltage of the bridge may be demodulated, filtered, and linearized to produce a DC output proportional to the target displacement. The bridge oscillation frequency may be as high as 1 MHz. High frequencies allow the use of thin targets and provide good system-frequency response.

Probes are commercially available with full-scale ranges from 0.25 to 30 mm with a nonlinearity of 0.5% and a maximum resolution of 0.1 μm. For nonconductive targets, conductive materials of sufficient thickness must be attached to the surface by means of commercially available adhesives. Because the target material, shape, and the like influence the output, it is necessary to calibrate the system statistically for a specific target. The recommended measuring range of a given probe begins at a standoff distance equal to about 20% of the stated range of the probe. In some cases, a standoff distance of 10% is recommended and so this is calibrated as the standard for the system. A distance greater than 10% of measuring range can be used as long as the calibrated measuring range is reduced by the same amount.

3.4 Piezoresistive displacement sensors

Piezoresistive displacement sensors are essentially semiconductor strain gauges with large gauge factors. High gauge factors, *see* **Article 192, Strain Sensors, Volume 3**, are obtained because the material resistivity depends primarily on the stress, not on dimensions only. Most piezoresistive sensors use two or four active gauges arranged in a Wheatstone bridge. Extraprecision resistors are used as part of the circuit in series with the input to control the sensitivity, balancing, and offsetting temperature effects. The mechanical construction of a piezoresistive displacement sensor is similar to the installation in Figure 1. In some applications, overload stops are necessary to protect the gauges from high-amplitude inputs.

3.5 Synchros

The term *synchro* is associated with a family of electromechanical devices. The construction of synchros is similar

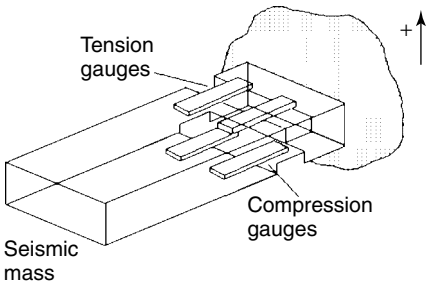


Figure 1. Bonding of piezoelectric and piezoresistive elements on to an inertial system.

to that of wound-rotor induction motors. The rotation of the motor changes the mutual inductance between the rotor coil and the three stator coils. The three voltage signals from these coils define the angular position of the rotor. Synchros are used in connection with a variety of devices, including control transformers, Scott T transformers, resolvers, phase-sensitive demodulators, and AD converters.

In some cases, a control transformer is attached to the outputs of the stator coils such that the output of the control transformer produces a resultant magnetomotive force (mmf) aligned in the same direction as that of the rotor of the synchro. In other words, the synchro rotor acts as a search coil in detecting the direction of the stator field of the control transformer. When the axis of this coil is aligned with the field, the maximum voltage is supplied to the transformer.

In other cases, AC signals from the synchros are first applied to a Scott T transformer, which produces AC voltages with amplitudes proportional to the sine and cosine of the synchro shaft angle. It is also possible to use phase-sensitive demodulation to convert the output signals to make them suitable for digital signal processing.

3.6 Induction potentiometers

A version of the rotary-type linear inductor is the induction potentiometer. Two concentrated windings are wound on stator and rotor. The rotor winding is excited with alternating current, thus inducing voltage in the stator windings. The amplitude of the output voltage is dependent on the mutual inductance between the two coils, where the mutual inductance itself is dependent on the angle of rotation. For concentrated coil-type induction potentiometers, the variation of the amplitude is sinusoidal, so linearity is restricted to the region of the null position. By carefully designing coils, a linear variation over $\pm 180^\circ$ may be obtained.

Standard commercial induction pots operate in the 50 to 400 Hz frequency ranges. They are small in size, from 1 to 6 cm, and their sensitivity can be in the order of 1 V/1° of rotation. Although the range of induction pots is limited to less than 60° of rotation, it is possible to measure displacements in angles from 0° to full rotation by suitable arrangement of a number of induction pots. As in the case of most inductive sensors, the output of the induction pots may need phase-sensitive demodulators and suitable filters. In many inductive pots, additional dummy coils are used to improve linearity and accuracy.

3.7 Interferometric sensors

There are many different types of interferometric sensors including optical fiber, laser, holographic, and monochromatic light types. In these sensors, Michelson, Sagnac, or Fizeau effects of interference of two or more light beams is used by suitable optical arrangements. The most commonly used optical arrangements are beam splitters, polarizing prisms, and diffraction gratings.

As two light beams interfere, interactions of light cause a series of bright and dark lines. If, for example, two interfering lights with wavelengths 1.5 μm are interfered, the spacing of bright and dark lines will be 0.75 μm. A suitable arrangement for an interferometric sensor is given in Figure 2. In many cases, the dark and bright points are counted in relation to the displacement of an object located within the interference. Many different designs based on different displacement mechanisms are available. Some of these include fringe counting interferometers, frequency modulation interferometers, heterodyne interferometers, phase-locked interferometers, laser Doppler and laser feedback interferometers, and fiber interferometers.

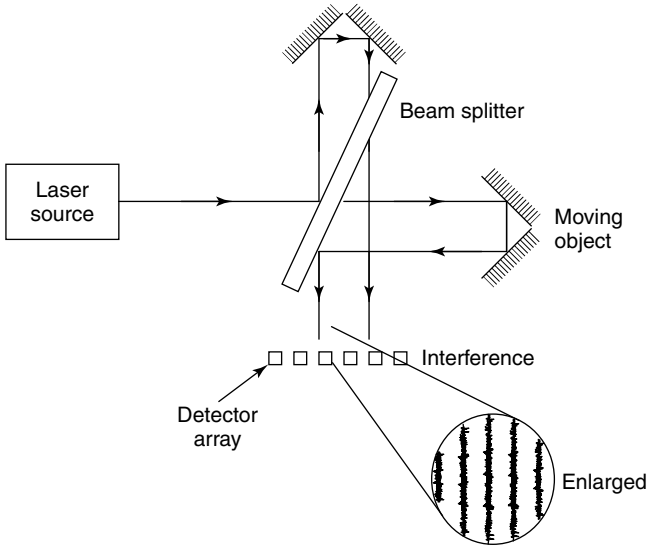


Figure 2. Interferometric displacement sensor.

4 SENSITIVITY AND INTERFERENCE ISSUES

Sensitivity and interference of displacement sensors depends on the type of sensor being used and the nature of the application environment. In the case of potentiometric displacement sensors, the sensitivity depends on the characteristics of resistive elements, the way that resistive elements are constructed, mechanical mounting methods, the linearity of the device, and the electrical loading. In these sensors, the electric, magnetic, and electromagnetic interference is minimal compared to others.

In magnetic displacement sensors, the influence of external magnetic fields and the internal magnetic fields due to coil orientation is significant as they directly effect the operation ranges and the sensitivity. These devices may need careful magnetic shielding. In many cases, multiple shields may be necessary.

Capacitive sensors may be prone to external and internal electric fields. In the design and construction stages, the fringing effect of the electric field must be taken care of carefully. The external electric field must carefully be eliminated. For this, carefully designed signal-processing techniques are used by selecting appropriate charge and current amplifiers. In many cases, appropriate liberalization techniques are employed to improve sensitivity and eliminate interference.

Piezoelectric sensors find diverse applications. They generally need to be mechanically robust. Since these transducers basically operate on electric field principles, they need to be shielded against the external fields. In some cases, instead of using ceramic type piezoelectric sensors, ferromagnetic polymer types that have high voltage sensitivity may be selected, thus minimizing the effect of external fields. In many cases, high-input impedance amplifiers are used to eliminate the influence of connecting cables.

As discussed above, the optical and laser sensors need stable power sources and good components to achieve maximum sensitivities. In these devices, the electric, magnetic, and electromagnetic interferences are minimal.

Sonic displacement sensors require a uniform transmission medium. If the transmission medium is not uniform, appropriate signal-processing techniques must be employed to maximize the sensitivity.

RF and microwave sensors are generally used for long-distance displacement measurements. They are prone to static charges in the atmosphere and variations in transmitted medium, and many other factors. Careful signal-processing techniques can eliminate most of the problems.

5 PROPRIETARY SPECIALTY DISPLACEMENT AND ANGLE SENSORS

Proprietary and lesser-used displacement sensors are constructed for two main reasons:

1. Owing to specific applications in which conventional sensors are not suitable.
2. Because of the availability of new technology and supporting software, with which companies seek competitive edge in their products.

It is not possible to cover a whole range of proprietary sensors in this article; however, some examples will be given.

Example 1 In the construction industry, traditionally, large concrete and steel structures must be monitored for years, not days. The environment is often hostile, and some changes (such as corrosion) may occur slowly, whereas seismic disturbances can occur quickly. As a result, the cost of using conventional sensors to achieve sustainable useful results under these conditions may not be desirable. A family of specially designed proprietary displacement sensors can monitor the construction continuously and can be connected to a digital network to enable remote monitoring. The technology behind such proprietary sensors can be based on various principles. The most common types are strain gauge, GPS, laser, and optical technologies.

Example 2 A new displacement sensor technology is appearing as a byproduct of CCD (charged couple device) technology. State-of-the-art CCD sensors, in combination with digital signal processors, make use of laser technology, offering many advantages in signal processing and precision aspects. These sensors are basically pixelized array detectors with 1024 discrete voltages representing the amount of light on each element of the detector. A CCD element detector can carry 1024×1024 light intensity information.

Example 3 There are many different types of proprietary microsensors offered by many companies. For example, Micro-Electro-Mechanical-Systems, MEMS-based sensors are available for a diverse range of applications from monitoring to treat paralyzed limbs of humans to orientation determination of spacecrafts. MEMS sensors became increasingly available for sensing parameters, such as acceleration, angular rate, low level magnetic fields, and so on.

The literature supporting these various devices does not contain material under specific devices but gives it in a general manner. It is, therefore, necessary to scan through texts to find material. The following bibliography lists sources.

FURTHER READING

Brignell, J. and White, N. (1996) *Intelligent Sensor Systems*, Rev. ed., Institute of Physics, Bristol, CT; Philadelphia, PA.

Buchla, D. and McLachlan, W. (1992) *Applied Electronic Instrumentation and Measurement*, McMillan Publishing Company, New York.

Bentley, J.P. (1988) *Principles of Measurement Systems*, 2nd edn, Longman Scientific and Technical, Burnt Mill.

Coombs, C.F. Jr. (2000) *Electronic Instrument Handbook*, 3rd edn, McGraw-Hill, New York (pp. 47.1–47.26).

Doebelin, E.O. (2001) *Measurement Systems: Application and Design*, 5th edn, McGraw-Hill, Singapore.

Dyer, S.A. (ed.) (2001) *Survey of Instrumentation and Measurement*, John Wiley & Sons, New York.

Gopel, W., Hesse, J. and Zemel, J.N. (1989) *Sensors–A Comprehensive Survey*, WCH, Weinheim.

Holman, J.P. (1989) *Experimental Methods for Engineers*, 5th edn, McGraw-Hill, Singapore.

Nachtigal, C.L. (1990) *Instrumentation & Control, Fundamentals and Applications*, *Wiley Series in Mechanical Engineering Practice*, Wiley-Interscience, Washington, DC (p. 890).

Soloman, S. (1999) *Sensors Handbook*, McGraw-Hill, New York.

Webster, J.G. (ed.) (2000) *Mechanical Variables Measurement-Solid, Fluid, and Thermal*, CRC Press, Boca Raton, FL.

Webster, J.G. (ed.) (1999) *Wiley Encyclopedia of Electrical and Electronic Engineering*, John Wiley & Sons, New York.

Webster, J.G. (ed.) (1999) *The Measurement, Instrumentation, and Sensors Handbook*, CRC Press, Boca Raton, FL.

194: Large-scale Metrology

Stephen Kyle
University College London, London, UK

1 Definition and Range of Large-scale Metrology 1287
2 Summary of Methods Used in Large-scale Metrology 1288
3 Theodolites and Total Stations 1290
4 Tracking Laser Interferometers 1291
5 Cameras and Photogrammetry 1292
6 Calibration of Instruments 1295
7 Orientation Techniques 1297
8 Standards and Good Practice 1299
9 Sample Applications 1300
References 1302

1 DEFINITION AND RANGE OF LARGE-SCALE METROLOGY

Large-scale metrology (LSM) concerns itself with precise 3-dimensional industrial measurements at 10 micron to submillimeter accuracies over ranges of 10s of meters. It makes considerable use of optical techniques, which are the subject here – *see* **Article 89, Optical Materials, Volume 2**; **Article 90, Optical Elements, Volume 2**; **Article 91, Light Sources and Detectors, Volume 2**; and **Article 92, Optical Measuring Instruments, Volume 2**. Testing of optical measuring equipment itself is found in **Article 93, Testing Optical and Other Radiation Systems, Volume 2**. Other LSM techniques are referenced below. Fine mechanical measuring devices are covered in **Article 80, Principles of Fine Mechanics – Kinematic and Elastic Designs, Volume 2**; **Article 81, Principles of Fine Mechanics – Systems Considerations, Volume 2**; **Article 84, Kinematical Regime – Rotary Motion, Volume 2**; **Article 85, Elastic Regime of Design – Design Principles, Volume 2**; **Article 86, Elastic Regime of Design – Spring Systems, Volume 2**; **Article 87, Elastic Regime of Design – Plates and Bimorphs, Volume 2** with testing of these in **Article 88, Error Sources in Fine Mechanics, Volume 2**.

The application spectrum ranges from high accuracy aerospace and automobile manufacturing to lower accuracy ship construction and as-built process plant. LSM methods developed out of a need for mobile, on-site measurement. This contrasts with more established techniques typically represented by CMM (coordinate measuring machine) technology.

CMMs have a probe that can slide along three mutually perpendicular axes, see Figure 1. The probe, often a mechanical touch probe, contacts the surface of an object placed on the machine and so defines a point on the object. Encoders on the slides record the probe's position. A simple tutorial is available at http://www.brownandsharpe.com/introduction/intro_en.pdf.

CMMs can offer accuracies of micrometers (microns) but have a fixed volume, a fixed location, often a few meters maximum dimension, and usually need a carefully controlled operating environment. They cannot readily deal with measurement tasks such as

- operating in a hazardous environment,
- measuring dynamic robot performance,
- checking dimensions on the root end of an aircraft wing.

Geometric methods and optical equipment originally designed for making topographical maps can solve these

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.



Figure 1. Coordinate measuring machine in use at NPL. (Crown Copyright, courtesy of NPL.)

problems and they form the core of LSM today. The instrumentation in LSM systems is inherently transportable and therefore applicable to on-site measurement. Fully electronic and automated features ensure that 3D data can be delivered online or within a short timescale.

LSM literature is limited except in photogrammetry (Section 5). Good texts are Atkinson (1996) and, in German, Luhmann (2000), Schwarz (1995) and Löffler (2002).

Not discussed here, owing to lack of space, are several closely related and complementary technologies:

- *Optical tooling* is based on the use of telescopes, prisms, and mirrors to define straight, parallel, and orthogonal lines for component assembly and alignment. Manufacturers Brunson have introductory tutorial information at http://www.brunson.us/p/LearnCenter.htm.
- *Laser radar scanners* can digitize large volumes in 3D by scanning a laser beam across the scene. Although accuracies are appropriate for, say, architecture, and civil engineering, systems offering metrology precision also exist. For an example, see www.metricvision.com.
- *Portable arm CMMs* are manually operated, mechanical arms that define a limited 3D measurement volume, extendable by additional techniques. They can avoid the line-of-sight requirement for optical systems. Web sites of some major manufacturers can be found at www.faro.com, www.romer.com, and www.romer.fr.

2 SUMMARY OF METHODS USED IN LARGE-SCALE METROLOGY

2.1 Triangulation

Triangulation locates a target using two or more directions measured from separate locations known relatively to one another. The directions are typically defined by camera views, telescope pointings, and laser beams. One direction places the target on a line in space; the second intersects this to complete the 3D location. Figures 2 and 3 illustrate triangulation by theodolite (see Section 3)

Multiple intersections can improve accuracy. This is a particular feature of photogrammetry (see Section 5), where camera views define target directions and where it is simple to take views from many different locations. Key features are as follows:

- For online work, a setup phase is required to determine relative instrument positions.
- Narrow intersection angles give poor results – 90° intersections are good.
- Measured scale bars of accurately known length provide true scale.

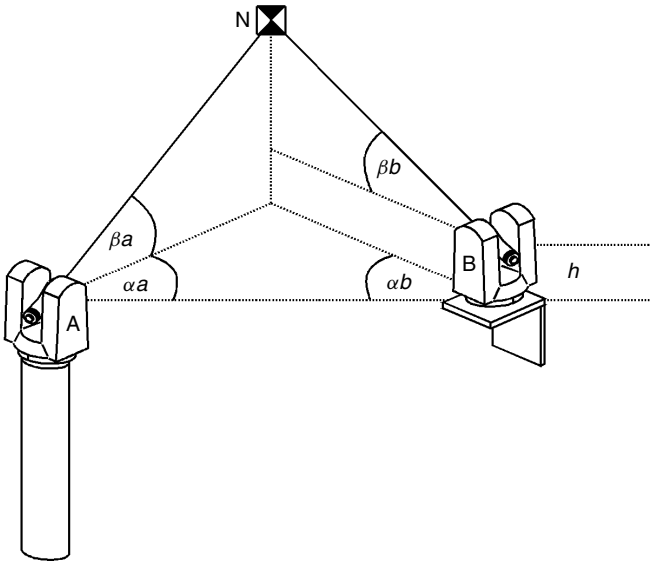


Figure 2. Triangulation by theodolite. (Reproduced by permission of Leica Geosystems.)

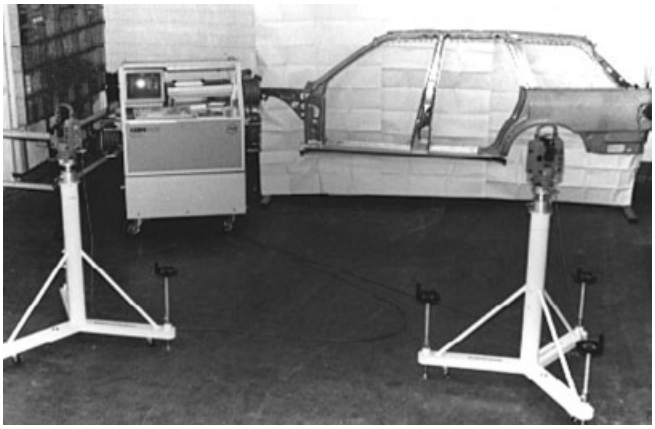


Figure 3. Audi panel measurement, 1996. (Reproduced by permission of Leica Geosystems.)

2.2 Polar measurement

Polar measurement locates object features in 3D space by direct measurement of spherical coordinates, that is, horizontal angle, vertical angle and range. Figures 4 and 5 show polar measurement by laser tracker (see Section 4).

Unlike triangulation, 3D measurement can be made from a single instrument and a setup procedure is not required. Scale is provided by range values in every measurement.

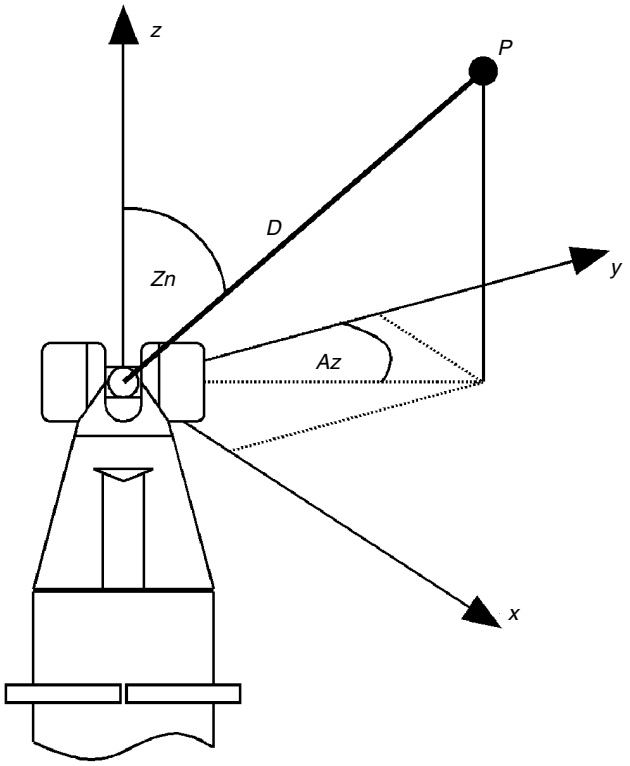


Figure 4. Polar measurement by laser tracker. (Reproduced by permission of Leica Geosystems.)



Figure 5. Robot tracking from a single position. (Reproduced by permission of Leica Geosystems.)

2.3 Multistation networks and bundle adjustment

Multiple instrument positions are a feature of triangulation measurements. Since optical instruments require line of sight, polar measurement generally also requires multiple instrument positions to view all sides of an object.

The network is commonly defined by means of a bundle adjustment. This was originally devised for photogrammetric triangulation in which the 'bundles' of target rays from each camera position were mathematically fitted together in order to obtain target coordinates as well as camera locations and orientations. The technique is easily extended to deal with pointings based on angles rather than image coordinates, and to include range measurements.

2.4 Trilateration

If range measurements *only* are made to a target from a minimum of 3 instruments with known relative positions, the target can be fixed in 3D space as the intersection of 3 spheres whose radii are the 3 range values. This is known as *trilateration*. Multilateration implies the use of multiple ranges.

Trilateration is not widely used in metrology but has application where refraction significantly disturbs accuracy. Angular pointings are then more affected than distance measurements and trilateration can achieve higher 3D accuracies than triangulation.

3 THEODOLITES AND TOTAL STATIONS

Theodolites and Total Stations are commonly used in land surveying but versions offering geodetic accuracies (arc second pointing) and optimized range measurement (submillimeter) can easily achieve 3D coordinate accuracies, which are submillimeter or better at industrial scales of measurement.

Two manufacturers explicitly address the LSM market:

1. See www.leica-geosystems.com/ims for the TPS 5000 system offering industrial theodolites and Total Stations.
2. See www.sokkia.net for Sokkia's MONMOS system based on the NET2100 Total Station.

The following comments are illustrated by Leica systems, which cover a wide application spectrum.

Theodolites are manually operated angle-measuring instruments based on a telescope that can be rotated about two intersecting orthogonal axes. When the telescope's crosshairs are centered on a target, readings from the angle encoders on these axes define the pointing.

The instrument has 3-foot screws that enable it to be precisely leveled. The pointing is then defined by a horizontal angle and a vertical angle (Figure 6). However, leveling is not a requirement for industrial measurement and the instruments can be operated on floating objects.

A Total Station (Figure 7) is a theodolite equipped with a coaxial EDM (Electronic Distance Meter) which can measure range to retroreflecting prisms and tape, or even directly to object surfaces, using the principle of phase measurement. By modulating an infrared or visible laser beam, the phase difference between the transmitted and reflected beams indicates the fractional part of the distance to the target. Different modulation frequencies enable the whole number of wavelengths to be determined and hence the absolute distance.

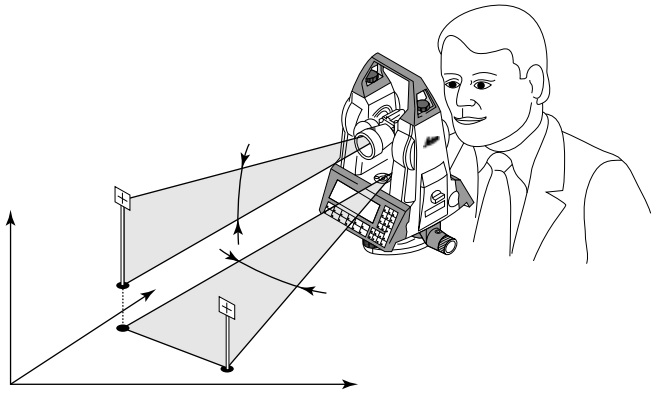


Figure 6. Theodolite angles. (Reproduced by permission of Leica Geosystems.)

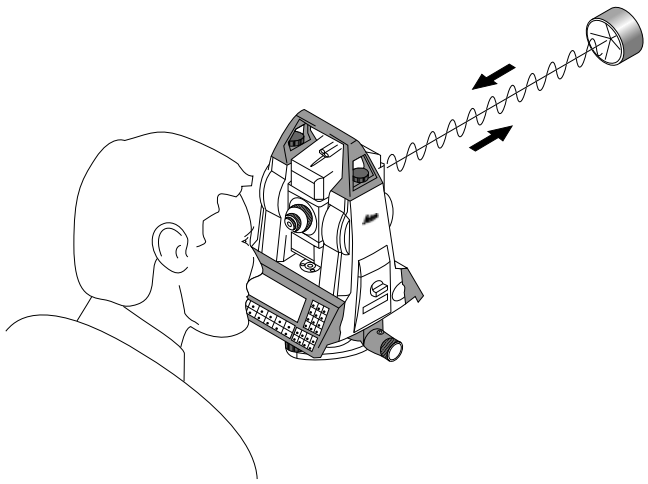


Figure 7. Range measurement by Total Station. (Reproduced by permission of Leica Geosystems.)



Figure 8. Theodolite and Total Station targets. (Reproduced by permission of Leica Geosystems.)

A wide range of targets and adapters are available for the direct and indirect measurement of points of interest. These include high-contrast adhesive targets for use with theodolites, reflective tape targets for use with Total Stations and targets in special adapters to measure hole centers (Figure 8). Hidden points can be located using rods with multiple targets offset from the contact point of interest at one end (Figure 9). Laser target spots, projected using a laser eyepiece, can provide fully noncontact targets for inaccessible or delicate surfaces (Figure 10).

Motor drives, target sensing, and radio control all provide mechanisms that automate and accelerate the measurement process, for example, by enabling target tracking.

Note that the theodolite intersection method is currently in decline, with much stronger interest being shown in the

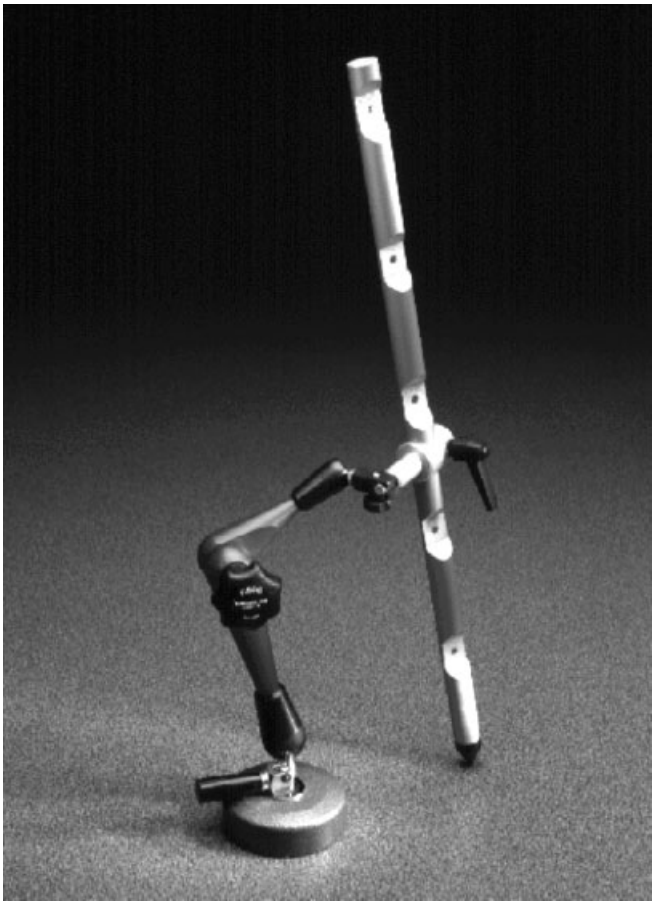


Figure 9. Hidden point rod. (Reproduced by permission of Leica Geosystems.)

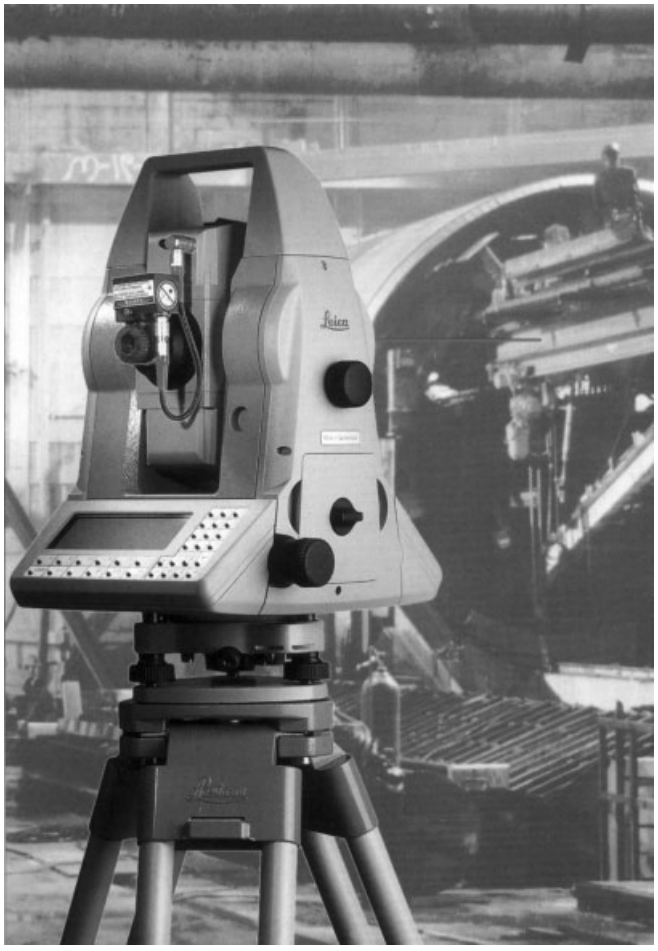


Figure 10. Laser eyepiece. (Reproduced by permission of Leica Geosystems.)

application of Total Stations, laser trackers (Section 4) and digital photogrammetry systems (Section 5).

4 TRACKING LASER INTERFEROMETERS

The tracking laser interferometer or laser tracker locates a moving reflector in 3D. Current manufacturers are Leica Geosystems (*see* Kyle (1999) and www.leica-geosystems.com/ims), Faro (www.faro.com) and API (www.apisensor.com) – *see* **Article 92, Optical Measuring Instruments, Volume 2** and **Article 91, Light Sources and Detectors, Volume 2** for more on interferometers.

A conventional interferometer (Figure 11) reflects a fixed laser beam off a retroreflector (see below). The reflector slides along the beam, returning it to the interferometer, which uses the wave property of interference to detect the reflector's movement to submicrometer resolution.

A tracking interferometer can point the interferometer beam in different directions, for example, by reflecting off a rotating mirror or by rotating the housing. It can therefore follow a retroreflector moving in 3D. The following description is based on the Leica laser tracker (see Figure 12 and Figure 13).

Tracking operates as follows (Figure 14):

- A mirror in the head is rotated about two orthogonally intersecting axes.
- An interferometer beam is reflected off the mirror and onto a retroreflector.
- The reflector returns the beam along the same path.
- Reflector movement along the beam is detected by the interferometer.
- Lateral reflector movement causes a lateral beam displacement, which is detected by a PSD (Position Sensing Device).
- The PSD signal moves the mirror to keep the beam on target.

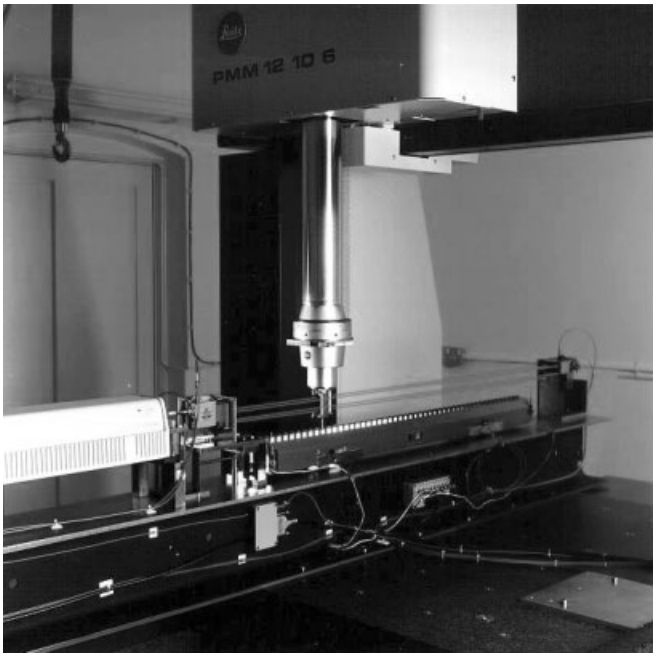


Figure 11. Conventional interferometer in use at NPL. (Crown Copyright, courtesy of NPL.)

- The mirror pointing is detected by angle encoders.
- Interferometer and angle encoders are read 3000 times a second.

The mirror angles (*Zn, Az*), and distance (*D*) locate the reflector at P by spherical coordinates (Section 2).

Since an interferometer can only measure a change of distance, not an absolute value, the tracker has a fixed point on its housing where measurements can be initiated and whose absolute distance is known by prior calibration. Interferometer changes are added to this fixed value to generate absolute distances. Alternatively, a separate coaxial ADM (Absolute Distance Meter) can initialize the distance at any temporarily stable reflector position. This uses phase measurement (Section 3) to calculate absolute distance but the process is not itself fast enough for tracking.

Retroreflection can be provided by a corner-cube reflector made up of 3 orthogonal plane mirrors. Figure 15 illustrates the principle. Figure 16 shows an actual reflector. Other types of retroreflectors are possible.

Since the reflector is not necessarily the point of interest, target probes that have offset probing points and that can be tracked in all 6 degrees of freedom (6DOF) have been developed, see Loser and Kyle (2003).

5 CAMERAS AND PHOTOGRAMMETRY

With knowledge of a camera's geometry, and precise measurement of the *xy* image coordinates of an object target

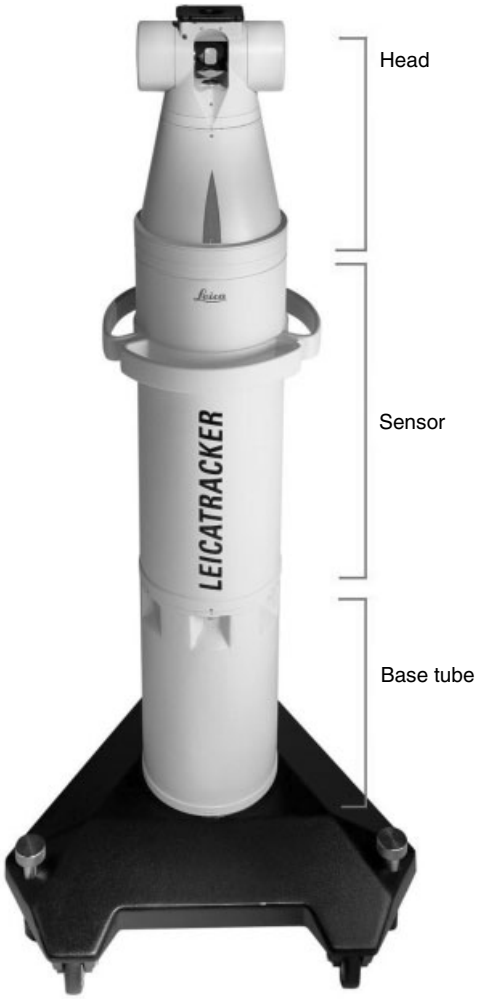


Figure 12. Leica tracker LTD 800. (Reproduced by permission of Leica Geosystems.)

point, the imaging ray from camera to target can be accurately reconstructed. Figure 17 shows how intersecting rays from two or more images locate targets in 3D space using triangulation (Section 1).

The techniques and equipment involved in making precise 3-dimensional measurements from photographic images are covered by the science of photogrammetry. Here, the reference model for cameras is the pinhole camera whose simple geometry is easy to handle mathematically. It implements the collinearity condition that target, pinhole, and image lie on a straight line and ensures that straight lines in object space create straight lines in the image (Figure 18). Its internal geometry is defined by the Principal Point, which is the foot of the perpendicular from pinhole to image plane, and Principal Distance, which is the length of this perpendicular (Figure 19).

A real camera has a complex glass lens in place of the pinhole and its geometric properties are only a close

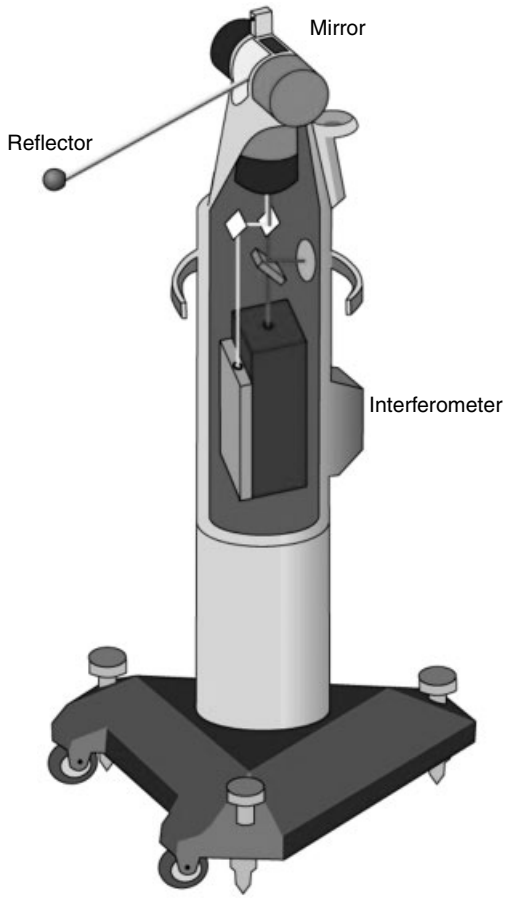


Figure 13. Leica tracker cutaway diagram. (Reproduced by permission of Leica Geosystems.)

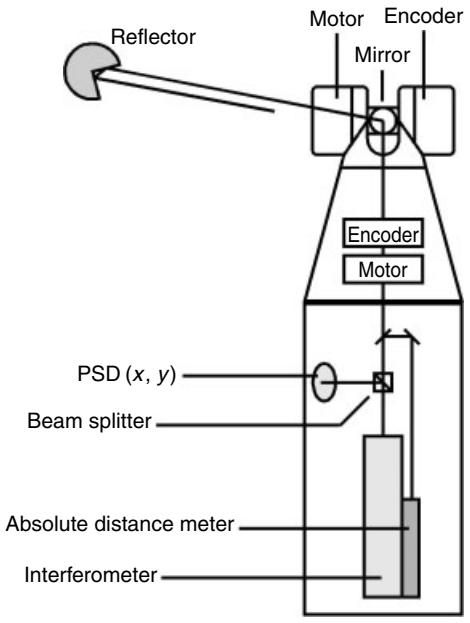


Figure 14. Leica tracker components. (Reproduced by permission of Leica Geosystems.)

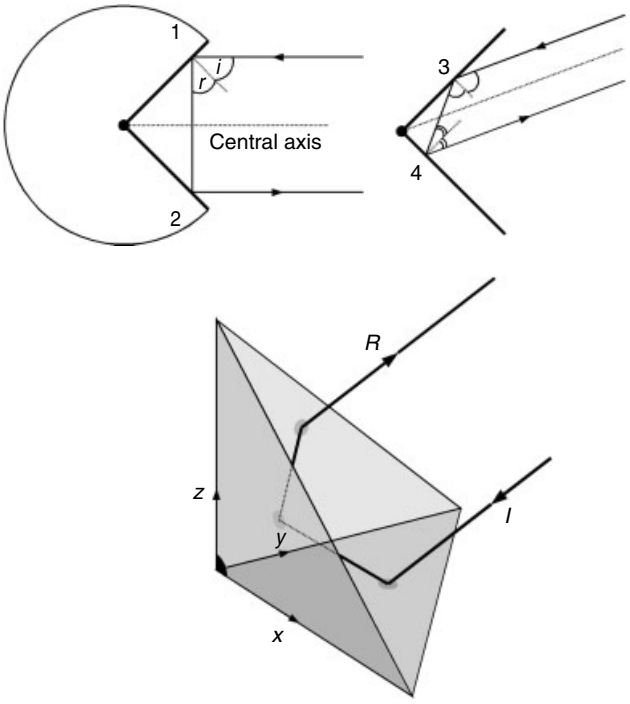


Figure 15. Principle of corner-cube reflector. (Reproduced by permission of Leica Geosystems.)

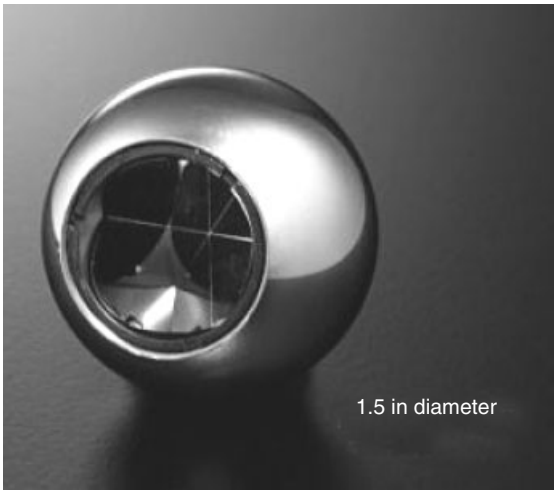


Figure 16. Example of corner-cube reflector. (Reproduced by permission of Leica Geosystems.)

approximation to a pinhole camera. To achieve metrological accuracy, real images must therefore be corrected to equivalent pinhole camera images using some process of camera calibration (Section 6).

Unlike everyday cameras, cameras for use in metrology have a more rugged and stable design, often with fixed focusing. With calibration, they then function like pinhole cameras of known internal geometry.

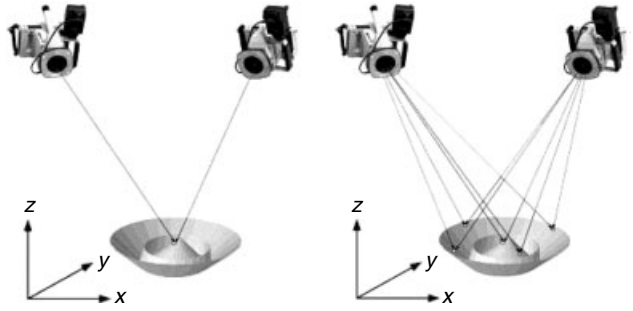


Figure 17. Single and multipoint triangulation by camera. (Reproduced by permission of Geodetic Systems.)

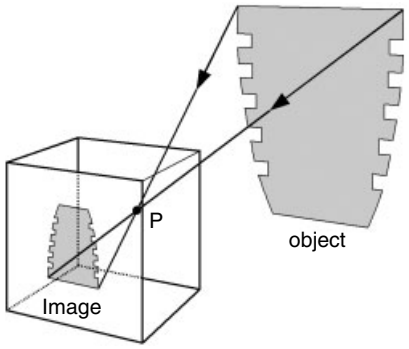


Figure 18. Pinhole camera operation. (Reproduced by permission of Leica Geosystems.)

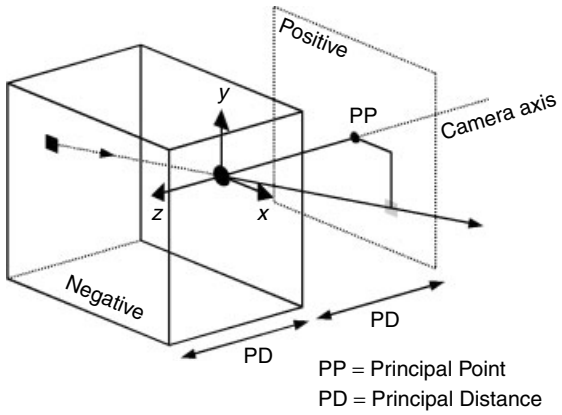


Figure 19. Pinhole camera parameters. (Reproduced by permission of Leica Geosystems.)

Early applications of photogrammetry in metrology had to accommodate the chemical processing of film and plate, and the separate measurement of the image outside the camera.

Modern digital photogrammetric systems use electronic cameras in which high resolution imaging chips of 2×3 K pixels are not unusual. With integrated image processing, they enable highly automated and online measurement procedures. Figure 20 shows the latest INCA3 camera from Geodetic Systems Inc. (www.geodetic.com).



Figure 20. INCA3 metric camera. (Reproduced by permission of Geodetic Systems.)

Object targeting typically makes use of highly reflective targets illuminated by camera flash. Images then resemble star fields, which simplifies image processing and improves image coordinate accuracy, leading to more accurate 3D data. Targets can be on adhesive tape for general surface measurement or on adapters for locating specific object features. Coded targets, coupled with appropriate image processing, provide unique identifiers for further automation. Figure 21 shows a range of GSI targets.

Target probes as in Figure 22 are used for online measurement.

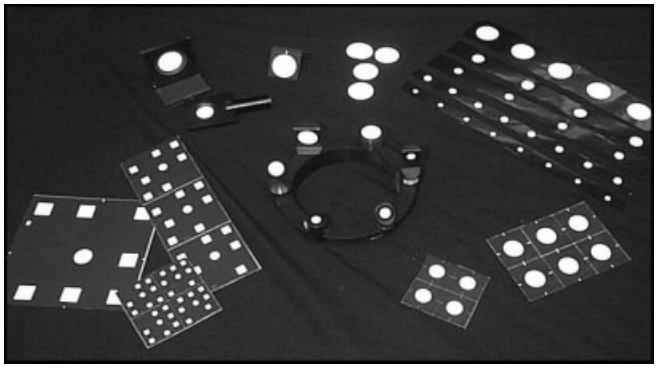


Figure 21. Photogrammetric targets. (Reproduced by permission of Geodetic Systems.)



Figure 22. Photogrammetric target probe. (Reproduced by permission of Geodetic Systems.)



Figure 23. ProSpot target projection. (Reproduced by permission of Leica Geosystems.)

Optically projected targets enable fully noncontact measurement as illustrated in Figure 23.

6 CALIBRATION OF INSTRUMENTS

Manufacturing tolerances and optical properties ensure that instruments do not conform to an idealized design model. Calibration is needed in order to achieve optimal object coordinate accuracy and some of principal features are presented here. Instrument calibration in general is discussed in **Article 217, Instrument Evaluation, Volume 3**. Test facilities are covered in **Article 216, Test Facilities, Volume 3**.

6.1 Reversal technique

Telescope reversal is illustrated by a theodolite telescope (Section 3), which can be rotated into two different positions in order to sight a target (Figure 24).

The following three examples show how this can detect errors and their sources by changing the sign of the error.

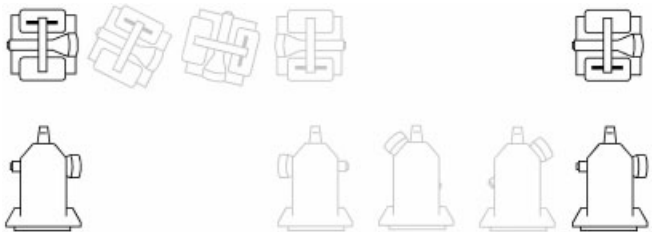


Figure 24. Telescope reversal. (Reproduced by permission of Leica Geosystems.)

6.1.1 Transit axis tilt

The transit axis is the telescope's horizontal axis of rotation. Transit axis tilt is present when this is not perpendicular to the vertical axis of rotation (Figure 25).

6.1.2 Horizontal collimation error

Horizontal collimation error is present when the telescope's line of sight is not perpendicular to the transit axis (Figure 26).

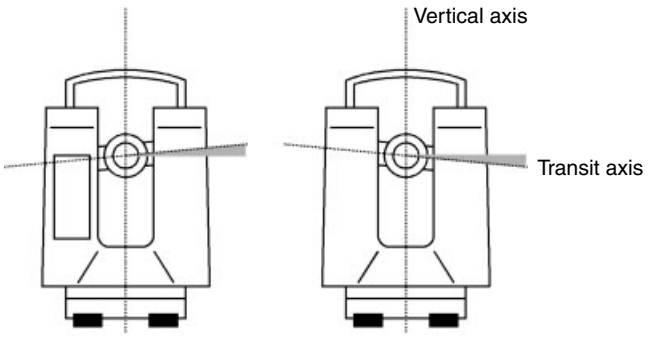


Figure 25. Transit axis tilt – sign change on reversal. (Reproduced by permission of Leica Geosystems.)

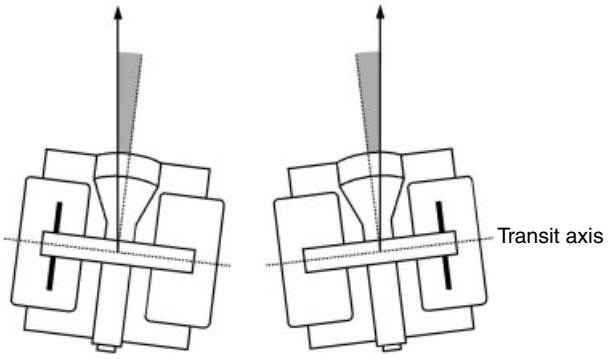


Figure 26. Horizontal collimation error – sign change on reversal. (Reproduced by permission of Leica Geosystems.)

6.1.3 Vertical index error

The zero mark on the vertical angle encoder may not be identical with a horizontal or vertical pointing, but this vertical index error can be found by summation (Figure 27).

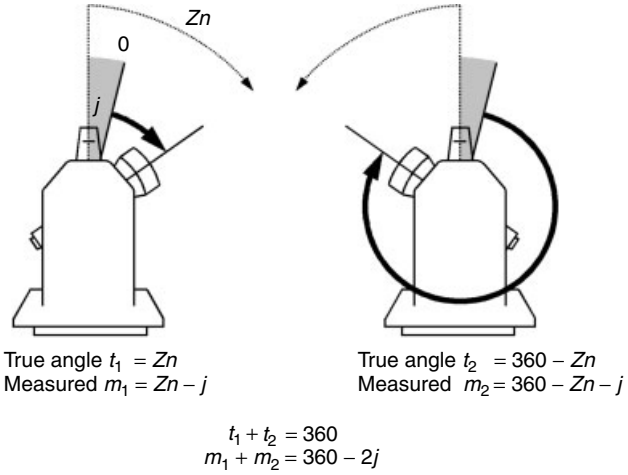


Figure 27. Vertical index error calculated by summation of pointings. (Reproduced by permission of Leica Geosystems.)

6.2 Mechanical model

A laser tracker (Section 4) has a number of subcomponents, which should have an ideal alignment and position relative to one another. Deviations from the ideal are illustrated by the Leica tracker.

Figure 28 shows three possible tilt parameters.

Figure 29 shows four possible offset parameters.

Figure 30 shows three possible encoder errors.

On-board reference distances are also required to initialize absolute range measurement. Determining the full parameter set requires measurement throughout the tracker's working range, including mirror reversal and measurements of accurate reference circles generated by attaching the reflector to a precision ball bar.

6.3 Optical model

In the pinhole camera (Section 5), which is the reference model for a real camera, straight lines in object space

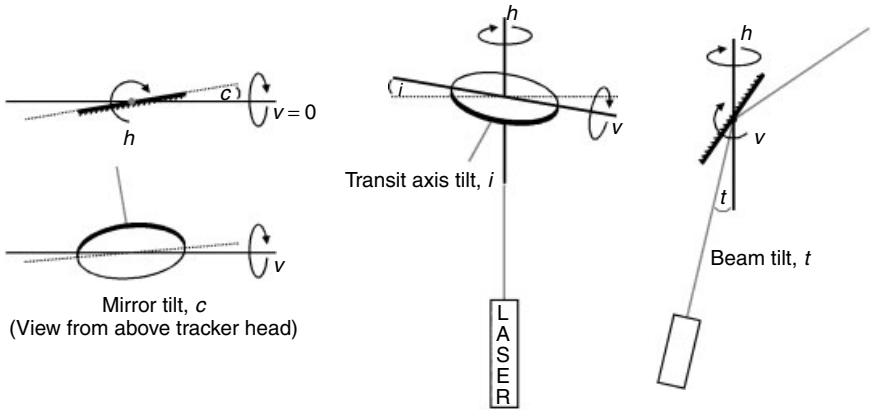


Figure 28. Leica tracker tilt parameters. (Reproduced by permission of Leica Geosystems.)

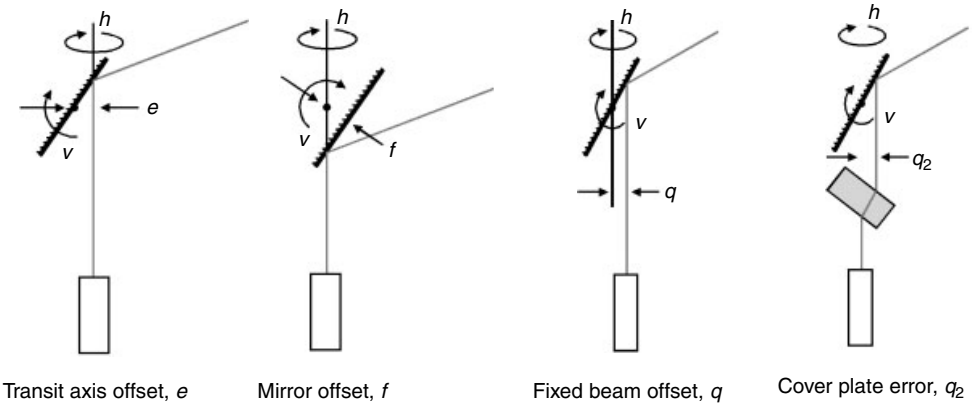


Figure 29. Leica tracker offset parameters. (Reproduced by permission of Leica Geosystems.)

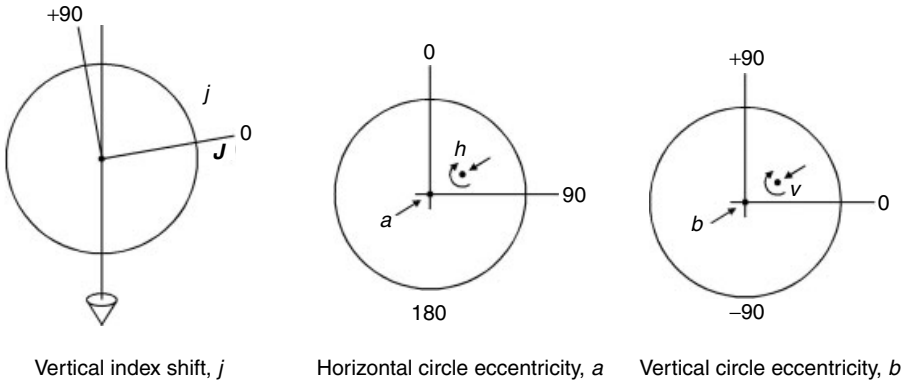


Figure 30. Leica tracker encoder parameters. (Reproduced by permission of Leica Geosystems.)

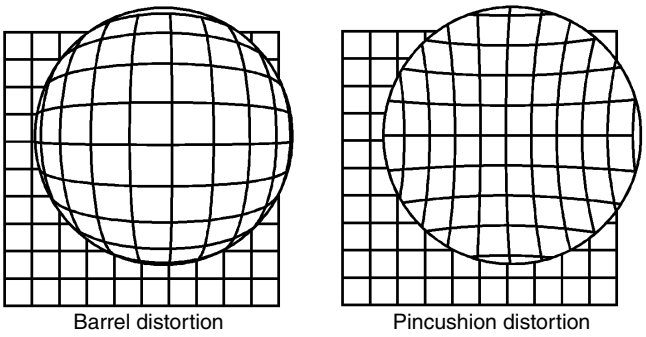


Figure 31. Lens distortion – barrel and pincushion. (Reproduced by permission of Leica Geosystems.)

generate straight lines in image space. However, lens distortion in a real camera will cause the imaged lines to bend, as illustrated in Figure 31.

A four- or five-parameter model is typically required to remove the distortion effects. In addition, parameters for Principal Point and Principal Distance must also be calculated. Only then can the rays in object space be accurately regenerated from image measurements.

Cameras can be calibrated by multiple images of a set of accurately positioned reference targets. However, a common alternative includes the parameters of a camera model as additional unknowns when calculating the coordinates of an unknown set of target points. This technique is known as *self-calibration*.

7 ORIENTATION TECHNIQUES

This section describes some common ways of directly determining the relative positions and angular orientations of measured objects and measuring instruments, which is a fundamental requirement in LSM. These use minimal data and can be further optimized by techniques that cannot themselves initialize the calculation. The methods require either manual or automatic identification of instruments and targets.

7.1 Collimation

Theodolites and Total Stations (Section 3) can be oriented to one another by first sighting along the line joining their rotation centers. This collimation procedure makes their telescope axes parallel (and superimposed). Since the rotation centers cannot normally be targeted, an indirect procedure requiring an external target, and telescope reversal (Section 8) is employed. The average pointing is the one required (Figure 32).

Given collimation pointings between instruments S1 and S2, intersecting an offset target at T1 provides sufficient data to calculate relative orientation, apart from a scale value D. Intersecting a second target at T2, a known distance L from T1, enables correct calculation of D (Figure 33).

7.2 Space resection

A method particularly used for camera/object orientations requires the image of four target points known in an object coordinate system (Figure 34).

From image coordinates and calibrated camera geometry, the space angle, a, between any two targets can be measured. When combined with target separation, D, the camera's possible position is restricted to a set of circles of specific radius and common chord (Figure 35). Three target images provide three circle sets from which the possible positions can be reduced to a maximum of four specific locations. Measurement of a fourth target resolves the problem. This technique is known as *space resection* and many possible solutions exist. See, for example, Smith (1965), Kyle (1990) and Munjy and Hussain (1996).

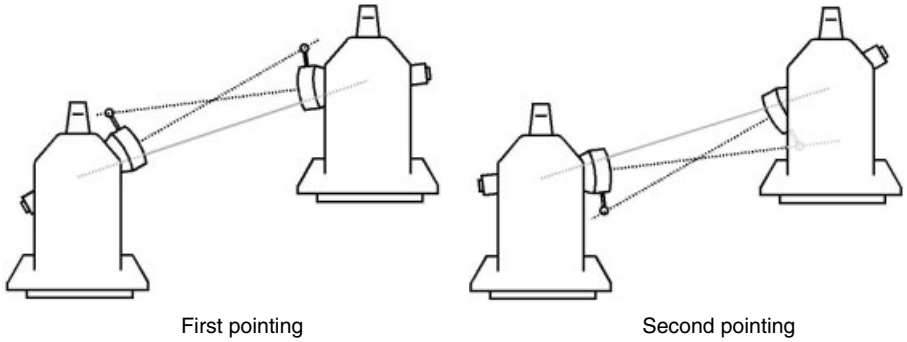


Figure 32. Collimation pointings. (Reproduced by permission of Leica Geosystems.)

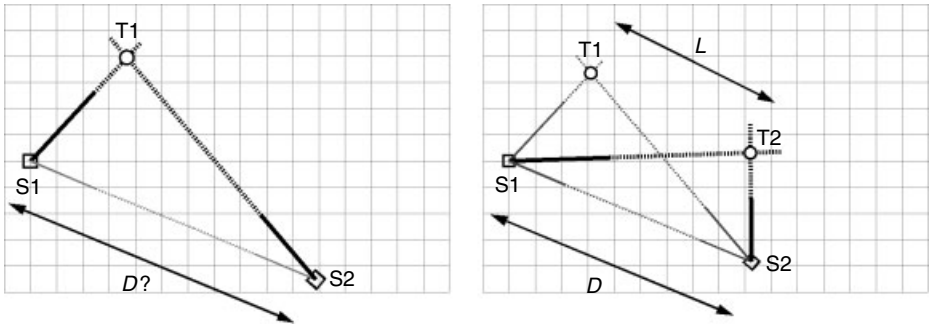


Figure 33. Orientation by collimation and offset targets. (Reproduced by permission of Leica Geosystems.)

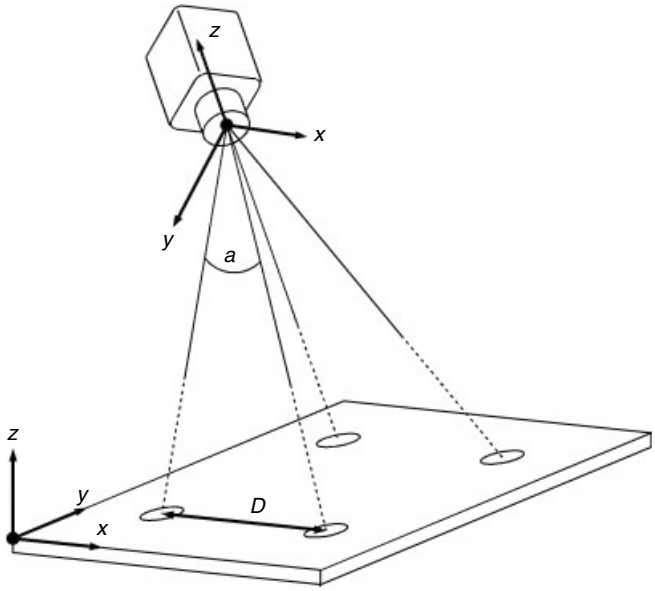


Figure 34. Space resection of camera relative to targeted object. (Reproduced by permission of Leica Geosystems.)

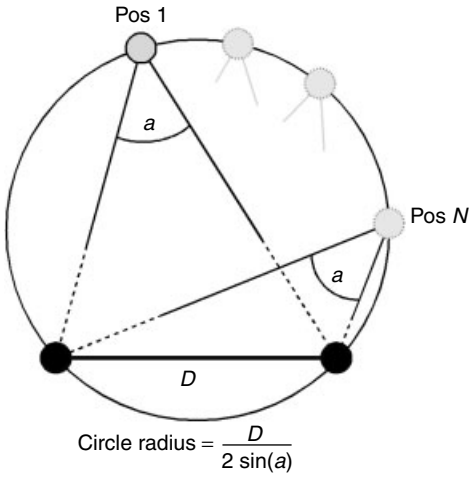


Figure 35. Resection circle. (Reproduced by permission of Leica Geosystems.)

7.3 3D transformation

Total Stations (Section 3) and laser trackers (Section 4) can be oriented with respect to one another if they have measured three common points forming a triangle. The same technique is applied to objects where three or more locations are measured in an arbitrary coordinate system but known, say, in a CAD coordinate system. By applying the parameters of a 3D transformation, illustrated in Figure 36, measurements ($P_1 \ldots P_3$) can be converted to CAD values ($R_1 \ldots R_3$).

The calculation requires a shift, S, between two corresponding points (Figure 37).

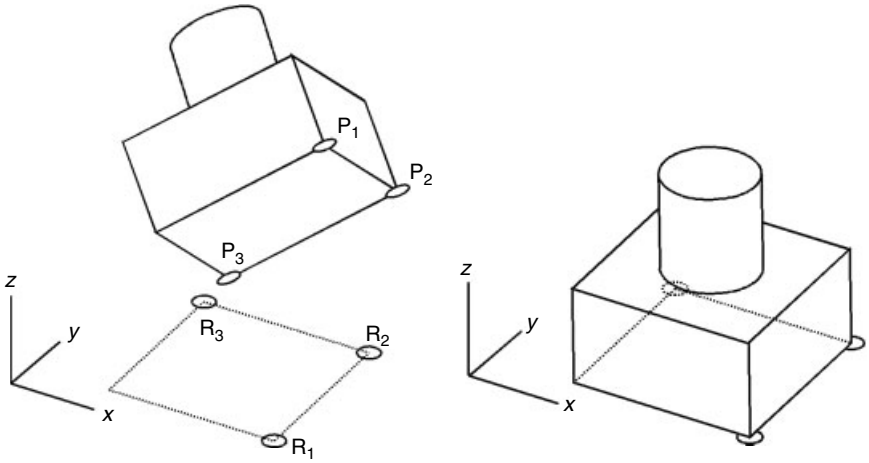


Figure 36. 3D transformation – overview. (Reproduced by permission of Leica Geosystems.)

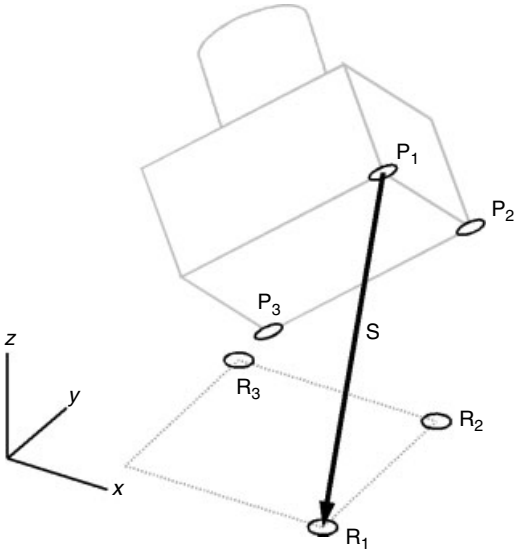


Figure 37. 3D transformation – shift vector. (Reproduced by permission of Leica Geosystems.)

Transformation is completed by a 3D rotation, **Ro**, derived from corresponding unit vectors, (Figure 38).

8 STANDARDS AND GOOD PRACTICE

Standardized procedures enable comparison between measurement systems and confirmation that a system is operating as specified. Good practice procedures enable users to optimize measurement results. Both are ultimately concerned with a system's coordinate accuracy.

Despite similarities with CMMs, for which standards exist, only very limited standards are available for LSM systems. CMMs have a fixed measurement volume and accuracy expressible in simple mathematical terms. In contrast, LSM networks have a volume dependent on task and an accuracy, which varies in a complex way within the network. Figure 39 shows a plan view of a dual theodolite triangulation network.

Uncertainties in mechanical systems are discussed in **Article 88, Error Sources in Fine Mechanics, Volume 2**. The theodolites have a fixed angular uncertainty indicated by the dotted lines on either side of the solid intersecting rays. This results in an area of positional uncertainty indicated by the gray quadrilaterals. Optimal accuracy is indicated by the 90° intersection, which applies to points on the dashed semicircle. No scale is indicated and the instruments could be 2 or 20 m apart, with a corresponding scale change in uncertainty. Features of an LSM system that contribute to the complex uncertainty analysis include

- very flexible choice in position, number, and type of measuring instruments, all of which can contribute to the measurement of a particular point;
- variable uncertainty due to intersection of two or more pointings to a target;
- different methods of introducing scale, from one or more scale bars to a range value in every measurement;
- an unrestricted measurement volume;
- simultaneous multiple point measurement in the case of photogrammetry.

In Germany, a proposal exists for photogrammetric systems in the form of VDI/VDE 2634. A common theme in developing possible standards for comparing and evaluating LSM systems is the use of reference lengths placed in different parts of the workspace. In fact, acceptance, and verification tests for CMMs are essentially length-measuring tasks. Reference length measurement is a simple

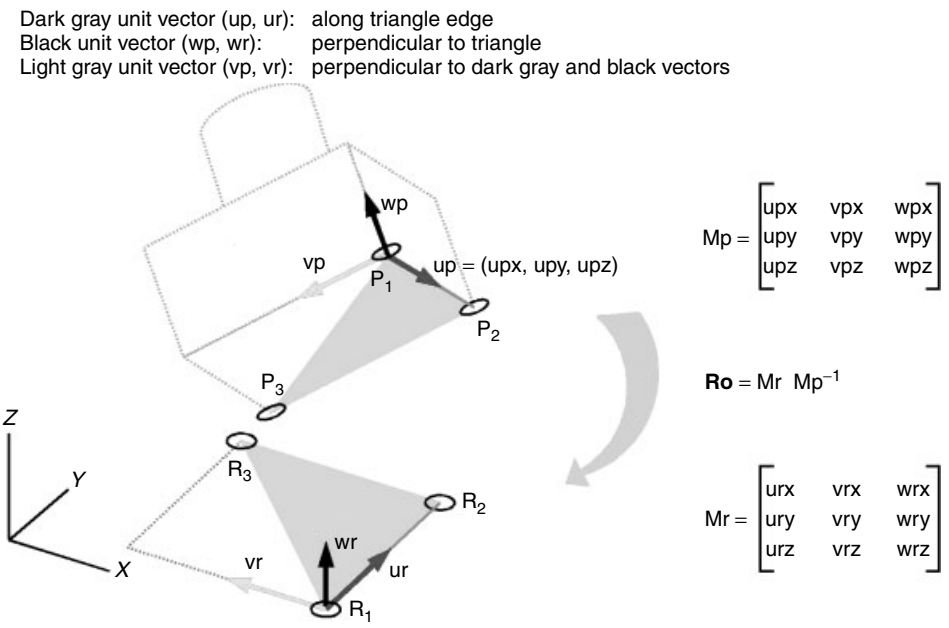


Figure 38. 3D transformation – rotation matrix. (Reproduced by permission of Leica Geosystems.)

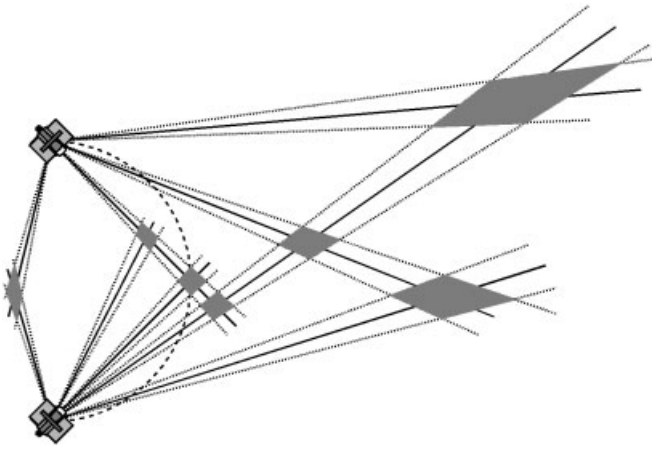


Figure 39. Effect of intersection error on accuracy. (Reproduced by permission of Leica Geosystems.)

and recommended procedure, which is frequently used on site to check system operation.

These comments lead on to good practice guides of which the most accessible example has been created at University College London, in cooperation with the UK's National Physical Laboratory and Leica Geosystems. See http://www-old.ge.ucl.ac.uk/research/projects/NMSPU_Best_Practice_Guide.pdf

This guide concentrates on laser trackers and to a lesser extent on theodolite systems. It does not explicitly consider photogrammetric systems although many comments are also relevant here. The project behind this guide is presented by Kyle *et al*. (2001).

9 SAMPLE APPLICATIONS

These examples provided by Leica Geosystems (www.leica-geosystems.com/ims) and Geodetic Systems Inc., (www.geodetic.com) indicate the scope of LSM applications and the advantages of using LSM methods.

9.1 Digital photogrammetry

Figure 40 shows a 2×3 m aircraft tooling jig being measured with a hand-held metric electronic camera, which demonstrated an efficient way of achieving the same results as theodolite intersection.

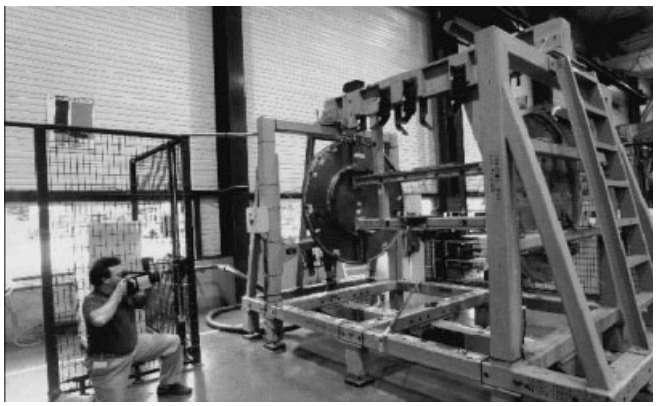


Figure 40. Using digital photogrammetry to measure an aircraft tooling jig. (Reproduced by permission of Leica Geosystems.)

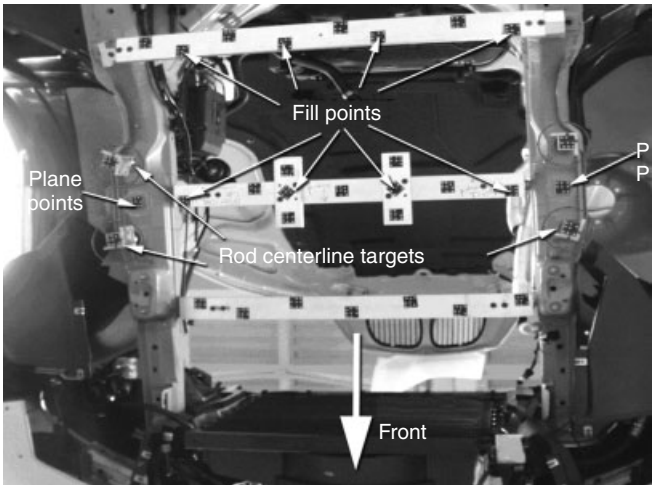


Figure 41. Photogrammetric targeting on axle carrier. (Reproduced by permission of Leica Geosystems.)

Total measurement time was 35 min as follows:

- Place 150 targets: 15 min.
- Record 60 images: 10 min.
- Process the images: 10 min.

Coordinate accuracy achieved: $<50\,\mu m$

BMW evaluated photogrammetry for production line measurement of an axle carrier (Figure 41) using special coded targets (Section 5).

Only 5 min were permitted for placing targets, taking images, and removing targets. 18 images were taken of 8 targets and 8 scales in a 2-min window, resulting in RMS accuracies in x, y, and z of $8\,\mu m$, $7\,\mu m$, and $17\,\mu m$ respectively.

9.2 Laser tracker

Daimler–Chrylser Aerospace had been measuring an aircraft rudder using a traditional mechanical gauge (Figure 42).

Manufacture of the gauges is expensive and their use is not efficient. They adopted an alternative approach using a laser tracker (Figure 43).

The advantages of using a tracker have been

- simpler, more compact system,
- faster data acquisition,
- more accurate data,
- immediate comparison with reference values,
- elimination of traditional gauge with associated cost and space savings.

NASA's X-38 space vehicle is a possible model for transporting astronauts and the atmospheric reentry requirement demands very precise assembly. Traditional alignment techniques were time consuming and so replaced by laser tracker (Figure 44)



Figure 42. Gauge for aircraft part checking at Daimler–Chrysler Aerospace. (Reproduced by permission of Leica Geosystems.)



Figure 43. Tracker measures aircraft component at Daimler–Chrysler Aerospace. (Reproduced by permission of Leica Geosystems.)

Advantages have been

- 70% reduction in inspection time with 'ring frame' inspection down from 2 weeks to under 3 days;

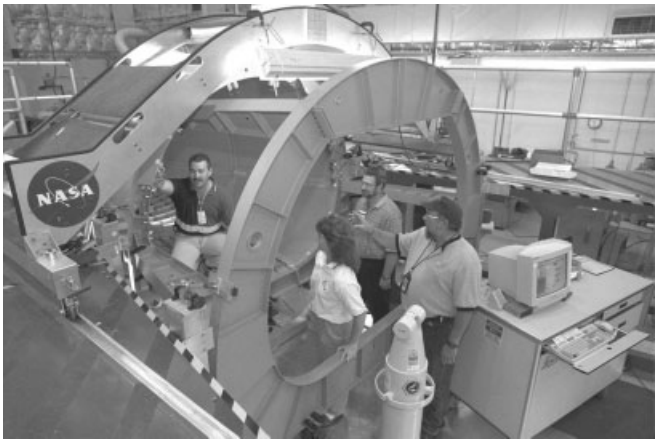


Figure 44. NASA X-38 measured by tracker. (Reproduced by permission of Leica Geosystems.)

- parts referenced in a single coordinate system within a tolerance of 0.05 inch;
- portability in bringing the system to and around the object.

9.3 Total Station

Leica's precision Total Stations are appropriate instruments for controlling the construction of a ship as they can achieve submillimeter accuracy at ranges up to 120 m.

One task is the positioning of bulkheads (Figure 45). These define cargo storage areas, which must be accurate in order to facilitate loading/unloading and prevent cargo movement.

Another task is roundness checks on submarine hulls. An older technique involving elaborate scaffolding and restricted mechanical measurement (Figure 46) has been considerably simplified by direct 3D measurement using the Total Station (Figure 47).

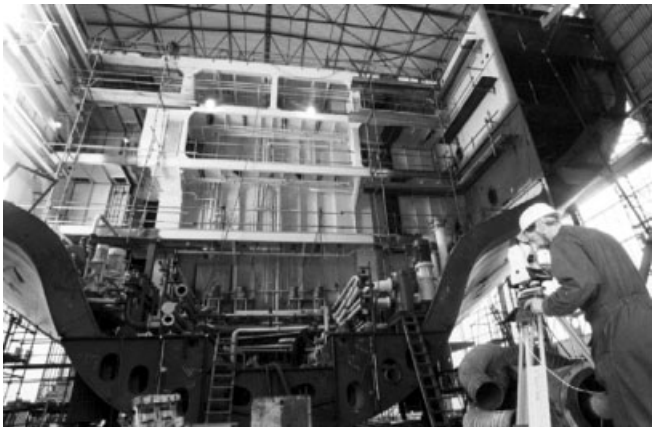


Figure 45. Bulkhead positioning at Thyssen–Nordseewerke. (Reproduced by permission of Leica Geosystems.)



Figure 46. Roundness check – old. (Reproduced by permission of Leica Geosystems.)

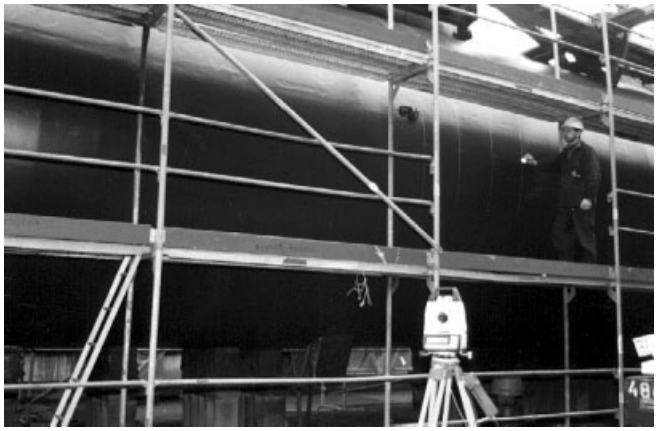


Figure 47. Roundness check – new. (Reproduced by permission of Leica Geosystems.)

REFERENCES

Atkinson, K.B. (ed) (1996) *Close Range Photogrammetry and Machine Vision*, Whittles Publishing, Scotland.

Kyle, S.A. (1990) A Modification to the Space Resection. *Allgemeine Vermessungsnachrichten*, **7**(90), 17–25.

Kyle, S.A. (1999) Operational Features of the Leica Laser Tracker, in *Business Improvement Through Measurement, IEE seminar ref. 1999/154*, (8/6/99) (Birmingham, England).

Kyle, S.A., Robson, S., Chapman, D.P., Cross, P.A., Iliffe, J.C. and Oldfield, S. (2001) Mathematics and best practice for triangulation and tracking methods in 3D metrology, in *National Measurement Conference*, (Nov. 2001) (Harrogate, England).

Löffler, F. (2002) *Handbuch Ingenieurgeodäsie, Machinen- und Anlagenbau*, 2nd edn, Wichmann, Heidelberg.

Loser, R. and Kyle, S.A. (2003) Concepts and Components of a Novel 6DOF Tracking System for 3D Metrology, in *Optical 3D Measurement Techniques VI*, Vol. 2, (eds. A. Grün and H. Kahmen), Druckerei Peter + Co., Zurich, Switzerland (pp. 5–62).

Luhmann, T. (2000) *Nahbereichsphotogrammetric*, Wichmann, Heidelberg.

Munjy, R.A.H. and Hussain, M. (1996) Closed-Form Space Resection Using Photo Scale Variation. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, **34**(B5), 390–393.

Schwarz, W. (ed) (1995) *Vermessungsverfahren im Maschinen- und Anlagenbau*, Konrad Wittwer, Stuttgart.

Smith, A.D.N. (1965) The Explicit Solution of the Single Picture Resection Problem, with a Least Squares Adjustment to Redundant Control. *Photogrammetric Record*, **5**(26), 113–121.

195: Characteristics of Temperature Measurement

Joseph McGhee
Formerly of University of Strathclyde, Glasgow, UK

1 Concept of Temperature 1307
2 Historical Development of Temperature Scales 1307
3 Thermodynamic Temperature Scale 1309
4 International Temperature Scales 1310
5 Guide to Types of Temperature Sensors 1313
Related Articles 1314
References 1314

1 CONCEPT OF TEMPERATURE

The concept of temperature is closely related to the physiological experience of hotness or coldness when some object is touched or approached. Indeed, *hotness* and *coldness* were regarded as two separate properties of matter in ancient Greece. Fuzzy qualitative descriptions, which are used when objects are described as relatively cold, cool, or tepid, or perhaps as hot, warm, or boiling, are intuitive ways of expressing notions related to personal comfort.

As soon as thermometers were invented, an immediate need arose for refinement of the concepts of temperature and hotness. Heat is transferred from hot to other cooler bodies in such a way that both bodies tend to equalize their temperatures, approaching a new common intermediate temperature.

This notion is contained in the definition of temperature given by the Scotsman James Clerk Maxwell. He stated that the temperature of a body is its thermal state, which is regarded as a measure of its ability to transfer heat to other bodies.

This assumption, that heat flows from warmer to cooler bodies, forms the basis of the International Temperature Scale of 1990, commonly referred to as ITS-90. The agreement to internationally adopt this *Temperature Scale* is the latest step in the long, difficult, and tortuous route, which error-prone science has taken toward a contemporary definition of temperature.

2 HISTORICAL DEVELOPMENT OF TEMPERATURE SCALES

During the second century B.C., the Greek physician C. Galen associated four degrees of coldness with the effects of different medicines upon human physiology during his medical treatments. The purpose of the medicines was to warm his chilled patients or to cool his feverish patients.

Building upon Galen's basic assumption, later 'scientists' tried to credit the effects of these medicines to different bodies by giving them different degrees of warmth or coldness.

The first device, which was used to measure the degree of warmth or coldness, seems to have been invented by Galileo Galilei sometime between the years 1592 and 1603 A.D. This instrument, which is shown in Figure 1(a), consisted of an air-filled glass bulb connected to a long tube immersed in a colored liquid. The liquid column rose or fell as a function of the ambient temperature, which influenced the entrapped air, thus indicating the degree of warmth or coldness. In the absence of any evidence that the instrument had graduations, it is better to call it an *air thermobaroscope*. Attempts to measure the length of the column were probably made using dividers or compasses. As the indicated values were also a function of the atmospheric pressure, its precision must have been quite poor.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

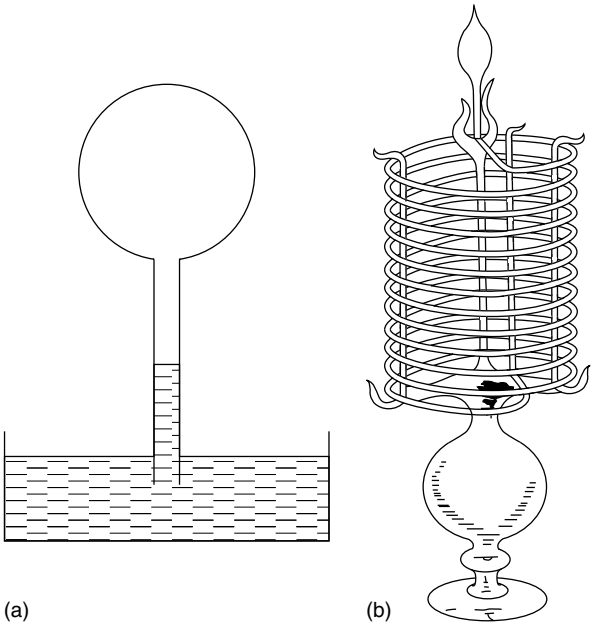


Figure 1. Early types of thermometer (a) Galileo's air thermoscope and (b) thermometer of Florentine Academy of Sciences.

The concept of atmosphere was unknown at that time. Subsequently, these atmospheric dependencies were removed by applying suitable corrections. It is apparent from these observations that although the response of a thermometer to variables other than temperature causes errors in the measurement, these errors may be compensated for by using appropriate corrections.

About the year 1650 A.D., members of the Florentine Academy of Sciences made the first thermometer, which is represented in Figure 1(b). This consisted of a spiral-shaped tube with a closed end. Although a graduation was a part of the device, numbers were not associated with it.

In due course, confidence in the stability and repeatability of physical states led to their association with defined temperature fixed points. It was realized that intercomparison of existing thermometers could only be performed by tests conducted in systems, which duplicated these states.

One of the first proposals, from H. Fabri from Leida in 1669, was for two fixed points. The lower point should be the temperature of snow and the higher point, the temperature of the hottest summer day.

A later proposal, which was made by C. Rinaldi from Padua in 1693, suggested that the fixed points should be the temperatures of the melting point of ice and the boiling point of water. Between these two points, 12 divisions should be introduced. In the same year, and for the first time, the British scientist E. Halley applied mercury as the thermometric liquid.

Rømer, a thermometrist working in Copenhagen at the end of the seventeenth and beginning of the eighteenth century, developed a scale in which zero degrees was associated with the coldest day, while the normal temperature of the human body was associated with 24 degrees. This made the temperature of boiling water equivalent to ≈ 50 to 55 degrees on this unusual scale, which was influenced by the predominant use of thermometers for meteorological purposes at that time. Hence, if the freezing point of water had been taken as zero, the repeated use of negative values for winter temperatures would have occurred. Winter temperatures of $-16\,^{\circ}\mathrm{C}$ ($\approx 0\,^{\circ}\mathrm{F}$) are quite common in continental Europe.

To avoid this, it seemed obvious to use the lowest attainable temperature of those days as zero. Shortly after Rømer proposed his scale, his laboratory was visited by D. G. Fahrenheit from Danzig (now the city of Gdañsk in Poland).

Fahrenheit's description of the mercury-in-glass thermometer in 1724 was a notable milestone in the development of thermometry. Fahrenheit, evidently influenced by Rømer's scale, proposed his own, very well known scale, which associates 32° with the freezing point of water and 212° with its boiling point. Why did he choose these divisions for this scale? Although the reasons are unclear, the upper fixed point, the temperature of boiling water ($\approx 212\,^{\circ}\mathrm{F}$), was probably rejected as being unnecessarily high for meteorological purposes. Instead, he merely changed Rømer's 24 degrees for body temperature to 96. This change, which was equivalent to four subdivisions on each degree of the Rømer scale, was also probably made because 96 is divisible not only by 2 but also by multiples of 3 and hence 12 – the decimal system commonly used now was not in general use at that time. The Fahrenheit scale, which persists today, is essentially the same as that described by him to The Royal Society in 1724.

Further development of the mercury-in-glass thermometer, in 1742, was due to the Swedish astronomer and physicist A. Celsius. He assigned 100° to the temperature of melting ice and 0° to the temperature of boiling water. The region between these two points was divided into 100 equal steps. Subsequently, after the death of Celsius in 1744, M. Strømer, his friend and scientific collaborator, reversed these values.

Eventually, as science developed, a need to measure temperatures above the melting point of glass arose. Prinsep's air thermometer, which used a gold bulb to measure temperatures of $1000\,^{\circ}\mathrm{C}$, in 1828, was followed soon after, in 1836, by a platinum bulb in a similar thermometer by Pouillet.

A true Thermodynamic Temperature Scale, described in Section 3, had been the unconscious aim of all the previous efforts. Such a scale was not possible until 1854 when its foundations were laid by William Thompson,

later afforded the title Lord Kelvin of Largs, and other contemporary scientists.

3 THERMODYNAMIC TEMPERATURE SCALE

The aim of any scale of temperature, especially the thermodynamic scale, is the representation of hotness and hotness relations between objects and events in the real physical world by real numbers. As numerical values are correlated to some defined temperatures, temperature fixed points are required to attribute certain temperature values. Interpolation then allows the definition of temperature between these temperature fixed points.

A thermometric working substance is needed one of whose properties and a correlating function must be assumed. The chosen function provides the means of attributing the chosen property of the working substance to a certain temperature. Because of the diversity of materials and their properties, there is an unlimited number of these temperature scales.

The material that most closely approximates this ideal thermometric working substance is an ideal gas. Indeed, it was the work of Robert Boyle and his coworkers in the middle of the seventeenth century that led to the conviction of many later scientists that there was such a thing as an *absolute zero* of temperature. These eminent individuals included G. Amontons, in Paris in 1699; J. H. Lambert, in 1770; and Gay-Lussac, in 1790. Gay-Lussac gave credit to J. A. C. Charles for the latter's previously unpublished researches. All of their efforts resulted in what is now called the *ideal gas law*, which is written in the usual form

$$PV = nkT \tag{1}$$

where P is the pressure, V is the volume, n is the number of moles of gas, k is Boltzmann's constant ($= 1.3807 \times 10^{-23}$ J K^{-1}) and T is the absolute temperature.

When the temperature is held constant, (1) corresponds to *Boyle's Law*.

Similarly, *Charles' Law* is obtained from (1) when the pressure is held constant.

Since there are no direct methods for measuring temperature, as there are with, say, length measurement, some difficulties are associated with temperature measurement. As only indirect temperature measurements are possible, any temperature scale depends upon the chosen thermometric working substance and its chosen property. Although any working substance may be applied in principle, its use will be restricted to some finite range determined by its thermal behavior.

For example, the application of mercury-in-glass thermometers is limited on the low-temperature side by the solidification of the mercury as it freezes and on the high-temperature side by the inability of the glass to expand indefinitely as well as by its melting temperature.

Melting of the glass was responsible for the Prinsep and Pouillet thermometers.

An ideal solution to the problem of proposing a suitable temperature scale would be to find one that is valid in any temperature range and is totally independent of the working substance.

The Kelvin Thermodynamic Temperature Scale, based upon the efficiency of the ideal reversible Carnot cycle, is such a scale. A reversible Carnot Cycle consists of a reversible heat engine operating between two isotherms at the temperatures T_2 and T_1, with $T_2 > T_1$, and two adiabatic processes. A reversible heat engine absorbs the heat Q_2 from the high-temperature source, at the temperature T_2, and discharges the heat Q_1 to the low-temperature source, at the temperature T_1. The difference between the absorbed heat Q_2 and the discharged heat Q_1, which is the external work, W, performed by the engine, may be written as

$$W = Q_2 - Q_1 \tag{2}$$

Reversing the engine action indicates that it may be driven by a second identical engine working between the same heat sources. The effect of such an action may be the heat flow from the lower to the higher temperature source. Using the properties of reversible processes, it may be proven that the ratio Q_2/Q_1 is a function of only the two source temperatures, so that

$$\frac{Q_2}{Q_1} = f(T_1, T_2) \tag{3}$$

Following Kelvin's proposal, it may be assumed that the functional relation in (3) is

$$\frac{Q_2}{Q_1} = \frac{T_2}{T_1} \tag{4}$$

Equation (4) is the basis of the thermodynamic temperature scale. From (2) to (4), the efficiency of a reversible heat engine is defined as

$$\eta = \frac{W}{Q_2} = \frac{Q_2 - Q_1}{Q_2} = \frac{T_2 - T_1}{T_2} = 1 - \frac{T_1}{T_2} \tag{5}$$

This efficiency and the definition of temperature, which is based upon the former, may be shown to be independent of the working substance. As a result, it may be used to

define the thermodynamic temperature scale.

$$T_1 = T_2(1 - \eta) \tag{6}$$

By means of this scale, any chosen thermal state, such as the melting point of ice, may be assigned a certain value of thermodynamic temperature. The Thermodynamic Temperature Scale may thus be founded upon a defined temperature difference between two temperature fixed points or a defined value of one temperature fixed point.

In the course of the development of technology, the manner of defining the thermodynamic temperature scale has changed. Until 1954, it was assumed that 100° represented the difference between the boiling point of water and the melting point of ice. Since then, there has been a return to the original and older proposals of Kelvin, in 1848; Mendeleiev, in 1874; and Giauque, in 1939. Thus, since 1954, the thermodynamic temperature scale is based upon the temperature fixed point, which is the triple point TP of water.

Triple points of physical materials are stable, repeatable temperatures where the solid, liquid, and gaseous forms of the material exist in thermal equilibrium. The triple point of water occurs at that temperature where ice, water, and water vapor exist in thermal equilibrium. A temperature of 273.16 has been assigned to this temperature fixed point.

In 1967, the 13th Conférence Général des Poids et Mesure (CGPM) introduced a new definition for the scale and a new symbol for the unit of thermodynamic temperature. This unit is called the *kelvin*, denoted by the symbol K. Notice that when the unit is being referred to, it is spelled with a small letter to avoid confusion with the person Kelvin. It is defined as 1/273.16th part of the thermodynamic temperature of the triple point of water.

However, an ideal Carnot cycle is impossible to realize in practice. Nevertheless, it can be demonstrated that the thermodynamic scale may be reproduced by a gas thermometer with an ideal gas as the working substance, in accordance with the ideal gas law given in (1). Either pressure difference at constant volume or volume difference at constant pressure can be chosen as the measure of temperature.

Here again, although the ideal gas is quite fictitious, it could be replaced by a noble gas at very low pressure. When the readings at constant volume, T_v, and the similar readings at constant pressure, T_p, are extrapolated to zero they tend to the same value, $T_v = T_p = T$, independently of the properties of the gas. Thus, the thermodynamic temperature scale may be reproduced using gas thermometers, which have an application range up to about 1350 K.

Another simple method of reproducing the scale at thermodynamic temperatures above 1337 K is allowed by means of thermal radiation from heated bodies. When this radiation is in thermodynamic equilibrium with the radiating body, some properties of this radiation are directly linked to the temperature of the body. The concepts of black-body radiation are essential for proper utilization of the method. These concepts, which are beyond the scope of this article, use the relation known as *Planck's law*.

4 INTERNATIONAL TEMPERATURE SCALES

Temperature measurement requires a

- primary standard
- practical measuring system
- system or body whose temperature is to be measured.

A primary standard system for measuring temperature is the *Kelvin Thermodynamic Temperature Scale* referred to above. Because of the difficulties that are involved in realizing this primary standard system, realizations are based upon *boiling points, freezing/melting points*, and *triple points*.

Boiling points correspond to characteristic temperatures where the liquid and gaseous states of a material exist in equilibrium.

Freezing/melting points are temperatures where a material undergoes an equilibrium change in its physical state from liquid to solid or solid to liquid respectively. They are preferred to boiling points as they are less sensitive to pressure changes.

Triple points are temperatures at which the solid, liquid, and gaseous forms of the material exist in equilibrium.

Practical realizations of temperature scales have been disseminated by previously adopted resolutions of the CGPM in 1889, 1927, 1948 (revised in 1960), 1968 (supplemented in 1976), and 1990. Figure 2 summarizes all of these scales for comparative purposes. The fixed points of these various scales, also given for comparative purposes, appear in Table 1.

The Normal Hydrogen Scale, or NHS, which was based upon the work conducted by Chappius, was proposed in 1887. Using hydrogen gas as the thermometric material, Chappius built a gas thermometer calibration facility covering the range −25 to +100 °C. This early scale, which was used to calibrate mercury-in-glass thermometers, was a true centigrade scale as its fixed points were the ice-point, at 0 °C, and the boiling point of water, at 100 °C.

Although work on the International Temperature Scale of 1927, or ITS-27, commenced as early as 1913, its

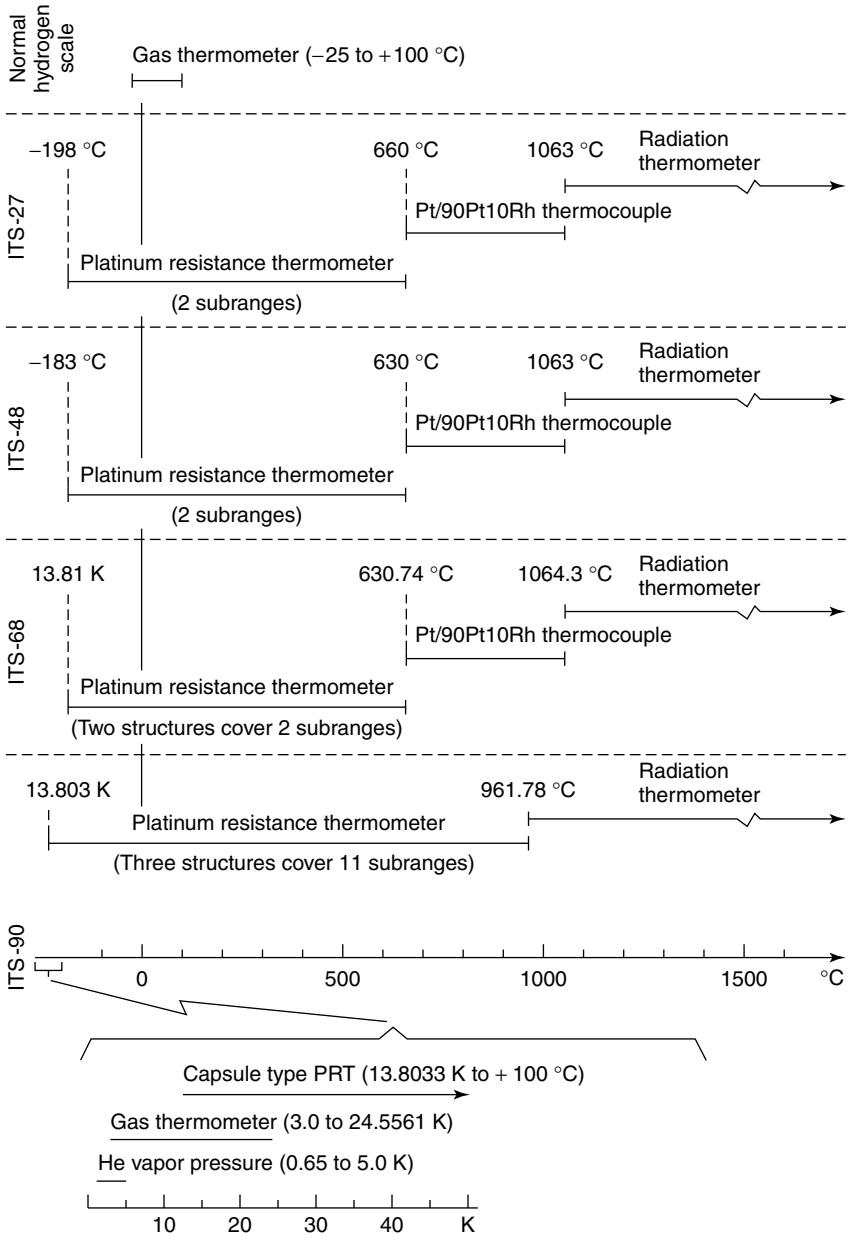


Figure 2. Comparison of the various temperature measurement scales and the measuring ranges of their standard interpolating instruments/sensors.

completion was delayed by the outbreak of World War I. It was defined with the assignment of six defining or fixed points. Development of thermometers using the noble metal Platinum gave rise to the Platinum Resistance Thermometer, or PRT, followed the pioneering groundwork of Siemens, in 1871, and Callendar, in 1887. By the end of World War I, PRTs were acknowledged as precision thermometers. This confidence provided the basis for their specification as one of the standard interpolating instruments of ITS-27. Over the range −190 to +660 °C, in the subranges −190 to 0 °C and 0 to 660 °C, the interpolating instrument was specified as the PRT made from platinum with defined properties exhibiting resistances at three temperatures, expressed as ratios with respect to the resistance at 0 °C. From 660 to 1063 °C, the scale was to be interpolated using a platinum/90% platinum-10% rhodium (Pt/90Pt10RH) thermocouple made from materials with specified properties. The Wien Law of Radiation defined temperatures above 1063 °C. ITS-27 was a major step forward in the universality of thermometry as it removed previously observed ambiguities in the specification of temperature.

Table 1. The temperature fixed points of various temperature scales (decimal point given to level of uncertainity).

Fixed point of scale \ Scale	Normal hydrogen scale	ITS-27	IPTS-48	IPTS-68	ITS-90
Vapour pressure point of helium					3–5
Triple point of equilibrium hydrogen				13.81	13.8033
Boiling point of hydrogen				17.042	17
Gas thermometer point of helium				20.28	20.3
Triple point of neon					24.5561
Boiling point of neon				27.102	
Triple point of oxygen				54.361	54.3584
Triple point of argon				83.798	83.8058
Condensation point of oxygen		90.18	90.18	90.188	
Triple point of mercury					234.3156
Freezing point of water	273	273.15	273.15		
Triple point of water			(273.16)	273.16	273.16
Melting point of gallium					302.9146
Boiling point of water	373	373.15	373.15	373.15	
Freezing point of indium					429.7485
Freezing point of tin				505.1181	505.078
Freezing point of zinc			(692.655)	692.73	692.677
Boiling point of sulfur		717.75	717.75		
Freezing point of aluminum					933.473
Freezing point of silver		1233.65	1233.95	1235.08	1234.93
Freezing point of gold		1336.15	1336.15	1337.58	1337.33
Freezing point of copper					1357.77

The tortuous path in the development of a temperature scale, which truly represented the thermodynamic scale, soon uncovered the inadequacies of ITS-27. Thus was born the International Temperature Scale of 1948 (ITS-48), which possessed the same number of fixed points as ITS-27 but with the freezing point of silver now specified as 960.8 °C instead of 960.5 °C, as in ITS-27. The lower PRT interpolating limit was also raised to −183 °C to coincide with the oxygen boiling point of −182.970 °C. Otherwise, the PRT standard interpolation subranges remained the same, as did that of the Pt/90Pt10Rh interpolating thermocouple.

In the case of the interpolating thermocouple, a quadratic interpolating equation was introduced with new constraints placed upon the acceptable values and tolerances of the emf. Above 1063 °C, the Wien Radiation Law was replaced by the Planck Law to improve the thermodynamic consistency of the temperatures in this range and also to allow the use of ITS-48 at higher temperatures than ITS-27.

In 1960, a revision of ITS-48 became known as the International Practical Temperature Scale of 1948(60), or IPTS-48(60), to avoid confusion with ITS-48.

The changes, which specified the water triple point temperature as 273.16 K, that created the present Kelvin Thermodynamic Scale also included the adoption of the triple point temperature instead of ice-point temperature as a fixed point of the scale. The name of the unit of temperature was changed to Celsius in place of centigrade. ITS-47 was a true centigrade scale as it had 100 degrees as the fundamental interval between the ice-point and the water boiling point. As the freezing point of zinc, at 419.505 °C, was more precisely realized, it was proposed as a replacement for the sulfur boiling point at 444.60 °C. New restrictions were placed upon one of the PRT ratios and upon the standard thermocouple emf.

The International Practical Temperature Scale of 1968, or IPTS-68, which was based upon *boiling points, melting/freezing points*, and *triple points*, arose from the need to extend IPTS-48 to lower temperatures because of the availability of improved measurement methods.

A total of 13 fixed points were used to define this scale. Although the interpolating instruments were the same as for IPTS-48, the PRT range was extended to cover the lower-temperature region down to 13.8 K using four wire resistance connections in two different sensor structures. The scale was also more closely defined in terms of a reference function, with four different deviation functions defined to provide correction in the four different temperature subranges for the particular PRT being calibrated.

In the original statement of IPTS-68, the same thermocouple covered the same range as in IPTS-48(60), with the

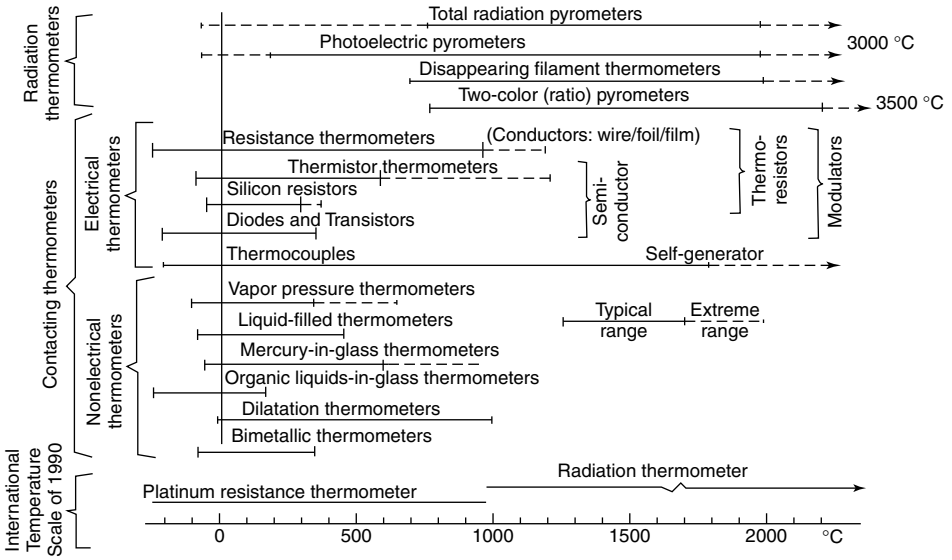


Figure 3. (Repeated in **Article 111, Classification of Sensors, Volume 2**) Classification of temperature-measuring instruments/ sensors by measuring range.

same quadratic form for the emf defining equation. The range, subsequently adopted by the 15th CGPM in 1975, was changed to 1337 K in IPTS-68(75) with a commensurate tightening of the emf specifications. Above 1337 K, the Planck Radiation Law defined the scale. An Extended Practical Temperature scale of 1976, or EPT-76, which includes revisions to IPTS-68, allowed IPTS-68 to be extended at low temperatures with the addition of 11 fixed points in the cryogenic range from the superconducting transition point of cadmium at 0.519 K to the boiling point of neon at 27.102 K.

IPTS-68 and EPT-76 have now been superseded by ITS-90, the International Temperature Scale of 1990. This scale, which has most boiling points eliminated, allows an extension of the lower range to 0.65 K. It agrees better with the thermodynamic scale and has greater continuity, precision, and reproducibility with appropriate subranges. As indicated in Figure 2, four classes of thermometers are used to establish the scale. The Helium vapor pressure thermometer covers the range from 0.65 to 5 K, where the saturated vapor pressure temperature relations of ^{3}He and ^{4}He hold. Constant volume gas thermometers, in which the gas pressure is a function of temperature, are standard thermometers over the range from 3 K to the triple point of neon at 24.5561 K. Over the temperature range from the triple point of equilibrium hydrogen, at 13.8033 K, to the freezing point of silver, at 961.78 °C, the standard instrument is a platinum resistance thermometer calibrated at specified sets by defining fixed points and using specified interpolating procedures. The Planck Radiation Law is used to define ITS-90 above the freezing point of silver. Overall, ITS-90 represents Thermodynamic Temperature with an uncertainty of ±2 mK from 1 to 273 K, increasing to ±7 mK at 900 K.

As already stated, ITS-90 is established using a total number of 17 fixed points. These comprise 1 vapor pressure point, 6 triple points, 8 freezing/melting points, and 2 boiling points. A comparison between the fixed points of the ITS-90 with the previous temperature scales, described above and illustrated in Figure 2, appears in Table 1.

5 GUIDE TO TYPES OF TEMPERATURE SENSORS

It is also possible to classify temperature sensors on the basis of their temperature range of application. Such a classification is repeated for completeness in Figure 3 – *see* **Article 111, Classification of Sensors, Volume 2** for more on the classification of temperature sensors.

General accounts on temperature measurement and its practice are available in ASTM (1993), Childs (2001), Nicholas and White (2001), Michalski, Eckersdorf and McGhee (1991), TC (2004), and U.S. Sensor (2004). More information is available: Thermocouples in **Article 196, Thermocouple Temperature Sensors, Volume 3**; resistance temperature detectors (RTD) in **Article 197, Metallic Resistance Temperature Detectors (RTDs), Volume 3**; and standards and practice of calibration in **Article 198, Calibration and Standards in Temperature Measurement, Volume 3**.

RELATED ARTICLES

Article 41, Terminology and Classification of Measurement Systems, Volume 1; **Article 196, Thermocouple Temperature Sensors, Volume 3**; **Article 197, Metallic Resistance Temperature Detectors (RTDs), Volume 3**; **Article 198, Calibration and Standards in Temperature Measurement, Volume 3**.

REFERENCES

ASTM. (1993) *Manual on the Use of Thermocouples in Temperature Measurement*, Pcn: 28-012093-40 (ASTM MANUAL SERIES) ASTM Committee E20 On Temperature Measurement.

Childs, R.N. (2001) *Practical Temperature Measurement*, Butterworth-Heinemann.

Michalski, L., Eckersdorf, K. and McGhee, J. (1991) *Temperature Measurement*, Wiley, Chichester.

Nicholas, J.V. and White, D.R. (2001) *Traceable Temperatures: An Introduction to Temperature Measurement and Calibration*, Wiley, Chichester.

TC. (2004) *Free Guide to Temperature Measurement for Engineers*, www.thermocouplesensors.co.uk.

U.S. Sensor. (2004) *Precision Thermistors, RTDs, Probes and Assemblies*, www.ussensor.com.

Dr Joe McGhee unfortunately passed away before his material was finalised. He will be remembered by the Measurement community.

This article is based in part on a chapter by J. McGhee *et al* originally published in Scientific Metrology, J. McGhee *et al*, Akademickie Centrum Graficzno Marketingowe, Poland, 1998.

196: Thermocouple Temperature Sensors

Jacek Kucharski
Technical University of Lodz, Lodz, Poland

1 Physics of Thermoelectric Effect 1315
2 Thermocouples 1316
3 Thermocouple Temperature Sensors 1317
4 Reference Temperature 1318
5 Measuring Circuits 1319
6 Testing and Calibrating 1320
References 1320
Further Reading 1320

1 PHYSICS OF THERMOELECTRIC EFFECT

1.1 Thermoelectric force (emf)

It was Seebeck who discovered that a current flows in a closed loop of two dissimilar metals when their junctions are at two different temperatures.

In 1834, Peltier, discovered that a junction of two dissimilar metals is respectively heated or cooled when a current is passed through it either in one direction or the other.

Another effect discovered by Lord Kelvin (W. Thomson) in 1854 was that a homogeneous current-carrying conductor, lying in a temperature gradient, will absorb or generate heat, in addition to and independently of Joule heating.

Of these three thermoelectric effects, only the Seebeck effect is the real source of thermoelectric force, electromotive force (emf). This force results from the variation of electron density along a conductor subject to a nonuniform temperature distribution, and is given by

$$\mathrm{d}E = \sigma_{\mathrm{A}}(\vartheta)\,\mathrm{d}\vartheta \tag{1}$$

where σ_{A} is the Seebeck coefficient of metal A.

The Seebeck thermoelectric force, E_{A}, in a homogeneous conductor whose ends are at temperatures ϑ_1 and ϑ_2 respectively, has a value given by

$$E_{\mathrm{A}}(\vartheta_1, \vartheta_2) = \int_{\vartheta_2}^{\vartheta_1} \sigma_{\mathrm{A}}(\vartheta)\,\mathrm{d}\vartheta \tag{2}$$

or

$$E_{\mathrm{A}}(\vartheta_1, \vartheta_2) = E_{\mathrm{A}}(\vartheta_1) - E_{\mathrm{A}}(\vartheta_2) \tag{3}$$

In a thermoelectric circuit composed of two homogeneous conductors A and B, whose joined ends are at the temperatures, ϑ_1 and ϑ_2, the resultant sum of both thermoelectric forces is

$$E_{\mathrm{AB}}(\vartheta_1, \vartheta_2) = E_{\mathrm{A}}(\vartheta_1, \vartheta_2) - E_{\mathrm{B}}(\vartheta_1, \vartheta_2) \tag{4}$$

Summing the partial emfs in the circuit of Figure 1 gives the resultant emf, E, as

$$E_{\mathrm{AB}}(\vartheta_1, \vartheta_2) = e_{\mathrm{AB}}(\vartheta_1) + e_{\mathrm{BA}}(\vartheta_2) \tag{5}$$

It should be emphasized that the resultant thermoelectric force of the circuit of Figure 1 is only a function of the types of metal A and B and of the temperatures, ϑ_1 and ϑ_2.

In the application of a circuit of two metals, A and B, it will be assumed that the temperature, ϑ_2, of one of the

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

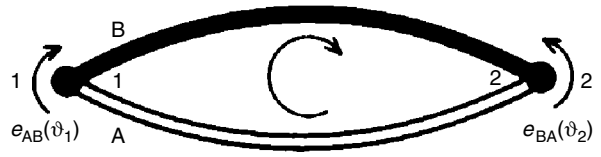


Figure 1. A closed thermoelectric circuit. (*Source*: Reproduced from 'Temperature Measurement, 2nd Ed', Michalski *et al*, 2001 © John Wiley & Sons Ltd.)

junctions, called the *reference junction* will be held constant at the reference temperature, ϑ_r. For $\vartheta_2 = \vartheta_r = \text{const}$, (5) may be written as

$$E_{AB}(\vartheta_1, \vartheta_r) = e_{AB}(\vartheta_1) - e_{AB}(\vartheta_r) = f_2(\vartheta_1) \qquad (6)$$

Equation (6) indicates that the emf of the circuit of two metals A and B is a function of the measured temperature, ϑ_1, of the measuring junction, which is junction 1 in Figure 1. It is of cardinal importance as it is the foundation of thermoelectric thermometry.

1.2 Law of the third metal

The practical application of a circuit with the two metals A and B, for the purpose of temperature measurement, also requires the introduction of an emf measuring device. Insertion of such a voltage measuring instrument necessarily involves the connection of a third metal, C, in the circuit. Fortunately, this addition is covered by the law of the third metal as shown in Michalski *et al*. (2001). This states that introducing a third metal, C, into a circuit of the two metals A and B, does not alter the resulting emf in the circuit, provided that both ends of the third metal C are at the same temperature. The third metal, C, may be introduced into a circuit at any point.

2 THERMOCOUPLES

Two dissimilar conductors, which may be metals, alloys or nonmetals, connected at one end is known as a thermocouple. In the notation shown in Figure 2, their point of connection is called the *measuring junction* and their free ends, the *reference junction*.

The thermal emfs of different materials used to form thermocouples, with platinum as the reference material are given in Michalski *et al*. (2001).

The International Standard International Electrotechnical Commission (IEC) 584 of 1995 covers standardized thermocouples. It states their emf values as a function of temperature, as shown in Figure 3.

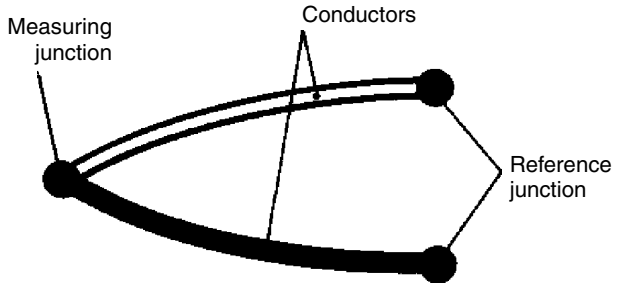


Figure 2. The physical structure and notation of a thermocouple. (*Source*: Reproduced from 'Temperature Measurement, 2nd Ed', Michalski *et al*, 2001 © John Wiley & Sons Ltd.)

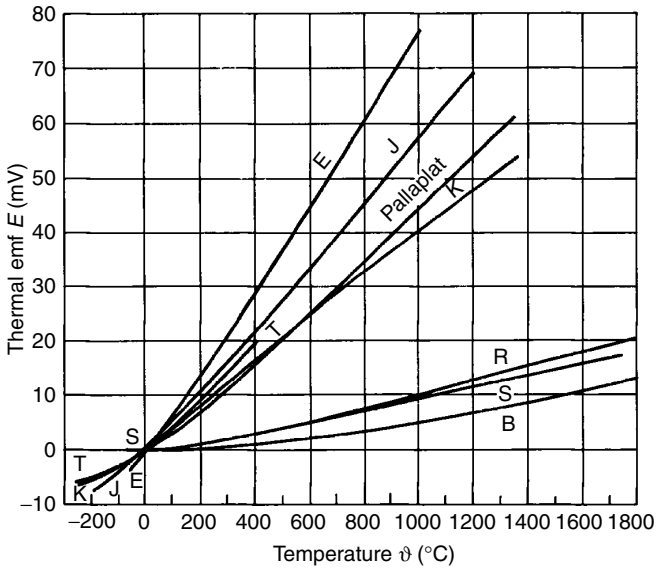


Figure 3. Thermal emf, E, of commonly used thermocouples as a function of temperature. (*Source*: Reproduced from 'Temperature Measurement, 2nd Ed', Michalski *et al*, 2001 © John Wiley & Sons Ltd.)

Some selected properties of standardized thermocouples are given below:

Platinum–rhodium/platinum, code (or type) S, (90% Pt–10% Rh/Pt), is the most popular of all the rare-metal thermocouples. It is specified for normal continuous use in the range −50 °C to +1300 °C. As its structure gives a high resistance to corrosion, thin wires may be used, resulting in low cost and low thermal inertia thermocouple.

Platinum–rhodium/platinum, code R, (87% Pt–13% Rh/Pt) has similar properties to the type-S unit described above.

Platinum–rhodium/platinum–rhodium, code B, (70% Pt–30% Rh/94% Pt–6% Rh), has the following advantages over the type-S unit

- a higher upper temperature range of 1800 °C,
- in the temperature range up to about 100 °C, the emf is negligibly small so that no reference temperature stabilization is needed.

Iron/copper–nickel, code J, (Fe/55% Cu–45% Ni), is standardized in IEC 584-l. It is a popular thermocouple, used up to 1000 °C, with a high emf and good resistance to oxidizing and reducing atmospheres.

Copper/copper–nickel, code T, (Cu/55% Cu–45% Ni) is rated for continuous use within the temperature range from −200 °C to +500 °C. It is commonly used in laboratory measurements.

Nickel–chromium/copper–nickel, code E, (90% Ni–10% Cr/55% Cu–45% Ni) corresponds to the Chromel/Constantan structure popular in the United States. With the highest emf of all metal thermocouples, it is used from −270 °C to +800 °C.

Nickel–chromium/nickel–aluminum, code K, (90% Ni–10% Cr/95% Ni, balance Al, Si, Mn) is a commonly used base-metal type in the temperature range of −200 °C to 1100 °C.

Nickel–chromium–silicon/nickel–silicon, code N, is similar to that of the code K type, but it is more resistant to corrosion.

Detailed descriptions of all these thermocouples and also special, nonstandardized high-temperature thermocouples are given in Michalski *et al*. (2001) and in Kinzie (1973).

The measuring junction of a thermocouple is formed by brazing, soldering, welding, twisting or rolling the two dissimilar materials together, as described by Baker, Ryder and Baker (1953).

3 THERMOCOUPLE TEMPERATURE SENSORS

3.1 Construction

The sensor is more than just the junction. A thermocouple sensor is a structure consisting of a thermocouple, which has associated wires, their insulation, and terminal head, mounted in a suitable protective tube, sometimes called a *well* or *sheath*.

The most common industrial thermocouple sensors are general purpose, straight sheathed units, for application at atmospheric pressure. As shown in Figure 4, its leads are insulated with ceramic tubes or beads. Protection from chemical and mechanical influences is provided by a tubular, metallic, or ceramic sheath. The terminal head, which contains a terminal block, is also equipped with a tight sealing gland through which the compensating or connecting cable enters the head. The unit is often made such that the thermocouple unit itself can be replaced without losing any liquid, in which the outer sheath is placed.

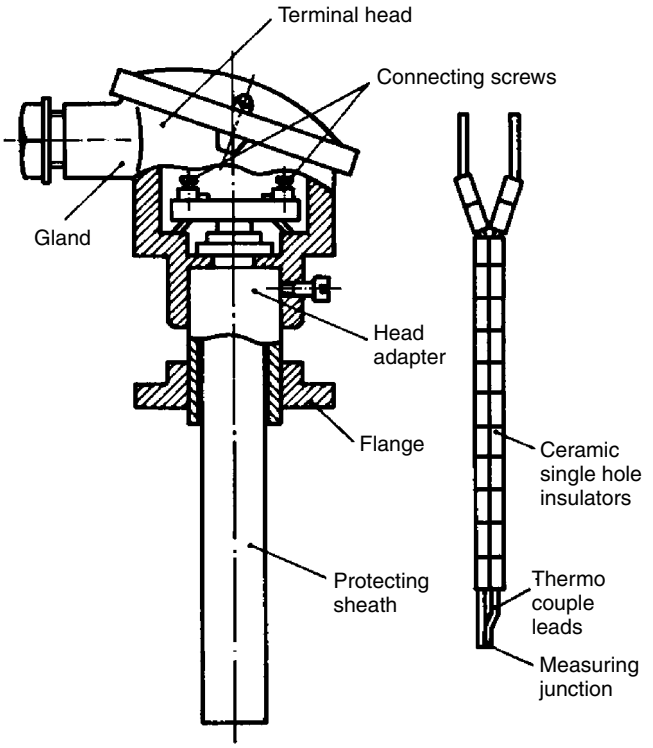


Figure 4. Industrial thermocouple assembly. (*Source*: Reproduced from 'Temperature Measurement, 2nd Ed', Michalski *et al*, 2001 © John Wiley & Sons Ltd.)

Standardized wire diameters for thermocouples may vary from 0.2 mm up to 4 mm for base metals and from 0.1 mm to 0.5 mm for rare metal units.

3.2 Sheath materials

Protecting sheaths are characterized by

- resistance to high temperatures,
- mechanical strength,
- corrosion resistance,
- chemical resistance,
- noncontamination of the given thermocouple,
- resistance to thermal shocks,
- imperviousness to gases.

Sheath materials influence the stability of calibration as well as the dynamic properties of a thermocouple; their properties and applications are given in Michalski *et al*. (2001).

3.3 Review of structures

A wide range of constructional forms, including the dimensions, sheath material, and applications for each of the thermocouple structures is given in Michalski *et al*. (2001).

The most popular industrial, straight thermocouple is shown in Figure 4.

Right-angle (corner) thermocouples are used for measuring the temperature of molten metals or salt baths.

High-pressure thermocouple sensors, as specified in DIN (Deutche Industrie Normen), are composed of a thermocouple insert placed in a thermocouple well. Thermocouple inserts, usually contain an easily exchanged short MI (mineral insulated) thermocouple.

Cylindrical screw-in thermocouples are normally mounted in tank or pipeline walls using welded fixing attachments.

Conical well thermocouples may be used for high-pressure measurements.

Mineral insulated thermocouples, which have a metal sheath, were originally designed in 1957 for temperature measurement in gas-cooled nuclear reactors. As they have important properties, they are now widely applied. A considerable advantage is their thin, robust, flexible structure, which permits easy bending combined with a short thermal response time.

Figure 5 illustrates the lateral cross-section of mineral insulated thermocouples. A continuous protective metallic sheath contains the wires of the thermocouple, which are placed in a densely compacted mineral powder that is electrically, but not thermally insulating. The sheaths may have diameters between 0.2 and 8 mm, while they may be many meters long.

4 REFERENCE TEMPERATURE

Compensating cables A′ and B′ described in Kinzie (1973) are used as an extension of the thermocouple from the terminal head, which is subject to big variations of temperature, to the new reference junction located at constant temperature as shown in Figure 6. As the temperature of the thermocouple terminal head is never very high, the emf versus temperature matching of the compensating cable to the thermocouple is only necessary in the temperature range 0–200 °C. When the compensating leads are made from the same material as the thermocouple wires, they are called *extension leads*.

If it is assumed that the characteristics of the thermocouple AB and of the compensating leads A′B′ are identical at the reference temperature, ϑ_r, it is apparent that

$$e_{A'B'}(\vartheta_r) = e_{AB}(\vartheta_r) \tag{7}$$

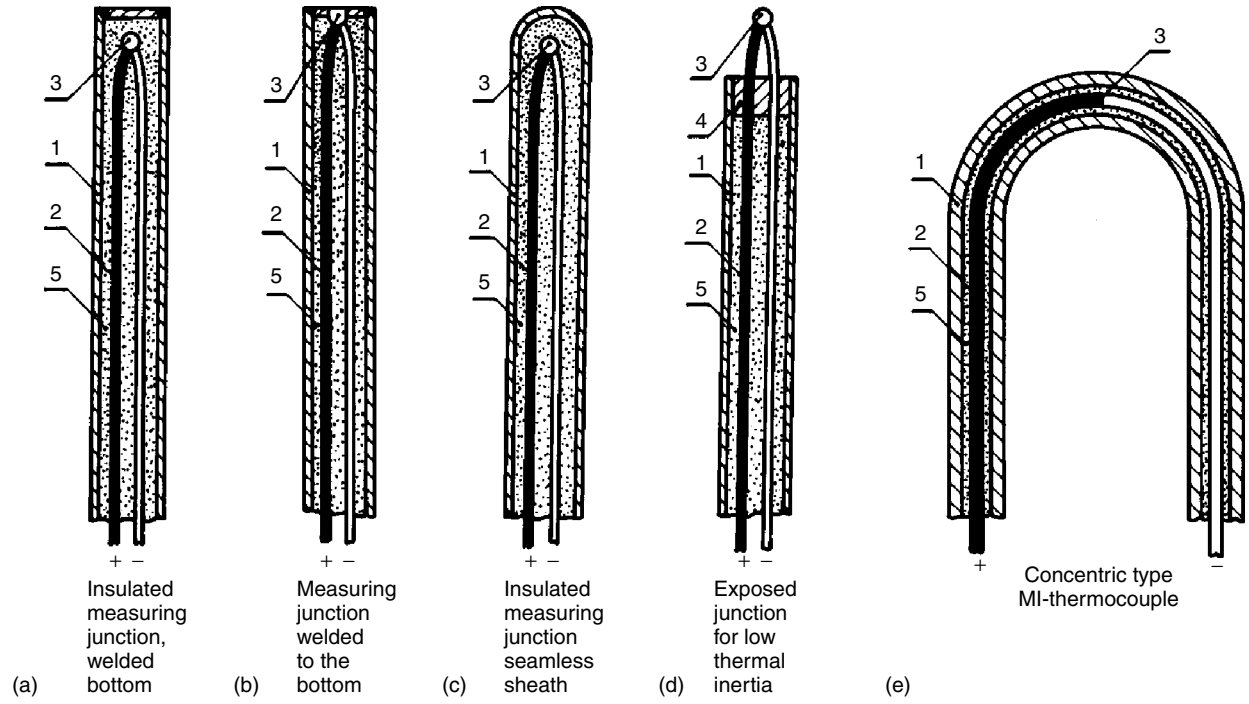


Figure 5. Cross-section of mineral insulated thermocouples, 1–metal sheath, 2–conductor, 3 – measuring junction, 4 – refractory cement, 5 – mineral insulant. (*Source*: Reproduced from 'Temperature Measurement, 2nd Ed', Michalski *et al*, 2001 © John Wiley & Sons Ltd.)

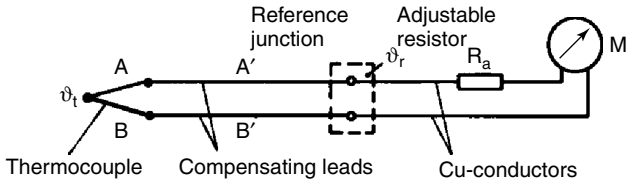


Figure 6. Thermocouple with compensating cable. (*Source*: Reproduced from 'Temperature Measurement, 2nd Ed', Michalski *et al*, 2001 © John Wiley & Sons Ltd.)

Equation (7) shows that the resulting emf equals that value of the emf exhibited by a thermocouple having its reference junction at the temperature, ϑ_r. It is then apparent that the temperature, ϑ_r, may be called the *reference temperature* in this case.

The outputs of compensating cables conform to IEC 584-1. The type of compensating cable can be identified on the basis of internationally standardized color-coding given by IEC Standard Project 60584-3.

The output emf of values of thermocouples and thus the thermometer indications are correct if the reference temperature ϑ_r equals its nominal value $\vartheta_{rn} = 0\,^{\circ}\mathrm{C}$, which rarely occurs in practice. Michalski *et al.* (2001) provides details of the following procedures for use when $\vartheta_{rn} \neq 0\,^{\circ}\mathrm{C}$:

- calculation of the true emf E at $\vartheta_r = 0\,^{\circ}\mathrm{C}$ based on an indicated value E' at $\vartheta_r \neq \vartheta_{rn}$, applicable for short time measurements;
- assumption that ϑ_r equals ambient temperature, for less precise measurement;
- reference junction placed in ice-point;
- reference junction placed in $0\,^{\circ}\mathrm{C}$ thermostat;
- reference junction placed in electrical thermostat operating at 50 or $60\,^{\circ}\mathrm{C}$;
- applying automatic correction devices, which add or subtract the necessary correcting signals to or from the measured emf E-value.

5 MEASURING CIRCUITS

Thermoelectric thermometers use the following measuring circuits, shown in a simplified way in Figure 6.

- deflection type circuits,
- variable voltage and variable current potentiometers,
- circuits with IC-amplifiers and transmitters,
- circuits with digital indicators,
- computer-based circuits.

The direct deflection method shown in Figure 6 is the oldest way of measuring the electrothermal signals generated in thermocouples. By virtue of its simplicity and low cost, this method is likely to continue to be used in many industrial applications when high precision measurement is not required. An adjustable resistor provides the facility to trim the loop resistance. Directly coupled (DC) amplifiers are being increasingly employed to boost electrothermal signals before they are measured by deflection type instruments.

The principle of operation of the potentiometer method is comparison of voltages. In the variable voltage potentiometer represented in Figure 7, the compensating voltage, V_c, is adjusted by setting the slide wire of the potentiometer, R_p, in such a way that zero is indicated by the zero current detector. The value of the emf that is to be measured is read from the slider position on an mV scale. The balance condition is then given by

$$E_x - I_c R_c = 0 \tag{8}$$

where E_x – measured emf, I – compensating current, R_c – potentiometer resistance.

Manually operated potentiometers, which are usually portable, find common use in laboratories and for more precise short-term or industrial control measurements.

Automatic variable voltage potentiometers are used for continuous operation. Electronic servo-balanced potentiometers, which are built mainly as potentiometric recorders, are also commonly used in the measurement of different physical quantities that might be transduced to electrical signals – *see* **Article 201, Voltage Measurement, Volume 3**.

Most often they are based upon the principle of the resistance bridge potentiometer. The thermoelectric force, E_x, to be measured, is compared with the off-balance

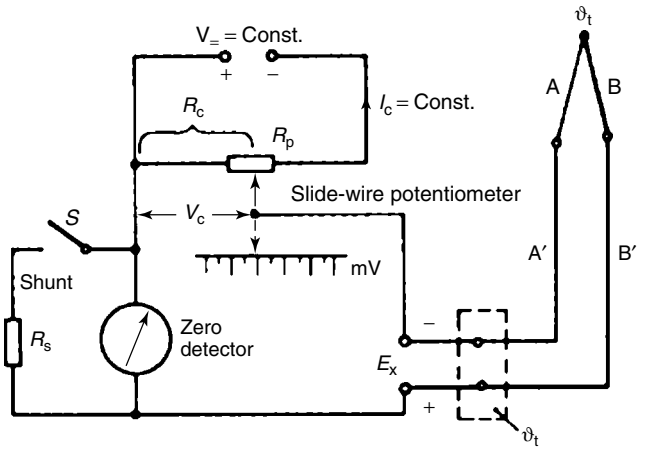


Figure 7. Manually operated variable voltage potentiometer. (*Source*: Reproduced from 'Temperature Measurement, 2nd Ed', Michalski *et al*, 2001 © John Wiley & Sons Ltd.)

voltage, of a resistance bridge – *see* **Article 126, Electrical Bridge Circuits – Basic Information, Volume 2**; and **Article 127, Unbalanced DC Bridges, Volume 2**. This voltage depends upon the slider position of a slide-wire potentiometer. An important facility when such potentiometers are calibrated in temperature degrees is automatic compensation of fluctuations in reference temperature, ϑ_r.

In *variable current potentiometers*, the thermoelectric force E_x to be measured is opposed by an adjustable voltage, generated as a voltage drop across a constant resistor. This voltage drop is adjusted by changing the compensating current I_c using a slide-wire resistor.

The value of the current, I_c, is a measure of the E_x, to be determined. A zero detector is used to indicate the null balance condition. They are also built as self-balancing potentiometers based upon an operational amplifier with feedback – *see also* **Article 113, Force-feedback Sensors, Volume 2**.

6 TESTING AND CALIBRATING

Thermocouple sensors used as working standards are of types S, R, and B. They are calibrated by the comparison method against a resistance thermometer and at the higher temperatures against standard optical pyrometers. They are also calibrated by the fixed points method. The reference junction is usually kept at 0 °C. They are used in the temperature range from 300 to 1300 °C for the calibration and testing of the same types of thermocouples of lower accuracy, as well as of other thermocouples.

Testing methods of industrial thermometers are

- the comparison method,
- the differential comparison method,
- the measurement of emf at fixed points,
- the measurement of emf at fixed points by the wire method.

The comparison method is used for all of the standardized thermocouples. In the range 300 to 1200 °C, the tested thermocouples are placed together with the standard thermocouple in an electrical furnace so that the temperature of their measuring junctions are equal.

For the thermocouples of non rare-metals, furnaces with metal blocks are used.

In the range below 300 °C, the thermocouples are tested in well-circulated baths using a mercury-in-glass thermometer as the standard. The emfs of the tested and standard thermocouples are measured by specialized digital voltmeters or calibrators.

The differential comparison method, following the American Society for Testing Materials (ASTM) E220-80 standard, is used only for rare-metal thermocouples. The same type of thermocouple is used both as the standard and as the tested thermocouples. Measuring junctions of all thermocouples are bound together by a platinum wire. The differential emf values between conductors of the same polarity of the standard and the tested thermocouples are measured. The same type of electric furnace is used as in the comparison method.

The measurement of emf at fixed points is mainly intended for testing rare-metal thermocouples. The tested thermocouple is immersed consecutively in crucibles with metals and salts of different freezing temperatures. The emf values are recorded as a function of time with the horizontal part determining the sought after emf value.

In the wire method mainly used for rare-metal thermocouples, the measuring junction of the tested thermocouple is cut in two with a short pure metal wire or plate and then soldered to both thermocouple conductors. The thermocouple is slowly heated up in the electric furnace till the inserted wire melts. The recorded emf versus time value clearly indicates the constant temperature part of the curve until the circuit is broken.

Thermocouple sensors should not be tested in a laboratory environment after a long working life or after becoming contaminated. Testing is made by the comparison method, in which the standard thermocouple is placed in situ alongside the tested one so that their measuring junctions are at the same temperature.

Testing methods, as well as the constructions of electric furnaces, fixed points crucibles and various baths are described in Michalski *et al*. (2001).

REFERENCES

Baker, H.D., Ryder, E.A. and Baker, N.H. (1953) *Temperature Measurement in Engineering*, Vol. 1, John Wiley & Sons, New York.

Kinzie, P.A. (1973) *Thermocouple Temperature Measurement*, John Wiley & Sons, New York.

Michalski, L., Eckersdorf, K., Kucharski, J. and McGhee, J. (2001) *Temperature Measurement*, 2nd edn, John Wiley & Sons, Chichester.

FURTHER READING

Several other titles cover temperature sensing. Their differently expressed explanations can be useful.

Bentley, R.E. (ed.) (1998) *Handbook of Temperature Measurement*, Springer, Singapore.

Childs, P.R.N. (2001) *Practical Temperature Measurement*, Butterworth-Heinemann.

Hofmann, D. (1976) *Dynamische Temperaturmessung (Dynamic Temperature Measurement)*, Verlag Technik, Berlin.

Hofmann, D. (1977) *Temperaturmessungen und Temperaturregelungen mit Berührungsthermometern (Temperature Measurement and Temperature Control with Contact Thermometers)*, Verlag Technik, Berlin.

Liptak, B.G. (1993) *Temperature Measurement*, CRC Press.

Leigh, J.R. (1988) *Temperature Measurement and Control*, Peter Peregrinus Ltd.

Nicholas, J.V. and White, D.R. (2001) *Traceable Temperatures: An Introduction to Temperature Measurement and Calibration*, John Wiley & Sons, Chichester.

Quinn, T.J. (1990) *Temperature*, Academic press, London.

197: Metallic Resistance Temperature Detectors (RTDs)

Dietrich Hofmann
Steinbeis Transfer Centre for Quality Assurance & Quality Measurement, Jena, Germany

1 Principles and Theory of RTDs 1322
2 RTD Temperature Sensors 1323
3 RTD Characteristics 1324
4 Practical RTD Setups 1325
5 Performance Factors of RTD Temperature Sensors 1326
6 Testing and Calibration of RTDs 1326
7 Reduction of Parasitic Lead Resistances 1327
8 Industrial RTD Probes 1328
9 Dynamical Characteristics of RTDs 1328
10 Measurement Assurance Program for Temperature 1329
References 1329

1 PRINCIPLES AND THEORY OF RTDs

High purity, chemical resistance, and its physical properties make platinum the ideal material for resistance temperature detectors (RTDs). Platinum resistance temperature detectors (PtRTDs) exhibit a number of important advantages compared to semiconductor sensors or thermistors (http://www.heraeus-sensor-nite.de/eng/technologie/frame_pttempsens.htm), for example,

- linear change of resistance with temperature,
- stable resistance temperature characteristics,
- stable resistance over cycling of temperature,
- usable over a large temperature range,
- long lifetime,
- broad temperature range from $-200\,^{\circ}\mathrm{C}$ to $1000\,^{\circ}\mathrm{C}$.

Unfortunately, PtRTD temperature sensors have a poor average sensitivity of 0.4%/K (e.g. a 100-Ω PtRTD changes only 0.4 Ω for every 1 K change).
(http://www.teamwavelength.com/downloads/newproducts/appnote-tc02.pdf)

1.1 The International Temperature Scale of 1990 (ITS-90)

The International Temperature Scale of 1990 (ITS-90) is the official International Temperature Scale since January 1990. It provides information on how the scale may be realized at different levels of accuracy. The ITS-90 extends upward from 0.65 K, is in close agreement with the Kelvin Thermodynamic Temperature Scale, and has subranges and alternative definitions in certain ranges that greatly facilitate its use (http://www.cstl.nist.gov/div836/836.05/papers/magnum90ITS90guide.pdf) p. 13.

Temperatures on the ITS-90 are expressed in International Kelvin temperatures T_{90} with the unit Kelvin (K) or expressed in the International Celsius temperatures t_{90} with the unit degree Celsius ($^{\circ}$C). The relation between T_{90} and t_{90} is

$$t_{90}/^{\circ}\mathrm{C} = T_{90}/\mathrm{K} - 273.15 \qquad (1)$$

(http://www.cstl.nist.gov/div836/836.05/papers/magnum90ITS90guide.pdf) p. 6.

Between 13.8033 K ($-259.3467\,^{\circ}$C) and 1234.93 K ($961.78\,^{\circ}$C), the ITS-90 is defined in terms of specified fixed

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

points to which temperature values have been assigned, by resistance ratios of platinum resistance thermometers (http://www.cstl.nist.gov/div836/836.05/papers/magnum90 ITS90guide.pdf) p. 11.

Standard platinum resistance thermometers (SPRT) are the instruments specified in the ITS-90 for performing measurements within the range 13.8033 K (the triple point of hydrogen) to 961.78 °C (the freezing point of silver).

SPRTs of long stem construction may be calibrated over various temperature ranges, depending on the application and required accuracy, by measurement at the appropriate ITS-90 fixed points (Figure 1).

The National Physical Laboratory (NPL) provides calibrations with uncertainties as indicated in Table 1.

Thermometers that are not suitable for measurement in fixed-point apparatus, on account of length, immersion needs, or for other reasons, will be calibrated by comparison with standard thermometers in appropriate baths, with larger calibration uncertainties.



Figure 1. Calibration service with long stem SPRT at NPL (http://www.npl.co.uk/thermal/publications/ds_pm020102.html).

2 RTD TEMPERATURE SENSORS

RTDs are wire wound and thin-film devices that work on the physical principle of the temperature coefficient of electrical resistance of metals. They require an electrical current to produce a voltage drop across the sensor that can then be measured by a calibrated read-out device (http://www.temperatures.com).

RTDs are usually included in books covering general temperature measurement (Hofmann, 1976, 1977; Leigh, 1988; Quinn, 1990; Bentley, 1998; Liptak, 1993; Childs, 2001; Michalski *et al.*, 2001; Nicholas and White, 2001).

The resistance change of platinum-temperature sensors basically depends on the length and diameter of the platinum wire. Categorized in various resistance classes from 100 to 1000 Ω, the sensors provide different signals as temperature increases. According to their design, two sensor types are distinguished:

2.1 Wire-wound Pt-temperature sensors

The measuring element of this sensor type is a wound platinum wire of defined length and diameter with two connecting wires at the end. Two different temperature sensors of this type on the market are given below.

Glass wire-wound Pt-temperature sensors (Figure 2) can be applied without protection in aqueous liquids, solvents, and oils. They consist of one or two platinum bands, wound on a glass core and homogeneously fused with a glass sleeve. They feature high shock and shake resistance and present high strength against external forces and pressure.

Ceramic wire-wound Pt-temperature sensors (Figure 3) are particularly suitable for applications in which a very

Table 1. Calibration ranges and uncertainties for SPRTs (http://www.npl.co.uk/thermal/publications/ds_pm020102.html).

Point	Temperature t_{90} (°C)	Uncertainty (°C)				
		Range 1	Range 2	Range 3	Range 4	Range 5
TP argon	−189.3442	–	0.002	0.002	0.002	–
TP mercury	−38.8344	0.0005	0.0005	0.001	0.002	–
TP water	0.01	0.0005	0.0005	0.001	0.002	0.003
MP gallium	29.7646	0.0005	–	–	–	–
FP indium	156.5985	–	0.001	–	–	–
FP tin	231.928	–	(0.001)	0.001	0.002	0.003
FP zinc	419.527	–	–	0.001	0.002	0.003
FP aluminum	660.323	–	–	–	0.004	0.005
FP silver	961.78	–	–	–	–	0.010

Note: TP = triple point, MP = melting point, FP = freezing point.



Figure 2. Glass wire-wound Pt-temperature sensor (http://www.heraeus-sensor-nite.de/eng/produkte/frame_pt_glas_gew.htm).



Figure 3. Ceramic wire-wound Pt-temperature sensor (http://www.heraeus-sensor-nite.de/eng/produkte/frame_pt_keram_gew.htm).

high temperature stability and high temperature shock resistance are required. They consist of a platinum coil that is embedded in an ultrapure ceramic body of aluminum oxide. They have an extremely low drift even in long-term use.

2.2 Pt thin-film sensors

Temperature sensors of Pt thin-film technology (Figure 4) are particularly suitable for mass applications. They feature an extremely small space requirement. They cover the large temperature range of $-200\,^{\circ}\mathrm{C}$ to $+800\,^{\circ}\mathrm{C}$.

Typical applications are in the automotive industry, white goods, HVAC, process technology, power stations, aviation and aerospace, equipment, and machines for medicine and industry.

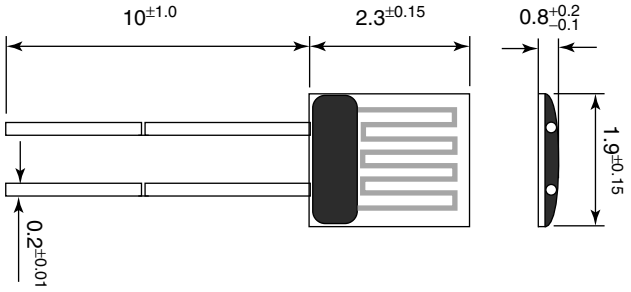


Figure 4. Pt thin-film sensor (http://www.heraeus-sensor-nite.de/eng/produkte/frame_pt_duenn_sens.htm).

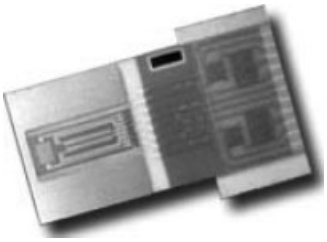


Figure 5. Pt sensor module (http://www.heraeus-sensor-nite.de/eng/produkte/frame_pt_sens_mod.htm).

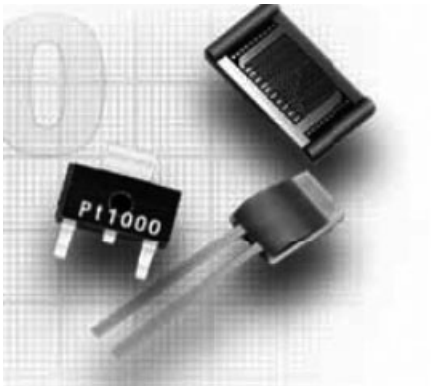


Figure 6. Primary enclosed Pt thin-film sensors (http://www.heraeus-sensor-nite.de/eng/produkte/frame_prim_geh.htm).

2.3 Pt sensor modules

Pt sensor modules (Figure 5) have a multifunctional design. They usually perform tasks such as air-volume determination (http://www.heraeus-sensor-nite.de/eng/produkte/frame_luftmassensensor.htm) or air-quality measurement in air-conditioned rooms (http://www.heraeus-sensor-nite.de/eng/maerkte/frame_klimatisierung.htm).

2.4 Primary enclosed sensors

For special applications, elementary sensors are assembled in standard housings or applied on personal computer boards (PCB) (Figure 6). These modules are created for direct use by the customers for further processing. Time-consuming and cost-intensive steps for production preparation are not needed.

3 RTD CHARACTERISTICS

Operating on the principle that resistance of any given metal changes with temperature, thermoelectric RTDs are the sensors of choice in applications in which accuracy, stability, and repeatability are critical. To maximize linearity and temperature ranges, thermoelectric RTDs are produced from

high-purity platinum. In addition, special high-temperature platinum delivers high stability and durability of RTDs at 650 °C (1202 °F). Thermoelectric RTDs are available in diverse configurations with two-, three-, and four-wire construction ranging from −200 °C to 650 °C (−328 °F to 1202 °F). Tolerances meet the standards of ASTM E1137 Grade A or B and IEC 751 Class A or B, depending on the customers needs.

(http://www.thaiinstrumentation.com/thermocouple_rtd.html).

3.1 Temperature circuits

2-wire circuit: A Wheatstone bridge, *see* **Article 126, Electrical Bridge Circuits – Basic Information, Volume 2** and **Article 127, Unbalanced DC Bridges, Volume 2**, is the most common approach for measuring an RTD (http://content.honeywell.com/sensing/prodinfo/temperature/technical/c15_139.pdf). As RT increases or decreases with temperature, V_{out} also increases or decreases. Use an operation amplifier to observe V_{out}. Lead wire resistance, L_1 and L_2 directly adds to the RTD leg of the bridge (Figure 7).

3-wire circuit: In this approach, L_1 and L_3 carry the bridge current. When the bridge is in balance, no current flows through L_2 so no L_2 lead resistance is observed. The bridge becomes unbalanced as RT changes. Use an operation amplifier to observe V_{out} and prevent current flow

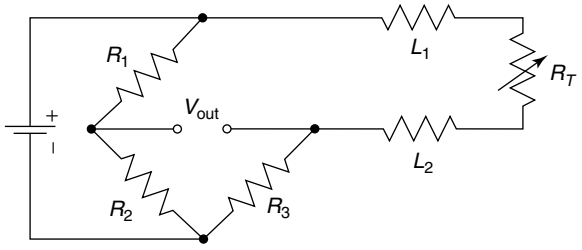


Figure 7. Two-wire circuit for RTDs (http://content.honeywell.com/sensing/prodinfo/temperature/technical/c15_139.pdf).

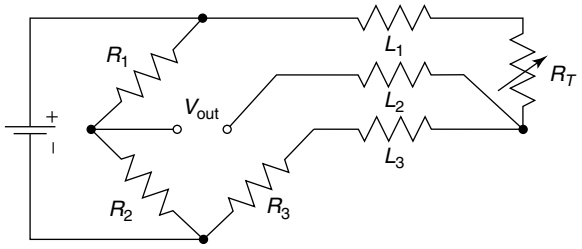


Figure 8. Three-wire circuit for RTDs (http://content.honeywell.com/sensing/prodinfo/temperature/technical/c15_139.pdf).

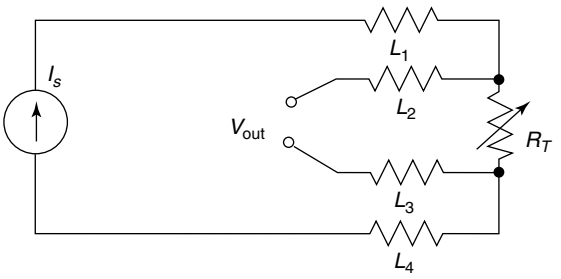


Figure 9. Four-wire circuit for RTDs (http://content.honeywell.com/sensing/prodinfo/temperature/technical/c15_139.pdf).

in L_2. The effects of L_1 and L_3 cancel when $L_1 = L_3$, since they are in separate arms of the bridge (Figure 8).

4-wire circuit: A 4-wire approach uses a constant current source to cancel lead wire effects even when $L_1 \neq L_4$. Use an operation amplifier to observe V_{out} and prevent current flow in L_2 and L_3 (Figure 9).

4 PRACTICAL RTD SETUPS

Sensor sensitivity affects a temperature control circuit's set point accuracy, measurement accuracy, and temperature stability operation. Parasitic resistances created by the wire leads of an RTD affect the set point and measurement accuracy (Figure 10).

The practical example in Figure 10 demonstrates the effect of lead resistance on the resistance measurement accuracy of a 100-Ω RTD. The sensor resistance measured in Figure 10 includes the parasitic lead resistances, which can be on the order of the sensor's sensitivity. The parasitic

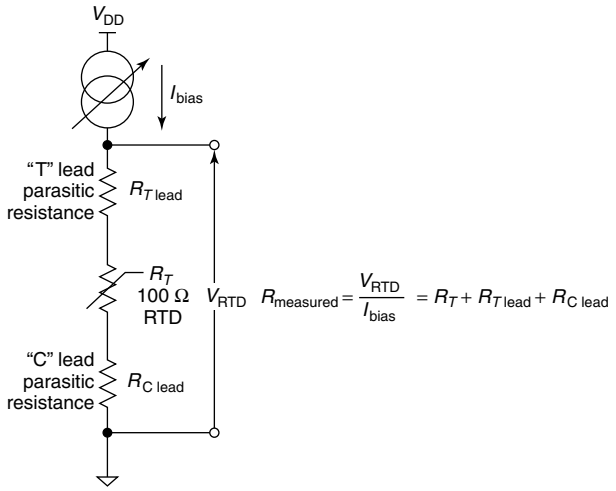


Figure 10. Effects of lead resistance on RTD measurement accuracy (http://www.teamwavelength.com/downloads/newproducts/appnote-tc02.pdf).

lead resistances are not easily measured since most RTD sensors are sealed against corrosion. A 100-Ω RTD with 0.1-Ω parasitic resistances in its leads will have an average temperature measurement offset of +0.5 °C.

5 PERFORMANCE FACTORS OF RTD TEMPERATURE SENSORS

Platinum's resistance versus temperature function is accurately modeled by the Callendar-Van Dusen equation. This equation uses constants A, B, and C, derived from resistance measurements at 0 °C, 100 °C, and 260 °C (Figure 11).

Callendar-Van Dusen Equation:

$$R_T = R_0(1 + AT + BT^2 - 100CT^3 + CT^4) \quad (2)$$

R_T = Resistance (Ω) at temperature T (°C)
R_0 = Resistance (Ω) at 0 °C
T = Temperature in °C.

For $T > 0$ °C, the quadratic formula can be used to solve for temperature as a function of measured resistance with the following result:

$$0 = R_0BT^2 + R_0AT + (R_0 - R_T) \text{ implies} \quad (3)$$

$$T_R = \frac{\{-R_0A + [R_0^2A^2 - 4R_0B(R_0 - R_T)]^{-2}\}}{2R_0B} \quad (4)$$

Platinum RTDs are specified by resistance at 0 °C, R_0, and alpha (α), a term related to the temperature coefficient of resistance (TCR). The Callendar-Van Dusen constants A, B, and C are derived from alpha (α) and other constants,

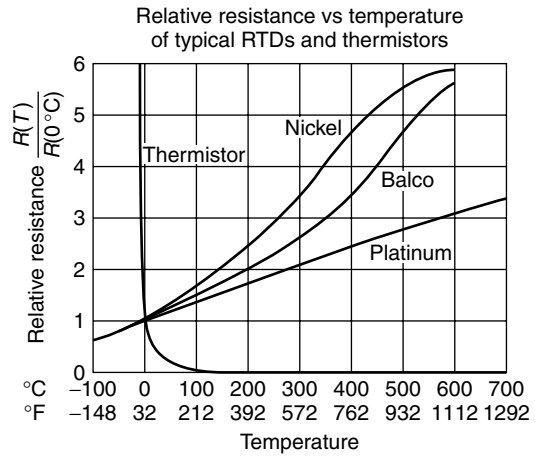


Figure 11. Platinum RTD resistance versus temperature function (http://content.honeywell.com/sensing/prodinfo/temperature/technical/c15_136.pdf).

Table 2. Callendar-van Dusen constants for platinum RTDs with $R_0 = 100\,\Omega$ and different temperature coefficients α (http://content.honeywell.com/sensing/prodinfo/temperature/technical/c15_136.pdf).

Alpha, α (°C)	0.003750 ± 0.00003	0.003850 ± 0.0001
Delta, δ (°C)	1.605 ± 0.009	1.4999 ± 0.007
Beta, β (°C)	0.16	0.10863
A (°C)	3.81×10^{-3}	3.908×10^{-3}
B (°C)	-6.02×10^{-7}	-5.775×10^{-7}
C (°C)	-6.0×10^{-12}	-4.183×10^{-12}

Note: Both $\beta = 0$ and $C = 0$ for $T > 0$ °C.

delta (δ) and beta (β), which are obtained from actual resistance measurements. Common Callendar-Van Dusen constant values are shown in Table 2.

The definitions of the Callendar-Van Dusen constants A, B, C, and alpha, delta, and beta (α, δ, and β), and their interrelationships are given by (5) to (10).

$$A = \frac{\alpha + (\alpha \bullet \delta)}{100} \quad (5)$$

$$B = \frac{-\alpha \bullet \delta}{100^2} \quad (6)$$

$$C_{T<0} = \frac{-\alpha \bullet \delta}{100^4} \quad (7)$$

$$\alpha = \frac{R_{100} - R_0}{100R_0} \quad (8)$$

$$\delta = \frac{[R_0(1 + 260\alpha) - R_{260}]}{[4.16\alpha R_0]} \quad (9)$$

$$\beta = \text{Constant for } T < 0\,^\circ\text{C} \quad (10)$$

The fundamental accuracy and repeatability performance of an RTD is determined by the repeatability of the empirically measured values $R_0 \pm \Delta R_0$; $R_{100} \pm \Delta R_{100}$ and $R_{260} \pm \Delta R_{260}$.

For tolerance standards and accuracy, *IEC 751*, the most commonly used standard for PtRTDs defines two performance classes for 100 Ω 0.00385 alpha PtTRDs, *Class A* and *Class B*. These performance classes (also known as *DIN A* and *DIN B* because of DIN 43760) define tolerances on ice point and temperature accuracy. These tolerances are also often applied to PtRTDs with ice point resistance outside of IEC 751's 100 Ω assumption.

Class C and *Class D* (each doubling the prior tolerance level) are also used.

(http://content.honeywell.com/sensing/prodinfo/temperature/technical/c15_136.pdf)

6 TESTING AND CALIBRATION OF RTDs

The National Institutes of Metrology provide calibration services for SPRTs from 13.8 K to 1235 K. From 13.8 K

to 83.8 K, SPRT calibrations are performed by comparison with a set of reference thermometers. From 83.8 K to 962 °C, calibrations are performed using the ITS-90 defining fixed points. Maximum expanded calibration uncertainties are given in Table 3 (http://ts.nist.gov/ts/htdocs/230/233/calibrations/thermodynamic/resistance%20_thermometry.htm#33010C).

The comparison calibration of rhodium–iron resistance thermometers (RIRTs) in the temperature range from 0.65 to 26 K, using the NIST-maintained ITS-90, is based on a set of reference capsule-type RIRTs. The expanded uncertainty of those calibrations is given in Table 4.

It is important that resistance thermometers be protected from any mechanical shock that could alter their calibration. For shipment, the thermometer should be softly supported within a case but not be free to rattle. This necessitates the use of packing material that does not become compacted. The thermometer case should be softly packed inside a shipping container. Suitable containers will be provided when a thermometer shipping container is not satisfactory for reuse (http://ts.nist.gov/ts/htdocs/230/233/calibrations/thermodynamic/resistance%20_thermometry.htm#33010C).

Table 3. Maximum expanded uncertainties of SPRTs at NIST (http://ts.nist.gov/ts/htdocs/230/233/calibrations/thermodynamic/resistance%20_thermometry.htm#33010C).

Temperature subrange (K)	Maximum expanded uncertainty (mK)
13.8033–273.16	0.54[a]
24.5561–273.16	0.40[a]
54.3584–273.16	0.29[a]
83.8058–273.16	0.39
234.3156–302.9146	0.20
273.15–302.9146	0.04
273.15–429.7485	0.32
273.15–505.078	0.37
273.15–692.677	0.54
273.15–933.473	0.68
273.15–1234.93	1.1

[a]The uncertainties used at the Ar, Hg, and H_2O triple points are based on the fixed-point realization.

Table 4. NIST expanded uncertainties for comparison measurements of RIRTs (http://ts.nist.gov/ts/htdocs/230/233/calibrations/thermodynamic/resistance%20_thermometry.htm#33010C).

Range (K)	Expanded uncertainty (mK)
0.65–2.0	0.5
2.0–5.0	0.08
5.0–8.8	0.17
8.8–13.8	0.21
13.8–17.0	0.34
17.0–20.3	0.18
20.3–24.5561	0.24
24.5561–54.3584	0.27
54.3584–83.8058	0.22

6.1 Efficient calibration methods for platinum resistance thermometers

Calibration methods for platinum resistance thermometers can be simplified by reducing the number of calibration points. For applications that do not require the low uncertainties of the ITS-90 definitions, several modifications are available to improve calibration efficiency. The proposed schemes utilize the ITS-90 reference functions, simplified deviation equations, and in some cases correlations between resistance ratio values. Calibration ranges of interest include 20 K to 273 K, 273 K to 933 K, and various subranges spanning the ice point from 77 K to 505 K. Thermometers are tested using two to three calibration temperatures with media, convenient for comparison calibration. The methods apply to a broad class of thermometers constructed with high-purity platinum. Several examples of fit error are presented for SPRTs and for more rugged aerospace sensor designs. Recommendations to minimize the influence of thermal hysteresis are also discussed (http://www.cstl.nist.gov/div836/836.05/thermometry/symposium/program.htm). The approximations introduce errors typically between 10 and 100 mK depending on thermometer type and temperature range. For many applications, these calibration methods satisfy the trade-off between cost and uncertainty. More information is available in **Article 198, Calibration and Standards in Temperature Measurement, Volume 3**.

7 REDUCTION OF PARASITIC LEAD RESISTANCES

Adding a third lead to the RTD and employing a balanced bridge circuit can reduce the effect of parasitic lead resistance (Figure 12).

The practical example in Figure 12 shows a 3-wire RTD connected to a compensated-lead bridge circuit. The lead resistances can be compensated for the situation in which the ratio resistors R_1 and R_2 are equal and R_X equals the resistance of the RTD. Changes in the bridge's voltage reference have little effect when the bridge is balanced.

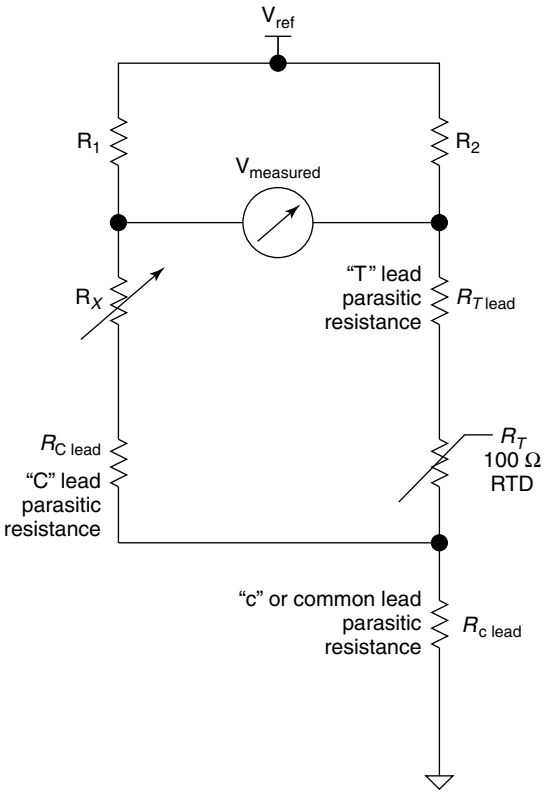


Figure 12. Compensated three-wire RTD bridge circuit (http://www.teamwavelength.com/downloads/newproducts/appnote-tc02.pdf).

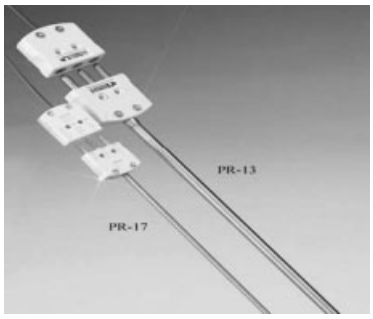


Figure 13. Quick disconnect RTD probe (http://www.omega.com/ppt/pptsc.asp?ref=PR-13&nav=temc03).

8 INDUSTRIAL RTD PROBES

For easier handling, quick disconnect RTD probes with standard size connectors (Figure 13) are used.

8.1 Technical data

$\alpha = 0.00385$ standard; 6-, 9-, 12-, 18-, and 24-inch lengths standard; $R_0 = 100\,\Omega$ at $0\,^{\circ}\mathrm{C}$; Three lead configuration standard.



Figure 14. Industrial RTD probe with protection head (http://www.omega.com/ppt/pptsc.asp?ref=PR-12&Nav=temc05).

Options: Two and four lead configurations; Teflon coating of probe; $\alpha = 0.00316$; Nonstandard probe lengths; other values of R_0.

For stronger environmental conditions, the industrial RTD probes with cast iron protection heads (Figure 14) are used.

8.2 Technical data

Three-wire configuration standard; $R_0 = 100\,\Omega$; 1/4-inch diameter; 18-inch length, European curve with $\alpha = 0.0385$.

9 DYNAMICAL CHARACTERISTICS OF RTDs

Industrial RTDs are enclosed in a heavy protective thermo well. The thermo well slows down the step response of the RTD significantly, increasing the time constant in the range of seconds.

To estimate the time constant, the RTD can be moved from a temperature bath Θ_{initial} to another temperature bath Θ_{final}. A typical step response is shown in Figure 15.

If at time t the indicated temperature is $\Theta_{\text{indicated}}$, then the time constant T may be estimated from the step response using

$$\Theta_{\text{final}} - \Theta_{\text{indicated}} = (\Theta_{\text{final}} - \Theta_{\text{initial}})\mathrm{e}^{-t/T} \qquad (11)$$

The transfer function of the inertial system with first-order characteristic is

$$F(p) = \frac{k}{Tp + 1} \qquad (12)$$

with p as the Laplace operator.

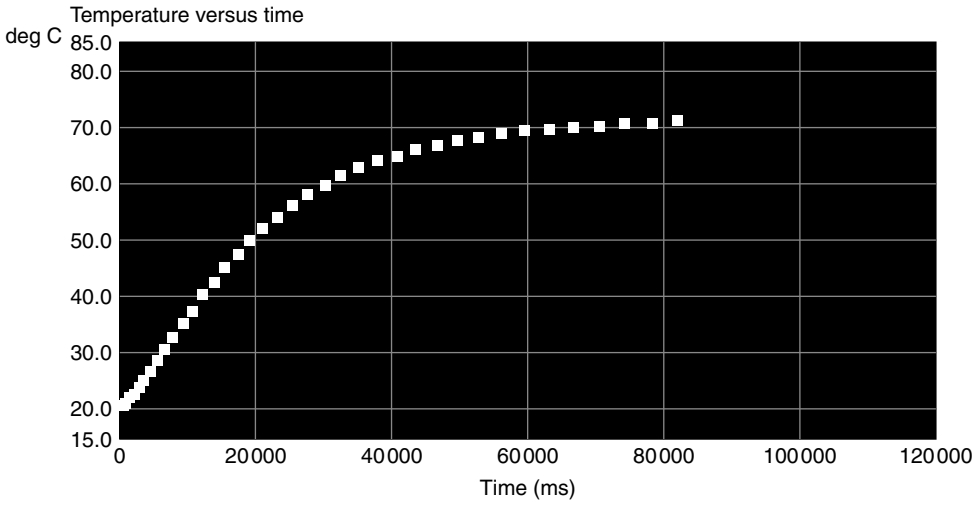


Figure 15. Transient function (step response function) of an industrial RTD (http://pasture.ecn.purdue.edu/∼abe591s/Temperature_Lab.pdf).

For quicker estimation, the time constant can be taken from the subtangent of the transfer function (Figure 15).

10 MEASUREMENT ASSURANCE PROGRAM FOR TEMPERATURE

The purpose of the Measurement Assurance Program (MAP) (http://ts.nist.gov/ts/htdocs/230/233/calibrations/thermodynamic/resistance%20_thermometry.htm#33010C) is to assure the accuracy of the calibration of temperature standards (83 K to either 420 °C or 661 °C) conducted by participating laboratories when using platinum resistance thermometry. The MAP transport standard consists of sets of three commercial SPRTs packaged in a special shock-proof shipping container. These SPRTs are used to assess both the reproducibility and the uncertainty of calibrations performed by the participating laboratory.

In order to achieve high accuracy, SPRTs used as standards should be either metal sheathed or of the matte-finish glass-sheathed type to avoid systematic errors arising from light-pipe effects in the glass sheath. The participant must have a triple point of water cell and an appropriate resistance bridge.

The National Institute of Standards and Technology (NIST) (http://www.nist.gov) provides worksheets on which the participant records data. The participant calculates the thermometer constants from experimental data, records them, and prepares tables of either resistance ratio or resistance versus temperature. These completed worksheets and the participant's calibration reports are sent to NIST with the return of the MAP SPRTs.

The SPRTs are recalibrated upon return to NIST and the participant's data are compared with NIST's calibration results. NIST provides a plot of the participating laboratory's temperature deviation from NIST values and a written analysis of the data, including any pertinent observations. In a typical MAP transfer, the participant records measurements over a period of 1 to 2 months. A typical turnaround time from the date NIST receives the participant's data until a test report is sent to the participant is 4 to 6 weeks. Former participants in the MAP have had expanded uncertainties that ranged from about 1 mK to several tens of mK.

No rigid recommendations can be given concerning how often a participant should utilize the temperature MAP service. Experience has indicated that when temperature measurements are in a state of statistical control, as evidenced by in-house check standards and control charts used to monitor the measurement process, the participant should be able to go at least 3 years between transfers from NIST without significantly degrading the confidence in the correctness of the measurements.

Among other NIST services for temperature calibration laboratories, NIST provides standard reference materials for use as defining fixed points of the ITS-90 and as secondary reference points (http://ts.nist.gov/ts/htdocs/230/233/calibrations/thermodynamic/resistance%20_thermometry.htm#33010C).

REFERENCES

Bentley, R.E. (ed.) (1998) *Handbook of Temperature Measurement*, Springer, Singapore.

Childs, P.R.N. (2001) *Practical Temperature Measurement*, Butterworth-Heinemann, Jordan Hill, Oxford, UK.

Hofmann, D. (1976) *Dynamische Temperaturmessung (Dynamic Temperature Measurement)*, Verlag Technik, Berlin, Germany.

Hofmann, D. (1977) *Temperaturmessungen und Temperaturregelungen mit Berührungsthermometern (Temperature Measurement and Temperature Control with Contact Thermometers)*, Verlag Technik, Berlin, Germany.

Leigh, J.R. (1988) *Temperature Measurement and Control*, Peter Peregrinus Ltd., London, UK.

Liptak, B.G. (1993) *Temperature Measurement*, CRC Press, Boca Raton, FL.

Michalski, L., Eckersdorf, K., Kucharski, J. and McGhee, J. (2001) *Temperature Measurement*, John Wiley & Sons, Chichester, UK.

Nicholas, J.V. and White, D.R. (2001) *Traceable Temperatures: An Introduction to Temperature Measurement and Calibration*, John Wiley & Sons, Chichester, UK.

Quinn, T.J. (1990) *Temperature*, Academic press, London, UK.

198: Calibration and Standards in Temperature Measurement

D.R. White

Measurement Standards Laboratory of New Zealand, Lower Hutt, New Zealand

1 Temperature Scales and Standards 1331
2 Calibration of Low-temperature Sensors 1333
3 Calibration of Mid-range Temperature Sensors 1333
4 Calibration of High-range Temperature Sensors 1334
5 Reference Thermocouple Tables 1335
References 1336

1 TEMPERATURE SCALES AND STANDARDS

Although temperature is defined thermodynamically, it is impracticably difficult to measure thermodynamic temperature. Instead, the international measurement community, through the Conférence Générale des Poids et Mesures (CGPM), provides an approximation to the thermodynamic scale based on practical thermometers and interpolation between fixed points with defined temperatures. The fixed points are the melting-, freezing-, and triple-point temperatures of various pure substances, and the interpolating thermometers include platinum-resistance thermometers, radiation thermometers, vapor-pressure thermometers, and constant-volume gas thermometers. This scale is sufficiently accurate and reproducible, of low cost, and accessible to satisfy almost all trade, safety, and research needs. At intervals of approximately 20 years, the scale is updated to match the latest thermodynamic measurements and measurement techniques. The current scale is called the *International Temperature Scale of 1990* or ITS-90 (Preston-Thomas, 1990), and all temperature measurements should be traceable to ITS-90. Previous scales were defined in 1927, 1948, and 1968, and where necessary, scale temperature is distinguished by a subscript according to the year of the scale, for example, T_{68} (K) or t_{68} (°C) for temperatures measured on the 1968 scale.

The ITS-90 allows measurement from 0.65 K to the highest temperature measurable in terms of Planck's radiation law. A recent addition is the Provisional Low Temperature Scale of 2000, PLTS-2000, (Rusby *et al*., 2001), a low-temperature extension that allows measurement down to 1 mK. Figure 1 shows a schematic diagram of the elements of ITS-90 and PLTS-2000. Notably, the important temperature range from 13 K to 962 °C is defined in terms of standard platinum resistance thermometers, SPRT, which are the easiest to use and calibrate of the ITS-90 thermometers. The accuracy of the ITS-90 ranges from about 0.2 mK at 0.65 K, 1 mK near room temperature, about 5 mK at 960 °C for SPRTs, to about 0.1 °C at 960 °C for radiation thermometers and increases as T^2 at higher temperatures.

The Bureau International des Poids et Mesures, BIPM (www.bipm.fr), publishes guidelines with extensive references to scientific papers (BIPM, 1990a) to assist in the realization of ITS-90. A second guide (BIPM, 1990b) provides advice on calibration and techniques for approximating ITS-90 using other thermometers and methods. Secondary fixed points (Bedford *et al*., 1996) can be used to calibrate thermometers at temperatures other than those defined by the ITS-90 fixed points.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

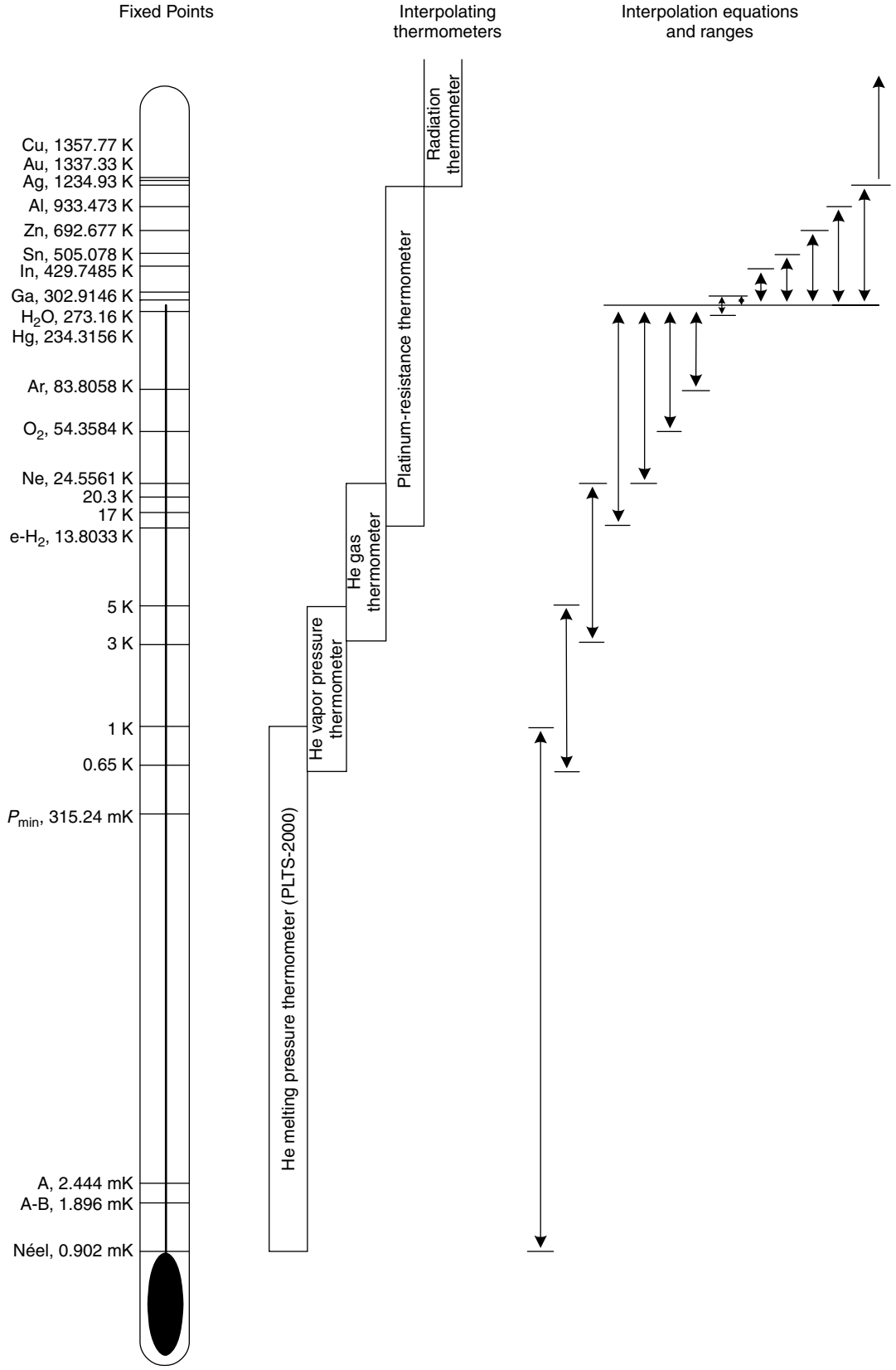


Figure 1. A guide to the main features of ITS-90. The temperatures are the temperatures of the fixed points defined by ITS-90 or PLTS-2000.

The science underlying the temperature scale is described in Quinn (1990) and Schooley (1986), while the principles of calibration and practical temperature measurement are described in Nicholas and White (2001) and Bentley (1998). More general texts providing extensive overviews of different temperature measurement techniques include Childs (2001), McGee (1988), and Michalski *et al*(2001).

2 CALIBRATION OF LOW-TEMPERATURE SENSORS

At temperatures below 150 K, material properties change significantly from their room temperature values. In particular, the thermal conductivity of good conductors, such as metals, increases and the heat capacity of all materials decreases as the temperature is reduced. Consequently, the elimination of temperature gradients and heat leaks becomes increasingly difficult. Knowledge of low-temperature techniques (White and Meeson, 2002) is necessary to carry out measurements at low temperatures.

For temperatures between 13 K and 273 K, the ITS-90 is defined in terms of capsule SPRTs (Figure 2). At lower temperatures, the ITS-90 is defined in terms of a gas thermometer and a vapor-pressure thermometer. These are not so convenient, and the realization of the ITS-90 is normally carried out simultaneously with the calibration of more convenient sensors such as rhodium–iron resistance thermometers. These are electrically similar to SPRTs but have much improved sensitivity (dR/dT) for temperatures down to 0.5 K.

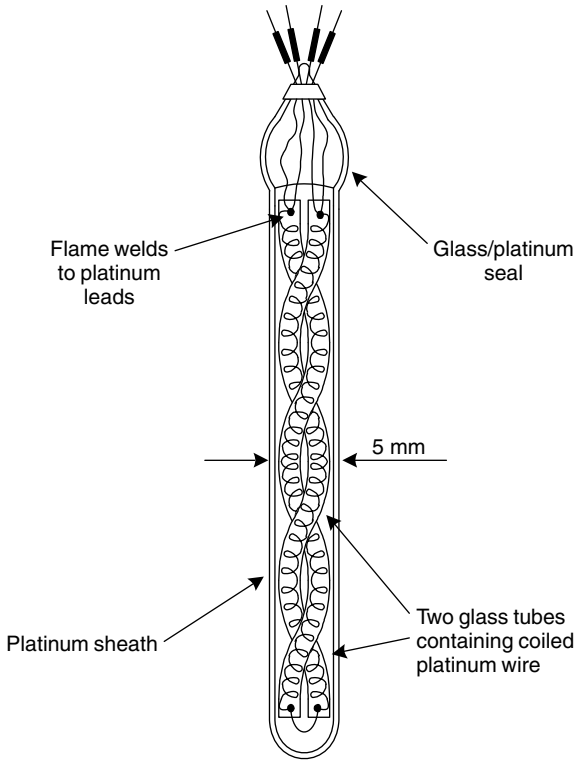


Figure 2. The construction of a typical capsule standard platinum-resistance thermometer used for temperatures below 156 °C. Rhodium–iron resistance thermometers have a similar construction.

For some temperature ranges above 0.5 K, vapor-pressure thermometers may be convenient reference thermometers. They are accurate, simple to make, relatively immune to systematic effects except for impurities, and tables have been published for a number of different fluids (BIPMb). They are, however, limited to quite narrow temperature ranges, and for some ranges there is no suitable fluid.

A copper block cooled using nitrogen or helium vapor and surrounded by radiation shields is a commonly used calibration medium. The copper block has several holes for mounting platinum resistance or rhodium–iron resistance thermometers and the thermometers under calibration. Vacuum grease or helium gas is used to ensure good thermal contact. For temperatures above 4 K, it is also possible to use closed cycle refrigerators, which are more convenient and faster than liquid helium refrigeration. The BIPM Guidelines for ITS-90 (BIPMa, BIPMb) describe a number of successful designs of cryostats suitable for calibration systems.

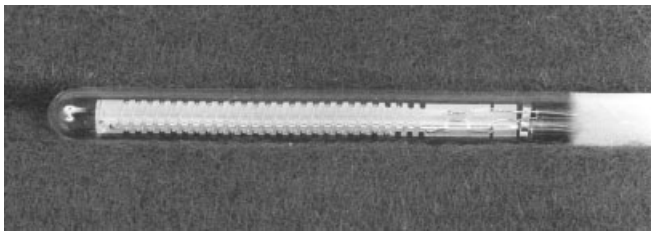
Many different sensors are available for use at low temperatures including germanium resistance thermometers, carbon resistance thermometers, and diode thermometers. Because semiconductors are nonlinear, these devices require a relatively large number of calibration points (15 to 30 depending on temperature range).

3 CALIBRATION OF MID-RANGE TEMPERATURE SENSORS

In the range from 150 K to 350 °C, temperature is defined in terms of the long-stem SPRT (Figure 3), with an accuracy of about 1 mK. At the next level of accuracy, 5 mK to 50 mK, the most useful reference thermometers are industrial platinum-resistance thermometers. These must employ a 4-wire resistance measurement to eliminate lead resistances and have the platinum wire partially supported to prevent strain and hysteresis effects. Other suitable reference thermometers include liquid-in-glass thermometers and thermistors, which are better suited to narrow temperature ranges, and rare-metal thermocouples (*see* **Article 196, Thermocouple Temperature Sensors, Volume 3**). Other thermometers including base-metal thermocouples (nickel based) are generally not suitable due to a lack of stability and reliability.



(a)



(b)

Figure 3. A long-stem standard platinum-resistance thermometer for use at temperatures above 150 K: (a) the full assembly in carrying case and (b) a close-up of the sensing element.

Because this temperature range includes a large fraction of temperature measurement activity, most equipment is available commercially. Several companies specialize in the supply of thermometry equipment and can outfit a complete calibration laboratory for any level of accuracy ranging from a full ITS-90 implementation at ± 1 mK to a general-purpose laboratory at ± 1 K.

A wide choice of calibration media is available with stirred fluid baths (e.g. Figure 4) being the easiest to use. Depending on temperature, the fluids range from light paraffins, methanol, light mineral oils, water (with and without antifreeze), heavy mineral oils, and silicon oils, through to mixtures of salts enabling access to temperatures up to 500 °C. Generally, each fluid enables a temperature range of about 80 °C between the temperature at which it stirs well (30 centistokes max. kinematic viscosity) and the temperature at which evaporation is excessive or decomposition is likely. Silicon oils, although expensive, have a useful range of about 160 °C, and molten salts enable a range of about 200 °C.

In recent years, dry-block calibrators have become popular for calibrations. However, their thermal design is such that they are prone to temperature gradients of up to 1% of the temperature difference with respect to ambient temperature. They are suitable for some industrial calibrations but not for laboratory work. Fluidized sand beds can also be used for higher temperatures but they can be dusty and prone to large temperature gradients.

Nicholas and White (2001) provide extensive advice for thermometer calibrations in this range. They detail

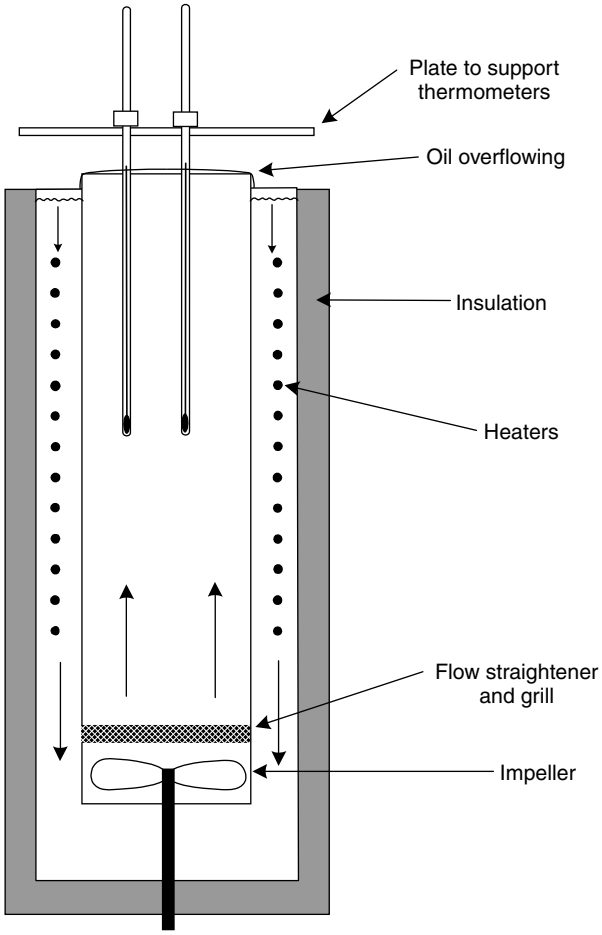


Figure 4. An overflow style calibration bath. The concentric design provides good control, nonturbulent flow, and horizontal viewing to ensure easy reading of liquid-in-glass thermometers.

important effects in the calibration of liquid-in-glass thermometers, platinum-resistance thermometers, thermocouples, and radiation thermometers, and include example procedures and uncertainty analyses. Other books describing the negative aspects of thermometers in addition to the operating principles include the BIPM Guide (1990b) and the three-volume handbook of Bentley (1998).

4 CALIBRATION OF HIGH-RANGE TEMPERATURE SENSORS

At temperatures above 350 °C, there are two thermometers defined by ITS-90 – SPRTs up to 960 °C and radiation thermometers for all temperatures above 960 °C. Radiation thermometers can also be used at temperatures down to −50 °C, but their accuracy is typically 0.1 °C at best, whereas SPRTs are accurate to a few millikelvins. Rare-metal thermocouples are also useful reference thermometers for temperatures above 450 °C where industrial platinum-resistance

thermometers become unreliable. The gold–platinum and platinum–palladium thermocouples perform almost as well as SPRTs with accuracies of 5 mK or so up to 1000 °C, and for platinum-palladium, with accuracies of 0.15 K or better up to 1500 °C.

At temperatures above 350 °C, thermal conductivity is often dominated by electromagnetic (infrared) radiation. For example, surfaces near 500 °C radiate approximately 20 kW m^{-2}. Obtaining calibration media with temperature gradients of less than 1 °C in these conditions is difficult. For the highest accuracy, heat pipes are used. These are open (pressure-controlled type) or closed vessels containing a pure liquid under vacuum (Figure 5). When the vapor-pressure of the liquid is sufficient and all surfaces in the vessel are wet, the effective thermal conductivity of the vessel can exceed 10^5 times that of copper by virtue of the absorption and release of latent heat at the surface of the fluid. Mercury, cesium, potassium, and sodium are all used for heat pipes with each covering approximately 250 °C to 400 °C temperature range. Heat-pipe furnace liners are available commercially.

Calibration of Rare-metal thermocouples (Types B, R, and S) are best carried out by welding the junctions of the thermocouples together. This satisfies simultaneously the requirements of ensuring that the reference junctions are in close thermal contact, at the same temperature, and each junction is isothermal to prevent the generation of spurious voltages. Because the Seebeck voltage is generated at the temperature gradients in the wires and not at the junction, thermocouple calibrations are only meaningful if the temperature gradients can be reproduced or the wires are homogeneous. For the highest accuracy work, a homogeneity test is essential to demonstrate that the wire is homogeneous (Bentley 1998, ASTM 1993).

Base-metal thermocouples are not usually calibrated except as representative samples of a batch. Heat treatment of the base-metal alloys introduces inhomogeneities of up to 8 °C (in Type K, less in other types), which often makes calibration of used thermocouples futile.

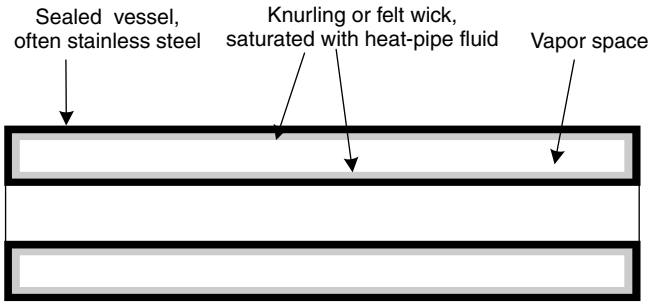


Figure 5. A section along the main axis of a closed heat-pipe furnace liner. The wick is saturated with the fluid so that evaporation at hot spots and condensation at cold spots moves heat via the latent heat of vaporization with almost zero change in temperature.

Radiation thermometers are calibrated using blackbody cavities, which simulate perfect emitters according to Planck's radiation law. For the highest accuracy, fixed-point black bodies employing the ITS-90 defined freezing points should be used. Variable temperature blackbodies using rare metal thermocouples as reference thermometers are better suited for the calibration of general purpose radiation thermometers with accuracies of about 1 °C. Both fixed-point and variable temperature blackbodies are available commercially. Quinn (1990) is a key reference for primary radiation thermometry, DeWitt and Nutter (1988) provide extensive information on the theory and practice of radiation thermometry, and Nicholas and White (2001) provide guidelines on use and calibration of general purpose radiation thermometers.

5 REFERENCE THERMOCOUPLE TABLES

The letter-designated thermocouples are no longer defined by alloy composition but by mathematical equations. This provides manufacturers with the opportunity to design alloys for different environments and temperatures. Therefore, to ensure specified accuracy, thermocouple wires should only be bought and used as a pair. The mathematical definitions are often presented as tables, and assume that the reference junction is at 0 °C (the ice point). The temperature range, composition, and manufacturing tolerance of a range of thermocouples types are summarized in Table 1.

The base-metal thermocouples (Types E, J, K, N, and T) all have nickel in one or both of the wires. They have a high Seebeck coefficient (>30 μV/°C above 0 °C), a modest accuracy (approximately 0.5 °C ± 0.5% of reading), and are of low cost. Noble-metal or rare-metal thermocouples (Types B, R, and S) all have platinum in both wires. They have a low Seebeck coefficient (approximately 10 μV/°C) but have a higher accuracy (approximately 0.2 °C ± 0.05%), and are relatively expensive. The tables for the letter-designated thermocouples are available on the National Institute for Standards and Technology, NIST, Web site (http://srdata.nist.gov/its90/main/).

Users should be aware that the same letter designation might be used for different thermocouples. For example, European equipment manufactured before 1993 using Type J wire may have been made to the older (Deutsches Institut für Normung e. V.) DIN Type J standard, which at some temperatures is several degrees different from the (International Electrotechnical Commission) IEC Type J standard. Since 1990, thermocouple standards have become

Table 1. Summary of the temperature ranges and compositions of some common and useful thermocouples.

Type designation	Composition, positive leg	Composition, negative leg	Temperature range (°C)	Tolerance	Reference	Comment
B	Pt 30% Rh	Pt 6% Rh	0 to 1820	0.5 °C or 0.4%	IEC 584	Suited for use above 1000 °C
C	W 3% Re	W 25% Re	0 to 2400	5 °C or 1%	ASTM E988	Suits reducing environment, high temperatures, brittle
D	W 5% Re	W 26% Re	0 to 2400	5 °C or 1%	ASTM E989	Suits reducing environment, high temperatures, brittle
G	W	W 26% Re	300 to 2400	5 °C or 1%	ASTM E1751	Suits reducing environment, high temperatures, very brittle
E	Ni 10% Cr	Cu 45% Cr	−270 to 1000	1.5 °C or 0.4%	IEC 584	Highest sensitivity, good at low temperatures
J	Fe	Cu 45% Cr	−210 to 900	1.5 °C or 0.4%	IEC 584	Suits reducing environments
K	Ni 10% Cr	Ni 2% Al 2% Mn 1% Si	−270 to 1372	1.5 °C or 0.4%	IEC 584	General purpose, low accuracy
L	Fe	Cu 45% Cr	−210 to 900	1.5 °C or 0.4%	DIN 43710	DIN precursor to Type J
N	Ni 14.2% Cr 1.4% Si	Ni 4.4% Si 0.15% Mg	−270 to 1300	1.5 °C or 0.4%	IEC 584	General purpose, better than Type K
R	Pt 13% Rh	Pt	−50 to 1768	1 °C or 0.3%	IEC 584	High accuracy, susceptible to contamination
S	Pt 10% Rh	Pt	−50 to 1768	1 °C or 0.3%	IEC 584	High accuracy, susceptible to contamination
T	Cu	Cu 45% Cr	−270 to 400	1.5 °C or 0.4%	IEC 584	High accuracy, general purpose, suits low temperatures
U	Cu	Cu 45% Cr	−270 to 600	1.5 °C or 0.4%	DIN 43710	DIN precursor to Type T
–	Au	Pt	0 to 1000	–	ASTM E988	Most accurate for this range
–	Au	Pd	0 to 1500	–	ASTM E1751	Most accurate above 1000 °C

Notes: 1. Letter designation for Types C, D, and G are not formal, but are recognized by many manufacturers.
2. Tolerance values are given indicative of manufacturing tolerances. Actual values vary according to grade of wire and temperature range.

harmonized and the older DIN Type J is now known as Type L. The Type T and Type U thermocouples are similarly related.

There are a number of useful specialized thermocouples without a formal letter designation. The American Society for Testing and Materials, in the ASTM E988 and ASTM E1751 standards define additional thermocouple types including three tungsten–rhenium thermocouples, which have an informal letter designation recognized by many manufacturers (Types C, D, and G), and are useful at high temperatures and in reducing atmospheres. ASTM E1751 includes the definitions for the platinum–gold and the platinum–palladium thermocouples, both of which use pure metal wires to avoid the inhomogeneity problem afflicting other thermocouples, and are accurate to a few millikelvin for temperatures below 960 °C.

Good advice on the use of the different thermocouples can be found in the ASTM manual (ASTM 1993) and in Bentley (1998).

REFERENCES

American Society for Testing and Materials (1993) *Manual on the Use of Thermocouples in Temperature Measurement*, 4th edn, Race Street, Philadelphia.

Bedford, R.E., Bonnier, G., Mass, H. and Pavese, F. (1996) Recommended Values of Temperature on the International Temperature Scale of 1990 for a Selected Set of Secondary Fixed Points. *Metrologia*, **33**, 133–154.

Bentley, R.E. (ed.) (1998) *Handbook of Temperature Measurement*, Springer, Singapore.

Bureau International des Poids et Mesures (1990a) *Supplementary Information for the International Temperature Scale of 1990*.

Bureau International des Poids et Mesures (1990b) *Techniques for Approximating the International Temperature Scale of 1990*.

Childs, P.R.N. (2001) *Practical Temperature Measurement*, Butterworth-Heinemann.

DeWitt, D.P. and Nutter, G.N. (eds.) (1988) *Theory and Practice of Radiation Thermometry*, John Wiley & Sons, New York.

McGee, T.D. (1988) *Principles and Methods of Temperature Measurement*, John Wiley & Sons, New York.

Michalski, L., Eckersdorf, K., Kucharski, J. and McGhee, J. (2001) *Temperature Measurement*, 2nd edn, John Wiley & Sons.

Nicholas, J.V. and White, D.R. (2001) *Traceable Temperatures: An Introduction to Temperature Measurement and Calibration*, John Wiley & Sons, Chichester.

Preston-Thomas, H. (1990) The International Temperature Scale of 1990 (ITS-90). *Metrologia*, **27**, 3–10. (Copy also in BIPM 1990a, Sèvres, Paris).

Quinn, T.J. (1990) *Temperature*, Academic press, London.

Rusby, R.L., Durieux, M., Reesink, A.L., Hudson, R.P., Schuster, G., Kühne, M., Fogle, W.E., Soulen, R.J. and Adams, E.D. (2001) The Provisional Low Temperature Scale from 0.9 mK to 1 K, PLTS-2000, in *Proceedings of TEMPMEKO 2001* (eds B. Fellmuth, J. Seidel and G. Scholz), VDE Verlag, Berlin (pp. 365–370).

Schooley, J.F. (1986) *Thermometry*, CRC Press, Boca Raton, FL.

White, G. and Meeson, P. (2002) *Experimental Techniques in Low-temperature Physics*, 4th edn, Oxford University press, Oxford.

199: Characteristics of Time and Frequency Measurement

Michael A. Lombardi
National Institute of Standards and Technology, Boulder, CO, USA

1 Definition of Time, Time Interval, and Frequency 1341
2 Coordinated Universal Time (UTC) 1342
3 Measurement Terminology 1342
4 Measurement Techniques 1342
5 Oscillators and Clocks 1344
Acknowledgment 1345
Related Articles 1345
Further Reading 1345

1 DEFINITION OF TIME, TIME INTERVAL, AND FREQUENCY

Time and frequency metrologists measure three basic quantities: time of day, time interval, and frequency. Time of day is usually provided in units of years, months, days, hours, minutes, and seconds. Devices that display or record time of day are called *clocks*. Synchronization is the process of setting multiple clocks to the same time.

Time interval is the duration or elapsed time between two events. The standard unit of time interval is the *second* (s). However, many applications require measuring shorter time intervals, such as *milliseconds* ($1\,\text{ms} = 10^{-3}\,\text{s}$), *microseconds* ($1\,\mu\text{s} = 10^{-6}\,\text{s}$), *nanoseconds* ($1\,\text{ns} = 10^{-9}\,\text{s}$), and *picoseconds* ($1\,\text{ps} = 10^{-12}\,\text{s}$). The second is one of seven base units in the International System of Units (SI). The second was once defined on the basis of the earth's rotational rate or as a fraction of the tropical year. That changed in 1967 when the era of atomic time keeping formally began. The second is now defined as follows:

> The duration of 9 192 631 770 periods of the radiation corresponding to the transition between two hyperfine levels of the ground state of the cesium-133 atom.

Frequency is the rate of occurrence of a repetitive event. If T is the period of a repetitive event, then the frequency f is the period's reciprocal, $1/T$. Conversely, the period is the reciprocal of the frequency, $T = 1/f$. Since the period is a time interval expressed in seconds (s), the close relationship between time interval and frequency is easy to see. The standard unit for frequency is the *hertz* (Hz), defined as events or cycles per second. The frequency of electrical signals is often expressed in multiples of hertz, including *kilohertz* ($1\,\text{kHz} = 10^3\,\text{Hz}$), *megahertz* ($1\,\text{MHz} = 10^6\,\text{Hz}$), or *gigahertz* ($1\,\text{GHz} = 10^9\,\text{Hz}$). A device that produces frequency is called an *oscillator*. *Syntonization* is the process of setting multiple oscillators to the same frequency.

The three time and frequency quantities are closely related. As mentioned, the standard unit of time interval is the second. Seconds are counted to establish date and time of day. Events or cycles per second are counted to measure frequency and time interval. Time interval and frequency are measured with far less uncertainty and more resolution than any other physical quantities. Today (2002), the best time and frequency standards realize the SI second with uncertainties near 1×10^{-15}.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
Published in 2005 by John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

2 COORDINATED UNIVERSAL TIME (UTC)

The world's major metrology laboratories routinely send their measurement data to the Bureau International des Poids et Measures (BIPM) in Sevres, France. The BIPM averages data collected from more than 200 atomic time and frequency standards located at more than 40 laboratories. The BIPM uses the data to generate Coordinated Universal Time (UTC), a timescale that realizes the SI second as closely as possible. The BIPM maintains UTC as a 'paper' timescale. National metrology laboratories use data published by the BIPM to steer their clocks and oscillators, and to generate and distribute real-time versions of UTC.

When necessary, leap seconds are added to UTC on either June 30 or December 31. Leap seconds keep atomic time (UTC) within ± 0.9 s of an older timescale called UT1, which is based on the Earth's rotation frequency. Leap seconds have been added to UTC at a rate of less than once per year since 1972.

UTC is the ultimate standard for time of day, time interval, and frequency. Clocks synchronized to UTC display the same hour, minute, and second all over the world (and remain within one second of UT1). Oscillators syntonized to UTC generate signals that serve as time interval and frequency standards.

3 MEASUREMENT TERMINOLOGY

The oscillator or clock being measured is called the *device under test* (DUT). A measurement compares the DUT to a standard or reference. The test signal for time measurements is usually a pulse that occurs once per second. Pulse width and polarity varies, but amplitude is often at TTL levels. The test signal for frequency measurements is usually a sine or a square wave that produces one cycle (360° or 2π radians of phase) in one period. The signal amplitude is expressed in volts (Figure 1).

Measurement results are stated in terms of accuracy and stability. Accuracy relates to the offset from an ideal value – *see* **Article 54, Explanation of Key Error and Uncertainty Concepts and Terms, Volume 1**. For example, time offset is the time interval between a measured on-time pulse and an ideal on-time pulse that coincides with the UTC second. Frequency offset is the difference between a measured frequency and an ideal frequency with zero uncertainty. This ideal frequency is called the *nominal frequency*.

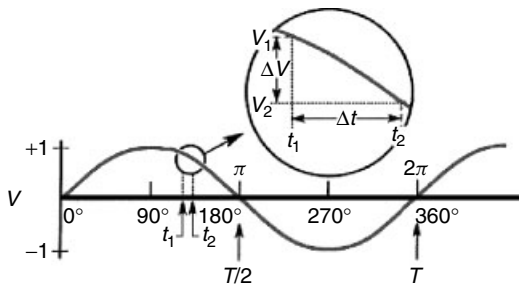


Figure 1. Oscillating sine wave.

Stability indicates how well a device produces the same time or frequency offset over a given time interval. It does not indicate whether the time or frequency is 'right' or 'wrong', but only whether it *stays the same*. Accuracy depends upon how well a device has been synchronized or syntonized, whereas stability is an inherent device characteristic.

Stability estimates are calculated from a set of either frequency offset or time offset measurements, and show time or frequency fluctuations with respect to a mean offset. The offset values are obtained by sampling or averaging over a stated interval, called tau (τ). Short-term stability refers to fluctuations over short intervals (often where $\tau = 1$ s, or at least where τ is less than 100 s). Long-term stability can refer to any τ value greater than 100 s, but usually refers to values longer than 1 day. An oscillator specification sheet usually lists stability estimates at τ values increasing in length until the point, called the *noise floor*, where the stability stops improving. The most common statistics used to estimate stability are the Allan deviation ($\sigma_y(\tau)$) for frequency stability, and the Time deviation ($\sigma_x(\tau)$) for time stability. Further information on the uncertainty of measurements is found in **Article 55, Uncertainty Determination, Volume 1**.

4 MEASUREMENT TECHNIQUES

If a single time interval or frequency measurement is made, the uncertainty is often limited by measurement system noise, or a lack of resolution. Averaging measurements usually leads to more resolution and lowers the measurement system's noise floor. To obtain a correct result, the measurement interval must be long enough to ensure that both the measurement system and reference have lower noise floors than the DUT.

Time interval is usually measured with a time interval counter (TIC). A TIC requires two signal inputs. One signal starts the counter and the other stops it, and the TIC measures the interval between the signals. The best TICs have a single shot resolution near 20 ps, which sometimes averages down to near 1 ps.

Simple frequency measurements are made with a frequency counter (Figure 2). The measurement reference

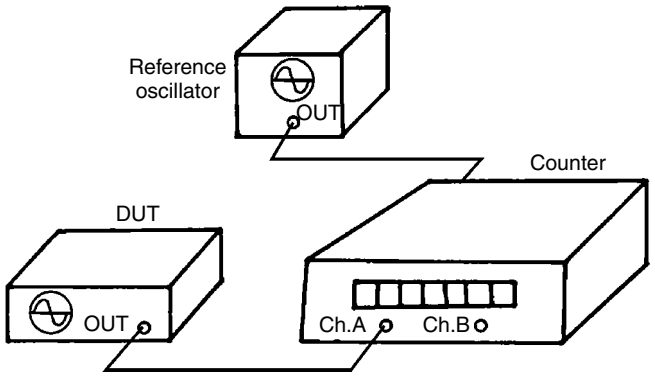


Figure 2. Frequency measurement with frequency counter.

is the counter's time base oscillator. Frequency offset (dimensionless) is determined as

$$f(\text{offset}) = \frac{f_{\text{measured}} - f_{\text{nominal}}}{f_{\text{nominal}}} \tag{1}$$

where f_{measured} is the reading displayed by the frequency counter, and f_{nominal} is the frequency labeled on the oscillator (the nominal frequency it is supposed to produce).

More advanced frequency measurements are phase comparisons between the DUT and reference signals. The concept of a phase comparison is best illustrated on an oscilloscope where the waveforms are visible (Figure 3), but higher resolution comparisons are made with a TIC, using the DUT to start the counter, and the reference to stop the counter. The time interval change is used to estimate the frequency offset:

$$f(\text{offset}) = \frac{-\Delta t}{T} \tag{2}$$

where Δt is the change in time interval, and T is the measurement period. This requires only two time interval readings, and Δt is simply the difference between the two readings. Typically, multiple readings are taken, least squares linear regression is applied to the data set, and Δt is obtained from the slope of the least squares line. If desired, the dimensionless frequency offset can be multiplied by the nominal frequency and converted to hertz.

Since standard test frequencies have short periods (100 ns in the case of 10 MHz), keeping track of cycles during a phase comparison is difficult. Therefore, phase comparisons normally test low frequencies with long periods, such as 1 or 10 Hz. The low frequency is obtained by dividing the test frequency, or by mixing it with another frequency to generate a beat frequency.

Divider systems (Figure 4) are more versatile, and can be built or programmed to work with nearly any test frequency. They can convert any test frequency to 1 Hz, for example, by simply removing $n - 1$ cycles of the test frequency. In this fashion, a 5-MHz signal can be divided and compared to a 1-Hz signal. Divider systems normally cannot detect frequency changes smaller than parts in 10^{11} in 1 s, since the single shot resolution and short-term stability of a TIC/Divider system is limited to tens of picoseconds.

Mixer systems (Figure 5) are less versatile and more complex. They are often designed for a single test frequency, they cannot measure 1-Hz timing signals or similar low frequencies, and they usually require the DUT and reference to have the same nominal frequency. The dual mixer time difference measurement system shown in the figure mixes the DUT and reference signals with a signal from a transfer oscillator offset from the nominal frequency. If the nominal frequency is 10 MHz, and the offset frequency is 10.000010 MHz, the result is 2 beat frequencies near 10 Hz that are compared using a TIC. The ratio of the nominal frequency to the beat frequency is the heterodyne factor ($10^7/10 = 10^6$ in this example). The system's resolution is the period of the TIC's time base oscillator divided by the heterodyne factor. For example, if the TIC

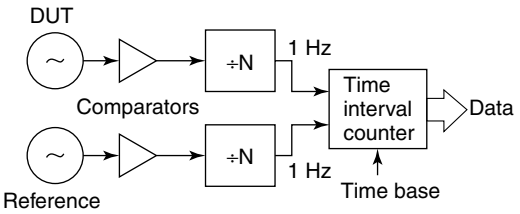


Figure 4. Time interval counter measurement system with frequency dividers.

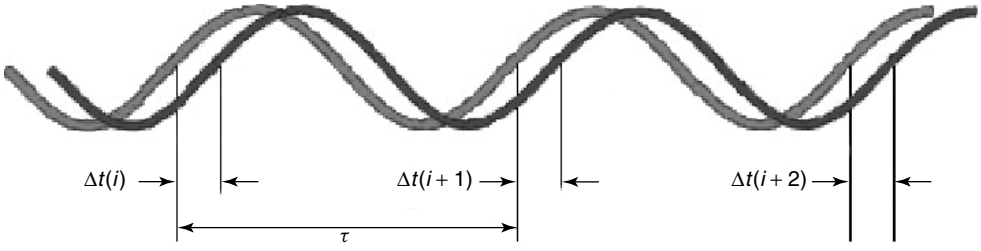


Figure 3. Phase comparison between two sine waves.

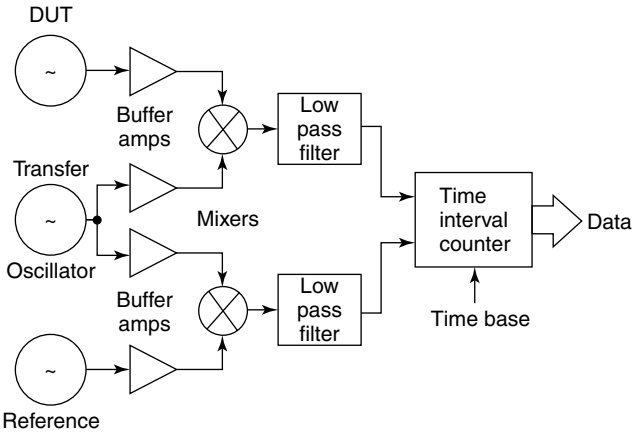


Figure 5. Dual mixer time difference measurement system.

has a 10-MHz time base, the resolution equals 100 ns/10^6 or 0.1 ps. This high resolution makes it possible to detect frequency changes of parts in 10^{13} in 1 s, and makes a dual mixer system well suited for measuring short-term stability. However, divider and mixer systems should produce the same answer when measuring long-term stability, since the noise floor of both systems eventually drops below the DUT noise floor.

5 OSCILLATORS AND CLOCKS

Oscillators consist of a resonator that produces a periodic event, and an energy source that sustains oscillation. The natural frequency of an oscillator is called the *resonance frequency*. A clock is a device that counts oscillator cycles and displays or records time. For example, a quartz wristwatch contains an oscillator that runs at 32 768 Hz. After counting 32 768 cycles, the watch circuitry records that one second has elapsed. Seconds are counted to establish longer time intervals such as minutes, hours, and days.

The Q of an oscillator is its resonance frequency divided by its resonance width, or the range of possible frequencies at which the oscillator can oscillate. High-Q resonators generally lead to more stable oscillators, since they will not resonate at all unless they are operated near their resonance frequency.

Quartz oscillators achieve the highest Q of any mechanical-type oscillators, but atomic oscillators have much higher Q factors. In theory, the atom is a perfect resonator. Atomic oscillators use the quantized energy levels in atoms and molecules as their resonance source. An electromagnetic field at a particular frequency boosts an atom from one energy level to a higher one. Conversely, an atom at a high energy level can drop to a lower level by emitting energy. The resonance frequency (f) of an atomic oscillator is the difference between the two energy levels divided by Planck's constant (h):

$$f = \frac{E_2 - E_1}{h} \tag{3}$$

Table 1 provides a summary of oscillators used as time and frequency standards.

5.1 Quartz oscillators

Quartz oscillators use a quartz crystal as a resonator. The crystal strains (expands or contracts) when a voltage is applied. When the voltage reverses, the strain reverses. This is known as the *piezoelectric effect*. The rate of expansion and contraction is the resonance frequency, and

Table 1. Summary of oscillator types.

Oscillator type	Quartz (TCXO)	Quartz (OCXO)	Rubidium	Commercial cesium beam	Hydrogen maser
Q	10^4–10^6	3.2×10^6 (5 MHz)	10^7	10^8	10^9
Resonance Frequency	Various	Various	6.834682608 GHz	9.192631770 GHz	1.420405752 GHz
Leading cause of failure	None	None	Rubidium lamp (life expectancy > 15 years)	Cesium beam tube (life expectancy of 3 to 25 years)	Hydrogen depletion (life expectancy > 7 years)
Stability, $\sigma_y(\tau)$, $\tau = 1$ s	1×10^{-8} to 1×10^{-9}	1×10^{-12}	5×10^{-11} to 5×10^{-12}	5×10^{-11} to 5×10^{-12}	1×10^{-12}
Noise floor, $\sigma_y(\tau)$	1×10^{-9} ($\tau = 1$ to 10^2 s)	1×10^{-12} ($\tau = 1$ to 10^2 s)	1×10^{-12} ($\tau = 10^3$–10^5 s)	1×10^{-14} ($\tau = 10^5$–10^7 s)	1×10^{-15} ($\tau = 10^3$–10^5 s)
Aging/year	5×10^{-7}	5×10^{-9}	1×10^{-10}	None	$\sim 1 \times 10^{-13}$
Frequency Offset after warm up	1×10^{-6}	1×10^{-8} to 1×10^{-10}	5×10^{-10} to 5×10^{-12}	5×10^{-12} to 1×10^{-14}	1×10^{-12} to 1×10^{-13}
Warm-up period	<10 s to 1×10^{-6}	<5 min to 1×10^{-8}	<5 min to 5×10^{-10}	30 min to 5×10^{-12}	24 h to 1×10^{-12}

is determined by the crystal's size and shape. Q factors for quartz oscillators range from 10^4 to 10^6.

Quartz oscillators are sensitive to changes in temperature, humidity, pressure, and vibration, but several designs reduce these environmental effects. The oven-controlled crystal oscillator (OCXO) encloses the crystal within a temperature-controlled chamber. The temperature-compensated crystal oscillator (TCXO) generates frequency changes equal and opposite to frequency changes produced by temperature. TCXOs are less expensive, but less stable than OCXOs.

Quartz oscillators have excellent short-term stability but poor long-term stability. An OCXO might be stable ($\sigma_y(\tau)$, at $\tau = 1$ s) to 1×10^{-12}. Long-term stability is limited by aging, or a frequency change with time caused by internal oscillator changes. The aging rate of a high quality OCXO might be parts in 10^9 per year, while a TCXO might age 100 times faster. Since quartz oscillators change frequency substantially over time, atomic oscillators are needed for applications requiring long-term stability.

5.2 Atomic oscillators

Rubidium oscillators are the lowest priced atomic oscillators. They operate at 6 834 682 608 Hz, the resonance frequency of the rubidium atom (^{87}Rb), and use the rubidium frequency to control the frequency of a quartz oscillator. The Q of a rubidium oscillator is about 10^7. Stability ($\sigma_y(\tau)$, at $\tau = 1$ s) is typically 1×10^{-11}, and about 1×10^{-12} at one day. Frequency offset ranges from 5×10^{-10} to 5×10^{-12} after a warm-up period of a few minutes or hours, so they meet the accuracy and stability requirements of most applications without adjustment.

Cesium oscillators physically realize the SI second, since the second is defined on the basis of the resonance frequency of the cesium atom (^{133}Cs), which is 9 192 631 770 Hz. A cesium oscillator should stay close to its nominal frequency without adjustment, and there should be no frequency change due to aging. The Q of a commercial cesium standard is a few parts in 10^8. Stability ($\sigma_y(\tau)$, at $\tau = 1$ s) is typically 5×10^{-12}, and reaches a noise floor near 1×10^{-14} in about one day, extending to weeks or months. The frequency offset is typically less than 1×10^{-12} after a 30-min warm up.

ACKNOWLEDGMENT

Contribution of United States Government, not subject to copyright.

RELATED ARTICLES

Article 54, Explanation of Key Error and Uncertainty Concepts and Terms, Volume 1; **Article 55, Uncertainty Determination, Volume 1**; **Article 200, Calibrations and Standards in Time Measurement, Volume 3**.

FURTHER READING

Agilent Technologies (2003a) *Fundamentals of Time Interval Measurement*, Agilent Application Note 200-3 (p. 68), URL: http://cp.literature.agilent.com/litweb/pdf/5965-7663E.pdf.

Agilent Technologies (2003b) *Fundamentals of the Electronic Counters*, Agilent Application Note 200 (p. 44), URL: http://cp.literature.agilent.com/litweb/pdf/5965-7660E.pdf.

Carr, J.J. (1996) *Elements of Electronic Instrumentation and Measurement*, Prentice Hall, NJ.

Kamas, G. and Lombardi, M.A. (1990) *Time and Frequency User's Manual*, National Institute of Standards, Special Publ. 559 (p. 160), URL: http://tf.nist.gov/general/pdf/461.pdf.

Sullivan, D.B., Allan, D.W., Howe, D.A. and Walls, F.L. (eds) (1990) *Characterization of Clocks and Oscillators*, National Institute of Standards, Technol. Tech Note 1337 (p. 352), URL: http://tf.nist.gov/general/pdf/868.pdf.

Webster, J.G. (1999) *The Measurement, Instrumentation, and Sensors Handbook*, CRC Press, Boca Raton, FL.

200: Calibrations and Standards in Time Measurement

Michael A. Lombardi
National Institute of Standards and Technology, Boulder, CO, USA

1 Time Standards 1346
2 Calibration of Short Time Interval Sensors 1346
3 Calibration of Medium and Long Time Interval Sensors 1348
4 Time Synchronization 1349
Acknowledgments 1349
Related Articles 1349
References 1349
Further Reading 1350

1 TIME STANDARDS

Time measurements include time interval measurements, which determine the elapsed time between two events; and time synchronization measurements, which determine the time offset between the test signal and the Coordinated Universal Time (UTC) second. Time standards typically produce 1 pulse per second (pps) signals. The period of these signals is a standard second that serves as a time interval reference. If the 1-pps signal is 'on time', or synchronized to UTC, it can also serve as a time synchronization reference. If the 1-pps signal is labeled, or time tagged, it can be used to synchronize time-of-day clocks. Some devices produce time tags at frequencies higher than 1 pps, with resolutions measured in milliseconds (ms), in microseconds (μs), or in nanoseconds (ns).

The second is defined as 'The duration of 9 192 631 770 periods of the radiation corresponding to the transition between two hyperfine levels of the ground state of the cesium-133 atom.' This definition was added to the International System (SI) of units in 1967, and therefore the 1-pps output from a cesium oscillator serves as an internationally recognized time interval standard. Commercially available cesium standards (Figure 1 provides a block diagram) routinely replicate the SI second with uncertainties near 1×10^{-13}. Even so, it is interesting to note that the 1-pps output from a cesium oscillator is not on time unless it has been synchronized. Prior to synchronization, the cesium timing pulse could be fast or slow by as much as 0.5 s with respect to UTC.

Cesium oscillators are expensive, and can cost $50 000 (USD) or more. They are usually owned only by facilities with stringent measurement uncertainty requirements. A more commonly owned time standard is a disciplined oscillator that receives radio broadcasts containing UTC information. The received information is often used to discipline a local quartz or rubidium oscillator, display the time-of-day, and to synchronize a 1-pps signal to agree with UTC. By applying continuous time and frequency corrections to the local oscillator signal, often through the use of a phase-locked loop, these radio-controlled devices can sometimes meet or exceed cesium oscillator performance over averaging periods of more than 1 day.

2 CALIBRATION OF SHORT TIME INTERVAL SENSORS

Measurements of time intervals less than 1 s are often made with oscilloscopes. Since oscilloscopes allow us to view the pulse waveforms, it is possible to select the

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
Published in 2005 by John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

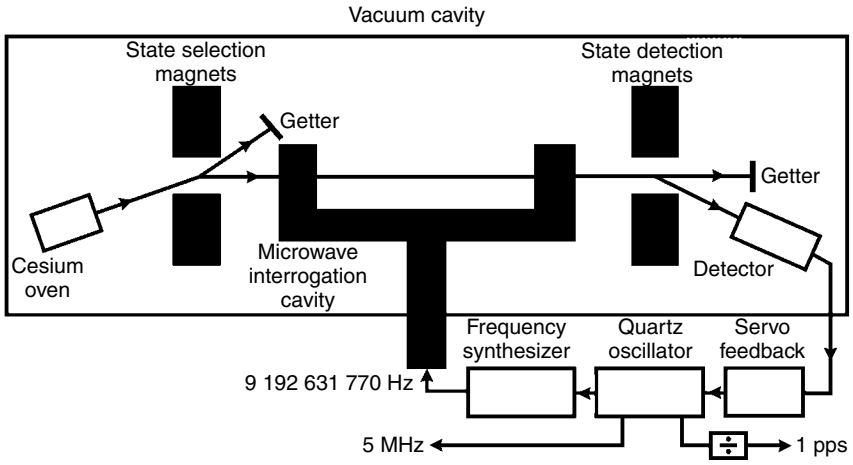


Figure 1. Block diagram of cesium oscillator.

precise start and end points that define a time interval, a major advantage when dealing with odd-shaped or noisy waveforms. Sometimes one pulse is displayed, while a different pulse is used to trigger the oscilloscope. Another method is to display both time pulses on a dual-trace oscilloscope. Triggering can be achieved from either of the two displayed pulses or from a third external pulse. The oscilloscope displayed in Figure 2 is triggering from channel A, and shows a 4-μs time interval between the A and B pulses.

The disadvantage of oscilloscopes for time interval measurements is lack of resolution and range. Oscilloscope range varies, and some models might be able to scale from 50 ps to 10 s per time division. However, the resolution

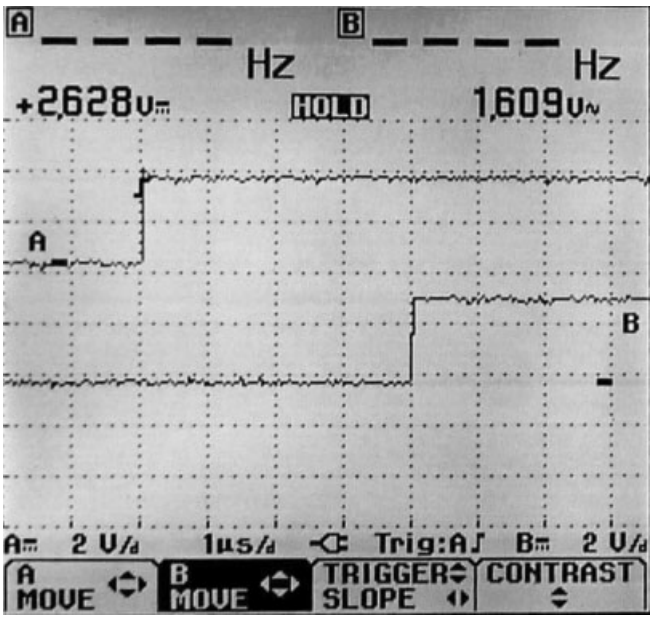


Figure 2. Measuring time interval with an oscilloscope.

of the time interval measurement is proportional to the length of the interval. If the time pulses are 10 ms apart, for instance, the best oscilloscopes can only resolve the interval to about ±1 μs. However, if the pulses are only 100 μs apart, the resolution is ±10 ns. In both cases, the relative uncertainty is 1×10^{-4}.

Time interval counters (TIC) are better suited for measuring all but the shortest time intervals. A start pulse gates the TIC (starts it counting) and a stop pulse stops the counting. While the gate is open (Figure 3), the counter counts zero crossings from time base oscillator cycle, and the resolution of the simplest TICs is equivalent to the period of a time base cycle (for example, 100 ns for a 10 MHz time base). However, many TICs use interpolation schemes to slice the cycle period into smaller parts. For example, slicing the 100-ns period into 5000 parts allows 20-ps resolution, a feat achieved by the best units. This translates to measurement uncertainties of a few parts in 10^{11} over a 1-s interval, often averaging down to parts in 10^{15} or 10^{16} after a few hours. Since these uncertainties are so low, TICs are routinely used for long-term comparisons between cesium oscillators.

The uncertainty of TIC measurements is, however, potentially limited by count ambiguity, trigger errors, or time base errors. A TIC generally has a +1 count or −1 count ambiguity in the least significant digit of its display. However, this ±1 count error usually averages out when multiple readings are taken, and is usually a problem only with single-shot measurements. Trigger errors are a more significant problem. They occur when the counter does not trigger at the expected voltage level on the input signal. Trigger errors can be caused by incorrectly set trigger levels or by input signals that are noisy (Figure 4), too large, or too small. Figure 4 illustrates that while triggering is desired on point a, noisy signals can cause triggering on points b or

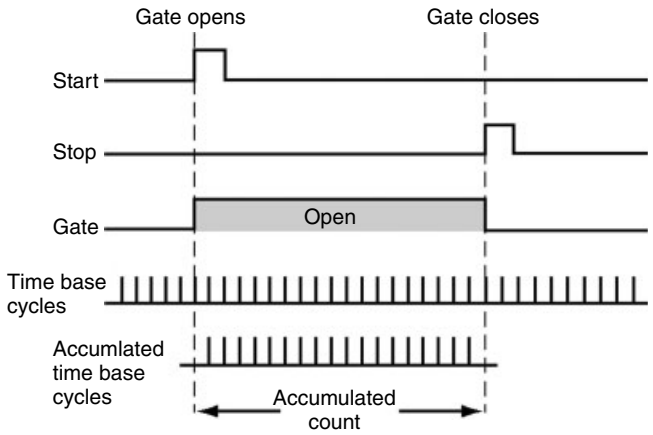


Figure 3. Gate and counting function of time interval counter.

c. This leads to incorrect measurements, so make sure the TIC is triggering at the proper point on both the start and stop waveforms (if necessary, view the waveforms on an oscilloscope). Time base errors are generally not a problem when measuring short time intervals. For example, consider a typical counter with 1-ns resolution and a time base stability of 1×10^{-8}. When a 1-ms interval is measured, the time base uncertainty is $10^{-3}\,\text{s}/10^{8} = 10^{-11}\,\text{s}$, or 10 ps. This is 100 times smaller than the 1-ns uncertainty contributed by the ± 1 count ambiguity. Even so, use the most stable oscillator available as an external time base, because time base errors are significant when measuring long time intervals.

Application notes are available for more detail, Agilent Technologies (2003a, b).

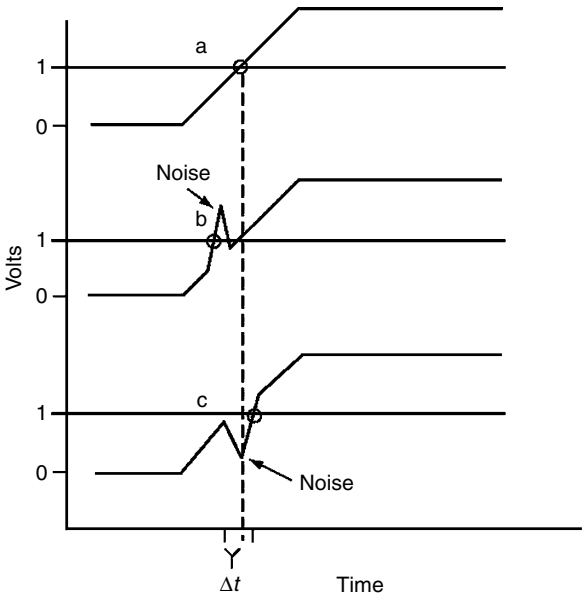


Figure 4. Potential effects of trigger errors on time interval measurements.

3 CALIBRATION OF MEDIUM AND LONG TIME INTERVAL SENSORS

A TIC is probably the best device for medium and long time interval measurements if it has the necessary range. For example, a TIC that counts 100 ns periods with a 32-bit counter will overflow after $2^{32} \times 100$ ns, or about 429 s. However, many TICs keep track of counter overflows, and some have the ability to measure intervals of about 1 day in time interval mode. If the TIC range is insufficient, most universal counters have a totalize mode that simply counts cycles from an external frequency source. Some commercial counters can count up to 10^{15} cycles. The period of the external gate frequency determines the resolution of this type of measurement. For example, if 1-μs resolution is desired, use a 1-MHz gate frequency. In this case, the time interval range would be $10^{15}/10^{6} = 10^{9}$ s, or nearly 32 years! In either time interval or totalize mode, a properly calibrated and triggered counter will probably exceed any measurement uncertainty requirements.

Low cost, simple timers are commonly used for coarse measurements of time intervals ranging from several seconds to many hours. These timers are typically found in manufacturing, industrial, medical, or laboratory facilities; or are sometimes used in commercial applications where consumers pay for services such as a ride in a taxicab or the use of a parking space. These timers are usually calibrated outside the laboratory, with field standard stopwatches. The stopwatches themselves require calibration, and this is often done by manually triggering the stopwatch while listening to an audio time signal. The legally required uncertainty for stopwatch calibrations is often 1 or 2 parts in 10^{4}, a large uncertainty limited mostly by human reaction time.

Stopwatch calibration methods exist that reduce or eliminate the problem of human reaction time. One method is to use a universal counter in totalize mode. Gate the counter with a signal from a calibrated signal generator. The gate frequency should have a period at least one order of magnitude smaller than the resolution of the stopwatch. For example, if the stopwatch has 0.01-s resolution (10 ms), use a 1-kHz frequency (1-ms period). This provides the counter with one more digit of resolution than that of the stopwatch. Start and stop the totalize counter by rapidly pressing the start–stop button of the stopwatch against the start–stop button on the counter. Then, compare the two readings and use the equation $\Delta t/T$ to get the results, where Δt is the difference between the counter and stopwatch displays and T is the length of the measurement run. For example, if $\Delta t = 100$ ms and $T = 1$ h, the time uncertainty is roughly 2.8×10^{-5}.

Table 1. Coordinated universal time (UTC) broadcast signals.

Station name or signal source	Nation responsible for signal	Coverage	Carrier frequency
JJY	Japan	Asian Pacific Rim	40 kHz 60 kHz
WWVB	United States	North America	60 kHz
MSF	United Kingdom	Europe	60 kHz
BPC	China	China	68.5 kHz
HBG	Switzerland	Europe	75 kHz
DCF77	Germany	Europe	77.5 kHz
WWV	United States	Worldwide[a]	2.5, 5, 10, 15, 20 MHz
WWVH	United States	Worldwide[a]	2.5, 5, 10, 15 MHz
CHU	Canada	Worldwide[a]	3.33, 7.335, 14.670 MHz
Indian National Satellite System (INSAT)	India	Indian subcontinent	2599.675 MHz
Global Navigation Satellite System (GLONASS)	Russia	Worldwide	1602.5625–1615 MHz (L1), 1240–1260 MHz (L2)
Global Positioning System (GPS)	United States	Worldwide	1575.41 MHz (L1) 1227.42 MHz (L2)

[a]Under optimal conditions.

A second method involves measuring the frequency offset of the stopwatch time base, and then converting the frequency offset to a time offset. Commercially available systems perform this measurement for both mechanical and electronic stopwatches (with 32.768 kHz or 4.19 MHz time bases). If an appropriate sensor is available (for example, an ultrasonic acoustic sensor for 32.768-kHz devices), this measurement can also be performed with a frequency counter. For example, if a 32.768-kHz time base has a 0.1-Hz frequency offset, the relative frequency offset ($0.1/32768 = 3 \times 10^{-5}$) translates to a time offset of about 108 ms in 1 h, or near the value for Δt in the previous example.

4 TIME SYNCHRONIZATION

Using the time interval measurement methods described above, a timing device can be synchronized by comparing it to a UTC reference, and by adjusting the time offset until it is as near zero as possible. Radio-controlled clocks are convenient UTC references. Models exist to receive signals scattered throughout the radio spectrum (Table 1), and they are commonly used for the synchronization of timing pulses and time-of-day clocks. The most common units receive low-frequency radio signals from stations such as WWVB (60 kHz) in the United States, MSF (60 kHz) in the United Kingdom, DCF77 (77.5 kHz) in Germany, or JJY (40 and 60 kHz) in Japan. The power levels from these broadcasts are large enough to cover their respective countries and a few neighboring countries, but no single broadcast is available internationally. The on-time pulse is delayed by the path between the transmitter and receiver, and if uncorrected for path delay, the synchronization uncertainty is often measured in milliseconds.

Better performance and true worldwide coverage can be obtained using a Global Positioning System (GPS) satellite receiver. Most GPS receivers receive the L1 carrier frequency (1575.42 MHz) broadcast by an orbiting constellation of at least 24 satellites. The long-term performance of a GPS receiver often meets or exceeds that of a cesium oscillator (2×10^{-13} uncertainty is typical at 1 day), and the 1-pps timing pulse of a calibrated receiver is usually within 100 ns of UTC. A properly operated GPS timing receiver is an excellent choice as a laboratory standard for time interval and time synchronization measurements.

ACKNOWLEDGMENTS

Contribution of United States government, not subject to copyright.

RELATED ARTICLES

Article 54, Explanation of Key Error and Uncertainty Concepts and Terms, Volume 1; **Article 55, Uncertainty Determination, Volume 1**; **Article 199, Characteristics of Time and Frequency Measurement, Volume 3**.

REFERENCES

Agilent Technologies (2003a) *Fundamentals of Time Interval Measurement*, Agilent Application Note 200-3 (p. 68), URL: http://cp.literature.agilent.com/litweb/pdf/5965-7663E.pdf.

Agilent Technologies (2003b) *Fundamentals of the Electronic Counters*, Agilent Application Note 200 (p. 44), URL: http://cp.literature.agilent.com/litweb/pdf/5965-7660E.pdf.

FURTHER READING

Carr, J.J. (1996) *Elements of Electronic Instrumentation and Measurement*, Prentice Hall, NJ.

Gust, J.C., Graham R.M. and Lombardi, M.A. (2004) *Stopwatch and Timer Calibrations*, National Institute of Standards and Technology, Special Publ. 960-12 (p. 60), URL: http://tf.nist.gov/general/pdf/1930.pdf.

Kamas, G. and Lombardi, M.A. (1990) *Time and Frequency User's Manual*, National Institute of Standards and Technology. Special Publ. 559 (p. 160), United States Government Printing Office, URL: http://tf.nist.gov/general/pdf/461.pdf.

Lombardi, M.A. (2002) *NIST Time and Frequency Services*, National Institute of Standards and Technology, Special Publ. 432 (p. 80), United States Government Printing Office, URL: http://tf.nist.gov/general/pdf/1383.pdf.

Lombardi, M.A., Nelson, L.M., Novick, A.N. and Zhang, V.S. (2001) Time and Frequency Measurements Using the Global Positioning System. *Cal Lab: The International Journal of Metrology*, **8**, 26–33, URL: http://tf.nist.gov/general/pdf/1424.pdf.

Zhang, V.S., Davis, D.D. and Lombardi, M.A. (1994) High Resolution Time Interval Counter, in *Proceedings of 26th Annual Precise Time and Time Interval (PTTI) Meeting* (pp. 191–200), (Reston, Virginia) URL: http://tf.nist.gov/general/pdf/870.pdf.

201: Voltage Measurement

Halit Eren

Curtin University of Technology, Perth, Western Australia, Australia

1 Introduction 1353
2 Basic Theory 1354
3 Measuring Networks 1354
4 Voltage Transformers 1355
5 Potentiometers 1356
6 Converters 1357
7 Rectifiers 1358
8 Indicators 1359
9 Digital Voltmeters 1359
10 Voltage and Voltage Measurement Standards 1360
Further Reading 1361

1 INTRODUCTION

The output signal form to be measured in the majority of sensors will require the measurement of either a voltage (this article), a current (**Article 202, Current Measurement, Volume 3**), or a resistance/impedance (**Article 203, Resistance Measurement, Volume 3**). Thus, their explanation is fundamental to most measurements.

So far as voltage measuring devices are concerned, voltage measurement can be classified as

1. low-voltage measurements, such as those generated by sensors;
2. medium-voltage measurements, such as those that exist in the power mains, laboratories, and industrial operations;
3. high-voltage measurements, such as a rise in power generators, transmission lines, and with plasma effects.

Consequently, there is a wide range of voltage measuring techniques and devices in use. This article concentrates on medium-voltage and high-voltage measurements.

Low-voltage measurements for sensors require sophisticated signal processing schemes – *see* **Article 121, Signals in the Presence of Noise, Volume 2**; **Article 176, Signals and Signal-to-noise Ratio, Volume 3**; **Article 178, Noise Matching and Preamplifier Selection, Volume 3**; **Article 179, Input Connections; Grounding and Shielding, Volume 3**; and **Article 181, Amplitude Modulated Signals: The Lock-in Amplifier, Volume 3**.

Common to all techniques are three major aspects that characterize the measurements:

1. *Amplitude*: If the voltage is smaller than a few millivolts, we may need to use suitable electronic components to amplify the signals. If the voltage is large, in the kilovolts and megavolts region, we may need to attenuate the magnitude in order to bring it to manageable levels.
2. *Frequency*: The frequency of a voltage signal plays an important role in configuring the appropriate components of a voltage measuring device. The frequency of interest can range from DC to a few gigahertz. If digital techniques of measuring are used, sampling of the voltage signals must conform to the Nyquist sampling criteria. The signal frequency waveform also needs careful consideration lest errors be generated.
3. *Duration*: The duration of a signal is significant in determining the appropriate technique to use for the measurement. Duration may vary from continuous signals; as in the case of power supplies, to impulses appearing for a few microseconds; as in the cases of surges, and corona effects in power transmission and distribution.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.

© 2005 John Wiley & Sons, Ltd.

ISBN: 0-470-02143-8.

2 BASIC THEORY

Voltage measurement is essential in electrical engineering as well as in many other disciplines of engineering and science. Voltage measurement involves determination of the electric potential difference between two points. The potential difference is the amount of work needed to move a unit charge located in an electric field from a reference point to another point. Hence, the potential difference is always relative to some reference point such as the Earth.

The potential difference is taken as the work per unit charge and the volt is related to the unit of work (joule – J) and the unit of charge (coulomb – C) by

$$1\text{volt} = \frac{1\,\text{joule}}{\text{coulomb}} \tag{1}$$

Although the concept of electric potential is useful in understanding electrical phenomena, it is worth noting that only the differences in potential energy are measurable. Therefore, it is commonly understood that the term *voltage* refers to potential differences. The measurement unit for the voltage is, in the International System of Units (SI), the volt (symbol: V).

Measurement of voltage is of utmost importance and extensively used in the electrical and electronic engineering, especially in the power industry. Moreover, when electronic devices for signal processing are involved, such as those used in telecommunication systems, control systems, and informatics, the majority of signals are in the voltage and current form. Therefore, voltage measurement constitutes an important area in industrial and scientific measurements, and in a diverse range of sensors of chemical, biological, and physical variables.

3 MEASURING NETWORKS

Since most common signals are in voltage form, there are many different techniques for processing the signals generated by a particular variable. However, some instruments called *voltmeters* are deliberately designed to measure voltages. There are five common types of voltmeters; these are

1. electromechanical instruments,
2. thermal type instruments,
3. electronic instruments,
4. Cathode Ray Oscilloscopes (CRO) or vacuum tube instruments (VTI),
5. virtual instruments.

1. *Electromechanical instruments*: These instruments are based on the mechanical interaction between various currents, between currents and magnetic fields, or between electrified conductors. Such interactions generate mechanical torque proportional to the voltage or to the square of the voltage under investigation. The generated torque is then balanced by a restraining torque, usually obtained by the use of suitably arranged mechanical springs. The balancing action causes the instrument pointer to be displaced by an angle proportional to the driving torque, and hence indicates the voltage to be measured. The value of the input voltage is therefore given by the reading of the pointer displacement on a graduated scale.
2. *Thermal type instruments*: These instruments are based on the thermal effects of a current flowing in a conductor. The reading is proportional to the square of the input voltage. These instruments are not used as widely as the others but are suitable for high frequency voltage measurement.
3. *Electronic instruments*: These instruments are based on purely electronic circuits, and attain the required measurement by processing the input signal by means of electronic semiconductor devices. The method employed to process the input signal can be either analog or digital. In the first case, analog electronic instruments are obtained, while in the second case digital electronic instruments are obtained.
4. *Oscilloscopes or vacuum tube instruments:* These instruments are basically voltmeters, and their main characteristic is to allow a graphic representation, on Cathode Ray Tubes (CRTs), or liquid crystal displays (LCDs). The time and amplitude characteristics of the signal can be determined directly from the displayed signal.
5. *Virtual instruments:* These are computer-based systems that can be programmed to function like voltmeters, ammeters, oscilloscopes, spectrum analyzers, and so on. They are equipped with Analog-Digital-Analog, converters – *see* **Article 139, Analog-to-Digital (A/D) Converters, Volume 3** and **Article 141, Digital-to-Analog (D/A) Converters, Volume 3** and are supported with software programs such as the LabView from National Instruments – *see* **Article 105, Modeling with LabVIEW™, Volume 2** and **Article 106, Virtual Instrumentation in Physics, Volume 2**.

These five basic types of voltmeters can be used for low-voltage, medium-voltage, and high-voltage measurements. In some cases, suitable arrangements such as amplification, attenuation, and conversions may be necessary to prepare the signals for measurement.

In addition, the voltage, current, resistance measurements can be made by a single instrument termed as *multimeters*

or *volt-ohm-multimeters*(VOMs). The VOM is a combination of the necessary circuits of a DC ammeter, DC voltmeter, AC ammeter, AC voltmeter, and a multirange ohmmeter. The typical DC voltage range of a common VOM is 0 to 1000 V, although with external resistor networks and voltage transformers, the range can be increased further, to say 5000 V. The DC current range is usually 10 A but with the use of external resistive shunt networks, or current transformers the range can be extended to much higher levels.

Particularly in power generation and transmission networks, precise, and adequate system measurements at the control centers are critical for high quality of operational decision making. The measurements taking place in remote locations are made available to the controllers by networks such as the Real Time Supervisory Control and Data Acquisition systems SCADA. The digital devices that collect the voltage measurements from the substation instruments and transducers are called *remote terminal units* (RTUs). The RTUs communicate with the SCADA over dedicated telephone lines, sometimes the transmission line itself and/or microwave channels.

The collection of analog measurements and the status of the circuit breakers from remotely monitored and controlled substations are telemetered by means of the cyclic scans of SCADA system. Typically, each scan lasts from 1 s to 10 s. This information has to be sufficient in number and evenly distributed across the network so that the observability of the system can be ensured.

4 VOLTAGE TRANSFORMERS

Transformers are extensively used in high-voltage measurements primarily to step the voltage down to a lower level that can be less expensively measured.

Transformers are devices that change the voltage level in the process of energy transfer from one AC system to another. The transformer has two coils, primary and secondary, both of which are wound on an iron core as shown in Figure 1.

The magnetic flux generated by the primary current links the secondary windings to generate a secondary voltage. Assuming 100% magnetic coupling, the ratio of the primary voltage (V_1) and secondary voltage (V_2) can be expressed by the turns ratio N_1 and N_2 as

$$\frac{V_1}{V_2} = \frac{N_1}{N_2} \tag{2}$$

This means the secondary voltage can be stepped up or down depending on the turns ratio. For example, if N_1 is 10 times greater than N_2, the secondary voltage is 10 times smaller than primary voltage.

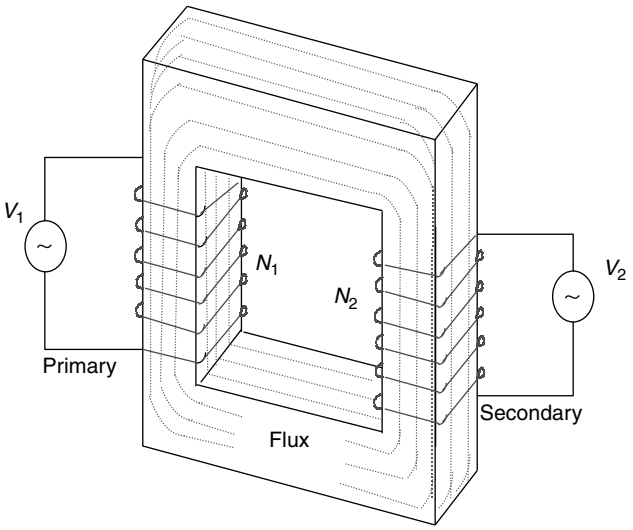


Figure 1. Transformer construction.

Voltage measurement instruments are used throughout the electric power system for monitoring and control purposes. Instrumentation components required to accomplish the measurements are the transducers, signal conditioners, and the analysis and monitoring equipment. Voltage transducers convert the system voltage to an acceptable level needed as the input to the signal conditioning equipment. The transducers are required to produce a scaled down replica of the input voltage to the accuracy expected for a particular measurement. Commonly used transducers are electromagnetic voltage transformer (VT), capacitive voltage transformer (CVT), and the cascade voltage transformer.

Voltage measurement transformers need to be specially designed to ensure the ratio is constant. Voltage transformers are much like small power transformers but they need to operate with secondary winding operating close to open circuit. The winding voltage drops are made small, and the rated flux density in the core is designed to be well below saturation density. These conditions maintain the calibrated voltage ratio.

Measuring very high frequency components in the voltage requires either a capacitive divider, or pure resistive divider. Special-purpose capacitor dividers can be obtained for accurate characterization of transients and harmonics up to at least 1 MHz. A disadvantage of such dividers is that they do not provide electrical isolation between the high voltage and the measuring systems and so impose special conditions for their use.

The cost of the electromagnetic VT tends to increase at an exponential rate with the rated voltage increase. An

alternative, more economic solution has been found in the CVT. This combines a capacitive potential divider with an auxiliary electromagnetic voltage transformer. This combination enables the insulation requirements of the electromagnetic unit to be reduced with associated savings in cost.

The CVT was developed to reduce the high cost of conventional VT by compromising on the frequency and transient response. Another possible solution to the problem is to use the cascade VT that is formed by connecting two or more electromagnetic transformer units in cascade. The primary windings of these transformers are connected in series and in this way the primary voltage is broken down in several distinct and separate stages. This solves the insulation problem and reduces costs. The secondary winding consists of a single winding on the last stage only.

Optical voltage sensors are also under development as replacements for VTs. They lower the costs compared to conventional CTs and VTs when mass produced. However, questions remain about their stability, sensitivity, and linearity. The sensitivity of the optical sensors to vibration and temperature changes must be minimized by either design or signal processing. The long-term stability of these sensors is under study in field trials, but techniques for calibrating them in the field and in the test laboratory have yet to be developed.

An important branch of voltage measurement is the determination of harmonics of such parameters as the fundamental power-frequency voltage, voltage dips, spikes, surges and sags, and other transient behaviors.

Specific reasons for taking harmonic measurements include confirming the presence of harmonics, evaluating the severity of the problem relative to acceptable limits, establishing compliance with standards and guidelines, harmonic filter design, providing input data for harmonic software analysis program, and designing an analytical model of the problem.

Instruments and systems for power quality measurements can help identify the sources of power quality degradation and protect customer equipment. Many power quality monitoring instruments are designed for input voltages up to 600 V rms and current inputs up to 5 A rms. In measurements, appropriate voltage transducers must be selected to provide these signal levels.

5 POTENTIOMETERS

There are many applications in which voltage measurements must be performed without drawing any current from the circuit to which the measurement device is connected. A typical case is the measurement of the electromotive force (emf), or no-load voltage, of a source with high internal resistance.

An important device that is capable of measuring voltages without drawing currents is the potentiometer. This device is not actually an instrument, but a set of electrical networks that are suitably arranged for the purpose. This arrangement also includes a galvanometer or indicator. The main purpose of the potentiometer is to perform an accurate comparison of an unknown emf against a standard one. There are many different types of potentiometers that can suit the requirements of a particular measurement, such as the slide-wire DC, slide-wire AC, Gall–Tinsley AC, and Drysdale–Tinsley AC Potentiometers. Owing to limited space, as an example, only the slide-wire DC potentiometer will be explained here.

1. *The slide-wire DC potentiometer*: The principle of operation of a slide-wire potentiometer can best be described by the slide-wire variety illustrated in Figure 2. The description of the various terms employed in this figure is as follows: E is the supply battery emf; R is the regulating resistance; i is the slide-wire current at balance conditions; r is the resistance per unit length of slide-wire; E_1 is the battery emf to be measured; E_2 is the emf of standard battery; l_1 is the length of slide-wire at balance conditions with emf E_1; l_2 is the length of slide-wire at balance conditions with standard emf E_2.

A battery of unknown emf E_1 is inserted in the circuit at a point shown in Figure 2. The galvanometer switch S is closed and the contact B moved along the slide-wire until the balance condition (zero reading on the galvanometer) is achieved. In this condition, no current flows through the galvanometer circuit. The current i supplied by battery E flows through the slide-wire. By Ohm's Law:

$$E_1 = irl_1 \tag{3}$$

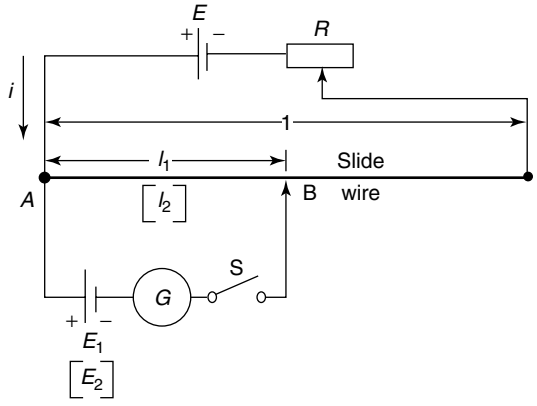


Figure 2. Slide-wire potentiometer.

Battery E_1 is replaced by a standard battery of known emf E_2. At the new balance condition, the following relationship can be written:

$$E_2 = irl_2 \tag{4}$$

From (3 and 4), the unknown emf can be obtained as

$$E_1 = \frac{l_1}{l_2} E_2 \tag{5}$$

A variable resistance is usually connected in series with the galvanometer to protect it against high currents during initial adjustments. This resistance is reduced to zero as balance is reached.

The simplest form of potentiometer consists of a supply battery E, a regulating resistor R and a slide-wire of uniform cross-section made of manganin, an alloy with a negligible temperature coefficient of resistance. The supply battery, usually of large capacity, provides the current in the slide-wire. The regulating resistance limits the value of this current to a reasonable quantity.

The many other forms of potentiometers are based on this simple form. The theory involved is akin to that of electrical Wheatstone bridges – *see* **Article 126, Electrical Bridge Circuits – Basic Information, Volume 2**; and **Article 127, Unbalanced DC Bridges, Volume 2**.

6 CONVERTERS

In many applications, conversion of voltage to another electrical parameter is essential in order to process the signals. In some, the voltages are already converted to meet the process needs, and the measurement of voltages is necessary to assure sound operation. Voltage conversion can be divided into groups:

1. *Voltage-to-voltage conversion of signals*: These may be amplifiers, attenuators, analog-to-digital converters, digital-to-analog converters, and so on. Voltage-to-voltage conversion of signals can be achieved by passive resistor networks, transistor circuits, operational amplifiers, and dedicated integrated circuits (ICs).
2. *Voltage-to-frequency conversion of signals*: These are special electronic devices that convert voltages into frequencies for the convenience of processing.
3. *Voltage-to-voltage conversion by power circuits*: These converters are essential in many domestic, laboratory, and industrial applications, for example, the electrical power drive. Conversion can be achieved by using power electronic components such as Darlington Transistors, power field-effect-transistors (FETs), silicon-control-rectifiers (SCRs), thyristor, triacs, and so on.

Voltage converters are the major subject of many books (see Further Reading section provided). Voltage-to-voltage conversion by power circuits are given here as examples.

AC-to-AC converters are used in many power-engineering applications. The conversion can be realized by a diverse range of equipment based on different operational methods. Some of these methods are

- phase controlled AC voltage controllers,
- pulse width modulation (PWM) AC voltage controllers (AC choppers),
- cycloconverters,
- matrix converters, and so on.

Basically, AC voltage controllers allow adjustment of the rms value of the output voltage. They are based on antiparallel connected power switches such as thyristors, triacs, diodes, and so on. There are two basic types:

1. Single-phase devices generally used in domestic and light load industrial applications.
2. Polyphase devices that are used in motor control, heating, and other applications.

Figure 3 illustrates a single-phase thyristor-based AC voltage controller. In this arrangement, one SCR conducts at a time, therefore, input and output currents are equal. S_1 is fired at an angle α and S_2 is fired at $(\pi - \alpha)$. The minimum firing angle must be equal to or greater than the load angle. If α_e is the extinction angle, using half symmetry, the rms value of output voltage for a sinusoidal input can be shown to be

$$V_{O_{rms}} = \sqrt{\frac{1}{\pi} \int_{\alpha}^{\alpha_e} [V_m \sin(\omega t)]^2 \, d\omega t} \tag{6}$$

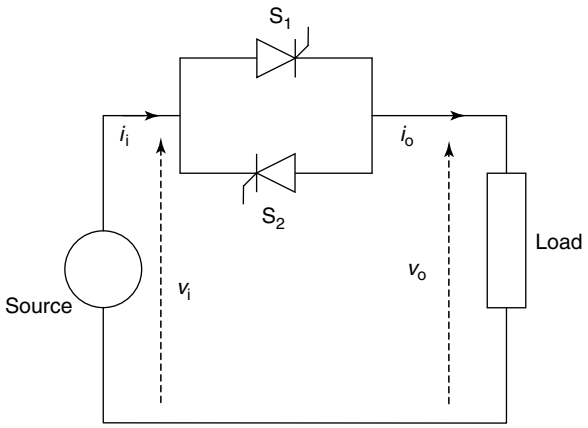


Figure 3. Single-phase voltage controller.

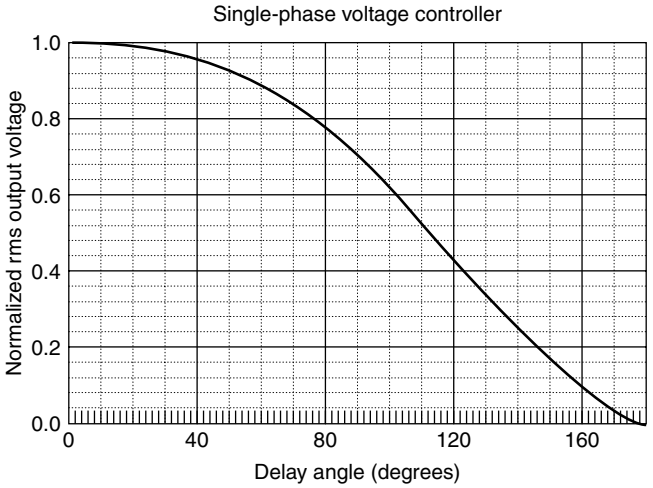


Figure 4. Normalized voltage versus delay (firing) angle.

With purely resistive load, from (6), the rms value output voltage can be determined to be

$$V_{o_{rms}} = V_m \sqrt{\frac{1}{2} - \frac{\alpha}{2\pi} + \frac{\sin 2\alpha}{4\pi}} \tag{7}$$

Equation (7) indicates that the output voltage is a function of the firing angle of the thyristors. The normalized rms load voltage versus firing angle for a single-phase controller is illustrated in Figure 4.

7 RECTIFIERS

For AC measurements, most instruments contain AC-to-DC conversion (or rectification), which are made from germanium or silicon diodes so that the voltages can be expressed in rms values. Depending on the signals, the rectification can be performed by transistor or operational amplifier circuits or by the SCRs.

AC/DC converters (rectifiers) operate on similar principles as the voltage controllers. These rectifiers can be configured in uncontrolled, semicontrolled, or fully controlled forms. Some of the configurations are

- single-phase half-wave rectifiers,
- single-phase full-wave rectifiers,
- polyphase half-wave rectifiers,
- polyphase full-wave rectifiers.

Rectifiers can be configured by using center-tapped transformer arrangements or by bridge forms. Since there are a variety of techniques available, here, as an example, we will give a single-phase rectifier using thyristors, as illustrated in Figure 5.

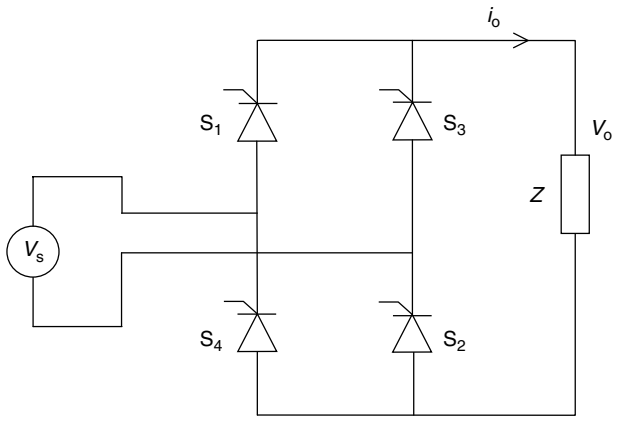


Figure 5. Single-phase fully controlled bridge rectifier.

The current flowing in the load impedance Z can be continuous or discontinuous. In the case of continuous currents, the mathematical analysis for currents and voltages uses Fourier series to show

$$V_o(\omega t) = V_o + \sum_{n=1}^{\infty} V_n \cos(n\omega t + \theta_n) \tag{8}$$

The DC average value is then

$$V_o = \frac{1}{\pi} \int_{\alpha}^{\alpha+\pi} V_m \sin(\omega t)\, \mathrm{d}(\omega t) = \frac{2V_m}{\pi} \cos\alpha \tag{9}$$

The amplitude of the AC terms can be calculated from

$$V_n = \sqrt{a_n^2 + b_n^2} \tag{10}$$

where

$$a_n = \frac{2V_m}{\pi}\left[\frac{\cos(n+1)\alpha}{n+1} - \frac{\cos(n-1)\alpha}{n-1}\right]$$

$$b_n = \frac{2V_m}{\pi}\left[\frac{\sin(n+1)\alpha}{n+1} - \frac{\sin(n-1)\alpha}{n-1}\right] \text{ for } n = 2, 4, 6\ldots$$

The Fourier series for the current is determined by superposition

$$I_{\text{rms}} = \sqrt{i_o^2 + \sum_{n=2,4,6\ldots}^{\infty} \left(\frac{I_n}{\sqrt{2}}\right)^2} \tag{11}$$

where

$$I_o = \frac{V_o}{R} \text{ and } I_n = \frac{V_n}{Z_n} = \frac{V_n}{|R + \mathrm{j}n\omega_0 L|}$$

The outputs of the controlled rectifiers can be improved by suitable arrangements of capacitive–inductive (L-C) filters.

8 INDICATORS

Indicators are electromechanical voltmeters that are based on the mechanical interaction between currents and magnetic fields, or between electrified conductors. Electromechanical instruments provide readings in analog form, by moving a pointer that indicates the measured value on a scale. The pointer can be real or virtual, usually a light beam reflected on a graduated ruler by a moving mirror. The scale is a linear scale if the pointer displacement is proportional to the measured quantity or a square-law scale if the pointer displacement is proportional to the square of the measured quantity. These instruments are classified according to their deflecting movements as

1. the moving-coil electromagnetic voltmeters (D'Arsonval Galvanometer);
2. the moving-iron electromagnetic voltmeters;
3. the electrodynamic voltmeters;
4. the electrostatic voltmeters;
5. electromechanical multimeters or VOMs.

Electromechanical instruments involve electromechanical forces. The controlling forces are either from the spring or from the force of gravity. They are aimed at balancing the deflecting forces so that the final pointer displacement is proportional to the measured quantity, or its squared value.

The spring control is achieved by the use of a spring arrangement, which is usually made from a phosphor–bronze spiral spring. The spring is attached to the moving system to provide a controlling torque that counterbalances the deflecting torque. It also provides the reading on the scale. An explanatory drawing is shown in Figure 6.

The spring has a large number of turns greatly reducing the deformation per unit length and giving an angle of deflection (θ) directly proportional to the deflecting torque. Occasionally, a long helical spring replaces the spiral spring. In some instruments, a single strip of phosphor–bronze provides the necessary spring control. Often two springs, wound in opposite directions are employed. This method is employed in order to reduce the reading errors.

Although the electromechanical instruments represent older technology in comparison to the electronic techniques, and the digital ones, in particular, they are still extensively used in many areas such as panel displays in industry, dashboard displays in motor vehicles, general-purpose low cost measurements and laboratories.

9 DIGITAL VOLTMETERS

The basic structure of a digital voltmeter (DVM) consists of the following three main stages:

1. analog signal processing,
2. analog-to-digital (A/D) converter,
3. digital signal processor (DSP).

The first stage of the device conditions the input signal, adapting it to the dynamics of the A/D. The A/D converter is responsible for sampling of the input signal and converting each sampled value into a digital form. The sequence of digital values obtained after the sampling and conversion operations is stored into the memory of the DSP and processed in order to attain the desired measured values. The detailed structure of a digital multimeter is given in Figure 7.

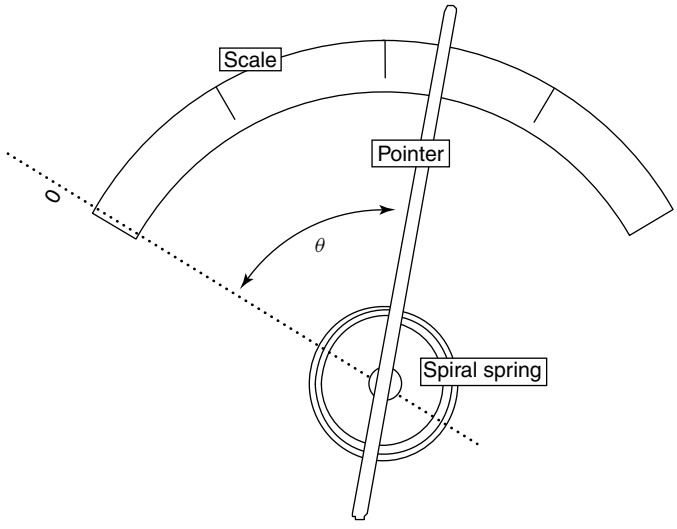


Figure 6. Schematic of a spring controlled voltage indicator.

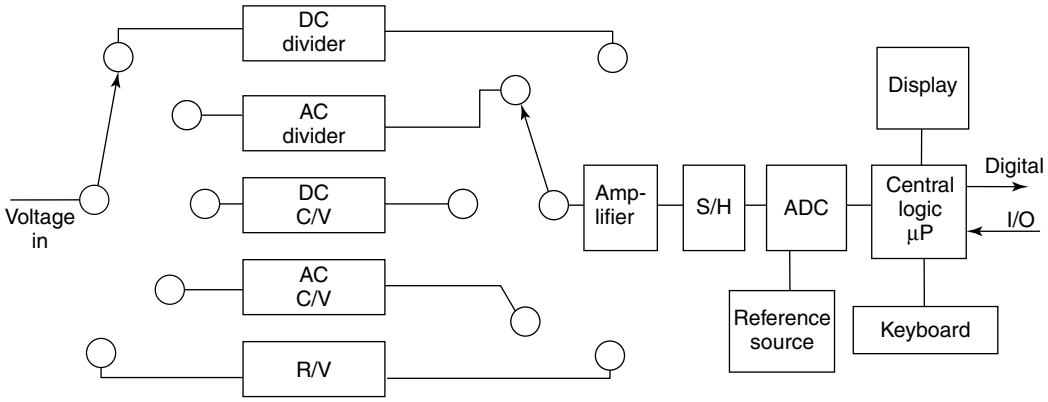


Figure 7. Block diagram of a typical digital multimeter.

DVMs may differ from each other in the following ways:

- number of measurement ranges,
- number of digits,
- accuracy,
- speed of readings,
- operating principles.

In a modern DVM, the basic measurement ranges to obtain full-scale values are as low as 0.1 μV. If an appropriate voltage divider is used, it is also possible to obtain full-scale values as high as 1000 V. The DVM, with a suitable input stage, can be used as an ammeter having very broad measurement ranges from 1 nA to 10 A.

When the measurement result is displayed on the instrument front panel, it is usually presented in decimal numbers, with a number of digits that typically range from 3 to 6. When the measurement result is sent to a DSP system, its representation takes the form of a binary-coded output signal. The number of bits of this representation typically ranges from 8 to 16, though 18-bit and 24-bit A/D converters are available.

The accuracy of a DVM is usually correlated to its resolution. This is quite obvious, since assigning an uncertainty lower than 0.1% of the range to a three-digit DVM makes no sense, since this is the displayed resolution of the instrument. Similarly, a poorer accuracy makes the three-digit resolution quite useless. Presently, a six-digit DVM can feature an uncertainty, for short periods of time in controlled environment, as low as the 0.0015% of reading or 0.0002% of the full range.

The speed of reading of a DVM can be as high as 1000 readings per second. When the A/D conversion is considered, the conversion rate is taken into account instead of the speed of reading. Presently, the conversion rate for 12-bit, successive approximation A/D can be as high as nominally 10 MHz. The conversion rate can be on the order of 1 GHz for lower resolutions when flash A/D converters are used – *see* **Article 139, Analog-to-Digital (A/D) Converters, Volume 3**.

10 VOLTAGE AND VOLTAGE MEASUREMENT STANDARDS

Voltage is an important property in electrical engineering and measurement technology. Hence, strict standards are developed in two major areas:

1. Voltage itself is governed by its own standards such as the Josephson junction standard – *see* **Article 43, Units, Volume 1**.
2. Voltage determination techniques are standardized to suit specific areas of application, such as the IEEE Std 1122 for digital recorders, IEC 60060 for high-voltage impulse calibrations.

The primary voltage standard is, today, the Josephson junction standard. Other voltage standards are nevertheless employed in the metrological laboratories as they are more convenient and, for example, are the time honored standard cells, although their accuracy is lower than that of the Josephson junction.

The Josephson junction standard describes the quantum standard of voltage. A Josephson junction consists of two super conductors separated by a thin insulation barrier. The junction is supercooled in helium cryostat and when it is irradiated by microwave energy (in the frequency range 10–100 GHz), the voltage–current characteristics get broken to the stepped form. The height of each voltage step can be found from

$$V = \frac{hf}{2e} \tag{12}$$

where h is the Plank constant, f is the frequency at which the junction is irradiated, and e is the electron charge.

Because only the irradiation frequency is a variable in the equation, the standard volt is related to the standard of time/frequency. The step voltages can be detected by the use of null detectors or resistors. The accuracy of this method is better than 1 part in 10^8. The uncertainty in the measurements is less than 1 part in 10^7.

The *standard cells* are made from saturated electrochemical Weston cells, which consist of electrodes of mercury (positive) and mercury cadmium amalgam (negative) placed in an electrolyte of saturated cadmium sulfate solution contained in an H-shaped glass container. There are two types of Weston cells – the saturated cell and the unsaturated cell.

The saturated cell has a voltage variation of approximately $-40\,\mu V$ per 1 K, whereas the unsaturated cell has a negligible temperature coefficient at normal room temperatures.

By using precision potentiometers, the standard-cell voltage can be compared with better than 1 part in 10^7 accuracy. However, the saturated cells are very sensitive to environmental conditions and the current they supply introduces further uncertainty in calibrations.

The various national standard laboratories refer to the Josephson junction standard as their primary standard for voltage. However, since the standard cells are much easier to employ, they also maintain a number of saturated cells, referred to by the Josephson junction standard, as the primary standards for voltage. Cells are kept in an oil bath to control their temperature within ± 0.01 K. The voltage of the Weston saturated cell at 293.5 K is 1.01858 V. They remain satisfactory as voltage standards for a period of 15 to 20 years with a drift voltage of about $1\,\mu V$ per year.

Electronic voltage standards, also known as the *transfer standards*, are based on electronic components such as diodes. When the voltage across a diode rises above a certain level, a current starts flowing (e.g. Zener breakdown) through it and the diode acts as a constant voltage device. When these diodes are fed by well-stabilized sources, they can provide a reference voltage that is stable by 1 part in 10^6. Often, they are used together with other electronic components such as transistors, operational amplifiers, and integrated circuits for stability and compensation of the environmental factors like temperature. They are widely used since they are not affected by environmental conditions, as is the case with standard cells.

Numerous standards define the procedures and methods for voltage determinations, particularly in power circuits. Some of these standards are

- IEC Std 60060-2 standard for high-voltage impulse calibrations;
- IEEE Std 519-1992 standard for harmonic control in electrical power systems;
- IEEE Std 4-1995 standard for high-voltage testing;
- IEEE Std 181-2003 standard for transition, pulses, and related waveforms;
- IEEE Std 1122 standard for digital recorders used with measurements in high-voltage impulse tests;
- IEEE Std 1451-1999 standard for smart sensors;
- IEEE Std 1459-2000 standard for the Measurement of Electric Power Quantities Under Sinusoidal, Nonsinusoidal, Balanced, or Unbalanced Conditions.

FURTHER READING

Dyer, S.A. (ed) (2001) *Instruments, Survey of Instrumentation and Measurement*, Wiley, New York.

Eren, H. (2002) Analogue and Discrete Input/Output, Costs, and Signal Processing, Chapter 1.9, in *Instrumentation Engineers Handbook*, 4th edn (ed. B. Liptak), CRC Press, Boca Raton, FL (pp. 123–141).

Eren, H. (2003) *Electronic Portable Instrument s- Design and Applications*, CRC Press, LLC, Boca Raton, FL.

Eren, H. and Ferraro, A. (2003) *Galvanometers and Electromechanical Voltmeters and Ammeters in Encyclopedia of Life Support Systems*, EOLSS/UNESCO, URL: http://www.eolss.net/E6-39A-toc.aspx.

Eren, H. and Ferraro, A. (2003) *Electronic Voltmeters and Ammeters in Encyclopedia of Life Support Systems*, EOLSS/UNESCO, URL: http://www.eolss.net/E6-39A-toc.aspx.

Hoeschelle, D.F. (1994) *Analog-to-Digital and Digital-to-Analog Conversion Techniques*, ISBN: 0-4715-7147-4 Wiley, New York.

Schlabbach, J. (2001) *Voltage Quality in Electrical Power Systems*, ISBN: 0-8529-6975-9, IEE, Stevenage.

Webster, J.G. (ed) (1999) *The Measurements, Instrumentation and Sensors Handbook*, ISBN: 0-8493-8347-1, CRC, IEEE Press, New York.

202: Current Measurement

Halit Eren
Curtin University of Technology, Perth, Western Australia, Australia

1 Introduction 1362
2 Basic Theory 1362
3 Measuring Networks 1363
4 Current Transducers 1365
5 Current Shunts 1366
6 Converters 1367
7 Rectifiers 1367
8 Indicators 1368
9 Digital Instruments 1368
10 Standards 1369
Further Reading 1369

1 INTRODUCTION

Knowledge of the voltage located at the terminal of each circuit element as well as the current flowing in each element gives full knowledge of the circuit behavior. Moreover, when electronic devices for signal processing are involved, such as those used in measuring systems, telecommunication systems, control systems, and informatics, the majority of involved signals are in the voltage and current forms. Therefore, the voltage and current measurements constitute an important area in industrial and laboratory measurements, sensors, and systems.

Electrical currents can be measured in many different ways. Once the current is converted to voltage form, the oscilloscope (strictly a voltage measuring device) is a typical device that can be used to measure currents. There are many analog and digital ammeters and multimeters that are offered by many vendors. In some applications, special current measurement techniques may have to be designed by using carefully selected transducers to meet the particular characteristics and specifications of the process. In this article, the operation of ammeters and multimeters is explained, current transducers are discussed, and examples are given.

2 BASIC THEORY

A charged body in an electric field is subject to a force. If the body is not restrained, it will start moving in the electric field. The result of the movement of charged bodies from one point to another point in the space is the electric current. In the metallic conductors used in the electric circuits, the charged bodies that can flow through the conductors are the electrons. A basic property of an electron is its charge (1.603×10^{-19} C).

Electric current is the rate of charge flowing in a cross-section of the conducting element in a second and its measurement unit is, in the SI, the ampere (symbol: A).

Since the current is taken as the charge flow per unit time, its measurement unit is related to the unit of charge and the unit of time (second s) by

$$1 \text{ ampere} = 1 \text{ coulomb/second} \tag{1}$$

A current will flow in any medium in which there are free charges to move. These conduction charges may be electrons, positively charged 'holes', or positive or negative ions, depending on the material. They are in continuous random motion and collide with each other and with the atomic structure of the material. When a conducting material is placed in an electric field, the conducting charges are accelerated in the direction of the field. The velocity acquired is small compared with the average value of the random velocity (typically in the order of 10^6 m s^{-1}).

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

The current flowing through a circuit element is related to the properties of the circuit element and to the voltage across its terminals by the well-known Ohm's law:

$$V = R\,I \tag{2}$$

where R is a quantity representing the electric behavior of the circuit element. Under DC conditions this quantity is called resistance and its measurement unit is, in the SI, the ohm (Ω).

In most of the usually met applications of electrical and electronic engineering, voltages and currents exist in continuous forms. This means that a generic voltage v and a generic current i can be represented as functions of time t:

$$v = v(t) \quad \text{and} \quad i = i(t) \tag{3}$$

In the DC systems, voltages and currents are constant, that is

$$v = v(t) = V_{\mathrm{DC}},\ \text{const.} \quad \text{and} \quad i = i(t) = I_{\mathrm{DC}},\ \text{const.} \tag{4}$$

The voltmeters and ammeters for DC systems are realized in order to measure V_{DC} and I_{DC} respectively.

In AC systems, the voltage and current are defined with respect to time as

$$v(t) = V_{\max} \sin(\omega t + \alpha) \tag{5}$$

where $V_{\max}$ is the maximal value of the sinusoidal voltage and α is the phase angle of the voltage:

$$i(t) = I_{\max} \sin(\omega t + \beta) \tag{6}$$

where $I_{\max}$ is the maximal value of the sinusoidal current and β is the phase angle of the current. ω represents angular frequency and t is time in (5) and (6). Quantities that are used to characterize sinusoidal voltage and current are

- maximal value
- average value
- root mean square value (rms).

The average and rms values are defined as follows:

$$\begin{aligned} V_{\mathrm{av}} &= \frac{1}{T}\int_0^T v(t)\,\mathrm{d}t = \frac{2}{\pi}V_{\max}, \\ I_{\mathrm{av}} &= \frac{1}{T}\int_0^T i(t)\,\mathrm{d}t = \frac{2}{\pi}I_{\max} \\ V &= \sqrt{\frac{1}{T}\int_0^T v(t)^2\,\mathrm{d}t} = \frac{1}{\sqrt{2}}V_{\max}, \end{aligned} \tag{7}$$

$$I = \sqrt{\frac{1}{T}\int_0^T i(t)^2\,\mathrm{d}t} = \frac{1}{\sqrt{2}}I_{\max} \tag{8}$$

where T is the period.

In the case of a circuit containing only a resistive load, the single frequency current wave is entirely symmetrical with the voltage wave. The current wave lags behind the voltage wave in the case of inductive loads. In a purely capacitive load, the current wave leads the voltage wave. In general, circuits contain elements of resistance, inductance, and capacitance in varying amounts and they must, therefore, assume some intermediate condition of phase angle between voltage and current.

Ammeters for AC systems are realized primarily by using AC-to-DC conversion techniques or by using suitable transducers, both of which will be explained in the following sections.

Different kinds of signals, such as pulse, random, and discrete signals can also be found in nature, especially at the atomic scale. Such signals, however, require measurement with different kinds of instruments, which are not covered here.

3 MEASURING NETWORKS

There exists a diverse range of methods and instruments available for current measurements, and some of these are

- electromechanical ammeters
- thermal type ammeters
- multimeters
- oscilloscopes
- virtual instruments.

Electromechanical ammeters provide readings in analog form by moving a pointer that indicates the measured value on a scale. The scale is a linear scale if the pointer displacement is proportional to the measured quantity. The energy required to displace the pointer is taken directly from the circuit to which the instrument is connected. This energy is the so-called instrument self-consumption and, in electromechanical instruments, can be nonnegligible. Electromechanical ammeters can be classified according to their deflecting movements as

1. moving-iron electromagnetic ammeters,
2. electrodynamic ammeters,
3. moving-coil electromagnetic ammeters (D'Arsonval galvanometers),
4. electromechanical multimeters or VOM.

Although electromechanical ammeters still find extensive applications, they are not used as much as their electronic counterparts, so they will not be explained further. However, because of its special features, the D'Arsonval galvanometers will be visited in Section 8 of this article.

Thermal type ammeters are based on the thermal effect produced by a current flowing in a conductor. These ammeters are also known as the *thermo-instruments* or *true rms devices*. Irrespective of the shape of the input signal waveform, the active power dissipated in the heated conductor is equal to the heat generated and is proportional to the true rms value of the input current.

A simplified construction of this kind of ammeters, also called the '*hot-wire instrument*' is shown in Figure 1. The heated element has a negligible temperature coefficient of resistance (that is, its resistance will remain essentially constant over its operating range) and a constant temperature coefficient of expansion. Since the heating effect is proportional to the square of the current, accurate effective (rms) values of AC currents are measurable irrespective of frequency and waveform. Also, since no magnetism is involved in providing scale deflection, stray magnetic fields have no effect on the operation.

The hot-wire element is made of platinum-iridium, which has a constant resistance and temperature coefficient of expansion over its operating range. The element is about 0.1 mm in diameter. Platinum-iridium has the added benefit of being able to withstand high temperatures without oxidation. Attached to the hot wire is a phosphor-bronze 'magnifying' wire. A silk thread, connected to the magnifying wire, passes around a pulley before being fixed to a spring, which keeps the system under tension. When the hot-wire element expands, its slack and that in the magnifying wire is taken up by the tension in the spring and the silk thread causing the pulley to rotate and the pointer to deflect on the scale.

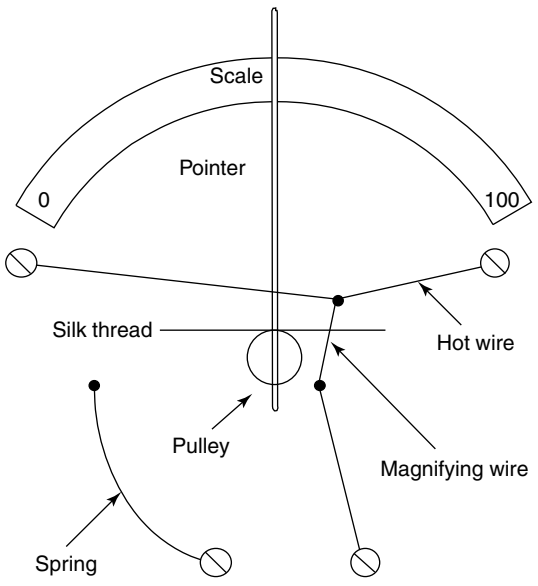


Figure 1. A hot-wire voltmeter.

Multimeters are instruments that can measure voltages, currents, and resistances. These instruments are also known as *volt-ohm-multimeters* (VOMs). The VOM is a combination of a DC ammeter, a DC voltmeter, an AC ammeter, an AC voltmeter, and a multirange ohmmeter with a switch to select the one to use. The DC voltage range of a common VOM is 0 to 1000 V, although with an external resistor the range can be increased to 5000 V. The DC current range is usually up to 10 A, although with an external shunt it can be extended to 20 A or more.

For AC measurements, the instrument contains a rectifying circuit made from germanium or silicon diodes. Because of the inertia of the moving coil, the meter indicates a steady deflection proportional to the average value of the current. Since AC currents and voltages are expressed in rms values, the meter is scaled to read the rms values of sinusoidal voltages through the form factor $\pi/2\sqrt{2}$. The indicated reading of a nonsinusoidal voltage may be erroneous, since its average value may differ considerably from the average value of a pure sinusoidal voltage.

Owing to the fact that the internal structure of the VOM cannot be optimized for a single measurement, the temperature and frequency ranges can be limiting factors. For example, a deterioration of 0.5% in the readings may be expected for every 1-kHz rise in frequency.

Further information on multimeters is provided in Section 8.

Oscilloscopes are probably the most versatile instruments for all forms of physical investigation, since almost any physical phenomenon can be converted into a corresponding electric voltage and the oscilloscope makes the waveform versus time visible for human observation. Measurements are performed on the horizontal axis and vertical axis since the screen features a graduated grid. Nowadays, digital storage oscilloscopes DSO are commonly used in voltage and current measurements. They sample the input signal and convert it into a sequence of data, which are stored in the DSO memory and displayed on cathode-ray tubes CRT or liquid-crystal displays (LCD). A typical structure of a digital oscilloscope is shown in Figure 2.

Virtual instruments use computers, interface electronics, and software to emulate the operational features of most traditional instruments. In this technology, plug-in data acquisition (DAQ) boards, PCMIA cards, and parallel port I/O devices are used to interface sensors and transducers of the system under investigation with computers. Once the signal is interfaced, the computer can be programmed to act just like a stand-alone instrument; but it can also

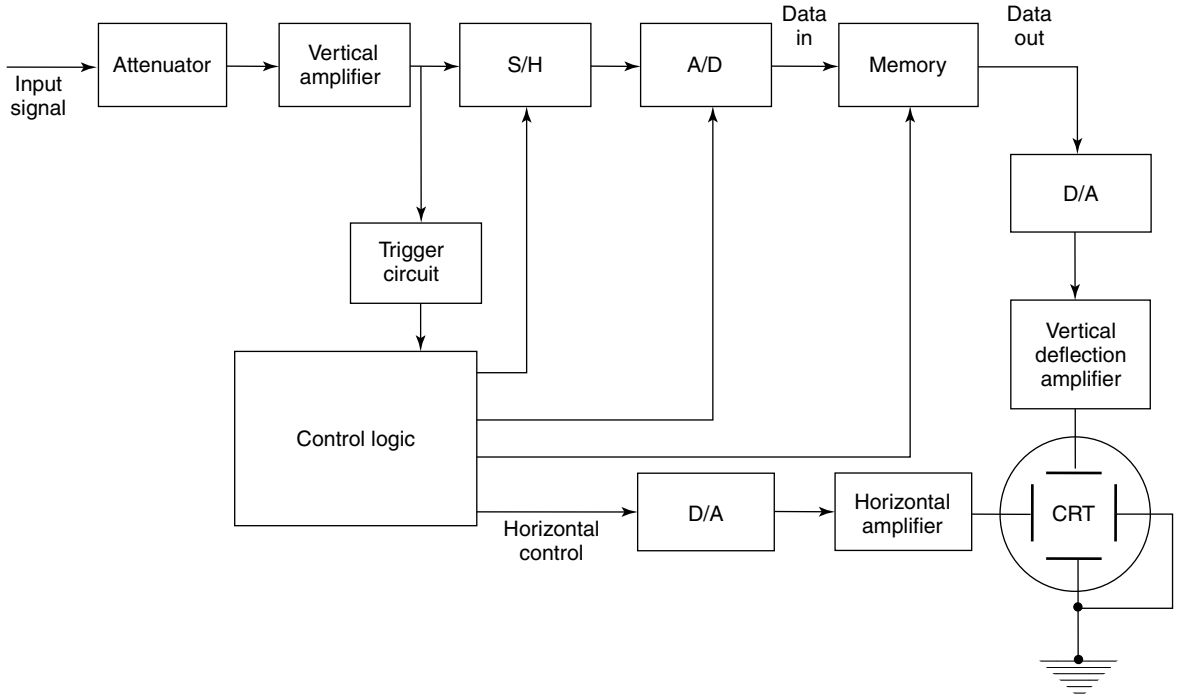


Figure 2. Structure of a digital storage oscilloscope.

offer the additional benefits of a computer by allowing flexibility in signal processing, reconfiguration of front panel displays, and providing extensive storage facilities. In virtual instruments, data can be saved or loaded in memory to be processed in popular spreadsheet programs and word processors, and a report generation capability complements the raw data storage by adding timestamps, measurements, user name, and comments.

4 CURRENT TRANSDUCERS

A transformer is an electrical device, which converts one AC alternating current into another one of the same frequency and having, in general, different characteristics in terms of voltage and current. It consists of two or more electric circuits coupled by a common magnetic circuit for the purpose of transferring electrical energy.

For current measurements in many circuits, it is not practical to allow primary current to flow directly through an instrument. One method is to divert a small current through a resistive shunt. A disadvantage of such a device is that, unlike transformers, they do not provide any insulation between the high-voltage side and measuring system. More commonly used transducers are

- current transformers (CT),
- Hall-effect transducers, and
- Rogowski Coils,
- optical current transducers (OCTs), and
- current shunts (explained in detail in Section 4).

Current transformers have two windings, designated as primary and secondary, which are insulated from each other. The primary winding is connected in series with the circuit carrying the line current to be measured, and the secondary winding is connected to instruments or protective devices. The secondary winding supplies a secondary current in direct proportion to the primary current almost without any difference in phase angle. Hence, the CT transforms line current into values suitable for measuring instruments, meters, protective relays, and other similar apparatus. CT also isolates the instruments and the relays from line voltages. The four common types of CT design are

- wound type
- bar type
- window type
- bushing type.

The current in any system changes more often and with greater magnitude than the voltage. Hence, selecting proper transducer for currents could be difficult. The right CT current rating and turns ratio to be used depends on the measurement objective.

In the case of disturbance recorders or protective devices, where fault or inrush currents are of concern, the CT

must be sized in the range of 20 to 30 times of normal load current. This will result in low resolution of the load currents and inability to accurately characterize load current harmonics.

Conventional current transformers are designed to operate within a frequency range from 15 to 100 Hz. With the need to measure harmonic content of the primary current, the frequency response of current transformers is essential to the measurement process. The frequency response of current transformers is effectively determined by the capacitance present in the transformer and its relationship with the transformer inductance. The standard metering class CT is generally adequate for frequencies up to 2 kHz. Phase angle shift between primary and secondary currents may start to become significant when the frequency is close to 2 kHz. For higher frequencies than 2 kHz, a window-type CT with a high ratio should be used. Desirable attributes for a CT, when harmonics are measured, include large ratio, small remanent flux, large core area (the more steel used in the core, the better the frequency response of the CT), secondary winding resistance and leakage impedance (as small as possible).

Rogowski Coils are more specialized transducers usually used for the measurements of very large currents, which last for a very short duration (hundreds of kiloamperes in less than a microsecond). The Rogowski Coils is a solenoid air core winding of a small cross-section looped around a conductor carrying the large primary current. The voltage induced on the terminals of the coil is proportional to the derivative of the primary transient current and number of turns. The coil is connected to an integrator. Integrated induced voltage gives a measure of the current in the primary conductor. The Rogowski Coils have the advantage of being free from saturation problems and have fast response time.

Hall-effect sensors operate on voltage difference across a thin conductor carrying current. The current depends on the intensity of the magnetic field applied perpendicular to the direction of current flow, as shown in Figure 3. An electron moving through a magnetic field experiences Lorentz force perpendicular to the direction of motion and to the direction of the field. The response of electrons to the Lorentz force creates a voltage known as the *Hall voltage*. If a current I flows through the sensor, the Hall voltage can mathematically be found by

$$V = R_H I B / t \tag{9}$$

where R_H is the Hall coefficient (cubic meters per degree of Celsius), B is the flux density (Tesla), and t is the thickness of the sensor (meters).

The value of R_H depends on the material used, temperature, and field magnitude. Its characteristics can be controlled to a certain extent by doping the base material with some impurities. For example, doping germanium with arsenic can reduce the temperature dependence at the expense of magnitude.

Semiconductor materials, such as gallium arsenide (GaAs), indium antimonide (InSb), and indium arsenide (InAs), produce the highest and most stable Hall coefficients. Because of its combined low temperature coefficient of sensitivity ($<0.1\%/^{\circ}$C), low resistance and relatively good sensitivity, InAs is the material favored by commercial manufacturers of Hall-effect devices.

Optical transducers represent one possible solution to the need for reliable and economical sensors in high-voltage measurement systems. Conventional transducers for voltage and current are expensive and require large volumes of electrical insulation when used on high-voltage lines. Optical current transducers are designed with all solid insulating materials and are therefore intrinsically safe. Because OCT is an electronic device, it differs fundamentally from CT with respect to the signal power involved. In a CT, the secondary signal has a power level of several watts, while power in the OCT secondary signal is typically a few microwatts.

A number of quite different OCTs are available. Diversity is present in all elements of the system. The sensor may be optical or electronic. The insulator may be ceramic or polymer and it may be used to support OCT or it may be suspended from it. All types of OCT use optics to isolate a high-voltage part of the system from a secondary side that carries information. The current being measured by an OCT is represented as modulated light. Typically, the link between sensor and user device (measuring instrument) is an optical fiber.

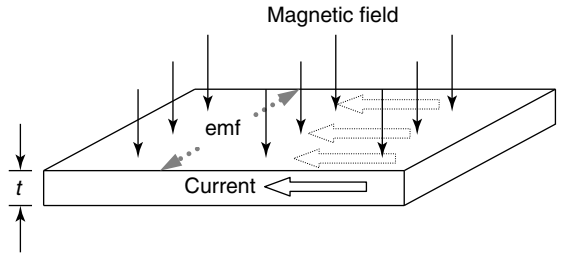


Figure 3. A Hall-effect sensor.

5 CURRENT SHUNTS

Apart from current transformers, CTs, and Hall-effect sensors, current shunts are a most common and cost-effective form of transducers that are used for current sensing.

It consists of a low-value resistor connected to the current path of the circuit, as shown in Figure 4. The waveform

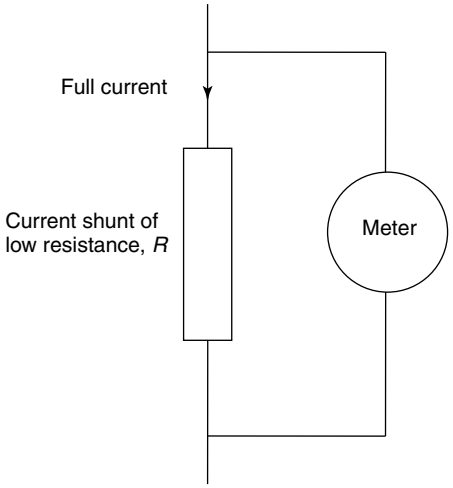


Figure 4. Current shunt arrangement.

of the voltage across the resistor is the replica of the current flowing through the shunt resistor. This voltage is then processed as a representative of the current through the circuit.

Shunt resistors are manufactured in various types and shapes to cater to the application needs. They can be divided into various groups:

- Sleeve shunts can be used for measuring impulse currents up to 500 A. The ohmic value of shunts can be changed by replacement of the active parts.
- Cage shunts can be used for currents up to 10 kA. The housing is usually filled with special sand to enhance operations and thermal characteristics.
- Tubular coaxial shunts can handle currents up to 100 kA for short durations.
- Special high current shunts are used for high currents in high-voltage applications, such as the arrestor testing.

Another way of using a current shunt is for rescaling of a meter. A suitable shunt resistor is connected across the meter. In this arrangement, majority of current flows through the shunt, but the full scale–deflection FSD of the meter is scaled to read the total current.

Shunt resistors usually have low resistance values, as low as 0.001 Ω giving a ratio 1/1000, that is, every 1/1000 (0.001) voltage drop across its length equals 1 A of current. A range of current shunts from 1 to 10 000 A with accuracy from 0.2 to 3% are available in compact flat plates, wires, strips, or brass-ended designs. Shunt resistors are manufactured by using Nichrome, copper, or similar materials of low resistivity. Mechanically, the current shunt is robust and can withstand both over-current and mechanical stress. NIST/NPL traceable calibration test certificates and certificates of conformance are available.

In applications where current shunts are selected for current measurements, attention must be paid to the inductive effects on shunts, terminal resistances of connections, and the types of leads used. This is especially so as the frequency content of the signal rises.

6 CONVERTERS

In many applications, signals are deliberately converted to currents for transmission of information (e.g. the 4–20 mA transmission method) to be connected back to voltage at the receiving end. Voltage-to-current (V–I) and current-to-voltage (I–V) converters are used extensively in the process industry to transmit information from one location to another. Current transmission is convenient because of its good immunity to electromagnetic interference and noise.

Figures 5(a) and (b) show how such converters are implemented using op-amp circuits. In the case of a V–I converter, the output is a current I_0 proportional to the input voltage V_i. Note that the portion of circuit between points A and B acts as a current source controlled by the input voltage V_i. In the case of the I–V converter, the output is a voltage V_0 proportional to the input current I_i. Also, note that whatever the load resistor placed at the output, the entire input current I_i will flow through resistor R_1. Thus, the portion of the circuit between point S and ground acts as a voltage source controlled by the input current I_i.

In some applications, it is important to note that the current-to-voltage converter can simply be a suitable terminating resistor that makes use of Ohm's Law to generate a voltage.

7 RECTIFIERS

In many applications, AC voltages and currents are measured by means of DC voltmeters and ammeters with an addition of a rectifying circuit in the input stage. Rectifiers

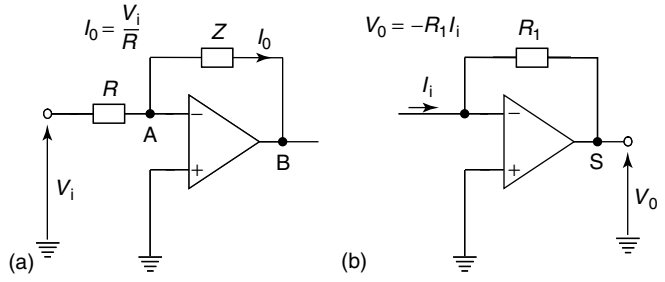


Figure 5. (a) Voltage and (b) current converters based on operational amplifiers.

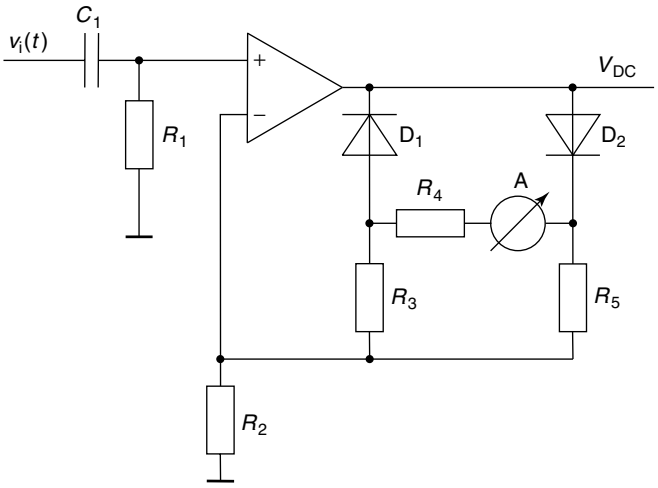


Figure 6. Rectifier-based AC ammeter.

can be classified as controlled or uncontrolled rectifiers, the majority of which are controlled types, which are obtained using thyristors and triacs.

Controlled rectifiers can vary the average value of the voltage applied to the load. They are suitable for use in rectification of single-phase or three-phase voltages and currents from constant AC supplies. A typical basic structure for a rectifier used in ammeters is shown in Figure 6.

The realized meter becomes an average voltage detector that can be used as an rms ammeter by labeling the scale to measure currents, provided that the input signal is a sinewave.

8 INDICATORS

Indicators are based on electromagnetic ammeter principles; one typical example is the D'Arsonval galvanometer. This galvanometer is an instrument specially designed for the measurement of very low currents (in the order of 10^{-8} A) and is therefore employed as null indicator in a majority of balance measurement methods (like the bridge and potentiometer methods). Because of the internal resistance of its coil, it can be also used to measure very low voltages, in the order of 10^{-7} V or even less.

The galvanometer structure comprises a rectangular coil of fine copper wire suspended between the pole faces of a permanent magnet. The magnet pole faces are curved to provide a radial flux. A fixed cylindrical iron core provides continuity of the magnetic circuit. The amplitude of the air gap between the fixed cylindrical iron core and the permanent magnet is about 3 mm. Each coil side lies halfway

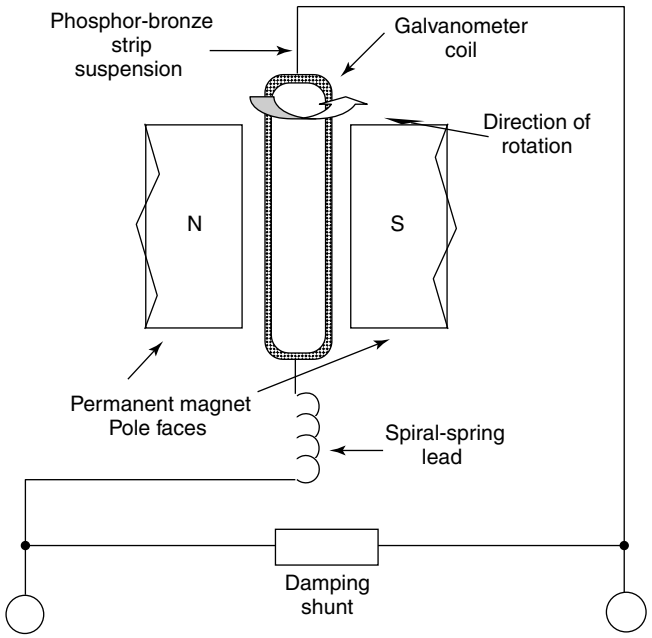


Figure 7. D'Arsonval galvanometer structure.

between the core and corresponding pole face. The suspension is a single fine strip of phosphor bronze and serves as a lead to the upper end of the coil. The lower end is connected to a lead consisting of a spiral spring. In the most accurate executions, the restraining torque is given by the torsion of the suspension strip. In this way, the friction torque is also practically removed. A small mirror, fixed to the suspension, reflects a narrow beam of light through a glass window in the outer case surrounding the galvanometer on to a scale placed a meter away, on which the deflection is measured. Simply winding the coil on aluminium frame provides eddy current damping. Resistive damping may also be obtained by connecting a variable resistor in parallel with the deflection coil. Proper adjustment of this resistance gives critical damping, thereby reducing measurement time. A typical galvanometer structure is shown in Figure 7.

9 DIGITAL INSTRUMENTS

Digital ammeters (multimeters) obtain the required measurements by converting the analog input signal into a sequence of digital samples uniformly spaced in time in the early stages of the signal processing. The input signals are therefore processed in the discrete-time domain and the measurement results are displayed in a digital form. It is worthwhile to note that the distinction between analog and digital meters is not because of the way the measurement results are displayed, but because of

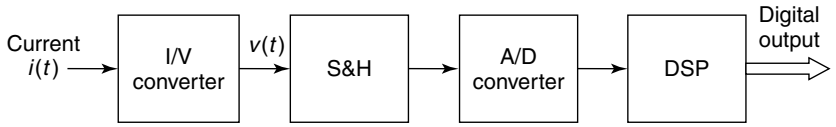


Figure 8. Structure of a modern digital meter.

the domain (continuous-time or discrete-time domain) in which the input signals are processed in the main body of the devices.

A more modern approach to this type of measurement exploits the capability of a fully digital structure as shown in Figure 8 that represents a voltmeter. The same structure can be adapted to an ammeter operation, provided that a current-to-voltage conversion input stage is inserted.

This structure samples the input signal $v_i(t)$ at constant sampling rate f_s and converts each sampled value into a digital code. The whole sequence of converted codes is stored in the memory of the DSP and then processed for the evaluation of information.

Assume that the input signal is periodic, with period T, and that the frequency spectrum is upper-limited by the harmonic component of order N. Digital signal processing theory, and in particular the sampling theorem, ensures that the information associated with the input signal can be totally retrieved from the sequence of the sampled data if at least $(2N + 1)$ samples are taken over a period T in such a way that $(2N + 1)T_s = T$, where the sampling period T_s being $T_s = 1/f_s$.

If $v_i(kT_s)$ is the kth sample, the rms value of the input signal is given by

$$V = \sqrt{\frac{1}{2N+1}\sum_{k=0}^{2N} v_i^2(kT_s)} \qquad (10)$$

As an example, a true rms AC ammeter based on this approach can feature an uncertainty as low as the 0.1% of the full-scale value with a 12-bit resolution analog-to-digital converter (ADC). According to the sampling theorem, the instrument bandwidth is limited to half the sampling frequency. This means that a 500-kHz bandwidth can be attained with modern devices. Although wider bandwidths can be obtained, this is paid for in terms of a lower resolution of the ADC devices.

In many cases, such as current measurements in high-voltage and power quality applications, special arrangements can be made by using appropriate current transducers and supporting equipment. In these cases, careful consideration is needed when sizing the transducers (e.g. CTs) required so that they take advantage of the full resolution of the instrument without clipping or distorting the measured signal. To obtain the most accurate representation of the signal being monitored, it is important to use as much of the full range of the ADC as possible.

10 STANDARDS

The unit of current is defined as, 'the ampere is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross-section, and placed 1 meter apart in vacuum, would produce between these conductors a force equal to 2×10^{-7} Newton per meter of length'.

There is no defined standard for electrical current as such as it is mainly derived from existing electrical standards of resistance and voltage, which are based on the quantum Hall effect and the Josephson effect respectively.

However, there exist numerous document standards for the current generation, harmonic contents, durations of currents in high-voltage applications, electrical and electronic circuits, batteries, grounding and safety aspects, circuit breakers and fuses, and so on. Some examples are

- IEEE C37.09–1979 for circuit breakers;
- ASTM D495-99 for test method for high-voltage, low-current, dry arc resistance of solid electrical insulation;
- IEC 60255-8 for high-voltage current-limiting fuses;
- AS 2024-1991 for high-voltage AC switchgear.

FURTHER READING

Dyer, S.A. (ed.) (2001) *Survey of Instrumentation and Measurement*, Wiley, New York.

Eren, H. (2003) *Electronic Portable Instruments-Design and Applications*, CRC Press LLC, Boca Raton, FL.

Eren, H. and Ferraro, A. (2003) Electronic Voltmeters and Ammeters, *Encyclopedia of Life Support Systems*, EOLSS/UNESCO, http://www.eolss.net/E6-39A-toc.aspx.

Eren, H. and Ferraro, A. (2003) Galvanometers and Electromechanical Voltmeters and Ammeters, *Encyclopedia of Life Support Systems*, EOLSS/UNESCO, http://www.eolss.net/E6-39A-toc.aspx.

Hoeschelle, D.F. (1994) *Analog-to-Digital and Digital-to-Analog Conversion Techniques*, Wiley, New York.

Rathore, T.S. (1996) *Digital Measurement Techniques*, Narosa Publishing, London.

Schnell, L. (1993) *Technology of Electrical Measurements*, Wiley, Chichester.

Webster, J.G. (ed.) (1999) *The Measurements, Instrumentation and Sensors Handbook*, CRC, Boca Raton, FL.

203: Resistance Measurement

Halit Eren
Curtin University of Technology, Perth, Western Australia, Australia

1 Characteristics of Resistance 1371
2 Basic Theory of Resistance 1372
3 Resistance Measuring Networks 1373
4 Bridges for Resistance Measurements 1373
5 Indicators for Use with Bridges 1375
6 Substitution Methods 1375
7 Variable Resistance Use 1375
8 Accuracy and Loading Effects 1376
9 Power Rating of Resistors 1376
10 Standards of Resistance 1377
Further Reading 1377

1 CHARACTERISTICS OF RESISTANCE

Resistance can be defined as the element in an electrical circuit that limits the flow of current in that circuit. Resistance of a material varies inversely proportional to the cross-sectional area of the material and directly proportional with its length and resistivity. Electrical energy flowing in a circuit containing resistance is converted to heat as the square of the current flowing through the resistance.

Materials that have low resistance are known as *conductors* and those that have very high resistance are known as *insulators*. At very low and specific temperatures, some materials exhibit zero resistance, a phenomenon termed *superconductivity*.

Resistance is one of the basic properties of electrical circuits and manufactured components called *resistors*. Although the primary function of resistors is to restrict current, they also find many other uses, such as pull-up, pull-down, impedance matching, voltage division, gain control, bias set, and so on.

The resistance of a resistor varies with manufacturing tolerances, temperature humidity, and voltage stress. Tolerances arise owing to manufacturing techniques, trimming of the material for the desired value, packaging, and soldering of the leads to resistive parts. The percentage variation of resistances quoted by the manufacturers is typically 20, 10, 5, and 1%. Precision resistors can have tolerance less than 1% but they are much more expensive.

Although resistors are manufactured in set values with some degree of tolerance, in many cases the exact values must be known. There are many methods of measuring resistances, as will be discussed in detail in this article. However, during measurements, careful consideration may be necessary due to numerous pitfalls. For example, when the resistance to be measured is small, then the resistance of the leads can cause a significant source of error. To deal with this problem, appropriate leads must be selected and the connection resistances must be minimized.

Apart from the direct measurement of the resistance of resistors and circuits, many sensors have resistive properties, thus requiring employment of appropriate signal processing techniques to monitor minute variations in resistances.

An example is the resistance temperature detector (RTD), which is based on the changes in electrical resistance with respect to temperature – *see* **Article 197, Metallic Resistance Temperature Detectors (RTDs), Volume 3**.

Another example is the strain gauge displacement sensor, which is based on resistance properties of electrical conductors – *see* **Article 192, Strain Sensors, Volume 3**. In this case, as the conductor is stretched or compressed,

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

its resistance alters owing to dimensional changes in the metal-film types.

Commercially available resistance measuring devices can be divided into three main groups: low-resistance meters, medium-resistance meters, and high-resistance meters. Many manufacturers supply meters ranging from few nano ohms to hundreds of tera ohms.

2 BASIC THEORY OF RESISTANCE

Resistance is related to current, voltage, and physical properties of materials. Current can flow in any medium in which there are conduction charges free to move. The conduction charges may be electrons, positively charged 'holes', or positive or negative ions, depending on the material. During current flow, the charges exhibit continuous random motion, colliding with each other as well as with atomic structure of the material.

When a conducting material is placed in an electric field, the conducting charges are accelerated in the direction of the field. The field adds a small average drift velocity to the random velocity. The magnitude of the drift velocity, v_d, is related to the strength of the field E by

$$v_d = \mu E \tag{1}$$

where μ is known as the *mobility* of the charge carriers.

The rate of flow of electric charge across a unit area of a plane normal to the direction of v_d is the current density, J, is given by

$$J = nqv_d \tag{2}$$

where n is the density of charge carriers and q is the charge of each particle. For electronic conductors, the electron charge, $e = q = 1.6 \times 10^{-19}$ C. Combining (1) and (2) gives

$$J = nq\mu E = \sigma E \tag{3}$$

where σ is the conductivity of the material, in Siemens per meter and J is the current density in coulombs per square meter per second. The inverse of conductivity is called the *resistivity* (ρ), which is measured in $\mathrm{VmA^{-1}}$ or Ωm. While the range of conductivity is somewhat arbitrary, it is very low in insulators, $\sigma < 10^{-15}\,\mathrm{S\,cm^{-1}}$, intermediate in semiconductors, $\sigma = 10^{-5}$ to $10^{3}\,\mathrm{S\,cm^{-1}}$, very high in conductors, $\sigma = 10^{4}$ to $10^{6}\,\mathrm{S\,cm^{-1}}$, and infinite in superconductors.

The mobility, μ, usually decreases with increasing temperature due to collisions between the moving electrons, that is, lattice vibrations. The number of charge carriers, n, remains constant for metallic conductors with increasing temperature, but increases exponentially for semiconductors and insulators. Thus, at very high temperatures some insulators become semiconducting, while at low temperatures some semiconductors become insulators.

Equations (1), (2), and (3) may be linked to basic electrical circuit theory by considering the flow of electric current in a section of a wire having a cross-sectional area of A and length l. Assuming a current I is uniformly distributed across the cross-section, the current density can be expressed as

$$J = \frac{I}{A} \tag{4}$$

Expressing the relationship between the electric field in the wire and the potential difference between its ends gives

$$E = \frac{V}{l} \tag{5}$$

Thus, from (3), (4), and (5), the relation between voltage and current becomes

$$V = \left(\frac{\rho l}{A}\right) I = RI \tag{6}$$

Therefore, the resistance R of a material can be expressed as

$$R = \frac{\rho l}{A} \tag{7}$$

R is a quantity representing the electric behavior of the circuit element under DC conditions. When a sinusoidal voltage is applied, the behavior of other circuit elements, such as capacitors and inductors, becomes significant.

In sinusoidal operation, the current flowing is determined by the reactance of the components of the circuit. Reactance is the ratio of the applied voltage to the applied current.

Inductors have inductive reactance $X_L = 2\pi f L\,\Omega$, where f is the frequency in Hz and L is the inductance in Henrys. Capacitance has capacitive reactance $X_C = (1/2\pi fC)\,\Omega$, where C is the capacitance in Farads.

The total effect of combined resistance and reactance in an AC circuit is called the *impedance*, Z, and its unit is ohm. Impedance is the vector sum of the resistance and reactance in the circuit. Impedance can be expressed in the vector form as

$$Z = R + j(X_L - X_C) \tag{8}$$

It can be seen in (8) that measurement of resistance may not be all that easy in circuits as the circuits will often contain some degree of inductive and capacitive properties, these often being the unpreventable imperfections of real systems.

In addition, resistors themselves may exhibit inductive and capacitive properties, which may have significant reactive properties at certain frequencies.

The impedance properties of circuits may be significant in many applications; hence, numerous measurement techniques have been developed. However, in this article, we will primarily concentrate on the measurement of resistance.

Thus, it can be seen that the simple resistance element is actually a complex component when its precise measurement is needed.

3 RESISTANCE MEASURING NETWORKS

Impedance in the AC, and resistance in the DC circuits are important parameters that characterize the circuit behavior, components, and materials used to make these components. In order to measure impedance, it is necessary to measure at least two values since it is a complex quantity, as expressed by 8. There are many modern impedance-measuring instruments that determine the total impedance and then determine the real and imaginary part for resistive and reactive components.

There exist many resistance and impedance measurement methods, each of which has advantages and disadvantages. Some of these methods are

- bridge method
- I-V method
- constant current source method
- network analysis method
- resonant method and more.

3.1 Bridge method

The unknown resistance or impedance forms one of the arms of an electrical bridge – *see* **Article 126, Electrical Bridge Circuits – Basic Information, Volume 2**. When no current flows through the detector, the value of the unknown component can be expressed from the relationship with the other bridge elements. The electrical bridge method is used extensively in resistance measurements; hence, Section 4 is dedicated to this method. This method gives typical accuracies of 0.1% and is applicable for a frequency range of DC to 300 MHz.

3.2 I-V method

The unknown impedance or resistance is calculated from the measured voltage and current values. Current is calculated using the voltage measurement across an accurately known resistor. In some cases, low-loss transformers are used to prevent the effects caused by the use of low-value resistor in the circuit. This is the most suitable for probe-type tests and the frequency range is typically from 10 KHz to 100 MHz.

3.3 Constant current source method

This is the method employed in most multimeters to measure resistances. The resistance measurements are made using a current source and a normal voltage input. The current flows through the unknown resistance and the voltage drop across the resistance is measured. For example, a 1-mA current source will provide a voltage of 100 mV across the 100 Ω of a Pt100 temperature measuring device at 0 °C. This method is applicable in the laboratory environment for a wide range of resistances.

3.4 Network analysis method

In AC applications, the reflection coefficient from the unknown impedance or resistance is obtained by measuring the ratio of the incident signal and the reflected signal. A directional coupler or bridge is used to detect the reflected signal and a network analyzer is used to supply and measure the signals. This is mainly used in high-frequency applications from 300 kHz and above. It gives good results when the unknown impedance is close to the characteristic impedance.

3.5 Resonant method

When a circuit is adjusted to resonance by tuning capacitors, the unknown impedance can be determined from the frequency, capacitance value, and Q-factor. The Q-factor can directly be measured by using a voltmeter across the tuning capacitor. It is accurate in low-impedance measurements in the range of 10 kHz to 70 MHz.

Some of these methods will be explained in greater detail in the sections to follow.

4 BRIDGES FOR RESISTANCE MEASUREMENTS

Bridges are commonly used to measure basic electrical quantities such as resistance, capacitance, inductance, impedance, and admittance. There are two groups of bridges, the AC bridge and the DC bridge. In each type, there exist numerous configurations such as the Wheatstone,

and Kelvin DC bridges, and Schering, Maxwell, Hay, and Owen AC bridges.

Bridges are essentially two-port networks in which the component to be measured is connected to one of the branches of the network. For a detailed explanation of electrical bridges, refer to **Article 126, Electrical Bridge Circuits – Basic Information, Volume 2** and **Article 127, Unbalanced DC Bridges, Volume 2**.

In a particular instrument, the selection of the bridge to be employed and the determination of values and tolerances of its components are very important. A typical example of an AC bridge, a series RC bridge, is given in Figure 1. We also offer some analysis briefly to illustrate their typical principles of operation.

From Figure 1, at balance when the detector D shows zero:

$$Z_1 Z_3 = Z_2 Z_x \tag{9}$$

Substitution of impedance values to this gives

$$\left(\frac{R_1 - j}{\omega C_1}\right) R_3 = \left(\frac{R_x - j}{\omega C_x}\right) R_2 \tag{10}$$

Equating the real and imaginary terms gives the values of unknown components as

$$R_x = \frac{R_1 R_3}{R_2} \tag{11}$$

and

$$C_x = \frac{C_1 R_2}{R_3} \tag{12}$$

Bridges can also be used in a deflection, or a null-balanced mode. A deflection bridge, known as the *unbalanced bridge*, is based on the detecting voltage across the bridge diagonal as illustrated in Figure 2. This figure also symbolizes a typical bridge amplifier. There are many other configurations of bridge amplifiers, depending on the grounding and the availability of grounded or floating reference voltages.

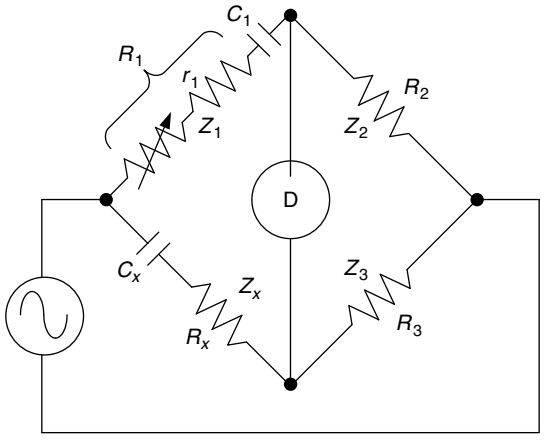


Figure 1. A series RC bridge.

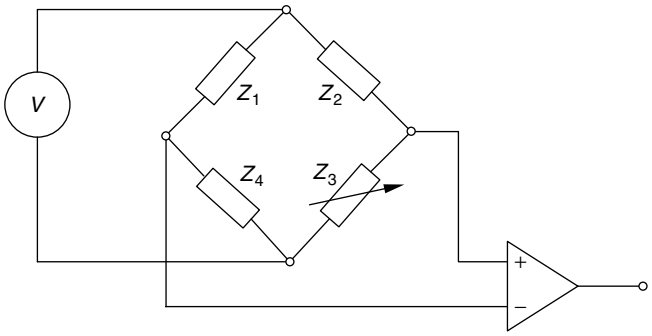


Figure 2. A deflection bridge.

In Figure 2, the output of the bridge is a nonlinear function of unbalance. In small ranges of measurements, the nonlinearity may be considered to be quasi-linear, but must be linearized by suitable techniques for large ranges.

The null-bridge requires it to be balanced at all times. This is achieved by adjusting one of the arms in response to variations in the other arm that carries the sensor. Null bridges have superior linearization characteristics compared to deflection bridges.

Wheatstone bridges, either in null or deflection modes, are commonly used for the conditioning of signals of the outputs of many sensors.

In some applications, the use of incorrect leads between the sensing elements and the signal conditioning circuits can result in severe errors. If care is not taken, some of these leads may have resistance values comparable to the resistances of the sensing elements thus resulting in poor stability. In such cases, a three-lead circuit, as shown in Figure 3, can be used.

In the three-lead circuits, the sensing elements are connected to the bridge using three wires made from the same

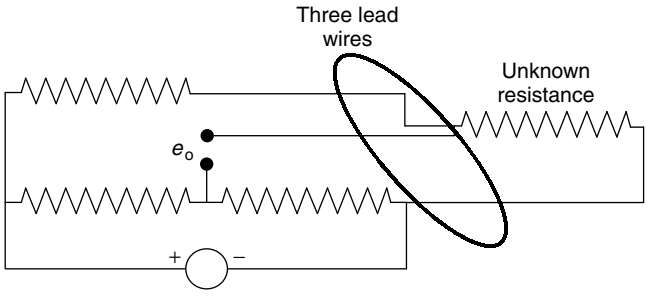


Figure 3. Three-lead configuration.

material and having identical lengths. Although the temperature changes the values of the resistances of the connecting wires, the effects of the one arm is canceled by the changes in the other arm. The third wire represents an additional resistance in series with the reading device and does not affect the measurements.

When very high accuracy is required, a four-wire or six-wire scheme can be adopted. In these cases, a voltage source excites the sensing device, and the corresponding output voltage is measured by a potentiometer. This device can be characterized by a virtual infinite input resistance. Neither the resistance of the connecting wires nor its variations can now affect the measurement results.

5 INDICATORS FOR USE WITH BRIDGES

Indicators for the bridge outputs can be simple voltmeters and galvanometers as well as sophisticated digital devices. In auto-balancing bridges, for example, a current is supplied to maintain a zero potential at the terminals. A detector detects potential at the low terminal and controls both magnitude and phase of an oscillator to maintain the detected potential at zero. When the bridge is unbalanced, the null detector detects an error current and the phase detector separates the vector components of this current. After having been filtered, the resultant signal is used to adjust the error current to zero in the null detector. This balancing operation can be performed automatically over a wide range of frequencies.

Most of the commonly used indicators are based on the principles of voltmeters. Once the error voltage is detected, the signal is processed either in the analog or digital form for suitable displays. Conventional resistance meters use needle displays based on galvanometers. Nowadays, displays are in digital form using LED arrays or LCDs. Once the digital platform is selected, the results may be displayed on a computer screen as indicators or numbers.

6 SUBSTITUTION METHODS

In some measurements, external effects such as temperature must be seriously taken into account. The resistance variation because of temperature can be of the same order of magnitude as the sensor-induced variations, thus prompting the selection of suitable compensating circuits. Best solutions for temperature compensation are obtained by using a dummy gauge as well as active gauges in the opposite legs of a Wheatstone bridge, as shown in Figure 4 (the so-called half bridge configuration). The bridge is balanced under the no-load condition and the sensors are mounted in such a way as to experience the same temperature fluctuations

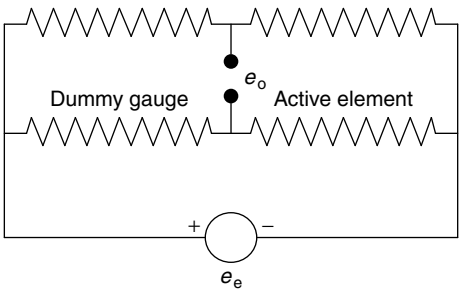


Figure 4. Wheatstone bridge by using dummy elements.

and hence undergo similar changes in resistance values. This assures that the bridge equilibrium is not altered due to temperature. In this arrangement, the sensor acts exclusively on the active gauge and the unbalance of the bridge becomes a factor of sensor resistance only.

The maximum value of the excitation voltage in Figure 4 is set by the self-heating phenomenon of the circuit owing to the currents flowing through the bridge arms. In fact, self-heating could destroy the sensing element and cause a series of alterations in the resistances.

7 VARIABLE RESISTANCE USE

Variable resistors, trimmers, and potentiometers find frequent use in all types of circuits including balancing of bridges. In electronic circuits, they are used for trimming op-amps, tapping reference voltages, setting trip points, adjustments voltages, currents, and frequencies, and so on. In electrical circuits, they are used to limit currents, and to adjust power flow in and out of the circuits. Variable resistors are available in carbon-composition and wire-wound types. Variable resistors tend to have good temperature stability. Nevertheless, they can be a source of externally induced noise, so it is advisable to provide good shielding for any possible electromagnetic interference (EMI).

Variable resistors are basically mechanical devices having inertia, friction, and their associated limitations. The sliding electrical contacts can be prone to oxidization and thus variable contact resistance as they move. In general, they can be sensitive to moisture, dirt, and wear and tear, and should be avoided whenever possible. Fixed resistors with the desired settings are preferable for precision operation.

Resistive potentiometers are basically resistance elements that have a movable arm that transform mechanical displacements into voltage forms. A simple potentiometer is illustrated in Figure 5.

In this Figure 5, e_e is the excitation voltage and $e_o(t)$ is the output voltage. The input signal is proportional to

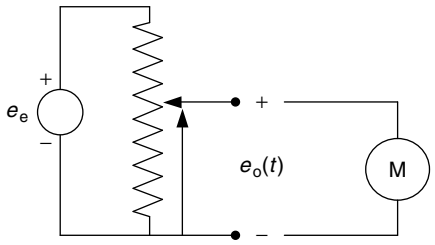


Figure 5. Schematic representation of a potentiometer.

the displacement $x(t)$ of the wiper. Under the no-load conditions, the output voltage is a linear function of the displacement. However, when a user device is connected to the output $e_o(t)$, a loading effect occurs that can make the device nonlinear. There are a number of techniques available to minimize the loading effects.

8 ACCURACY AND LOADING EFFECTS

For real measurements, it is reasonable to assume that results always contain some errors. Typical errors are

- loading effect
- instrument inaccuracies, such as DC biasing
- residuals in test fixtures and cables
- noise.

8.1 Loading effect

If the resistance of the measuring device shown in Figure 5 is comparable in size with the unknown resistance under investigation, the current drawn by the meter can distort the readings. One of the solutions to this problem is to select a high input–impedance meter that is at least 50 to 100 times higher than the impedance of the load. If loading effect is unavoidable, mathematical techniques can be used to make corrections.

8.2 Instrument inaccuracies

There may be many reasons for instrument inaccuracies. Many manufacturers design and produce precision instruments to measure high-resistances (called *Megger*) and low-resistances.

8.3 Residuals in test fixtures and cables

Typical examples are parasitic effects, cable impedances, test signal levels, guarding, contact resistances, and so on.

8.4 Noise

This is a discussion topic in its own right in such measurements. Noise can be generated by the measuring equipment, by the sample under test or imposed externally by means of electromagnetic interference – see the Section on **Noise and Interference**.

In applications in which sensitivity and precision are required, adequate compensation techniques and specially designed cables must be employed in highly controlled environments.

9 POWER RATING OF RESISTORS

The power absorbed (in the form of heat) by a resistor is a product of resistance and square of the current through it (i.e. $P = I^2R$). Therefore, the size, surface area, thermal properties, and composition of resistance are important for it to be able to dissipate the generated heat without changing its electrical properties, especially in the long term. Resistors, therefore, are manufactured for specific applications. The selection of an appropriate resistor for an application must meet the requirements of the circuit.

In electronic applications, resistors can be viewed under six different groups, these are: carbon-composition, carbon-film, thin-film, metal-film, wire-wound, and foil resistors.

9.1 Carbon-composition resistors

These are made from graphite, ceramic, and resin compressed at high temperatures. These resistors have poor temperature coefficients, stability, and shelf life. Carbon-film resistors generally are used in low power and higher-resistance applications, above 1 MΩ. They have low-noise properties.

9.2 Metal-film resistors

These are formed by sputtering metal onto ceramic tubes. They have good accuracy and low variability, therefore are most suitable for long term and continual use. Their noise figures are much lower than the carbon-composition resistors. Metal-film resistors have very low temperature coefficients and are suitable for linear circuits.

9.3 Thin-film resistors

These are made from tantalum nitrate sputtered and etched on a silicon substrate. They have low temperature coefficients and are suitable for resistance networks and surface-mount technology.

Table 1. Properties of resistor types.

	Carbon composition	Carbon-film	Metal-film	Thin film	Wire wound	Foil
Accuracy (±ppm)	10 000	5000	100	50	20	5
Maximum resistance (MΩ)	22	50	10	1	1	0.25
Temperature coefficient (±ppm/°C)	1200	1000	100	100	10	2
Maximum voltage (V)	500	500	350	200	1000	500
Humidity effects (%)	15	3	0.5	0.5	0.5	0.02
Power range	Low	Low	Low	Low	Any	Low
Noise	Low	Low	High	Low	Low	Low
Long-term stability	Poor	Low	Fair	Fair	Excellent	Excellent
Price	Very low	Low	Moderate	Moderate	High	High

9.4 Wire-wound resistors

These are made from nickel-chromium wire wound on a beryllium oxide core. They have excellent temperature coefficients. Wire-wound resistors find applications particularly in high-power circuits where accuracy is of prime importance. One drawback of the wire-wound resistors is that they may have substantial inductances, which can make them unsuitable in high-frequency applications (say above 50 kHz).

9.5 Foil resistors

These are made from nickel-chromium bonded onto a ceramic substrate and laser trimmed for accuracy. They have the best temperature coefficients and have very high precision. Foil resistors have low inductive and capacitive properties. They are good for high-frequency applications.

General-purpose, high-quality resistors are available in DIP packages, which look like ICs. They are used in many applications, particularly in digital circuits as pull-up or pull-down resistors.

The properties of the various kinds of resistors are summarized in Table 1.

It is worth noting that resistors for applications in electronic circuits are much smaller compared to their electrical and power applications counterparts, but their composition and thermal properties are similar.

10 STANDARDS OF RESISTANCE

The most accurate resistance standard presently available is based on the Hall-effect – *see* **Article 43, Units, Volume 1**. This is a quantum effect that takes place in a bidimensional electronic gas (that can be obtained on a very thin layer of semiconductor materials like the Si MOSFETs and GaAs thin structures) in the presence of a very high magnetic–induction field ($B \cong 10\,\mathrm{T}$) and at very low temperatures ($T \cong 1\,\mathrm{K}$). Under these conditions, the Hall resistance takes discrete values:

$$R_{\mathrm{H}} = \frac{h}{ie^2} \tag{13}$$

Where i is an integer, h is the Plank constant and e is the electron charge.

The quantity h/e^2 is called *von Klitzing constant* and has a value of 25812.807 Ω, known with a relative uncertainty of 2×10^{-7}. Moreover, experiments that lead to the determination of this constant has a reproducibility of about 10^{-9}. However, experiments require strict environmental conditions (low temperature, high magnetic–induction field) that can be obtained only by a few National Standard Organizations. For this reason, resistance standards are obtained by means of stable, low temperature–coefficient alloys that are usually maintained in the laboratories with these standard resistors being offered by many vendors. The resistance values of standard resistors are referred to as *Hall-effect-based standard for calibration*.

There are many standards for determining resistance under specific applications, such as

- IEEE 81-1983 *Measuring Earth Resistivity, Ground Impedance, and Earth Surface Potentials of a Ground System (Part 1)*
- IEEE 81.2-1991 *Measurement of Impedance and Safety Characteristics of Large Extended or Interconnected Grounding Systems (Part 2)*
- IEEE 118-1978 *Test Code for Resistance Measurements*

FURTHER READING

Eren, H. (2003) *Electronic Portable Instruments-Design and Applications*, CRC Press, LLC, Boca Raton, FL.

Eren, H. and Ferraro, A. (2003) Electronic Voltmeters and Ammeters, *Encyclopedia of Life Support Systems*, EOLSS/UNESCO, http://www.eolss.net/E6-39A-toc.aspx.

Doebelin, E.O. (1990) *Measuring Systems-Application and Design*, 4th edn, McGraw-Hill, New York.

Dyer, S.A. (ed.) (2001) *Instruments, Survey of Instrumentation and Measurement*, Wiley, New York.

Hoeschelle, D.F. (1994) *Analog-to-Digital and Digital-to-Analog Conversion Techniques*, Wiley, New York.

Schlabbach, J. (2001) *Voltage Quality in Electrical Power Systems*, IEE, Stevenage.

Webster, J.G. (ed.) (1999) *The Measurements, Instrumentation and Sensors Handbook*, CRC Press, New York.

204: Capacitance and Inductance Measurement

Consolatina Liguori
University of Cassino, Cassino, Italy

1 Measuring Networks 1379
2 Capacitance and Inductance Forms 1381
3 Selection of Test Frequency 1382
4 Measurement of Loss Angle 1383
5 Stresses Imposed by Measurement 1383
6 Single Arm Measurements 1384
7 Differential Measurements 1385
8 Excitation Sources 1385
9 Digital Instruments 1386
10 Standards 1387
Related Articles 1388
References 1388

1 MEASURING NETWORKS

There are many techniques for the measurement of capacitance and inductance. General principles and examples are dealt with in texts such as Osvàth (1992), Coombs (1995), and Morrison (1993); the most used ones can be grouped in the following categories:

- Current–voltage methods
- Resonance methods
- Time constant methods
- Bridge methods
- Reflection methods.

To choose the most appropriate method for a specific application some factors, such as frequency coverage, measurement range, measurement accuracy, and ease of operation must be taken into account, but it is necessary to make tradeoffs between them since there is not a single measurement method that completely fulfills all measurement needs. Background for the measurement of voltage is found in **Article 201, Voltage Measurement, Volume 3**, current in **Article 202, Current Measurement, Volume 3**, and resistance in **Article 203, Resistance Measurement, Volume 3**.

1.1 Current–voltage methods

A sinusoidal current, of pulsation $\omega = 2\pi \mathrm{f}$, is forced through the unknown component. The capacitance/inductance value is derived from the voltage and current measurements. Usually, the current that flows through the component is obtained from the voltage measurement across a known low-value resistor, R, as reported in Figure 1.

1.2 Resonance methods

These methods use the fact that a combination of inductance and capacitance resonates at certain frequencies. In particular, the component to be measured is placed in a tuned circuit with known components and the resonance frequency is measured. In Figure 2, a simplified circuit for the inductance measurement is reported: the resonance is attained by adjusting the tuning capacitor, C, and the unknown inductance value is evaluated from the test frequency, f, and the C value.

1.3 Time constant methods

The unknown component is placed in a resistive circuit and the time constant of the obtained circuit is evaluated.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

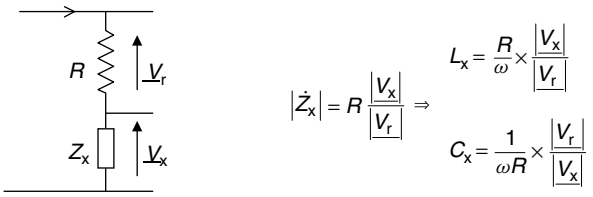


Figure 1. Current–voltage technique for impedance measurement.

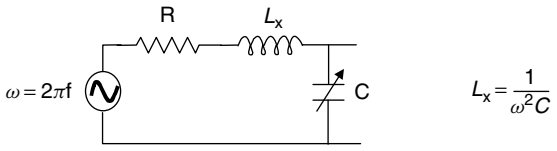


Figure 2. Example of inductance measurement with the resonance method.

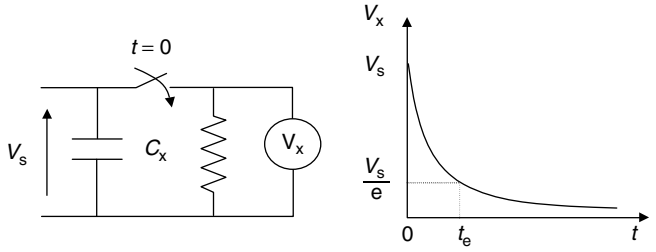


Figure 3. Example of capacitance measurement with the time constant method.

Figure 3 shows a measuring network for capacitance measurements. At time instant $t = 0$, the capacitor C_x, charged with voltage V_s, is shunted with the resistor R. Then the voltage across C_x becomes:

$$V_x = V_s e^{-\frac{t}{C_x R}} = V_s e^{-\frac{t}{\tau}} \tag{1}$$

The time constant, τ equal to the elapsed time from $t = 0$ to the time instant when V_x becomes equal to V_s/e, is measured, and the C_x value is obtained from it: $C_x = \tau/R$.

1.4 Bridge methods

The schematic of a 4-arm AC bridge network is reported in Figure 4. The voltage source is typically a sinusoidal oscillator whose frequency is adjusted to the required value.

Z_x is the component under measurement and the other three arms contain fixed and/or variable standard elements. The variable elements are adjusted until the detector, D, senses zero voltage across it (no current flows through it). In this condition, the value of the measurand can be obtained

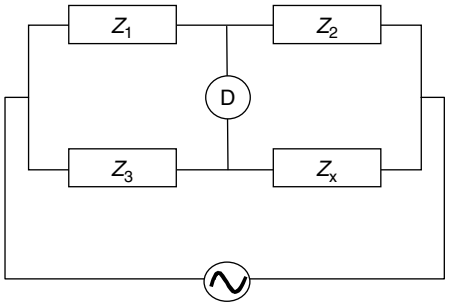


Figure 4. Schematic of impedance bridge.

by its relationship with the other bridge elements:

$$\dot{Z}_x = \frac{\dot{Z}_2 \times \dot{Z}_3}{\dot{Z}_1} \tag{2}$$

Bridges are used to make accurate measurements of unknown capacitances and of associated losses in terms of some known external capacitances and resistances.

Various types of bridge circuits, employing combinations of L, C, and R components as bridge elements, are used for various applications. In the following, the Schering and the Maxwell–Wien bridges used for capacitance and inductance measurement respectively are described as examples.

In the Schering bridge (see Figure 5), the unknown capacitance is modeled with the series equivalent circuit; the variable elements can be the reference capacitors (C_1, C_2), and this allows the use of the bridge in high frequency, or the resistance and the capacitance R_1 and C_1; this solution is preferred in high voltage measurements for safety reasons (the voltage drop on the variable components, adjusted by the operator, is low).

With these components, (2) becomes

$$C_x = C_2 \times \frac{R_1}{R_3}, \quad R_x = R_3 \times \frac{C_1}{C_2} \tag{3}$$

As you can see, the relationships are analytically independent of the frequency. It has to be noted that when both

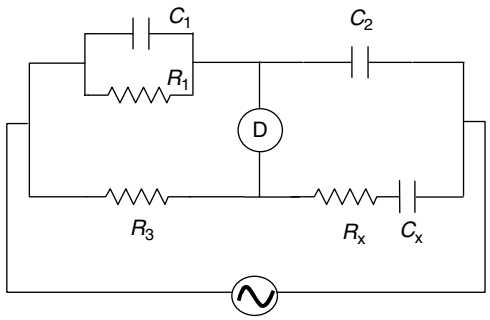


Figure 5. Shering bridge.

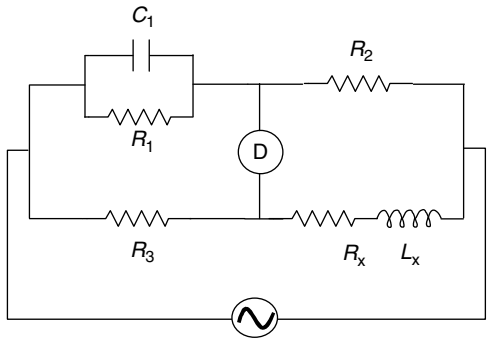


Figure 6. Maxwell–Wien bridge.

C_1 and C_2 are variable, the balancing procedure is difficult to achieve (sliding balance problem) since C_2 appears in both the C_x and R_x relationships. As you can see, the relationships are analytically independent of the frequency.

In the Maxwell–Wien bridge (see Figure 6), the unknown inductance is modeled with the series equivalent circuit and the equilibrium condition is obtained adjusting the R_1 and C_1 values. The relationships to obtain the unknown parameters are:

$$L_x = C_1 \times R_1 \times R_3,\ R_x = \frac{R_2 \times R_3}{R_1} \tag{4}$$

Also, this bridge is theoretically frequency independent and the balancing procedure is not affected by the sliding balance problem.

More detail in bridge design is found in **Article 126, Electrical Bridge Circuits – Basic Information, Volume 2**; and **Article 127, Unbalanced DC Bridges, Volume 2**.

1.5 Reflection methods

The inductance or capacitance value is derived by a reflection coefficient measurement. A step signal is applied to the component under test and the reflected signal is analyzed: measuring the ratio of the incident signal to the reflected one, the reflection coefficient is obtained.

2 CAPACITANCE AND INDUCTANCE FORMS

Capacitance is defined as the capability of a system, circuit, or device for storing electric charge. Namely, the capacitance, C, is defined as the ratio of stored charge, Q, to the impressed potential difference, V:

$$C = \frac{Q}{V} \tag{5}$$

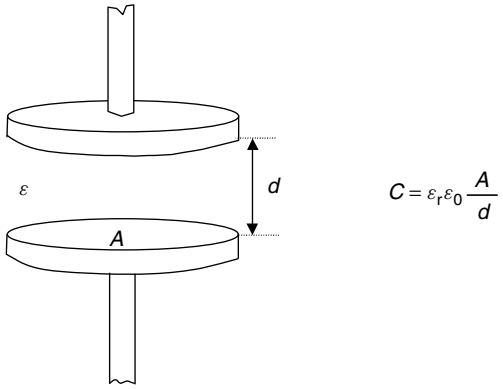


Figure 7. Example of a capacitor.

The device designed to store charge is called a *capacitor*. A simple capacitor (see Figure 7) consists of two plates made of an electrically conducting material and separated by a dielectric material, ε (e.g. mica, oil, paper, tantalum, or air).

Real capacitors present parasitics: dielectric absorption and leakage currents. Absorption introduces a time lag during the charging and discharging of capacitors. The leakage current prevents indefinite storage of energy in the capacitor.

In order to take into account these parasitics, a capacitor is represented with the equivalent circuit reported in Figure 8, where:

C is the capacitance;
R_s and L are the parasitics existing in the lead wires and electrodes;
R_p represents leakage between the two electrodes.

Usually, the simplified circuits shown in Figure 9 are used: the series circuit is chosen for low leakage capacitor (high R_p), while the parallel one is preferred for high leakage currents (low R_p). Measuring capacitor in series or parallel mode can provide different results, and how they differ depend on the quality of the capacitor.

Inductance is a quantity that measures the electromagnetic induction of an electric circuit component. The self-inductance, L, of a circuit component determines the

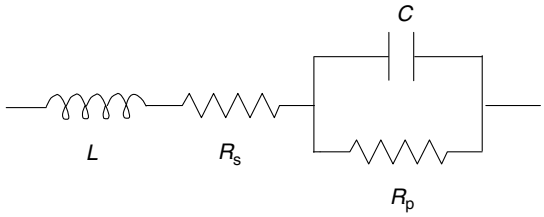


Figure 8. Equivalent circuit of a real capacitor.

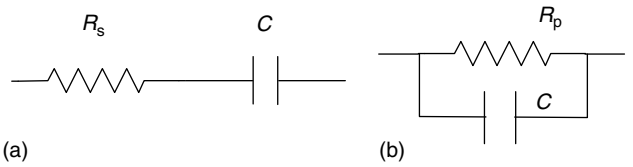


Figure 9. Simplified equivalent circuits of a capacitor: (a) series circuit and (b) parallel circuit.

magnitude of the electromagnetic force, V, induced in it as a result of a given rate of change of the current,

$$i, L = \frac{V}{\mathrm{d}i/\mathrm{d}t} \tag{6}$$

Similarly, the mutual inductance, M, of two components, determines the electromagnetic force that one of each may induce in the other for a given current change. An inductor consists of a wire, of good conducting materials, wound around a core. The simplest core material used is air, but for volumetric efficiency of the inductor, magnetic materials such as iron and ferrites are employed.

Real inductors include parasitic resistive proprieties, due to the resistance of the conductor used for the winds, and present distributed capacitance between turns and between layers of the winding, due to the imperfection of the insulation. The equivalent circuit of a real inductor is reported in Figure 10, where:

R_p represents the loss of the core;
R_s represents loss of the wire;
C is the distributed capacitance between the turns of the wire.

With DC and low frequencies, the effects of the latter two factors are negligible, and an inductor can be represented by the simple series circuit, as in Figure 11, where L is the inductance and R_s the resistance of the winding.

More details on the capacitance and inductance forms are presented in Eren and Goh (1999) and Kumar *et al*. (1999).

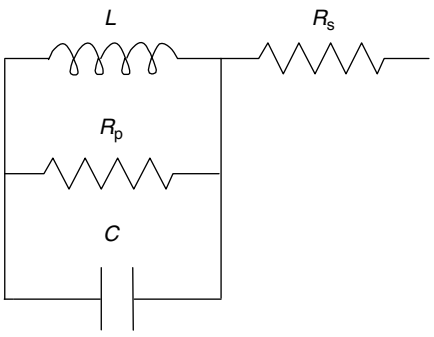


Figure 10. Equivalent circuit of a real inductor.

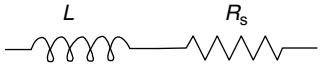


Figure 11. Simplified equivalent circuit of an inductor.

3 SELECTION OF TEST FREQUENCY

The actual value of capacitance and inductance takes into consideration the effects of the parasitic components; thus, it is frequency dependent. As an example, for a capacitor at high frequencies, the effects of lead inductance, series resistance, and leaking resistance become more pronounced. Consequently, in general, it can be said that the test frequency should be chosen close to the operating frequencies of the unknown component.

There are many measurement methods, see also Section 1, whose frequency range applicability is different.

3.1 Current–voltage methods

With this approach, the component reactance module is measured $Z = |V|/|I|$. Considering the components as ideal, because of the relationships between the reactance module and the test frequency (see Table 1), some considerations can be drawn. For low capacitance value (less than 100 pF), the supply frequency should be high enough, typical values ranging from 10 kHz to 100 MHz, in order to yield reasonable values for their impedances and vice versa for inductors.

3.2 Bridge methods

A class of bridges, the so-called resonance bridges (Wien), is structurally frequency dependent, namely, the relations involve the component reactance, and consequently the considerations for current–voltage methods are still valid.

Other kinds of bridges (e.g. Schering, Maxwell–Wien etc.) are theoretically frequency independent; but they are practically frequency dependent due to the presence of parasitics in the unknown device and in the bridge connections. In particular, at high frequencies the stray capacitance becomes a significant factor, while at low frequencies the magnetic effects have to be taken into account. Bridges

Table 1. Reactance of capacitance and inductance.

Component	Capacitance	Inductance
Reactance module	$\frac{1}{\omega C}$	ωL

are also dealt with in **Article 126, Electrical Bridge Circuits – Basic Information, Volume 2**; and **Article 127, Unbalanced DC Bridges, Volume 2**.

3.3 Resonance methods

The value of the unknown quantity is obtained by tuning a standard and variable component in the measuring network and depends directly on the test frequency (e.g. $L = 1/\omega^2 C$). Since the range variability of the standard component is limited, at a given frequency only a small range of the unknown can be measured. In order to extend the input range, the test frequency is variable. In particular, the frequency value is determined as a function of the unknown component magnitude order, and then the standard component is regulated to obtain the resonance condition.

Moreover, it is necessary to underline that at a high frequency for a low-value inductor, measurement may be impossible, since the test cable capacitance can cause other resonances.

Problematic issues concerning the frequency selection are dealt with in texts such as Agilent Technologies Inc (2000) and Bartlett (1992).

4 MEASUREMENT OF LOSS ANGLE

In sinusoidal condition, using a phasor notation, an ideal capacitor, the voltage across it, V, lags the current trough, I, by 90° (see Figure 12(a)). For a real capacitor, at the capacitive current a resistive current (in phase with the voltage), owing to the loss in the dielectric, is added; consequently the current through the capacitor is not 90° out of phase with respect to the applied voltage (see Figure 12(b)). The angle, δ, between the resulting current, I, and the capacitive current, I_c, is called the *loss angle* and depends on the used dielectric characteristic and on the capacitor structure. In order to measure this angle, traditional phase

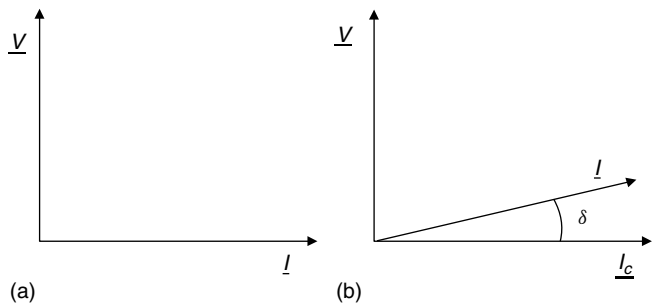


Figure 12. Current and voltage phasors for (a) ideal capacitor and (b) real capacitor.

measurement techniques (zero crossing, oscilloscope methods, three-voltmeters method etc.) can be employed; but since this angle is normally small the obtainable uncertainty is too high.

Often, in order to quantify the dielectric loss, the dissipation factor, (DF), also known as *loss tangent* is used, $DF = \tan(\delta)$ instead of the loss angle of course, for a good capacitor (small δ), the loss angle is approximately equal to the dissipation factor ($\delta \approx \tan(\delta)$).

The dissipation factor is defined also as the inverse of the quality factor and consequently can be obtained as the ratio between the energy lost and the energy stored in the capacitor

$$DF = \frac{energy\ lost}{energy\ storage} \tag{7}$$

Using a capacitor equivalent circuit, the relationships between the loss tangent and the equivalent capacitance and resistance can be determined analytically.

Considering a series equivalent circuit, (Figure 9(a)) we have:

$$\tan\delta = \frac{V_R I}{V_C I} = \frac{R_S I^2}{\frac{1}{\omega C_S} I^2} = \omega C_S R_S \tag{8}$$

For a parallel equivalent circuit (Figure 9(b)):

$$\tan\delta = \frac{V I_R}{V I_C} = \frac{\frac{V^2}{R_P}}{\frac{V^2}{\omega C_P}} = \frac{1}{\omega C_P R_P} \tag{9}$$

see also Krupka and Geyer (1999), and Eren (1999).

Consequently, in order to determine the dissipation factor/loss angle, impedance measurement methods (e.g. bridges, resonance techniques, LCR meter) that allow the value of both components of the capacitor equivalent model to be evaluated can be used.

Some digital instruments can directly display the loss angle of the component under measurement.

The loss angle depends on the frequency, but, normally, it is measured at 1 kHz.

5 STRESSES IMPOSED BY MEASUREMENT

In order to correctly measure inductance and capacitance, all the influence parameters have to be taken into account. In particular, the more typical dependency factors that affect measurement results are the following:

- Frequency
- Test signal level
- DC bias
- Physical environment.

The influence of these factors on the measurement process is explained in Agilent Technologies Inc (2000), Bartlett (1992), and Eren (1999).

5.1 Frequency

As said in Section 3, frequency dependency is common to all real-world components because of the existence of parasitics. Not all parasitics affect the measurement but the prominent ones that determine the component's frequency characteristics have to be quantified.

5.2 DC bias

Some capacitors are polarized; therefore, any applied alternating voltage should have a DC component that avoids any polarity violation. Furthermore, the capacitance of some types of dielectric ceramic capacitors varies depending on the DC bias voltage applied.

Analogously for cored-inductors, for which, owing to the magnetic flux saturation characteristics of the core material, the inductance varies according to the DC bias current flowing through the coil.

5.3 Test signal level

Since components are not completely linear, the test signal amplitude may affect measurement results and different considerations have to be made in the case of a capacitor or an inductor.

5.3.1 Capacitor

This dependency varies as a function of the dielectric constant used; consequently, the test signal level has to be chosen near to the working level. Furthermore, when the capacitor is subjected to high operating voltages, if the electric field in the dielectric exceeds the breakdown value, it damages the dielectric permanently. Usually, values of surge voltages are given to indicate the ability of capacitors to withstand high transients.

5.3.2 Inductor

Cored-inductors are current level dependent due to the electromagnetic hysteresis of the core material. This dependency is significant: in fact, when a high-level test signal is applied to an inductor, measurement may be impossible for a certain frequency range. This is because of the nonlinearity of the core material that causes harmonic distortion of the test signal current.

5.4 Physical environment

Physical environments, for example, humidity, magnetic fields, light, atmosphere, vibration, and time may affect the impedance value. Consequently, it is necessary to control these influence parameters.

6 SINGLE ARM MEASUREMENTS

The capacitor and inductor are two terminal components, but, depending on the value being measured, other configurations, allowing parasitic effects to be reduced, are used, namely, for low impedance (high-value capacitance or low-value inductance) measurement, lead resistance and lead inductance in the contact electrodes, test fixture, and cables must be minimized. Toward this aim a four-terminal or a four-terminal pair configuration has to be used to connect the unknown components to the measurement instrument (see Figure 13). The impedance value is defined as the ratio between the voltage across the voltage sensing points (PL, PH) and the current flowing in the supply leads when no current flows (IL, IH) in the voltage path.

Conversely, for high impedance (low-value capacitance or high-value inductance), stray capacitances become significant compared to the impedance under measurement. To constrain the effects of the stray capacitance, a three-terminal configuration is used. An electric shield enclosing the component constitutes the third terminal of the impedance. The effect of the shield is to keep the parasitic impedances constant in the presence of any change in its electric environment, but it does not reduce their values. There are several options available for the connection of the shield, as an example, the solution to Figure 14 is used, in order to prevent the flowing of the leakage current into the ammeter leads. As a consequence, parasitic effects do not influence the measured value.

The three-terminal and four-terminal connections can be combined resulting in a five-terminal connection. With

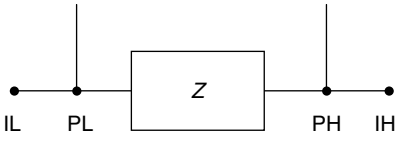


Figure 13. Four-terminal connection.

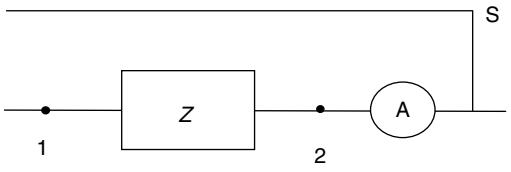


Figure 14. Three-terminal configuration.

this connection, benefits from both connections can be obtained.

For accurate measurements, open/short compensation should be properly performed, in order to compensate the effect of stray admittance and residual impedance respectively, namely, when the instrument input terminals are open, stray admittance is measured, whilst residual impedance is negligible; when the instrument input terminals are shorted, the measured impedance represents residual impedance. These measured parameters are used for correcting the unknown impedance measured value.

Useful suggestions can be found in Osvàth (1992), Coombs (1995), and Agilent Technologies Inc (2000).

7 DIFFERENTIAL MEASUREMENTS

Some capacitive and inductive sensors are differential: namely, the quantity to be measured determines a variation of the reactance value of the sensor. These sensors generally have three terminals; as an example, in some capacitive differential sensors, the central plate moves in response to a physical variable with respect to fixed plates. Consequently, for one measurand change, two capacitances vary: one increasing and one decreasing, obtaining a higher sensitivity (Webster and Pallas–Areny, 1991).

Normally, a bridge measurement configuration, *see also* **Article 127, Unbalanced DC Bridges, Volume 2**, is used with the two impedances placed in adjacent arms; the output from the central plate is zero at the central position and increases as it moves left or right. The major advantage of this configuration is the obtained linear relationships between the measured voltage and the impedance variations.

Considering the bridge of Figure 15, the measured voltage is equal to

$$V_m = \left[\frac{R_1}{R_1+R_2} - \frac{Z_1}{Z_1+Z_2}\right] \times E$$

$$= \left[\frac{R_1}{R_1+R_2} - \frac{\dfrac{1}{j\omega C_1}}{\dfrac{1}{j\omega C_1}+\dfrac{1}{j\omega C_2}}\right] \times E$$

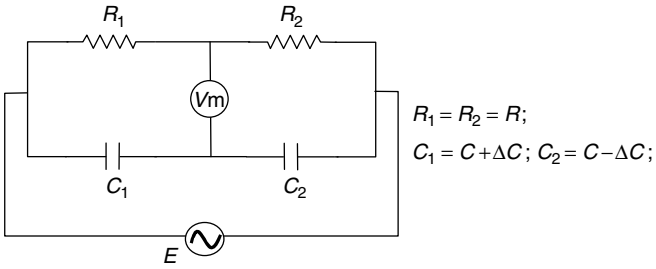


Figure 15. Example of bridge circuit for differential measurements.

$$= \left[\frac{R}{R+R} - \frac{j\omega C_2}{j\omega C_1 + j\omega C_2}\right] \times E$$

$$= \left[\frac{1}{2} - \frac{C+\Delta C}{2C}\right] \times E = \frac{\Delta C}{2C} \times E \qquad (10)$$

Consequently, the relationship between the measured voltage and the capacitance variation is linear. Moreover, all unwanted changes that simultaneously affect both sensors (e.g. change owing to temperature) are compensated thanks to the chosen mounting circuit. Similar circuits can be used for inductance sensors.

8 EXCITATION SOURCES

The methods used for capacitance and inductance measurement, both utilized in analog circuit and digital instruments, require sinusoidal (current/voltage, bridges, resonance) or step (time constant, reflection) excitation signal.

8.1 Sinusoidal test signal

Since the impedance value depends extrinsically and intrinsically on the test signal frequency, the first two requirements for the sinusoidal signal are the stability and accuracy of the selected frequency and the spectrum purity.

Quartz oscillators are normally used. Their oscillation frequency is stable but cannot be varied. Consequently, direct or indirect methods (see Figure 16) are used to implement a high precision frequency variation. Direct synthesis is accomplished by multiplying, dividing, and mixing the oscillator output; the indirect method derives its frequency from one or more phase locked voltage (PLL) tuned oscillators. The stability in both cases is derived from the reference source, but the indirect method allows a wider range of frequency and better accuracy to be obtained.

As for the signal purity, there are three principal contributors that degrade the ideal behavior (the signal spectrum constituted by a single line): harmonic distortion,

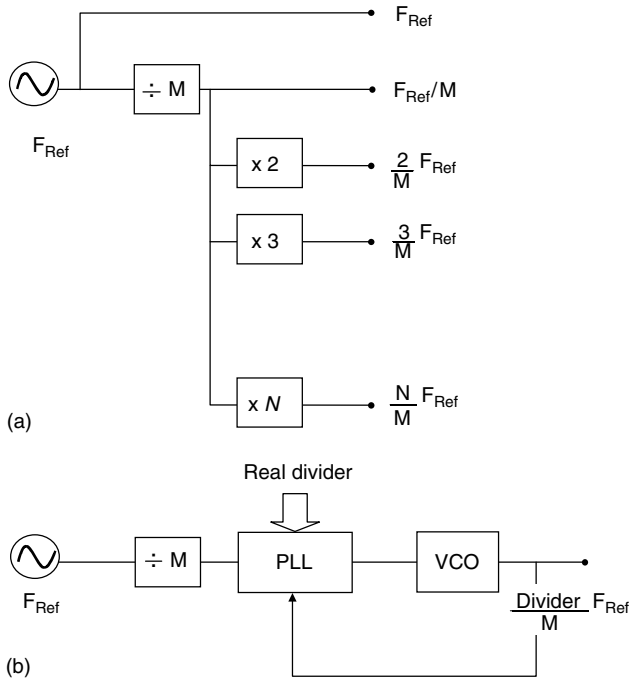


Figure 16. Oscillator: (a) direct synthesis and (b) indirect synthesis.

spurs, and phase noise. A 'harmonic' is a tone whose frequency is a multiple of the sine wave frequency: if harmonics are present, the resulting waveform is not sinusoidal. The harmonic distortion quantifies the harmonic presence (Figure 17(a)). There are several sources of spurs (nonharmonic tones). Examples are unwanted coupling among circuits, distortion produced in the mixer, and so on (Figure 17(b)). The phase noise refers to the sidebands present in the signal spectrum (Figure 17(c)).

A further requirement, for the signal source, is the possibility of adjusting its output power level. In fact, the signal level should be set as high as possible to obtain a good S/N ratio, since a high S/N ratio improves the accuracy and stability of the measurement; however, in some cases, the level should be decreased, such as when measuring cored-inductors. In addition, occasionally, to the AC test signal a DC voltage has to be added.

Generally, a frequency synthesizer is used to meet frequency accuracy, resolution, and stability needs. The amplitude of signal source output is adjusted for the desired test level by an output attenuator.

8.2 Step test signal

The main requirement for voltage step signal is a very low rise-time. In fact, the faster is the step signal, the finer will be the instrument resolution. Furthermore, a control on the output power and on the DC component allows more accurate measurement to be performed.

For more detail on signal generators, see books such as Doebelin (2003) and Coombs (1995).

9 DIGITAL INSTRUMENTS

There are a variety of digital instruments for impedance measurement, which provide the convenience of a direct digital readout and adequate accuracy (Agilent Technologies Inc, 2000; Bartlett, 1992). Moreover, almost all are able to evaluate the complex impedance (real and imaginary part) of the component. In addition, thanks to the presence of a processing unit, they can directly display the complex impedance, its module and phase, the dissipation factor, and the value of the series or parallel equivalent circuit components.

In the following section, the most used digital instruments, grouped on the basis of the measurement principle, are summarized.

9.1 Current/voltage methods

There are a number of instruments (e.g. LCR meter, impedance analyzer) based on the current/voltage method. All foresee the simultaneous sampling of both signals many times per cycle; but they differ on the technique used for the evaluation of the attenuation and the phase shift between current and voltage. Technique for the evaluation of the attenuation and the phase shift between current and voltage include sine correlation, phase sensitive detectors

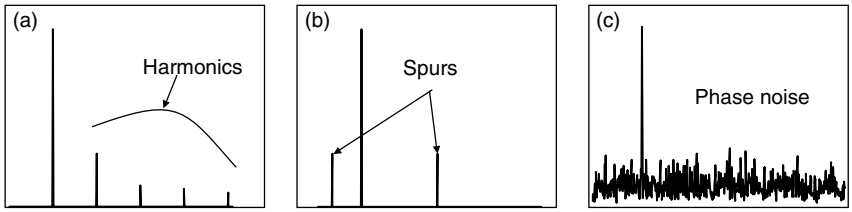


Figure 17. Sinusoidal signal impurities.

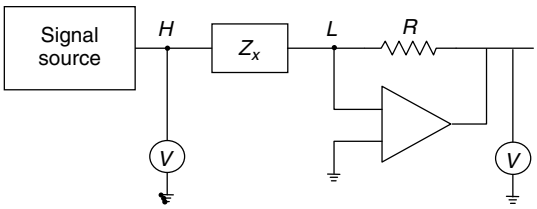


Figure 18. Block diagram of an autobalance bridge.

(lock-in amplifiers), fast Fourier transform FFT techniques, Lissajous figures, and so on.

The so-called auto balancing bridge method is commonly used in modern low frequency, usually within few hundreds of MHz, impedance measurement instruments. In Figure 18, a block diagram of a digital bridge is reported.

The current, flowing through the unknown component, Z_x, also flows through the resistor R. The potential at the L point is maintained at zero volts (thus called a *virtual ground*), because the current through R balances with the Z_x current by operation of the I-V converter amplifier. The unknown impedance is calculated using voltage measurement at the high terminal and that across R.

In practice, the configuration of the auto balancing bridge differs for each type of instrument. Generally, LCR meters in a low frequency range, typically below 100 kHz, employ a simple operational amplifier for its I-V converter. This type of instrument has a disadvantage in accuracy, at high frequencies, because of performance limits of the amplifier. Wideband LCR meters and impedance analyzers employ an I-V converter consisting of sophisticated null detector, phase detector, integrator (loop filter), and vector modulator to ensure a high accuracy for a broad frequency range over 1 MHz. This type of instrument can attain a maximum frequency of 110 MHz. *See also* **Article 202, Current Measurement, Volume 3** and **Article 201, Voltage Measurement, Volume 3**.

9.2 Time constant methods

A capacitance meter essentially measures the charge or discharge time of the capacitor under test. The capacitor is allowed to charge up to an accurate reference voltage, and a gate pulse is started when the capacitor begins to discharge to a lower accurate voltage level through an accurate known resistance. When the capacitor discharges to a lower accurate level, the gate signal is closed. The time interval of the gate voltage is a measure of the RC time constant of the unknown capacitor and the precision resistor. A clock frequency and counter circuit measure the time constant.

Many digital capacitor testers also have the ability to measure inductances. One of the digital L–C meters is unique in that inductances are measured by ringing the current wave shape. The number of cycles an inductor resistor circuit produces after being used is a measure of the inductance.

9.3 Reflection methods

The network analysis, based on reflection methods, is the most used solution at high frequency. A network analyzer can be divided into four main parts: a signal source; a signal separation device to separate the incident, reflected, and transmitted signals; a receiver to convert the microwave signal to a lower intermediate signal; and a signal processor unit to determine the impedance value starting from the measured signals.

10 STANDARDS

The term standard associated with a passive component refers to an object built to reproduce a physical quantity or parameter very accurately.

10.1 Capacitors

These capacitors are very stable, have very low dissipation factors, small temperature coefficients, and very little aging effect Eren (1999).

Standard capacitors can also be fixed or variable.

Fixed capacitors are constructed from interleaved metal plates. The area of the plates and the distance between them are determined and constructed with precision. National metrology institutes maintain a bank of primary standard air capacitors that can be used to calibrate secondary and working standards for laboratories and industry. Generally, smaller capacitance working standards are obtained from air capacitors, whereas larger working standards are made from solid dielectric materials (mica, polymeric, paper, and ceramic).

Variable capacitors are generally made of air or ceramic dielectric materials and consist of two assemblies of spaced plates positioned together by insulation members such that one set of plates can be rotated; usually they are available in decade units.

10.2 Inductors

There are two groups of inductance standards: namely, standards of self-inductance and those of mutual inductance; both types can be fixed or variable Krupka and Geyer (1999).

Standard inductors are made without iron cores and use glass, porcelain, marble, and wood impregnated with paraffin for the coil materials. They are nonmagnetic and provide good insulation and adequate stability. The basic coil forms are flat cylindrical coil of square winding cross-section and toroidal configuration.

Fixed mutual-inductance standards are similar to the self-inductance standards except that they consist of two windings instead of one.

Variable standards of single or mutual inductance are obtained by connecting in series some self-inductors or sections of the secondary winding of a mutual inductor. Furthermore, variometers/inductometers are realized to allow a continuous adjustment of the inductance.

RELATED ARTICLES

Article 73, Overview of Electrical and Electronic Technique, Volume 2; **Article 79, Testing Electronic Systems, Volume 2**; **Article 201, Voltage Measurement, Volume 3**; **Article 202, Current Measurement, Volume 3**; **Article 203, Resistance Measurement, Volume 3**.

REFERENCES

Agilent Technologies Inc. (2000) *The Impedance Measurement Handbook – A Guide to Measurement Technology and Techniques*.

Bartlett, E.R. (1992) *Electric Measurements & Testing: Tips and Techniques for Technicians and Engineers*, McGraw-Hill, New York.

Coombs, C.F. (1995) *Electronic Instrument Handbook*, 2nd edn, McGraw-Hill, New York.

Doebelin, E.O. (2003) *Measurement Systems*, 5th edn, McGraw-Hill, Singapore, New York.

Eren, H. (1999) Capacitance Measurement, in *Wiley Encyclopedia of Electrical and Electronics Engineering*, Vol. 3 (ed. J.G. Webster), John G., New York (pp. 1–15).

Eren, H. and Goh, J. (1999) Capacitance and Capacitance Measurements, in *The Measurement, Instrumentation, and Sensors Handbook* (ed. J.G. Webster), John G., New York (pp. 45.1–45.27).

Krupka, J. and Geyer G. (1999) Loss Angle Measurement, in *Wiley Encyclopedia of Electrical and Electronics Engineering*, Vol. 11 (ed. J.G. Webster), John G., New York (pp. 606–619).

Kumar V., Jagadeesh Kumar, V., Murti, V.G.K. and Sankaran, P. (1999) Inductance Measurement, in *Wiley Encyclopedia of Electrical and Electronics Engineering*, Vol. 9 (ed. J.G. Webster), John G., New York (pp. 722–733).

Morrison, A.S. (1993) *Principles of Measurement and Instrumentation*, Prentice Hall, New York.

Osvàth, P. (1992) Measurement of Impedances, in *Technology of Electrical Measurements* (ed. L. Schnell), USA (pp. 147–224).

Webster, J.G. and Pallas-Areny, R. (1991) *Sensors and Signal Conditioning*, John Wiley & Sons, New York.

205: Theory of Vibration Measurement

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Definition of Vibration 1391
2 Physical Features of Vibration 1391
3 Mathematical Description 1394
4 Amplitude and Frequency of Vibration 1394
5 Importance of Damping 1395
6 Application to Vibration Measurement 1396
7 Data Processing 1396
8 The Literature on Vibration 1397
Related Articles 1397
References 1397

1 DEFINITION OF VIBRATION

Vibration here is the oscillatory mechanical motion of an object. It is a dynamic state. Several other measurable parameters fit under the term vibration.

- Relative position
- Velocity
- Acceleration
- Jerk (the derivative of acceleration)
- Dynamic force.

For each parameter it may be the instantaneous value, average value, or some other descriptor that is needed.

Vibration, in the general sense, occurs as a periodic oscillation, a random movement, or as a transient motion, the latter more normally being referred to as *shock* when the transient is large in amplitude and brief in duration.

2 PHYSICAL FEATURES OF VIBRATION

Vibration can occur in linear or rotational forms of motion, the two being termed *translational* or *torsional vibrations* respectively. In many ways, the basic understanding of each is similar because a rotational system also concerns displacements in space. Translational forms are outlined in the following description. An equivalent rotational system usually exists for all arrangements described.

In vibration measurement, it is important to decide whether a mechanical sensor can be physically attached to the test object – called a *contacting* or *intrusive method*.

If mass cannot be added, then a noncontacting technique is necessary.

Detailed measurement of vibration can be a most complex measuring task. The basic requirement is to determine features of the motion of a point, or more practically an extended object in a three-dimensional space relative to a reference framework; see Figure 1.

A point in space has three degrees of freedom. It can translate in one or more of the three directions when referred to the x, y, z Cartesian coordinate system shown. Rotation has no meaning for a purely point measurement. To monitor free motion of a theoretical point object three measurement-sensing channels are required to capture all three motions so that the vector of direction can be reconstructed.

If the object of interest has significant physical size, it must be treated as an extended object in which the rotations about each of the three axes described above provide a further three rotational degrees of freedom. Thus, to monitor the free motion of a realistic object we may need up to six sensor channels, one for each degree of freedom. These may be built into a single sensor housing and processed to give absolute values.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

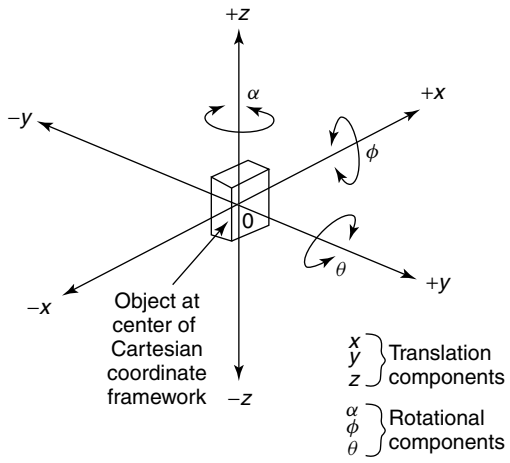


Figure 1. Possible motions of an extended object in space relative to a Cartesian framework.

In practice, some degrees of freedom can be nominally constrained (but are they really?) possibly eliminating the need for some of the six sensor channels. Practical installation should always contain a test that evaluates the degree of actual constraint possible because vibration sensors will usually produce some level of output for the directions of vibration they are not primarily seeking to measure. This is called their *cross-axis coupling*, transverse response, or some such terminology.

In many installations, the resultant direction of the motion vector might lie in a constant fixed direction with time. In such cases, in principle, only one sensor will be required provided it can be mounted to sense in exactly that direction. If not, as is often the case, more than one unit will be required, the collective signals then being combined to produce the single resultant.

The potentially available frequency spectrum of vibration parameters extends, as shown in Figure 2, from very slow motions through frequencies experienced in machine tools and similar mechanical structures to the supersonic megahertz frequencies of ultrasound.

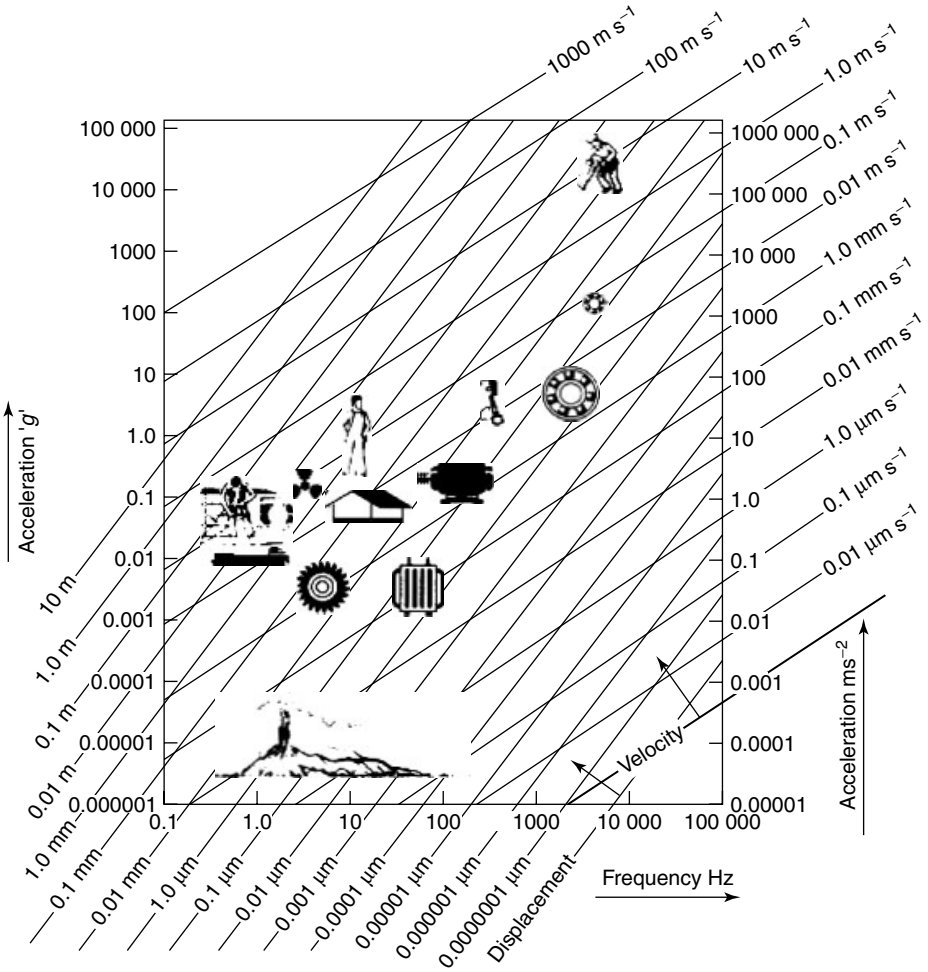


Figure 2. Frequency spectrum and magnitude of vibration parameters. (Reproduced by permission of Bruel & Kjaer.)

It is not possible to cover this range with a single general-purpose sensor. Each application will need careful consideration of its many parameters to decide which kind and model of sensor should be applied in order to make the needed measurement.

The complicating factor in vibration measurement is the distributed nature of mechanical systems that enables the energy causing a vibration to flow from a point. This leads to complex patterns of vibration, requiring care in the positioning of sensors.

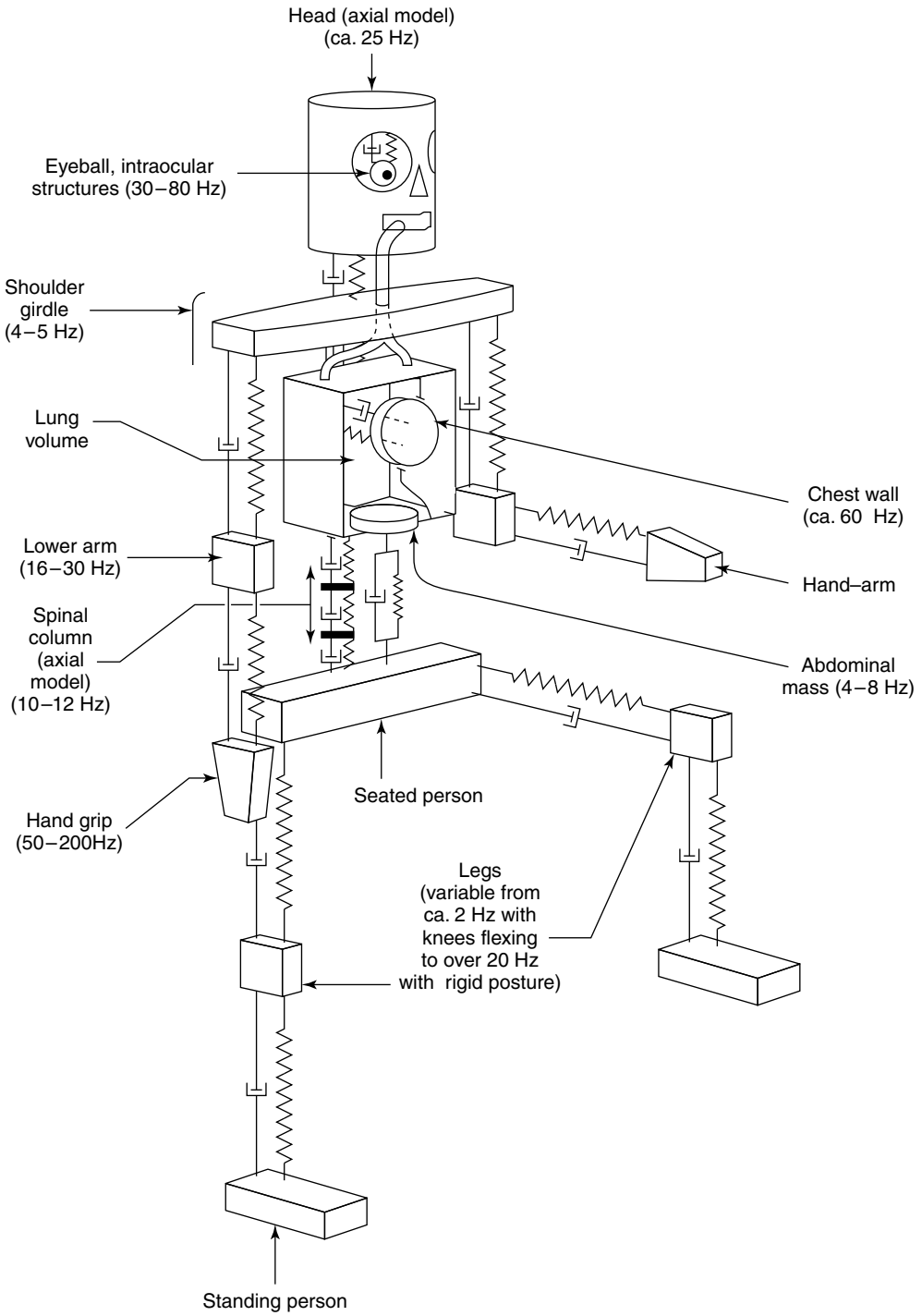


Figure 3. Mechanical systems can be modeled in terms of springs, masses, and dampers; model of the human body. (Reproduced by permission of Bruel & Kjaer.)

3 MATHEMATICAL DESCRIPTION

Mechanical systems, including the human body given as an example in Figure 3, comprise mass, spring compliance (or stiffness), and damping components.

In the simplest case, where only one degree of freedom exists, linear behavior of this combination can be well described using linear mathematical theory to model the time behavior as the result of continuous force excitation, or some initial position displacement – *see* **Article 60, Zero-order System Dynamics, Volume 1**; **Article 61, First-order System Dynamics, Volume 1**; **Article 62, Second-order System Dynamics, Volume 1** for a general analysis of linear motion.

Vibration can be measured by direct comparison of instantaneous dimensional parameters relative to some adequately fixed datum point in space. The fixed point can be on an independent measurement framework (fixed reference method) or can be a place that remains sufficiently stationary because of its high mechanical inertia (seismic system) that holds it relatively constant in position.

The accelerometer is the main vibration sensor – *see* **Article 207, Acceleration Measurement, Volume 3**. It is formed with a mass, and a spring having some damping. If operating in its linear range, it will exhibit a second-order linear system output response, q_o, that is related to an input function, q_i, by the generalized differential equation

$$\frac{a_2 \mathrm{d}^2 q_o}{\mathrm{d}t^2} + \frac{a_1 \mathrm{d}q_o}{\mathrm{d}t} + a_o q_o = q_i \tag{1}$$

For the specific mechanical vibratory system given in Figure 4, this becomes

$$\frac{m \mathrm{d}^2 x_0}{\mathrm{d}t^2} + \frac{c \mathrm{d}x_0}{\mathrm{d}t} - k_s x_0 = q_i \tag{2}$$

where m is the effective mass (which may need to include part of the mass of the spring element or be composed entirely of it), c the viscous damping factor, and k_s the spring compliance (expressed here as length change per unit of force applied).

In practice, these three components may be distributed to a varying extent across an object. Often, they can be adequately considered using the lumped model shown. Where the degree of distribution is significant, then analysis becomes much more complex and is studied using such computer-based methods as the Finite Element Method (FEM).

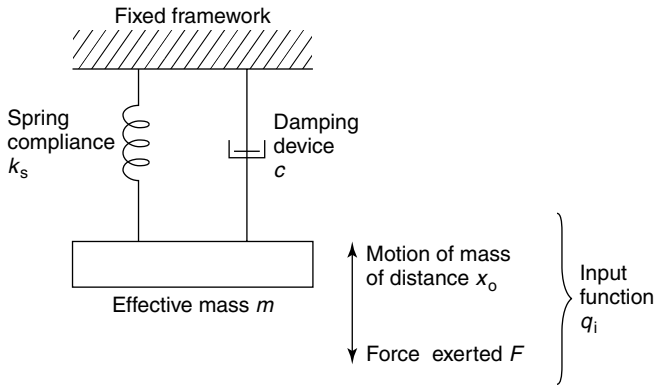


Figure 4. One degree of freedom, spring-mass-damper system model.

4 AMPLITUDE AND FREQUENCY OF VIBRATION

Where the damping effect is negligible, the system will have a frequency at which it will naturally vibrate if excited by a pulse input. This natural frequency ω_n is given by

$$\omega_n = \sqrt{\frac{k_s}{m}} \tag{3}$$

Presence of damping will alter this vibration frequency value, but as the damping rises the system is less able to provide continuous oscillation.

The static sensitivity is given by the spring constant, either as k_s, the spring compliance, or as its reciprocal term stiffness that is expressed as force per unit extension.

The influence of damping is readily described by a dimensionless number, called the *damping ratio*, given by

$$\xi = \frac{c}{2\sqrt{k_s \times m}} \tag{4}$$

It is usually quoted in a ratio form that relates its magnitude with respect to that at $\xi = 1$.

These three important parameters are features of a spring system as a whole formed from the object under test, plus any contributions from an added sensor, any protective cover, and its wiring. The parameters are independent of the input driving function.

Such systems have been extensively analyzed when excited by the commonly seen input forcing functions, the step, impulse, ramp, and sinusoid. A general theory for handling any input function other than these is also available.

In practice the step, impulse, and continuous sinusoidal responses are used in analyses, as they are reasonably easy to apply in theory and in practical use.

5 IMPORTANCE OF DAMPING

From the plots of different damping ratios for a step input given in Figure 5, it can be seen that as the damping factor ξ increases the response to a transient step force input (applied to the mass) can vary from a sinusoidal oscillation at one extreme (called *underdamped*) to a very sluggish climb to the final value (*overdamped*). The nominal value of $\xi = 0.7$ is called the *critical damping* state as this is the fastest rise to its final value.

In the case of a continuous sinusoidal force being applied to the input, the system frequency response varies as shown in Figure 6.

Note the resonance buildup point at ω_n, and that it is limited by the existing degree of damping. Thus, the damping of the system to be measured by addition of the sensor itself, if it is of the seismic kind, can be of importance as a modifier of likely system responses.

A well-known measurement error situation is where an oscillatory response is being plotted with a mechanical chart recorder. If the response of the chart drive mechanism (itself a second-order system) is not adequate, the record is attenuated as the oscillation frequency of the measured input rises. The same effect can arise with poorly set up vibration-measuring systems.

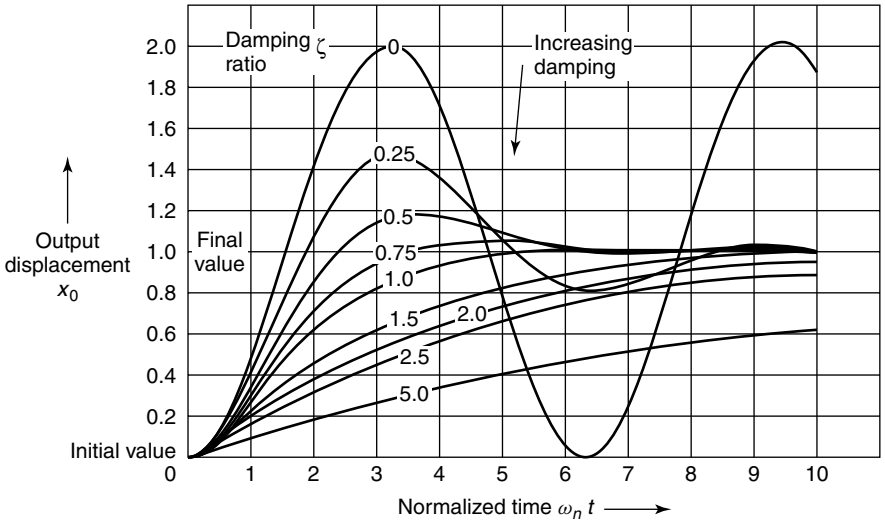


Figure 5. Displacement responses of a second-order system to an input force step.

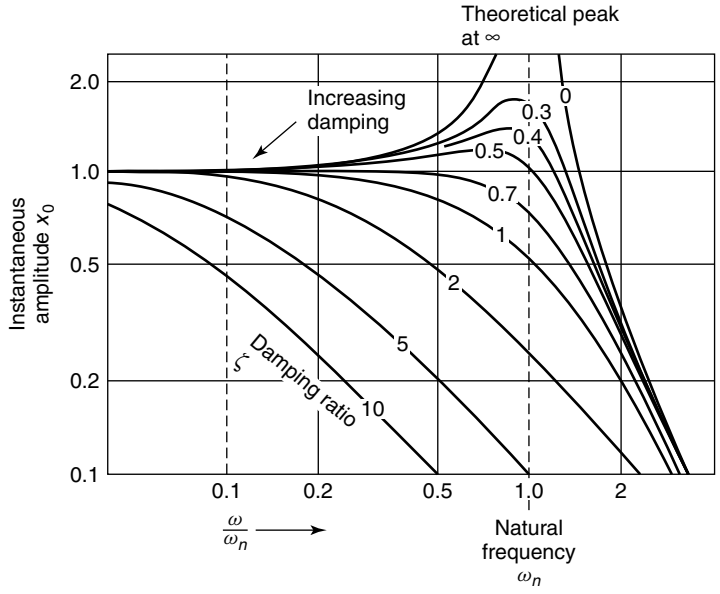


Figure 6. Displacement responses of a second-order system to a continuous sinusoidal force input.

As damping increases, the system response takes on the form of the lower first-order exponential response system, and it cannot oscillate.

The above discussion, with respect to vibration of the measurand is also the basis of understanding the operation of seismic vibration sensors – *see* **Article 208, Amplitude and Velocity Measurement, Volume 3**.

Second-order measuring systems, therefore, will have a natural frequency of vibration. This is the frequency at which they vibrate when given impulse energy that is not overridden by continuous forced vibrations. Thus, a sensing system that is second order and not damped will (because of noise energy inputs) produce outputs at its natural frequency that are not correlated with frequencies occurring in the system of interest. Use of vibration sensors must, therefore, recognize these limitations by selecting designs that position their resonant frequency correctly, and that possess the right degree of damping. Such issues are covered in **Article 206, Practice of Vibration Measurement, Volume 3**.

6 APPLICATION TO VIBRATION MEASUREMENT

In practice, it is also often more convenient to sense vibration by an indirect means and obtain the desired unit by mathematical processing. For example, accelerometers are conveniently used to obtain motional forces (from force = mass × acceleration) and hence stresses and strains. Acceleration signals can be twice integrated with respect to time to yield displacement. Sensors that operate as velocity transducers can yield displacement by single integration.

Integration of data is generally preferred to differentiation as the former averages random noise present to a smaller value compared to the signal, whereas the latter, in reverse, will deteriorate the signal-to-noise ratio.

7 DATA PROCESSING

Signals obtained in vibration measurements will take many forms depending on the vibration characteristics of the object under test. These will include

- repeating pulses,
- sine waves,
- complex waveforms,
- noise.

Their nature can be

- continuous, having the same signal waveform,
- varying waveforms with time,
- modulated amplitudes,
- modulated frequencies,
- short bursts,
- random.

The kinds of information to be extracted from the raw signal can

- be frequency and magnitude of an oscillation;
- be identification of the source of pulses and unwanted noises;
- impact behavior as objects collide or are deliberately impacted;
- be the frequency spectrum of a signal that establishes relative amplitudes and energy distributions; in vibration, spectral analysis is also taken on a spectrum;
- be a prediction of the future condition of equipment.

These signals are used to establish such things as

- wear in bearings,
- noise inside a vehicle,
- dangerous vibrations that can cause mechanical failure,
- response of loud speakers,
- where to attach balancing weights,
- machine health monitoring,
- earthquake research,
- building vibration in wind,
- where to apply noise dampening materials, and so on.

Thus, the signal processing of vibration signals often has to use the full power of signal processing in both the analog and digital domain.

The methods used for general analog and digital signal processing, in general, are applicable – *see* **Article 142, Z-transforms, Volume 3**; **Article 143, DFT and FFTs, Volume 3**; **Article 144, DSP Chip Sets, Volume 3**; and **Article 145, DSP Tools, Volume 3**.

Texts devoted to vibration measurement cover the special methods used, Harris and Piersol (2001), McConnell (1995), Thomsen (2003), and Wowk (1991).

Today, as much as possible this signal processing is carried out in the Digital Signal Processing, DSP domain. These methods are very powerful and make use of dedicated DSP chips and processors. Tool kits are commonly available and have become so widespread that many of the analysis tools are now routinely built into equipment, test sets, and into the vibration sensor system itself.

The sensor, however, is usually analog by nature so the extraction of signals in high noise environments is often essential – *see* **Article 176, Signals and Signal-to-noise Ratio, Volume 3**; **Article 178, Noise Matching and Preamplifier Selection, Volume 3**; **Article 180, Bandwidth Reduction of Baseband DC Signals, Volume 3**;

Article 181, Amplitude Modulated Signals: The Lock-in Amplifier, Volume 3; **Article 182, Boxcar and Signal Averagers, Volume 3**; **Article 183, Correlators in Signal Extraction, Volume 3**; and **Article 184, Photon Counting, Volume 3**.

8 THE LITERATURE ON VIBRATION

There exist many general books on the kinds of transducers that are used to measure the various vibration variables. Some of the titles published have become standard sources carrying forward considerable unchanging knowledge, Bruel and Kjaer (1982), Trampe-Broch (1980), and Harris and Piersol (2001).

Whereas the data processing has changed greatly over the past decades, the sensing principles and nature of vibration remain the same, making older works valid for many aspects of vibration.

Few general instrument reference texts, such as Webster (1999), Dyer (2001), address the subject of vibration as a distinct topic. Relevant material will be found there under such headings as velocity and acceleration measurement, accelerometers, position sensing, and piezoelectric systems.

There are, as would be expected, works entirely devoted to vibration and related measurands. The following will be of value to those who require more than the limited introduction that an article such as this can provide, Crandall (1959), Harris and Piersol (2001), Ling and Shabana (1996), McConnell (1995), Mobley (1999), Smith (1989), Steinberg (2000), Thomsen (2003), Timar (1989), and Trampe-Broch (1980).

The various trade houses that manufacture vibration-measuring and testing equipment often provide extensive literature and other forms of training aids to assist the uncertain user. Their application engineers also offer assistance in vibration system design and set up.

RELATED ARTICLES

Article 180, Bandwidth Reduction of Baseband DC Signals, Volume 3; **Article 206, Practice of Vibration Measurement, Volume 3**; **Article 207, Acceleration Measurement, Volume 3**; **Article 208, Amplitude and Velocity Measurement, Volume 3**.

REFERENCES

Bruel & Kjaer. (1982) *Measuring Vibration-Elementary Introduction*, Bruel & Kjaer, Naerum.

Crandall, S.H. (1959) *Random Vibration*, Wiley, New York.

Dyer, S.A. (2001) *Survey of Instrumentation and Measurement*, Wiley, New York.

Harris, C.M. and Piersol, A.G. (2001) *Harris' Shock and Vibration Handbook*, McGraw-Hill, New York.

Ling, F.F. and Shabana, A.A. (1996) *Theory of Vibration: An Introduction*, Springer Verlag, Berlin.

McConnell, K.G. (1995) *Vibration Testing: Theory and Practice*, Wiley Interscience, New York.

Mobley, R.K. (1999) *Vibration Fundamentals*, Butterworth-Heinemann, Oxford.

Smith, J.D. (1989) *Vibration Measurement and Analysis*, Butterworth-Heinemann, Oxford.

Steinberg, D.S. (2000) *Vibration Analysis for Electronic Equipment*, Wiley Interscience, New York.

Thomsen, J.J. (2003) *Vibrations and Stability: Advanced Theory, Analysis, and Tools*, Springer Verlag, Berlin.

Timar, P.L. (1989) *Noise and Vibration of Electrical Machines*, North-Holland, Amsterdam.

Trampe-Broch, J. (1980) *Mechanical Vibration and Shock Measurements*, Bruel & Kjaer, Naerum.

Webster, J.G. (1999) *The Measurement, Instrumentation, and Sensors Handbook*, CRC Press, Boca Raton, FL.

Wowk, V. (1991) *Machinery Vibration*, McGraw-Hill, New York.

206: Practice of Vibration Measurement

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Mass-spring Sensors 1398
2 Areas of Application 1402
3 Cross Coupling, Cabling, and Amplifiers 1402
4 Influence Effects on the Sensor 1404
5 Loading by the Sensor 1404
6 Duration of use of Sensor in Tests 1404
7 Amplitude Calibration 1405
8 Accelerometer Calibration 1406
9 Shock Calibration 1406
10 Force Calibration 1406
Related Articles 1406
References 1406

1 MASS-SPRING SENSORS

Many vibration measurements make use of the mass-spring, seismic, sensor system. The seismic sensor is mostly described in the literature as the accelerometer, but this use is not the general case.

Other methods of vibration measurement used include inductive, capacitive, and optical methods that are primarily noncontact displacement measuring methods having a fast response and small range capability – *see* **Article 191, Displacement and Angle Sensors Performance and Selection, Volume 3**. The strain gauge, a contact method of displacement measurement, is also able to deliver vibration data; this is discussed in **Article 192, Strain Sensors, Volume 3**.

Given the appropriate design of spring-mass-damping combination a seismic sensor attached to a vibrating surface can yield displacement, velocity, or acceleration data. The principles and theory of this kind of sensor are covered in **Article 205, Theory of Vibration Measurement, Volume 3**. Unfortunately, the conflicting response needs of the three do not enable a single design to be used for all three cases. It is, however, often possible to derive one variable from another by mathematical integration or differentiation of the data.

The design of vibration sensors can be most sophisticated. Fortunately, the user of vibration sensors rarely becomes involved in its design but must appreciate the design features in order to avoid incorrect use. Minor digressions from best practice can lead to metrological error that is not easy to detect.

Two forms of seismic sensor exist. The first, called the *open-loop sensor* makes use of the motion of a mass moving relative to the support case to operate either a displacement or a velocity sensing transducer. These possess some undesirable properties that often rule them out of practical use.

The second form closes the loop (and is, therefore, referred to as a *closed-loop* or *servo seismic sensor*) using the output signal to produce an internal force that retains the mass in the same relative position with respect to the case, the magnitude of the force needed to effect balance being a measure of the vibration parameter. As will be explained, this method while being more sophisticated is much more effective.

1.1 Open-loop sensors

The fundamental schematic arrangement of the open-loop seismic sensor form is as given in Figure 1. There the mass, spring, damper, and sensing element are clearly identified.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

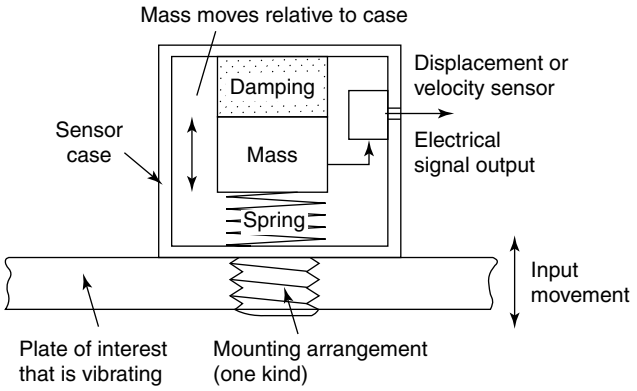


Figure 1. Schematic layout of open-loop, seismic form of vibration sensor.

In a real design, the case and the mass are also part of the system's stiffness and they have to be made adequately stiff. Actual construction can vary widely depending upon how the spring force and damping are provided and upon the form of the sensor used.

The spring element is sometimes produced as a distinct mechanical assembly. Figure 2 is an example made using a flexure-strip spring suspension. Alternatively, perforated membranes, helical coils, torsion strips, disks, helical springs, solid compliable pieces, and the like can be used, *see* **Article 85, Elastic Regime of Design – Design Principles, Volume 2**.

Often the compliance of the mass itself may be the spring element, for example, in the piezoelectric crystal, which also acts as the sensing element and contributes some of the damping needed.

Important design parameters of a spring element used for measurement are its

- compliance
- amplitude range
- fatigue life
- constancy of spring rate with time
- temperature
- influence effects
- suitability to be manufactured and packaged to produce a suitable sensor unit.

All masses have compliance to some extent. This has to be allowed for in the support packaging design or else the case can modify the measurement amplitudes obtained. Except for sensors with the higher natural frequencies, the masses used to form the support case can be regarded as completely rigid compared to the compliance of the spring element.

Design of these sensors uses the best of fine mechanical element design techniques as are covered in **Article 80, Principles of Fine Mechanics – Kinematic and Elastic Designs, Volume 2**; **Article 81, Principles of Fine Mechanics – Systems Considerations, Volume 2**; **Article 82, Kinematical Regime – Members and Linkages, Volume 2**; **Article 83, Kinematical Regime – Fasteners, Bearings, Volume 2**; **Article 84, Kinematical Regime – Rotary Motion, Volume 2**; **Article 85, Elastic Regime of Design – Design Principles, Volume 2**; **Article 86, Elastic Regime of Design – Spring Systems, Volume 2**; **Article 87, Elastic Regime of Design – Plates and Bimorphs, Volume 2**; and **Article 88, Error Sources in Fine Mechanics, Volume 2**.

Rotary equivalent forms of the linear arrangement shown in Figure 1 are also available.

Sensing methods that have been used in vibration sensors include the electrical-resistance sliding potentiometer, variable inductance (see, for instance, Figure 2), variable reluctance, variable capacitance, metallic strain gauges (metallic bonded and also unbonded use as shown in Figure 3), semiconductor strain gauges, piezoelectric crystal and magnetostrictive elements, position-sensitive optical detectors,

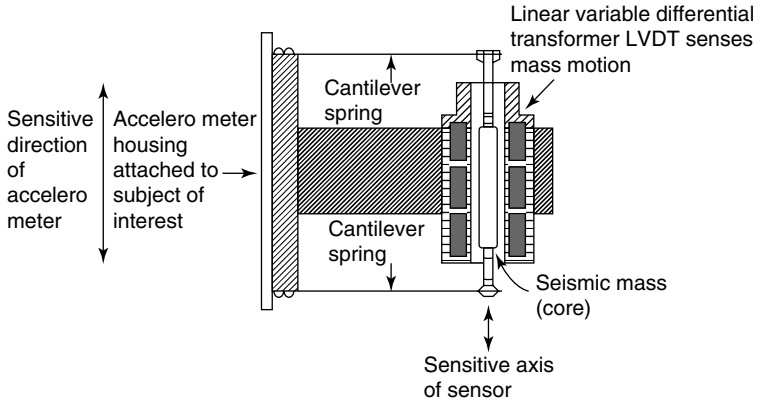


Figure 2. Spring vibration sensor using a parallel flexure-strip spring suspension and inductive sensor of mass displacement.

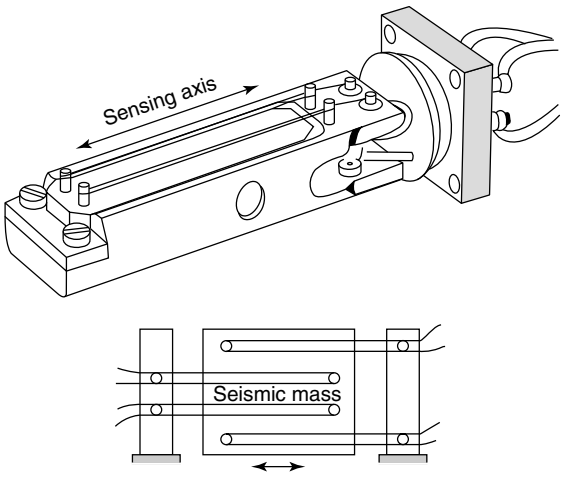


Figure 3. Displacement sensing of mass motion using unbonded strain gauges.

PSD, and electromagnetic principles that can provide direct velocity sensing.

Sensors of the mass-spring form are usually encapsulated and set up to be used as an accelerometer. The encapsulation takes many forms, ranging from miniature units of total weight around just 1 g through to 0.5 kg units where the sensing mass must be physically large to get the response down to low frequencies. Simultaneous measurements in two, or three, orthogonal directions are often required. To meet this need, seismic sensors are made that consist of two, or three, units that are mounted in different directions, as shown in Figure 4.

Temperature invariably affects the metrological performance of sensors and here also it changes, to a varying degree, the characteristics of the displacement sensing method, the spring rate, the damping coefficient, and

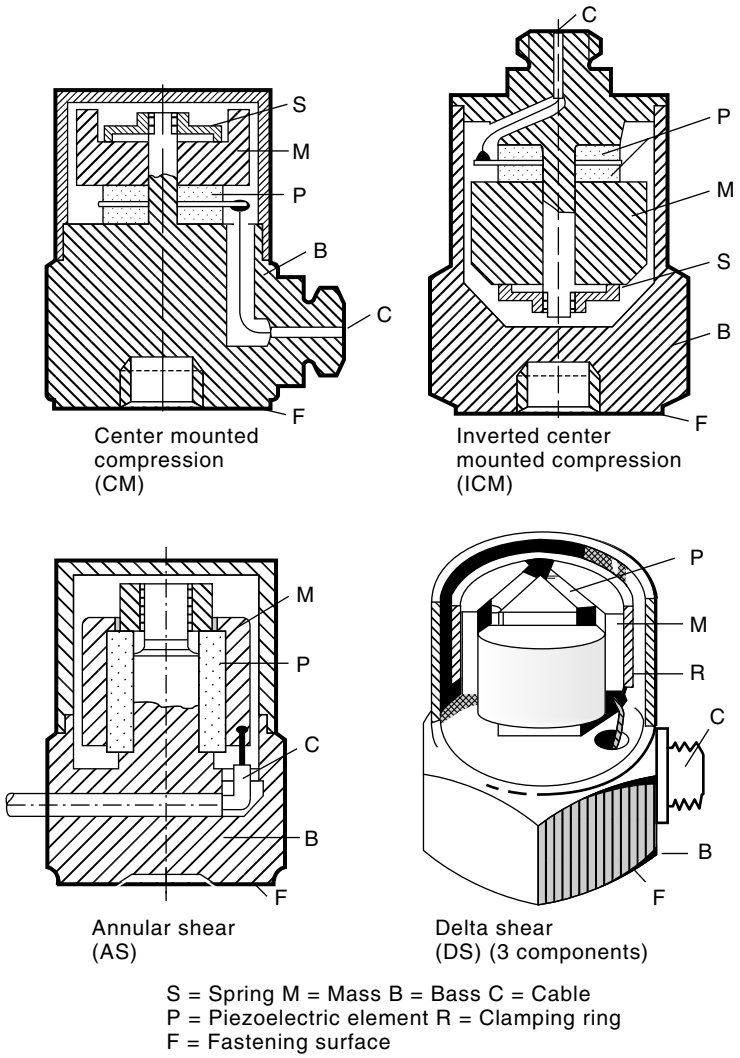


Figure 4. Examples of small size single- and three-axis accelerometers using piezoelectric displacement sensing. (Reproduced by permission of Bruel & Kjaer.)

the case dimensions. Effective design – *see* **Article 85, Elastic Regime of Design – Design Principles, Volume 2** and **Article 88, Error Sources in Fine Mechanics, Volume 2** – can reduce these effects considerably, but to get the best from a design of high performance vibration sensors can require compensation for temperature change.

This is either performed in the electronic circuitry as a correction to the data using a temperature sensor placed accordingly or by incorporating some form of thermomechanical device into the spring-layout to reduce the error at source.

1.2 Servo vibration sensors

The performance of open-loop seismic sensors can be improved with respect to their sensitivity, accuracy, bandwidth, and output signal amplitude by forming the design into a closed-loop operation.

Figure 5 gives the schematic diagram of a simple form of closed-loop system, which is based upon a moving-coil actuator and a capacitance position sensor. This particular kind is described here, as the elements of the closed-loop system are clearly evident.

The inertial mass upon which the acceleration is to be exerted is able to rotate on the end of a freely supported arm. This is attached to the electrical coil placed in the permanent magnetic field supplied by the magnet assembly. Acceleration applied to the mass attempts to rotate the arm thereby causing displacement of the arm. This unbalances the capacitive displacement sensor monitoring the relative position of the pendulous mass. The displacement signal produces an input to the difference-sensing amplifier. The amplifier drives a corresponding electric current into the coil, causing the arm to rotate back to the null displacement position. Provided the loop response is rapid enough, the mass will be retained in a nearly constant place relative to the displacement sensor. Acceleration variations are, thereby, converted to variations in coil current.

In this way, the displacement sensor is used in the preferred null-balance mode – *see* **Article 126, Electrical Bridge Circuits – Basic Information, Volume 2** and **Article 127, Unbalanced DC Bridges, Volume 2** – wherein error of linearity and temperature shift are much reduced. In this design, only the more easily achieved proportionality between coil current and force is important. Servo instruments are further described in **Article 113,**

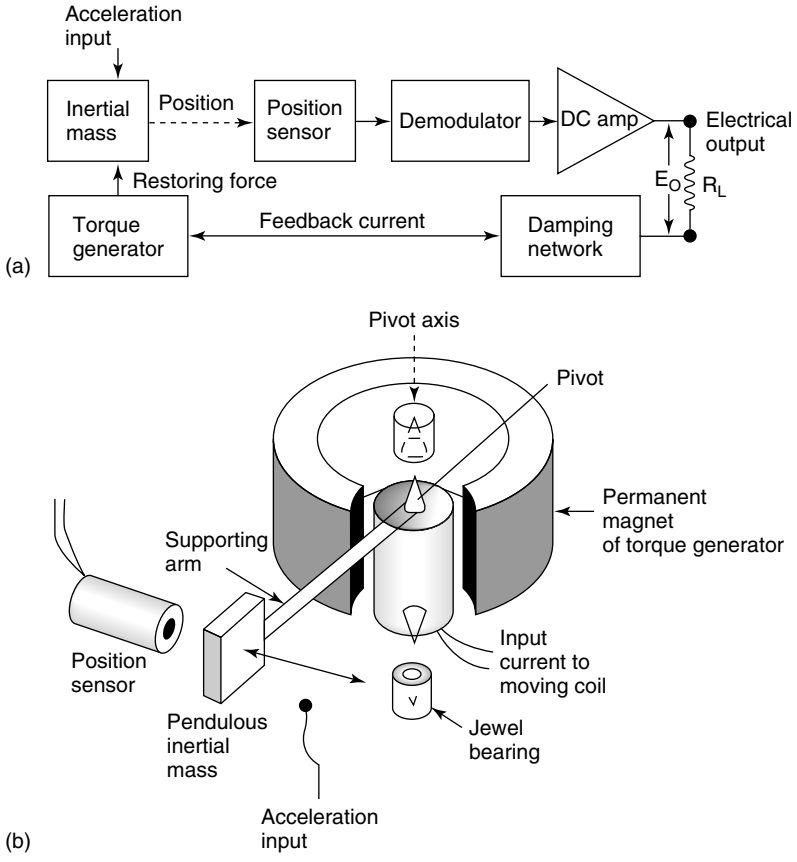


Figure 5. Component layout of a simple form of closed-loop accelerometer.

Force-feedback Sensors, Volume 2, in Jones (1982), and sections in Webster (1999).

Microminiaturization has provided considerable scope to form mass-spring systems in microelectro mechanical systems (MEMS) devices – *see* **Article 85, Elastic Regime of Design – Design Principles, Volume 2**; **Article 163, Uses and Benefits of MEMS, Volume 3**; and **Article 164, Principles of MEMS Actuators, Volume 3**. The complete seismic spring system can be formed at micrometer dimensions, meaning that the mass is very much less than constructions formed of mixed materials as shown in Figure 4.

The spring element is made in many forms – cantilever, double-ended supported beam, vibrating strings, and disks. With such small masses, these are able to operate at much higher vibration frequencies and shock rates than traditional forms.

They are used in many applications such as airbag sensors and also in accelerometer sensors for general-purpose use. Solid-state manufacture, however, is not able to fully compete with the performance of the best traditional forms of sensor because of production variations and parameter drift over time.

2 AREAS OF APPLICATION

When searching for information about a measurement technique, it is often helpful to have an appreciation of the allied fields that use the same equipment. Vibration will be of interest in many different applications, but some can be singled out as the main areas to which commercial marketing forces have been directed to enhance certain sensor characteristics.

2.1 Machine health monitoring

A significant field of interest is that of machine health monitoring, or condition monitoring. Failures in machines can often be avoided by 'listening' to the sounds and vibrations made by the system and making changes early enough. An example is shown in Figure 6, where the position of vibration sensors and their typical frequency spectra are given. Vibration and other forms of sensor are applied to the operating system, first while running in its early life to provide a base calibration for ongoing comparison, and then at periodic intervals during life.

If the frequency/amplitude data (the so-called signature) has changed from the baseline shape, then this can provide diagnostic information, suggesting which component is beginning to fail and when it will reach that state. It can then be conveniently replaced before a major untimely break down occurs. Figure 7 shows the set of equipment commonly needed to set up one of these systems.

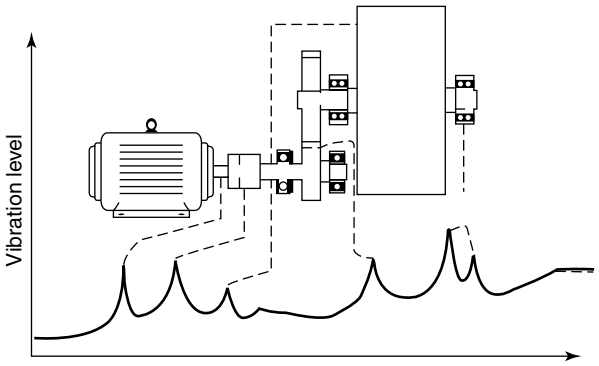


Figure 6. Vibration levels observed are recorded for comparison with a normal frequency spectrum signature.

Further information on machine health condition monitoring can be found in Reeves (1999), Rao (2000), and Mobley (2002).

2.2 Energy exploration

When a very robust sensor is needed, the energy exploration application field is a good place to start looking. The phrase Measurement While Drilling (MWD) has been coined for their in-process measurements.

For this market, ruggedized vibration sensors are needed to withstand the working conditions existing deep down drilling boreholes, wherein shock, high temperatures, and pressure demand much more protection than does a laboratory unit. There temperatures up to 185 °C and shock levels of 1000 g can arise in practice.

2.3 Space

In space navigation and control and inertial grade navigation, some key requirements are small in size and weight, high in reliability, and low in noise levels. As an example, one unit offered for this market has an RMS noise level up to 10 Hz of 0.6 μg.

3 CROSS COUPLING, CABLING, AND AMPLIFIERS

With vibration measurement, it is all too easy to produce incorrect data. This and the following section headings

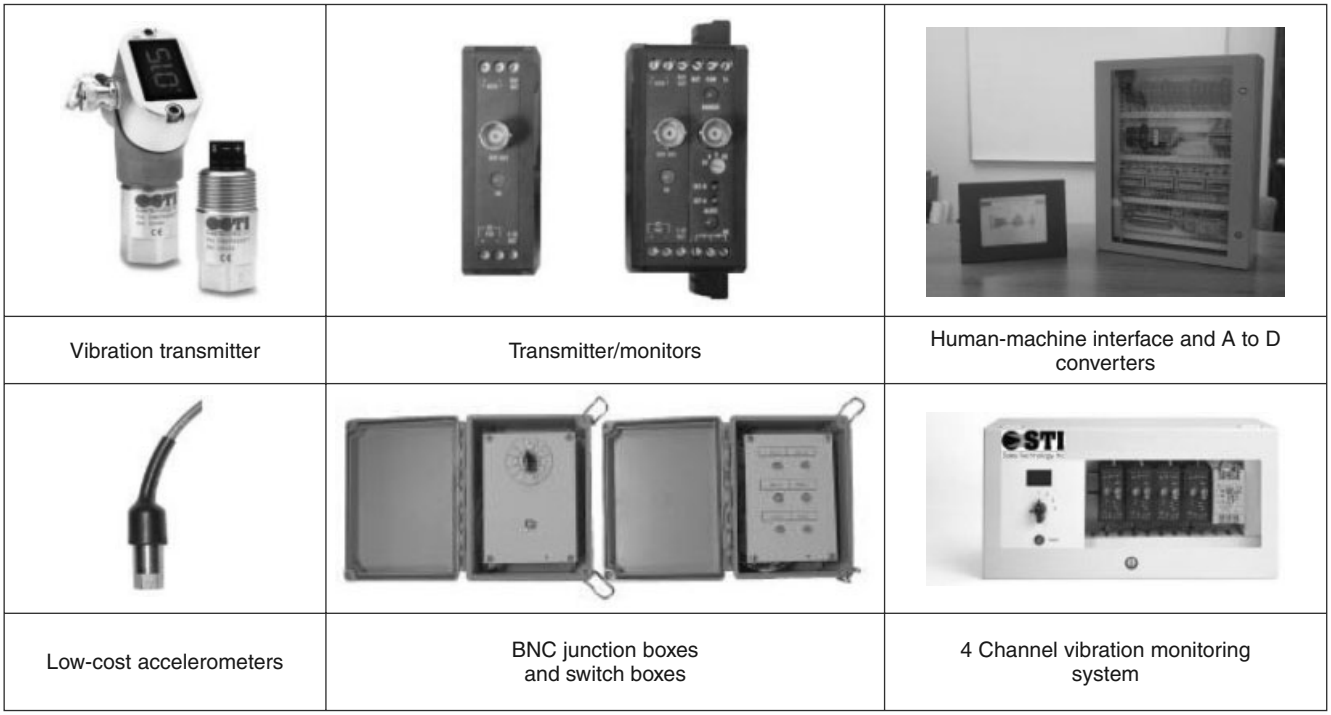


Figure 7. Range of equipment needed to instrument a machine health monitoring system. (Courtesy STI, Sales Technology Inc.)

address several important installation conditions that should be carefully studied for each new application.

3.1 Cross-coupling and cross-axis sensitivity

A vibration sensor is usually designed with one principal measuring axis that is perpendicular to its mounting surface.

Transducers will exhibit cross-axis coupling to some extent. This arises because the sensing system will also be partially sensitive to components of transverse motion. This means maximum sensitivity is not exactly aligned with the axis of symmetry of the mounting case but will be the resultant vector of the cross-axis sensitivity for two orthogonal planes through the center of the sensor.

Best practice measurements include a test that vibrates the sensor in a direction perpendicular to the direction of normal use.

Once the cross-axis sensitivity is established, a value judgment is needed to decide if this error exceeds the allowable limit for the application to hand; special care is needed where complex vibration is involved.

Rotational (torsional) sensitivity may also be important.

These tests can be avoided each time they are used if the sensors are precalibrated for this source of error and, of course, are still within calibration. An upper limit of 3 to 5% is offered by suppliers as the upper bound; most sensors will calibrate under this magnitude.

Sensors that have no cross-axis sensitivity quoted should not be used until it is established that it does not impact on the measurement needed. For example, conducting a comparison test of two objects placed into the same vibratory conditions will not be influenced much by any cross-axis sensitivity, unless its value is way above the 5% level, whereupon the signal could become saturated and thus limited.

In the case of a triaxial accelerometer (three units mounted together on each of the three-dimensional axes), any high magnitude cross-axis output sensitivity can corrupt data if the principal motion is at right angles to the base.

A quality sensor will be calibrated and usually marked on the sensor with a red spot placed at the point of maximum transverse sensitivity. A polar plot of the output might also be supplied that is tied into the position of the spot.

3.2 Coupling compliance

The compliance of the bond made between the sensor and the surface it is mounted on must be adequately stiff. If not, the surface and the sensor form a system that can vibrate in unpredictable ways. As an example, an insufficiently stiff

mounting can give results that produce much lower frequency components than truly exist. In extreme cases, the sensor can be shaken free as it builds up the unexpectedly low-resonance frequency of the joint to dangerous amplitude levels. As a guide, the joint should be at least ten times stiffer than the sensor mounting so that the resonant frequency of the joint is well above that of the sensor.

3.3 Cables and preamplifiers

Certain types of sensors, notably the piezoelectric (PZT) kind, are sensitive to spurious variations in capacitance and charge. Sources of such charges include the tribo-electric effect occurring in the bending motion of the vibrating connecting cables; mechanical stress in the material produces electrical signals as the cable bends. Special kinds of connecting cable are used that allow for movement of the cable.

Another source of potential error arises from any varying relative humidity, as this can alter electric field leakage; this becomes important when designing long-term installations.

The preamplifier stage is also important to get right for the PZT sensing element has very high output impedance. Such issues are largely overcome by integrating the preamp with the sensor unit.

4 INFLUENCE EFFECTS ON THE SENSOR

Ideally, the sensor should operate in a perfect environment wherein sources of external error, called *influence parameters*, are insignificant – *see* **Article 16, Reduction of Influence Factors, Volume 1**.

In vibration sensing, possible influence effect error sources include temperature variation of the sensor, possible magnetic field fluctuations (especially at radio frequency), and existing background acoustic noise vibrations. Each of these might induce erroneous signals.

A good test for influence parameters is to fully connect the sensor system, observing the output when the measurand of interest is known to be at zero level, such as by clamping it in a very stiff manner. The output will then comprise internal noise of the sensing channel and internal defects, such as system mains supply hum, plus contributions of any impacting influence effect. If the sensing noise is not observable, then more gain is needed to ensure the test is valid.

Where practical, the important error inputs can be systematically varied to see the sensor response. Many a vibration measurement has finally been seen to be worthless because some form of influence error turned out to be larger than the true signal from the measurand. Vibrations apparently occurring at electric mains frequency (50–60 Hz) and harmonics, thereof, are most suspect. Measurement of mechanical vibration at these frequencies is particularly difficult because of the need to separate true signal from influence error noise signals.

5 LOADING BY THE SENSOR

Vibration sensors contain mass. As this mass is made smaller, the sensitivity of the sensor usually will be reduced, so units with larger mass are more desirable. However, addition of mass to a vibrating system can load the mass of that system, causing shifts in frequency and amplitudes from the normal state of operation.

For this reason, manufacturers offer a wide range of attached type sensors. Provided the mass added is, say, 5% or less of the mass of interest, then the results will usually be acceptable.

Cables can reduce the overall mechanical compliance by forming a mechanical bridge from its holding down point to the sensor case. This can corrupt the true signal magnitude. They can also increase the damping, thereby shifting frequencies.

Where a system is particularly sensitive to external loading, the use of noncontact, fixed-reference methods may be the only way to make a satisfactory measurement.

6 DURATION OF USE OF SENSOR IN TESTS

When damping of a structure is very small, as in many mechanical structures, the time taken for a resonance to build up to its peak value is long. When using forced vibration frequency scanning to seek resonances, it is important not to sweep the excitation input frequency too rapidly.

As the Q Factor rises, the time to reach full amplitude of the resonance can take many minutes in large-size mechanical systems.

One often-used test is that for obtaining the vibration frequency spectrum of a machine tool as its tool post is excited, as it would be when machining metal. The time taken to sweep through the complete spectrum can take many tens of hours. The issue then is to be certain that the scanning and measuring equipment maintains its declared specifications over such long continuous periods of operation.

Some tests are best done in the time domain using an impulse or a step increase, the time domain data resulting being transformed into the frequency domain. Impulse tests need special care to avoid occurrence of zero shift and

ringing errors, caused by nonlinearities, in electrical aspects of amplifiers and the sensor for which residual mechanical strain can also arise if the spring element is overloaded during such shocks.

7 AMPLITUDE CALIBRATION

Here, attention is drawn to the issue of sensor calibration. Whereas the fixed-reference methods do have some relevance in the practical measurement of vibration, the need for a convenient fixed datum is very often not met.

Static amplitude (displacement) is easily calibrated using a standardized micrometer, displacement sensor, or optical interferometer. The motional element of the sensor is deflected a known amount and the output measured.

Note that some types of accelerometer, such as the PZT type, are not DC-connected and thus have zero static output at any position; there has to be motion present, at a relatively rapid rate, to generate an output.

Dynamic calibrations may be made either by comparison, using a technique of known accuracy and frequency response, or by using a calibrated vibration generator.

Dynamic amplitude of the non-DC-connected types requires use of a noncontacting method of measuring the

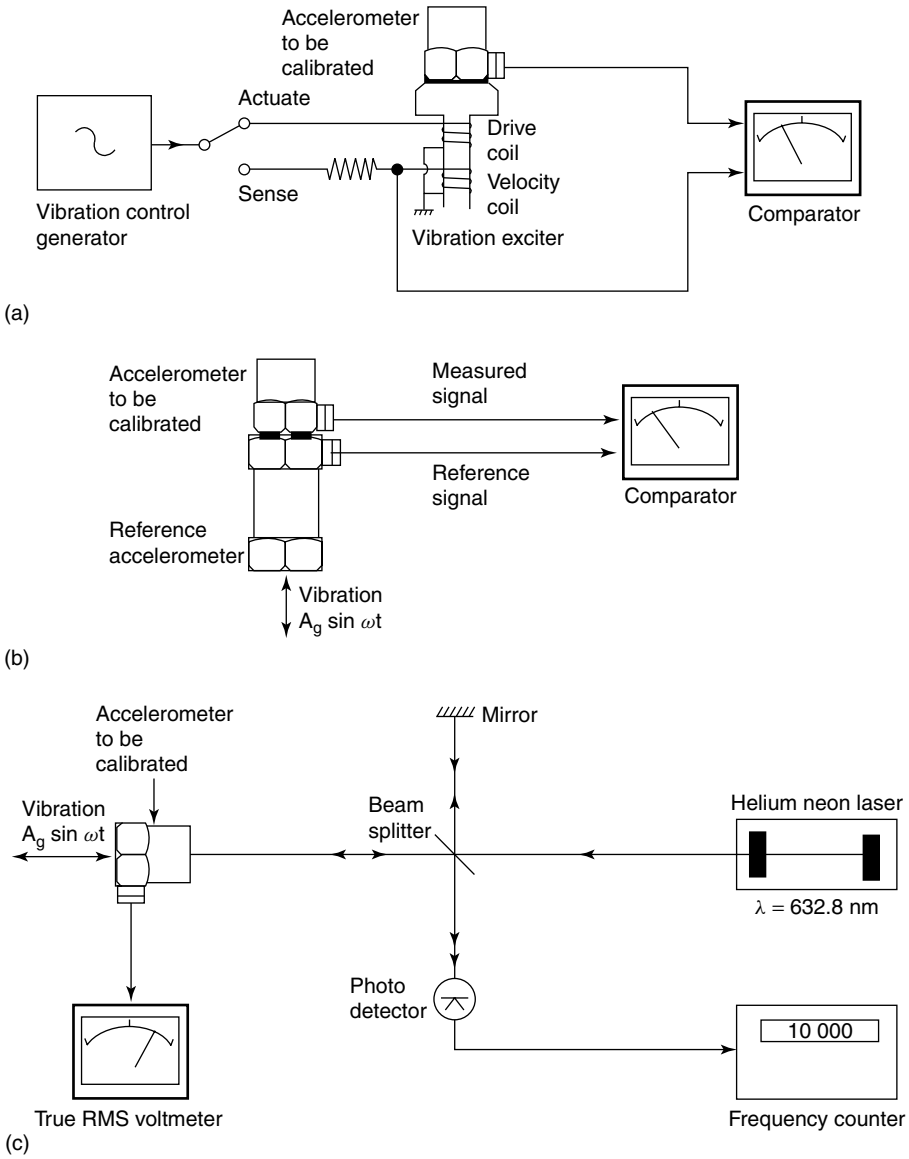


Figure 8. Calibrating accelerometers. (a) Calibrated vibration exciter shaking accelerometer at calibrated levels; reciprocity method. (b) Back-to-back calibration of a calibrated accelerometer against one to be calibrated; comparison method. (c) Absolute measurement using optical interferometry.

displacement taking place. This is viewed on a display screen along with the sensor output. Viewing the signal is recommended as this allows the wave shape to be checked. Once the nature of the signal has been validated, the data can be processed into more directly useful charts or numerical values.

8 ACCELEROMETER CALIBRATION

Figure 8 shows outlines of three methods of calibrating accelerometers and other vibration-measuring sensors. Calibration is normally performed at $500\,\mathrm{rad\,s^{-1}}$. Figure 8(a) is the reciprocity method. A calibrated vibration exciter shakes the accelerometer at calibrated levels. Figure 8(b) shows the comparison method. This uses back-to-back calibration of a calibrated accelerometer against the one to be calibrated. Figure 8(c) is a noncontact method using optical interferometry; this gives an absolute not relative determination.

Other methods that can be used are to subject the accelerometer to accelerations produced by the Earth's force. Simple pseudostatic rotation of an accelerometer in the vertical plane will produce accelerations in the 0 to ± 1 g range (g is used here for the Earth's acceleration value). Larger values can be obtained by whirling the accelerometer on the extremity of a rotating arm of a calibrating centrifuge. Alternatively, it can be mounted on the end of a hanging pendulum.

9 SHOCK CALIBRATION

Short-duration acceleration, as produced by impact, requires different approaches to calibration. Accelerations in such applications as explosive devices can exceed 10 000 g and last for only a few milliseconds.

A commonly used method is to produce a calibrated shock by allowing a steel ball to free-fall on to an anvil on which is mounted the sensor. This method provides an absolute calibration but, as with all of the methods described, has uncertainties associated with the practice of the method. In this case, one source of error is caused by the difficulty of releasing a ball to begin its downward path without imparting a small negative or positive velocity at time zero.

Another method uses a long pendulum carrying a large mass. This is allowed to swing down to impact on the object carrying the sensor under test. Physical consideration allows the rate of energy release to be calculated and thus the impact forces, but it is not that accurate. It is useful for relative calibrations where the sensor under test and a standardized unit are mounted in the same impacted object.

10 FORCE CALIBRATION

Static forces can be calibrated by applying dead weights to the force sensor, the weights being calibrated masses.

Dynamic forces arising in vibration can more easily be determined using the relationship force = mass × acceleration. A shaking table is used to produce known accelerations on a known mass. In this way, the forces exerted on the accelerometer can be determined along with the corresponding output voltage or current needed to produce transducer sensitivity constants.

At the completion of any calibration, the determination and support data about the method of calibration must be recorded according to the organizations policy. This ensures traceable results in the event of dispute arising later – *see* **Article 45, Calibration Process, Volume 1**. The period between calibrations is covered in **Article 46, Calibration Interval, Volume 1**.

Space does not permit greater explanation; detailed accounts of vibration sensors are available in the literature, Trampe-Broch (1980), Bruel and Kjaer (1982), Smith (1989), Timar (1989), Wowk (1991), McConnell (1995), Ling and Shabana (1996), Mobley (1999), Steinberg (2000), Dyer (2001), Harris and Piersol (2001).

RELATED ARTICLES

Article 57, Static Considerations of General Instrumentation, Volume 1; **Article 59, Introduction to the Dynamic Regime of Measurement Systems, Volume 1**; **Article 205, Theory of Vibration Measurement, Volume 3**; **Article 207, Acceleration Measurement, Volume 3**; **Article 208, Amplitude and Velocity Measurement, Volume 3**.

REFERENCES

Bruel & Kjaer (1982) *Measuring Vibration-Elementary Introduction*, Bruel & Kjaer, Naerum.

Dyer, S.A. (2001) *Survey of Instrumentation and Measurement*, Wiley, New York.

Harris, C.M. and Piersol, A.G. (2001) *Harris' Shock and Vibration Handbook*, McGraw-Hill, New York.

Jones, B.E. (1982) Feedback in Instruments and its Applications, in *Instrument Science and Technology*, (ed. B.E. Jones), Adam Hilger, Bristol.

Ling, F.F. and Shabana, A.A. (1996) *Theory of Vibration: An Introduction*, Springer Verlag, Berlin.

McConnell, K.G. (1995) *Vibration Testing: Theory and Practice*, Wiley Interscience, New York.

Mobley, R.K. (1999) *Vibration Fundamentals*, Butterworth-Heinemann, Oxford.

Mobley, R.K. (2002) *An Introduction to Predictive Maintenance*, Butterworth-Heinemann.

Reeves, C.W. (1999) *The Vibration Monitoring Handbook*, (Coxmoor's Machine and Systems Condition Monitoring Series), Coxmoor Publishing Co., Oxford.

Rao, J.S. (2000) *Vibratory Condition Monitoring of Machines*, CRC Press.

Smith, J.D. (1989) *Vibration Measurement and Analysis*, Butterworth-Heinemann, Oxford.

Steinberg, D.S. (2000) *Vibration Analysis for Electronic Equipment*, Wiley Interscience, New York.

Timar, P.L. (1989) *Noise and Vibration of Electrical Machines*, North-Holland, Amsterdam.

Trampe-Broch, J. (1980) *Mechanical Vibration and Shock Measurements*, Bruel & Kjaer, Naerum.

Webster, J.G. (1999) *The Measurement, Instrumentation, and Sensors Handbook*, CRC Press, Boca Raton, FL.

Wowk, V. (1991) *Machinery Vibration*, McGraw-Hill, New York.

207: Acceleration Measurement

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Use of Seismic Sensors 1408
2 Optimal use of Seismic Sensor for Acceleration Measurement 1408
3 Typical Accelerometers 1409
4 Response to Complex Waveforms 1409
5 The Piezoelectric Sensor 1410
6 Amplifiers for Piezoelectric Sensors 1411
7 Microaccelerometers 1411
8 Signal Conditioning in Microaccelerometers 1411
9 Measurement of Shock 1412
Related Articles 1413
References 1413

1 USE OF SEISMIC SENSORS

Of the many measurands that appear under the heading of vibration – *see* **Article 205, Theory of Vibration Measurement, Volume 3** – the most measured is acceleration.

The fixed-reference method of measuring acceleration is rarely used, determinations being made more easily with the mass-spring, seismic form of sensor. This method uses the inertial mass as a short-term fixed reference: the test object moves whilst the mass stays sufficiently stationary in space for short periods of time.

For the seismic sensor system, the mass and the spring compliance are both fixed entities. Consideration of the $F = m \times a$ (force = mass × acceleration) law with spring compliance being linear shows that displacement of the mass relative to the sensor case is proportional to the acceleration of the case.

The curves, plotted in Figures 5 and 6 of **Article 205, Theory of Vibration Measurement, Volume 3** for sinusoidal input of force to a second-order system, are applicable for obtaining the output response curves of accelerometers using displacement sensing. In this case, the vertical axis is interpreted as the relative displacement of the mass for a given acceleration.

2 OPTIMAL USE OF SEISMIC SENSOR FOR ACCELERATION MEASUREMENT

The curves, Figure 6 in **Article 205, Theory of Vibration Measurement, Volume 3**, show that a seismic sensor will provide a relatively constant output sensitivity representing sensor acceleration from very low vibration frequencies up to around 0.5 of the natural frequency of the spring-mass arrangement used. Further, the damping ratio can be optimized at around 0.5–0.6 and electronic compensation added (if needed) to raise the upper limit a little further than the undamped resonance point.

At first sight it might, therefore, appear that a single, general-purpose design could be made having a very high resonant frequency. This, however, is not the case for the deflection of the spring (which is a major factor deciding the system output sensitivity) is proportional to $1/\omega_n^2$. In practice, this means that as the upper useful frequency limit is extended, the sensor sensitivity falls off rapidly. Electronic amplification allows low signal output to be used but with additional cost to the total measuring system. As the signal level falls, added amplification will also increase the noise level.

At the very low frequency end of the accelerometer response, the transducers become ineffective as

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

accelerometers because the accelerations produce too small a displacement to be observed against the background noise level. By increasing the mass, the units can be made to respond to lower frequencies – *see* **Article 208, Amplitude and Velocity Measurement, Volume 3**.

3 TYPICAL ACCELEROMETERS

As a guide to the range of capability available, one major manufacturer's catalog for traditional accelerometers not made in the more recent microminiature form, offers accelerometers with sensitivities ranging from a small $30\,\mu V\,ms^{-2}$ through to $1\,V\,ms^{-2}$ with corresponding sensor masses of 3 g and 500 g and useful frequency ranges of 1000 Hz and 0.2 Hz. As the mass increases, the frequency range falls or, conversely, to get low frequency response the mass has to increase. Sensors have been constructed for higher frequencies using the MEMS, microelectro mechanical system technology as is discussed in Section 7. A selection of traditional accelerometers is shown in Figure 1.

The many constraints placed upon the various performance parameters of a particular seismic sensor can be represented in numerical terms on a single chart, Figure 2, Harris and Crede (1976).

It is not always appreciated that all materials exhibit some degree of mechanical compliance and that the modulus of elasticity is a property of the material, not the way it is used. Elastic structures, such as helical springs, are used to increase the compliance to suit a particular task – *see* **Article 85, Elastic Regime of Design – Design Principles, Volume 2**.

To obtain the stiffness needed, accelerometer springs may occur as stamped rigid plates, as flat cusped spring washers, or even as a sufficiently compliant clamping bolt. In applications of piezo-sensitive materials, use is often made of the compliance of the material itself.



Figure 1. A range of accelerometers is required to cover the full needs of vibration measurement. (Courtesy, Honeywell Sensing and Control.)

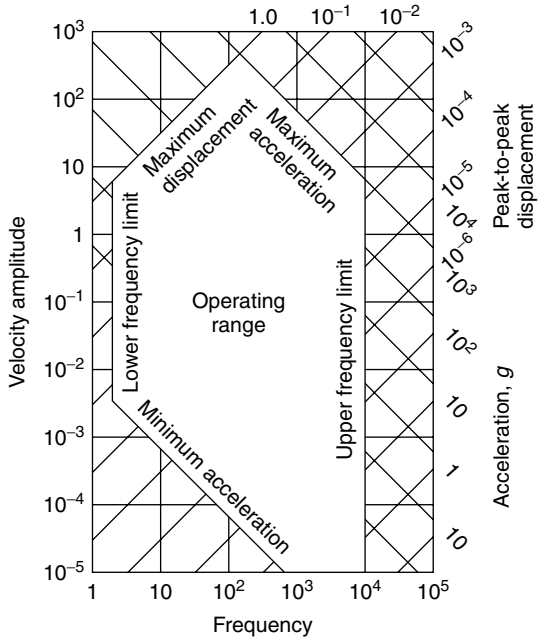


Figure 2. Useful linear operating range of an individual seismic vibration sensor can be characterized with this type of chart.

4 RESPONSE TO COMPLEX WAVEFORMS

The response curves given in Figures 5 and 6 of **Article 205, Theory of Vibration Measurement, Volume 3** relate to seismic sensors excited by sinusoidal signals. To predict the behavior of a certain sensor, such as an accelerometer, used to measure other kinds of continuous or discrete waveforms, it is first necessary to break down the waveform into its Fourier components – *see* **Article 29, Relationship Between Signals in the Time and Frequency Domain, Volume 1**. The response, in terms of amplitude and phase, to each of these is then added to form the resultant response.

It has been stated above that damping can be added to extend the useful bandwidth of a seismic sensor. However, where this is done it can increase the phase-shift variation with frequency; that is, the components are sensed with different phase shifts to the undamped case and thus the reconstructed signal is no longer a fully faithful replica of the original motion.

A signal comprising many frequencies will, therefore, produce an output that depends largely on the damping and natural frequency values of the sensor. A number of responses can be plotted, such as shown in Figure 3 from Harris and Crede (1976). Generally, the damping value for best all-round results is that near the critical damping value of 0.7.

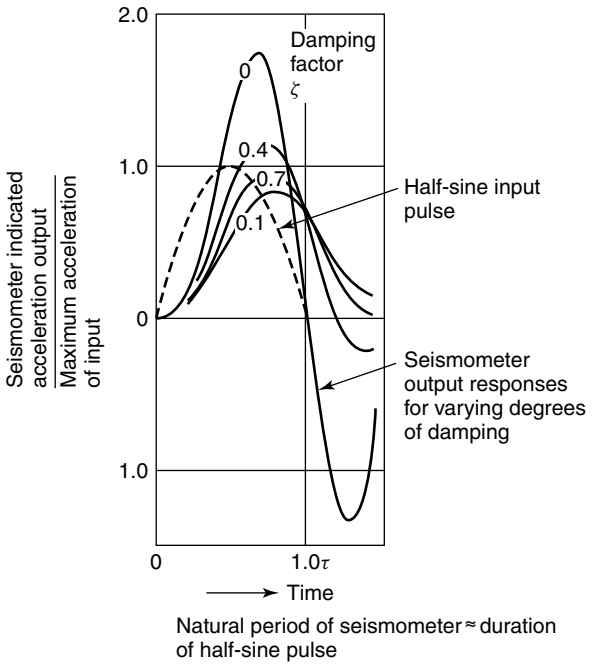


Figure 3. Example of response, at various damping-factor levels, of a seismic accelerometer to a complex forcing input-half-sine wave of similar period to that of the natural resonance period of the sensor.

5 THE PIEZOELECTRIC SENSOR

Numerous sensing methods have been devised to measure the motion of the mass in a seismic sensor. Discussed here is the most commonly used PZT, piezoelectric transducer method: others are described in Norton (1969), Oliver (1971), Webster (1999).

Force applied to certain crystalline substances such as quartz produces, between two surfaces of a suitably shaped crystal, an electric charge that is proportional to that force. This charge is contained in the internal electrical capacitance formed by the electrically insulating, high-dielectric material and its two conducting surfaces deposited on each face.

The first approximation descriptive mathematical relation for this effect is

$$q = a \times F \times K_s \tag{1}$$

where q is the electrical charge generated by force, F (in Newtons), applied across the faces of a piezoelectric device having a mechanical compliance of spring rate K_s (mN^{-1}) and a material constant a (of dimensions $C\,m^{-1}$).

The constant a, depends on many factors including the geometry of the crystal, position of electrodes, and material used. Typical materials now used (natural quartz is less sensitive and, therefore, less applicable) include barium titanate with controlled impurities, lead zirconate, lead niobate; many others are available but their composition is a trade secret.

The usual PZT sensing material is made from loose powder that, after shaping, is fired at a very high temperature. While cooling, the blocks are subject to an electric field that traps a polarization.

Sensitivity of PZT materials is temperature-dependent on the charge sensitivity and the capacitance value, both of which alter with temperature. These effects do not follow simple linear laws. Such materials have a critical temperature, called the *Curie point*. They must never be taken above it for it will then destroy the properties as calibrated.

The Curie point varies from 120 °C for the simpler barium titanate forms ranging up to values close to 600 °C. For the interested reader, detailed explanation can be found in **Article 118, Ultrasonic Sensors, Volume 2**, Bruel & Kjaer (1976), Endevco (1980), Harris and Crede (1976), Trampe-Broch (1980), and in the detailed information provided by the makers of PZT materials in application notes, and many Internet Web sites.

To read the charge of a PZT sensor as it is vibrated by the subject under test, an electronic amplifier is used that converts charge magnitude to a voltage equivalent. It is the nature of this sensing system that it provides no DC response; when stationary there will be no output other than system noise.

In practice, the PZT sensors used to measure acceleration can be operated down to around 0.1 Hz, dependent on the amplitude to be measured. With a natural resonant frequency that can be made relatively very high (up to 100 000 Hz in some designs), PZT sensors provide a useful frequency range that can cover most vibration needs. The system response, however, relies not only on the sensor but also upon the cables and the preamplifier used with the PZT unit – *see* **Article 206, Practice of Vibration Measurement, Volume 3**.

To produce the charge, the PZT material can be used in its pure compression, shear, or bending modes. This method of sensor design is amenable to the combination of three units giving the three translation components of vibration. Figure 4 in **Article 206, Practice of Vibration Measurement, Volume 3** gives some examples of commercially available PZT accelerometers.

The PZT material itself contributes only of the order of 0.03 of the required critical damping factor. Without additional damping being added, PZT transducers should not be used too close to their resonant frequency – see Figure 6 in **Article 205, Theory of Vibration Measurement, Volume 3** – or an error will arise.

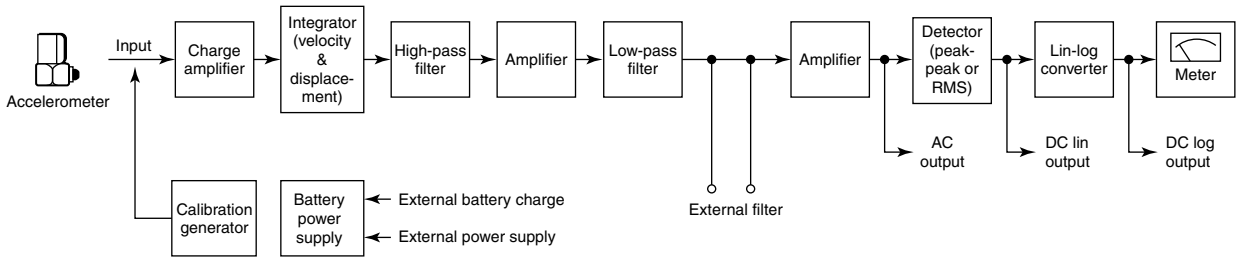


Figure 4. Block diagram of vibration measuring system showing functions that may be required. (Reproduced by permission of Bruel & Kjaer.)

Mounting arrangements within the case also can add additional damping. Some use an additional spring to pre-compress the PZT element so that it remains biased in compression under all working amplitudes: this makes for more linear operation.

Typical sensor element charge sensitivities range from 0.003 up to $1000\,\mathrm{pC\,m\,s^{-2}}$, implying that the following preamplifier units will also need to vary considerably.

6 AMPLIFIERS FOR PIEZOELECTRIC SENSORS

An amplifier for reading out the state of the PZT sensor is one that has very high input impedance, an adequate frequency response, and low output impedance. Adjustment of the gain, filtering action and integration time variously yield velocity and displacement sensing – *see* **Article 208, Amplitude and Velocity Measurement, Volume 3**.

Figure 4 is a typical analog conditioning system incorporating most features that might be needed. Some of the block functions often use digital processing methods, but the charge amplifier still needs to be analog.

The charge amplifier could be designed to see the sensor either as a voltage or a charge (current) source. The latter is preferred for, using modern electronic-feedback operational amplifier techniques, the effect of cable, sensor, and amplifier capacitances can be made negligible (which in the voltage-reading method, is not the case). Cable length is, therefore, of little consequence. This is justified as follows.

Figure 5 is the relevant equivalent circuit for a PZT accelerometer connected to an operational amplifier (the preamplifier) via a cable. It includes the dominant capacitances that arise in practice.

It can be shown (see Trampe-Broch (1980), for example) that the use of feedback in this way with a very high amplifier gain A gives

$$e_0 = \frac{S_q a}{C_f} \tag{2}$$

This shows that the user need only define the sensor charge sensitivity S_q and the feedback capacitance C_f in order to be able to relate output voltage from the preamplifier to the acceleration of the sensor.

If the preamplification shown here is carried out in the sensor housing, then a special connection cable having reduced triboelastic effect is not needed from the sensor to the fixed cable support. Many cases, however, cannot place the preamplifier in the sensor housing as that adds mass to the accelerometer, which may not be acceptable.

7 MICROACCELEROMETERS

Microelectronic forms of the accelerometer made by the MEMS technology – *see* **Article 162, Principles of MEMS, Volume 3**; **Article 163, Uses and Benefits of MEMS, Volume 3**; and **Article 164, Principles of MEMS Actuators, Volume 3** – can be remarkably small compared with traditional forms.

For example, one unit reported at the R&D stage by researchers at Seoul University, Kwang *et al*. (2003), has a proof mass of 15 μg that is $870 \times 1027 \times 50$-μm thick. The accelerometer's natural frequency at 14 kHz implies operation up to 7 kHz or a little more if compensation is used. The spring elements supporting the lateral capacitive comb type sensor are 4-μm thick and 250-μm long. The sensing gap is 3 μm.

Another development unit, Figure 6, shows one of the many forms of suspension used to support the proof mass, LLNL (2000).

8 SIGNAL CONDITIONING IN MICROACCELEROMETERS

The signal processing used in the MEMS forms of accelerometer can be very sophisticated as it allows integration of the system blocks needed to produce a ready to use robust output signal. These are

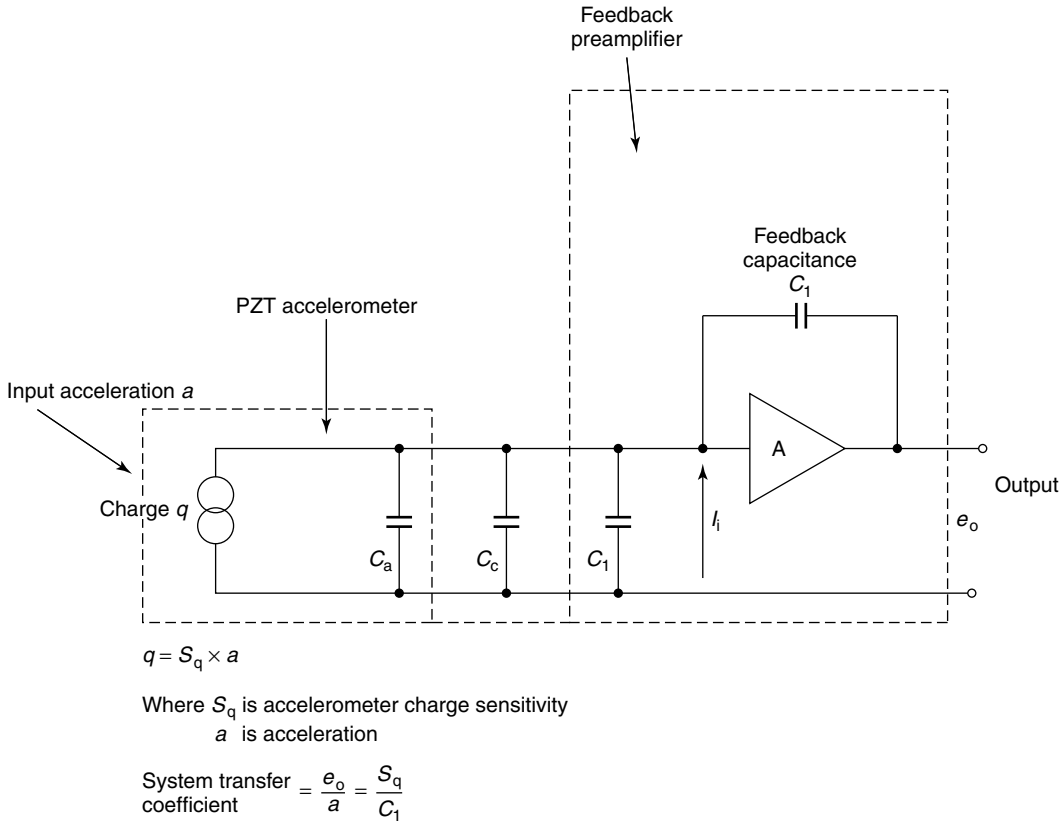


Figure 5. Equivalent circuit for piezoelectric sensor when interrogated, as a charge-generating device, by an operational amplifier technique.

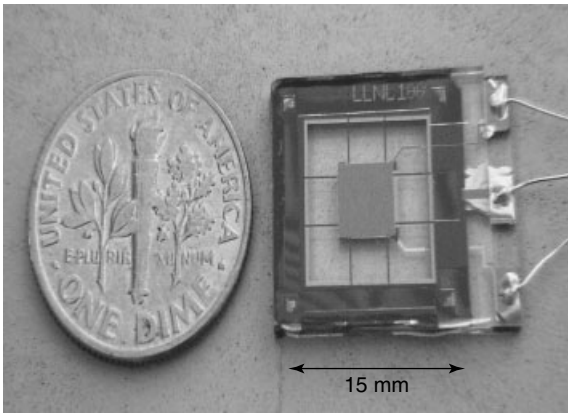


Figure 6. An experimental microaccelerometer sensor.

- seismic sensing element,
- voltage regulator,
- A–D converter,
- sample and hold,
- microprocessor,
- switching,
- noise reduction,
- computer controlled settings,
- smart features – *see* **Article 160, Smart Sensor System Features, Volume 3**.
- bias voltages,
- wireless link interface.

A good study, readily available to consult on the Internet is Kulah *et al*. (2004). Their account shows the intricacies and level of achievement possible with fully integrated MEMS units.

Further details of MEM systems are found in **Article 162, Principles of MEMS, Volume 3**; **Article 163, Uses and Benefits of MEMS, Volume 3**; and **Article 164, Principles of MEMS Actuators, Volume 3**.

Signal conditioning is covered in the Section on Data Acquisition Systems starting at **Article 132, Data Acquisition Systems (DAS) in General, Volume 3**.

9 MEASUREMENT OF SHOCK

Shock is a sudden short impulse of applied force that can generate very large acceleration (100 000 g can arise) that is not recurrent. It can be regarded as a once-only occurrence

of a vibration waveform, although sometimes it is used to describe a short burst of oscillation. Examples are objects subject to explosions and impacts arising because of a collision.

Understanding the formalized behavior of a given vibration sensor to a shock situation requires Fourier analysis of its response to a truncated wave shape, that is, a discontinuous wave. The mathematics becomes more complex.

Theoretical considerations lead to the generalization that as the waveform becomes more like a single pulse of high magnitude and very short duration, the frequency band of the sensor must be widened if the delivered output is to be a satisfactory replica of the actual input vibration parameter.

Thus, the accelerometer used to measure shock must preferably possess sufficient flat bandwidth below its natural frequency point. Special DSP, digital signal processing methods allow measurement signal extraction beyond this criterion using advanced algorithms.

Fidelity of output increases as the period of the natural frequency of the sensor becomes shorter than the pulse length. An idea of the variation of responses with natural frequency and damping is available in graphs given in Harris and Crede (1961). An example can be found in Figure 3.

Here also the very large force exerted on the transducer (force = g × mass of sensor) requires a total design that recognizes the need to withstand large transient forces without altering the mechanical residual strains in the sensor system during measurement. Mounting circuitry on the sensor itself adds more room for induced shock error, for the silicon chips can be shock sensitive.

Well-designed shock sensors can accurately measure single half-sine-wave pulses as short as 5 μs. Some amount of ringing in the output is usually tolerated in order to provide measurement of very short duration shocks. Compensation to increase the bandwidth is also useful in extending the range of operation.

RELATED ARTICLES

Article 59, Introduction to the Dynamic Regime of Measurement Systems, Volume 1; **Article 61, First-order System Dynamics, Volume 1**; **Article 62, Second-order System Dynamics, Volume 1**; **Article 205, Theory of Vibration Measurement, Volume 3**; **Article 206, Practice of Vibration Measurement, Volume 3**; **Article 208, Amplitude and Velocity Measurement, Volume 3**.

REFERENCES

Bruel & Kjaer. (1976) *Piezoelectric Accelerometer and Vibration Preamplifier Handbook*, Bruel & Kjaer, Naerum.

Endevco. (1980) *Shock and Vibration Measurement Technology*, Endevco Dynamic Instrument Division, San Juan Capistrano, CA.

Harris, C.M. and Crede, C.E. (1976) *Shock and Vibration Handbook Vol. 1, Basic Theory and Measurements*, McGraw-Hill, New York, (1961, reprinted in 1976).

Kulah, H., Salian, A., Yazdi, N. and Najafi, K. (2004) A 5V closed-loop second-order sigma-delta, micro-g micro-accelerometer, *Center for Wireless Integrated Microsystems*, University of Michigan, http://www.eecs.umich.edu/∼hkulah/pdfs/HH-02-Kulah.pdf.

Kwang, Y., Yeon, J.C., Kim, Y.H., Rhee, S.W. and Oh, S.H. (2003) High Resolution Silicon Accelerometer Using Eutectic Bonding, in *Nanotech Conference 2003*, abstract http://www.nsti.org/procs.

LLNL. (2000) *MEMS microaccelerometer*, Lawrence Livermore National Laboratory, http://microtechnology.llnl.gov/Devices.html.

Norton, H.N. (1969) *Handbook of Transducers for Electronic Measuring Systems*, Prentice Hall, Englewood Cliffs, NJ.

Oliver, F.J. (1971) *Practical Instrumentation Transducers*, Pitman, London.

Trampe-Broch, J. (1980) *Mechanical Vibration and Shock Measurements*, Bruel & Kjaer, Naerum.

Webster, J.G. (1999) *The Measurement, Instrumentation and Sensors Handbook*, CRC Press, Boca Raton, FL.

208: Amplitude and Velocity Measurement

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Measurement of Displacement with a Seismic Form of Sensor 1414
2 Theory of Displacement Measurement Using a Seismic Sensor 1414
3 Practice of Displacement Measurement Using a Seismic Sensor 1415
4 Measurement of Velocity 1415
5 Theory of Velocity Measurement 1415
6 Practice of Velocity Measurement 1416
Related Articles 1416
References 1417

1 MEASUREMENT OF DISPLACEMENT WITH A SEISMIC FORM OF SENSOR

Where a fixed reference method is inconvenient, one of several forms of seismic sensor system can be employed as follows.

It is sometimes more convenient to mathematically integrate the signal from a velocity transducer making use of a fixed reference framework.

Integration is a sound process, for the noise present is reduced by the integration, averaging process. The integral of an acceleration measurement is velocity and the second integral is then displacement. Thus, an accelerometer can be used to measure displacement under certain conditions.

2 THEORY OF DISPLACEMENT MEASUREMENT USING A SEISMIC SENSOR

The second-order equations of motion, given in **Article 205, Theory of Vibration Measurement, Volume 3**, for a mass that is moving relative to a fixed reference frame (the mode for studying the movement of vibrating objects) can be reworked to provide response curves relating the displacement amplitude of the seismic mass to that of its case.

Figure 1 is the family of response curves showing the effects of different operating frequency and the degree of damping. Given that the case is moving in sympathy with the surface of interest, it can be seen, from the curves, that for input vibration frequencies that are at least twice the natural frequency of the seismic sensor, the measured output displacements will be a true indication (within a percent or so) of the movements of the surface. This form of seismic displacement sensor is also often called a *vibrometer*.

It is possible to lower the frequency of operation by using a damping factor with a nominal value of 0.5. This, however, does introduce more rapidly changing phase-shift error with frequency, which may be important. The lowest frequency of use above which the response remains virtually flat is seen to be where the various damping factor curves approach the horizontal line equal to unity ratio, to within the allowable signal tolerance.

Given that the chosen damping remains constant and that a system does follow the second-order response, it is also possible to provide electronic frequency compensation that can further lower the useful frequency of operation by a useful amount.

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

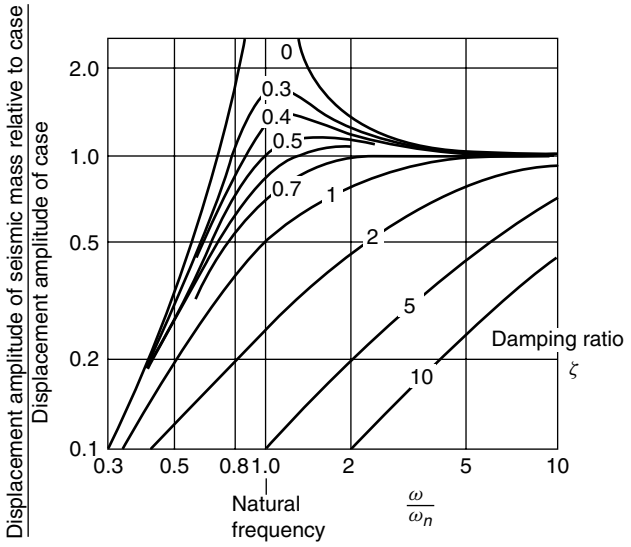


Figure 1. Responses relevant to displacement and velocity sensing with seismic sensors.

Thus, to directly measure displacement amplitudes with a seismic sensor, it must have a natural frequency set to be below the lowest frequency of interest in the subject's vibration spectrum. In this mode, the seismic mass virtually remains stationary in space acting as a fixed reference point. It is also clear that the seismic sensor cannot measure very low frequencies, for it is not possible to construct an economical system having a low enough resonant frequency, as that needs massive masses and very springy suspensions; these are the features of the large strong motion seismometers that measure ground and building motion and displacement during earthquakes.

3 PRACTICE OF DISPLACEMENT MEASUREMENT USING A SEISMIC SENSOR

The curves are theoretical perfections and would apparently indicate that the seismic sensor, in this case, will have flat response to infinite frequency. This is not the case in practice, for, as the frequency of vibration rises, the seismic sensor structure begins to resonate at other frequencies caused by such mechanisms as the spring vibrating in modes other than the fundamental of the system.

Given that the low frequency range of accelerometers can extend down to less than 1 Hz (*see* **Article 207, Acceleration Measurement, Volume 3**), it may often be more practical to twice integrate an accelerometer signal, in order to derive displacement amplitude, rather than to make use of a direct-reading displacement seismic sensor design.

4 MEASUREMENT OF VELOCITY

This section covers simple traditional forms of velocity sensor used to measure the short motions of solid objects that will possibly also need acceleration determination.

The many other forms for measuring velocity (i.e. flow speeds) of gases, liquids, solids, and mixed-phase flows are covered in the following articles:

- Laser Doppler velocimeters and other flowmeters – **Article 187, Flowmeter Selection and Application, Volume 3**
- Angular and displacement devices in general – **Article 191, Displacement and Angle Sensors Performance and Selection, Volume 3**; **Article 192, Strain Sensors, Volume 3**; **Article 193, Specialty Displacement and Angle Sensors, Volume 3**.

Other methods, such as digital encoders, Hall-effect toothed wheel sensing, tuning forks, magnetohydrodynamics, and optical correlation are variously covered in Webster (1999), Norton (1969), and Dyer (2001).

5 THEORY OF VELOCITY MEASUREMENT

The prime method used to generate a direct velocity signal makes use of the law of electromagnetic induction. This gives the electrical voltage e generated as N turns of an electric coil to cut magnetic flux ϕ over time t as

$$e = -N\frac{\mathrm{d}\phi}{\mathrm{d}t} \tag{1}$$

Velocity sensors are thus self-generating; they produce a voltage output that is proportional to the velocity at which a set of turns moves through a constant and uniform magnetic field.

Many forms of this kind of sensor exist. The commonly used, moving-coil arrangement comprises a cylindrical coil vibrating inside a magnetic field that is produced by a permanent electromagnet. A commonly seen arrangement, Figure 2(a), is that typified by the reversible-role, moving-coil, loudspeaker movement. For this form, the output voltage V_{out} is given by

$$V_{\text{out}} = -Blv \times 10^{-9} \tag{2}$$

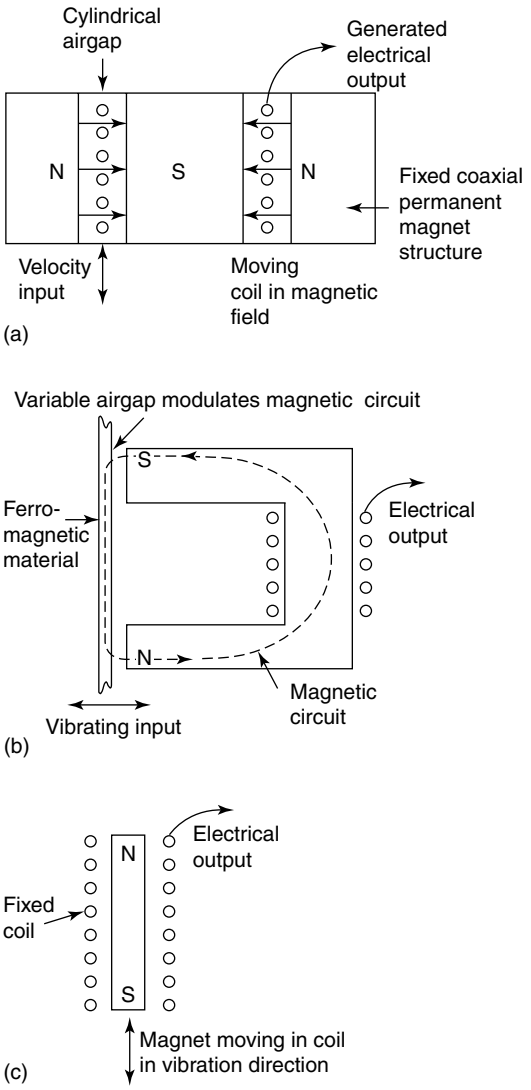


Figure 2. Forms of velocity sensor. (a) Moving coil, (b) variable reluctance, (c) moving permanent magnet.

where B is the flux density in Tesla, l the effective length of conductor contributing in total flux-cutting, and v the instantaneous velocity, expressed in ms^{-1}, of the coil relative to the magnet. Given that the design ensures that the magnetic field is uniform in the path of the fixed conductor length, the sensor can provide very linear output.

6 PRACTICE OF VELOCITY MEASUREMENT

These were the sensors adopted in early seismology studies because of their inherently high output at relatively low velocities. The coil impedance will generally be low, enabling signals with good signal-to-noise ratios to be generated along with reduction of error caused by variations in lead length and type. They are, however, large with resultant mass and rigidity problems. Many of the forms tend to have relatively low resonant frequencies (tens of hertz), which restricts use to the lower frequencies. It will be apparent that these sensors cannot produce signals at zero velocity because no relative movement occurs to generate flux-cutting.

A second variation of the self-generating velocity sensor, the variable-reluctance method, uses a series magnetic circuit containing a permanent magnet to provide permanent magnetic bias. A part of this circuit, the armature, is made so that the effective air gap is varied by the motion to be monitored. Around the magnetic circuit is placed a pickup coil. As the armature moves, the resulting flux variation cuts the coil, generating a signal that can be tailored by appropriate design to be linear with vibration amplitude. This form of design has the advantage that the armature can readily be made a part of the structure to be monitored, as shown in Figure 2(b). This version is not particularly sensitive, for the air gap must be at least as large as the vibration amplitude. As an example, a unit of around 12-mm diameter, when used with a high magnetic permeability moving disc set at a 2-mm distance, will produce an output of around $150\,mV\,ms^{-1}$.

A third method uses a permanent magnet as the mass supported on springs. One example is shown in Figure 2(c). Vibration causes the magnet to move relatively to the fixed coil, thereby generating a velocity signal. This form can produce high outputs, one make having a sensitivity of around $5\,V\,ms^{-1}$.

Where a fixed reference cannot be used, this form of sensor, instead of a displacement sensor, can be built into the seismic sensor arrangement. In such cases, the vibrating seismic sensor will then directly produce velocity signals. These will follow the general responses given in Figure 1. From those curves, it can be seen that there is a reasonably flat response above the natural frequency, which is inherently quite low.

RELATED ARTICLES

Article 187, Flowmeter Selection and Application, Volume 3; **Article 191, Displacement and Angle Sensors Performance and Selection, Volume 3**; **Article 192, Strain Sensors, Volume 3**; **Article 193, Specialty Displacement and Angle Sensors, Volume 3**; **Article 205, Theory of Vibration Measurement, Volume 3**; **Article 207, Acceleration Measurement, Volume 3**.

REFERENCES

Dyer, S.A. (2001) *Survey of Instrumentation and Measurement*, Wiley, New York.

Norton, H.N. (1969) *Handbook of Transducers for Electronic Measuring Systems*, Prentice Hall, Englewood Cliffs, NJ.

Webster, J.G. (1999) *The Measurement, Instrumentation and Sensors Handbook*, CRC Press, Boca Raton, FL.

209: Characteristics of Chemical Measurements

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 The Nature of Chemical Measurements 1421
2 Chemical Properties of Matter 1422
3 Measurable Characteristics of Chemical Variables 1423
4 Terminology 1424
5 Units 1424
6 Standards 1425
Related Articles 1426
Reference 1426
Further Reading 1426

1 THE NATURE OF CHEMICAL MEASUREMENTS

Chemical analytical instrumentation, used to determine the chemical properties of materials, relies heavily upon optical, chemical, and physical techniques combined with electronic signal control and processing, and also fine mechanics to structure the optical components in relation to the source, detectors, and materials under test.

This class of measuring system, although based upon the same physical and mathematical processes as other classes, incorporates a wide range of arrangements used in its own characteristic way. It, therefore, provides another facet of measurement system design that often involves concepts and methods unlike those found in physical measurements. Of particular relevance to chemical measurement is the fact that the initially sensed signal energy levels are usually very low and therefore require sophisticated signal processing to extract the meaningful information – see the articles under the section on **Signal Recovery in the Presence of Noise**.

This class of measurement system concerns considerable information about numerous variants. The field, overall, had adopted numerous types of transducers. This along with its varied application, makes a general comprehensive account virtually impossible in even one reasonably large text. This article, and **Article 210, Optical Transducers for Chemical Measurements, Volume 3**; **Article 211, Mass Spectrometry, Volume 3**; **Article 212, Chromatography, Volume 3**; and **Article 213, Electrochemical Measurements, Volume 3** introduce the main methods.

Texts on analytical measurements tend to concentrate on specific areas within the class: for example, on spectroscopy; on chromatography; on air or water purity; on process control instruments; and on principles of instrumental analysis. Given at the end of this article is a bibliography of representative works.

Chemical measurement presents significant problems of clear classification and definition. The discipline as a whole is large enough to support separation into areas that are themselves regarded as disciplines, examples being organic and inorganic chemistry. Locating information on how measurements are done can be difficult because all or part of many measuring methods overlap these divisions.

Also covered are those sensing systems that can distinguish and/or quantify different chemical species, that is, establish what chemicals are present in a sample and to what extent, usually as a percentage of the sample size. This is an empirical definition, and excludes, for instance, the chemical applications of density measurement.

Chemists classify their techniques partly by field of application, and partly by principle. The classification given here is restricted to the principle of the transducer; this will sometimes lead to problems. For example, one could

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

entertain a lengthy debate on the principle of mass spectrometry itself.

Space permits only a condensed summary to be provided on commonly available chemical transducers. Where possible, references are given to industrial measurement applications of chemical transducers, although these are not extensive compared with laboratory applications. A systems approach to chemical measurement will greatly assist in taking laboratory methods to the industry. Global markets for analytical instruments are the subject of regular surveys; they are a major sales line. The latest links to these services – they can be expensive – can be found on the Internet using search words 'analytical instrument market survey'.

2 CHEMICAL PROPERTIES OF MATTER

To provide a basis for considering chemical measurements, what is first needed is a quick overview of the basic nature of matter and its properties in the perspective of measuring their presence in a sample.

As is common knowledge, matter is composed of atoms formed into elements. The comparatively small size of these basic entities gives the first clue to the difficulties in sensing them. They are all very small in terms of the things into which they are built, and to which measurement is mostly needed. Typical atom sizes are in the subnanometer region. They are all but impossible to view via an image of them. They are investigated as a collection that is present in a sample and the measurement is expressed as a percentage of the whole. The electron microscope can just about detect some of them, but only in strictly controlled conditions of ultrahigh vacuum and after illuminating them in their steady state with an electron beam. Rarely do chemical detectors actually 'see' the individual mass parts at this level. The electron has a rest mass of 9.10×10^{-30} kg – it is particularly small as an entity to observe.

The mole is the unit of substance; it is commonly used in chemistry. It is defined as the amount of substance that contains as many elementary entities as there are atoms in 0.012 kg of carbon.

Atoms divide into a central mass called the *nucleus* that is formed by protons and neutrons, around which rotate a varying number of electrons. These elemental entities occur only as certain sets of combinations that are ordered according to the well-known periodic table of the elements. For example, the hydrogen atom has just 1 electron, 1 proton, and zero neutrons. The table is available from several sources using the search words 'chemical periodic table'; including http://www.chemicalcelements.com.

There are some 112 known elements; this number is still rising slowly because of the discovery of others in the nuclear isotope group. Since the discovery of radio isotopes in the early twentieth century, it has also been found that there are many subatomic particles other than those classically described above for the atom.

The number of electrons for an element follows a strict relationship as shown in the periodic table. The number is related to the assembly of nucleus parts, the neutron and the proton. The proton is 1836 times larger in mass than the electron, but that is still small for direct imaging or sensing as a lone item.

The manifestation of the electron is alternatively described as a charged particle spinning around the nucleus of the atom – the corpuscular theory. Alternatively, it is also considered as a statistically existing energy source that is present at a point only for some of the time – the wave theory. These alternative models of the electron are used as they best fit a particular situation; no single model has yet been developed that explains all of the known properties.

A key issue regarding the measurement of chemical properties is that all components of matter involve energy as well as mass. Again, the basic quantities involved are very small. Electrons possess the same fixed value of negative electrostatic charge of nominally 1.6×10^{-19}. As an ampere is a flow of one coulomb per second, it needs a vast number of them to cause measurable currents even when detection has only current levels less than a microampere available. The nucleus is positively charged to hold the element together by the forces of attraction.

The quantity of charge associated with electrically charged particles of known magnitudes therefore follow the discrete values for elements according to the number of electrons present in the element. The magnitude of forces that bind matter together follow known laws, making measurement feasible if the physical construction of material can be detected.

It is possible to temporarily add or remove electrons in the dynamic state of matter. These *ions*, as they are called, alter the charge levels again affording the opportunity to determine specific elements.

Moving up to large chemical entities, the elements combine in numerous ways to form molecules. With over 100 elements available in numerous combinations for forming molecules, and these in turn, forming sets of molecules and mixtures (those materials wherein the chemical components are not converted but retain their state), it is easy to see why there are innumerable chemical molecules and mixtures in existence.

To end this primer on the nature of matter, it is also important to remember that a material can exist in one or more of four states:

- solid
- liquid

- gaseous
- plasma.

The electrical properties of electrical conduction vary in these different states. With increase in temperature, most pass from solid to plasma state passing through each as the temperature rises. Some do not follow the ascendancy rule. Carbon, an element, for example, passes from solid to gas and cannot be liquified. These variations are often used to enable an analysis.

3 MEASURABLE CHARACTERISTICS OF CHEMICAL VARIABLES

Thus, measurement of chemical species in a sample can be carried out using one or more of many different principles that are based on energy, mass separation, and force sensing. The individual regions of the electromagnetic energy spectrum, ranging from longwave microwaves up through the infrared, visible, ultraviolet X-ray to the shorter radiation frequencies are also useful for their varying properties and effects on materials. It will be found that chemical instrumentation will fit one of the following classes:

- Makes use of direct measurement of an existing physical variable (passive measurement);
- Requires some kind of conversion to render the chemical parameter observable before a direct physical measurement can be employed (active measurement).

In general, a chemical measurement is usually made by detecting the differences between two or more entities; far less so is the absolute value determined. Given a baseline value as a standard, it becomes feasible to measure the makeup of a sample of matter. This leads to many possibilities. Table 1 lists some common methods.

Measurements can be made by observing the steady state levels of parameters of the sample, as with sensing the color of diluted blood to estimate the red cell level, this being done as a comparison against a reference material as the standard. Alternatively, they can be made by observing the properties of a chemical reaction that is taking place between the sample and a reacting medium, such as in the use of an ion sensitive probe. This kind of measurement consumes energy or mass in its operation, so it will have a given interval of usefulness unless the additional materials are replenished.

The time given to make an observation can often be very short; some being as short as a microsecond. To overcome this, if the reaction can be repeated continuously, the source is continuously reexcited and lockin amplifiers are used that retrieve the signal – *see* **Article 184, Photon Counting, Volume 3**.

Table 1. Examples of types of chemical measuring instruments.

Type of detection	Example of instrument class
Mass of atoms present in sample	Mass spectrometry
Energy of reradiation, after energy added at another spectrum location	Fluoroscopy
Energy emitted by sample as energy added	Emission spectroscopy
Energy absorbed as energy added	Absorption spectrophotometry
Conversion of energy spectrum by adding energy	Raman spectrometry
Migration rates of chemical species through a separation medium	Chromatography (liquid and gas)
Electrical charge variations in voltage and current terms	Voltametry, ion selective probes. Polarography.
Color value without addition of reagent	Colorimetry
Color change with addition of reagent	Colorimetry

The amplitude of the signals produced can be extremely small, some methods literally counting the photons (packets of energy) being emitted. Photons are energy packets given off when an electron jumps between orbits of an element. The threshold for this to happen is nominally 1.2 eV, meaning that with 1000 V as the electric potential the energy level is just 1000 electron charges per photon.

In general, a sound knowledge of the appropriate chemistry is needed to create chemical instruments and this has to be coupled with well-understood signal processing in the physical electrical regime. Interpretation of data then needs sound chemical appreciation. Crossing all of these discipline boundaries needs care and special experience.

Finally, in this overview is covered the need to appreciate that chemical species rarely exist in isolation from others. Even the purest materials will possess some level of impurity. The various methods available are rarely able to detect the required variable while ignoring all others present in the test. Many methods can give biased results depending on the type and magnitude of the contaminant present.

It is essential to first appreciate what spread of species are expected to be available in a sample, and then to set up a measurement system that will work properly with those of no interest being present. This also raises the converse problem, that of missing a key parameter because a suitable test was not done for it.

In the measurement of mineral assays, for instance, the analytical instruments chosen need to have their settings set to look for the species of interest. Not doing the right

tests can be a major failing; there was once the case of an assay for gold that did not recognize that there might be (and there was in high proportion!) antimony present in the sample whose economic value far exceeded that of the percentage of gold.

The usual activities of a chemical instrument are as follows:

- prepare the instrument to make it ready to measure;
- calibrate the instrument;
- prepare the sample ready for presentation/coupling to a sensing detector.

Up to this stage, the main techniques used are mechanical, hydraulic, chemical, and heat. Electrical methods and signal processing is not used other than to control any automation of these stages.

For example, a pollution gas monitor for a given gas will first purge its intake tubes and test chambers. A reference gas is then admitted to calibrate the sensor. The sample is then passed in and the values are recorded.

Then comes the following stages:

- presentation to the sensor;
- processing of signals in real time or off line.

These stages are more akin to physical instrument design.

In some cases, purging and calibration will be applied at the end of the test to set down drift reference points at each end of the time-varying process in order to allow removal of drift in the final determination.

A useful reference to have on hand when making these low-level chemical measurements is Armutat (2004).

4 TERMINOLOGY

Chemical entity description uses a notational method developed to describe the many forms of matter and is thus based in the physical description of energy. Some say it is all physics and nothing else! However, it has needed to constantly develop its own shorthand notation to cater for the great variety of structures and interactions involved. These groups are as follows:

- Names of chemical elements
  Example: The notation 'Ag' is the shorthand symbol for the element silver. This, and the other elements notations are found defined in the period table.
- Molecule description
  Example: The shorthand notation AgCl represents the molecule combining silver and chlorine in a given way and proportion – called '*silver chloride*'. Some molecules are so complex, possessing such a long name, that a shorthand form is used to reduce the writing needed, for example, 'polyvinyl chloride' has the acronym name of PVC.
- Relationships taking place in chemical reactions
  Example: When two or more chemical species are placed together in given circumstances they can react to form other species and by-products. An example reaction used as a standard reference half cell in electrochemical measurements is

$$\mathrm{AgCl} + \mathrm{e} \longrightarrow \mathrm{Cl} + \mathrm{Ag}$$

This means that each molecule of a large number of silver chloride molecules in the reference electrode has an electron added for which the result, after dynamic energy effects have died out, is that the silver and the chlorine are no longer combined as a molecule and are now free in their elemental forms. The electron provides the energy to break the bonding forces keeping the two together.

Interpretation of notations and equations, and what they represent, needs the skills of people truly experienced in chemistry. This discipline is not always taught in detail to engineers who design instruments. Conversely, chemists are not taught much about engineering design. A partnership between these two is essential for the creation and modification of chemical instrumentation.

Today, the computer is used extensively to model the structure of atoms, elements, molecules, and to compute complex reactions. Databases as handbooks and computer databases are maintained that contain the names and properties of chemical substances. A good start is http://chemfinder.cambridgesoft.com and the search words are 'chemicals databases'.

5 UNITS

Numerous different techniques are involved; many measurement units will be encountered in chemical measurements. These need care in interpretation for instrument designers are not always familiar with chemical quantities.

Commonly encountered basic units are as follows:

Mass: The kilogram is the preferred unit, but subdivisions are commonly used because of it being too large for most sample sizes. Often used are the microgram, milligram, and gram.

Amount of substance: The mole, symbol mol, is defined as the amount of substance of a system that contains as many elementary entities as there are atoms in 0.012 kg of carbon. This is met in reaction equations and in dilution proportions.

Electrical: Electrical field variables are also encountered – volts, amps, watts with generally very small magnitudes at the sensing stage. Capacitance is common but not so with inductance. Electric field variable are also commonly seen.

Radiation: The EM spectrum variables, such as wavelength, bandwidth, center frequency, and spectral graphs are used in many methods, as are optical energy levels as power, reflectance, absorption, and photon count.

Temperature: This is usually expressed in terms of the Absolute Temperature scale, with units of the kelvin, K.

Pressure: The SI metric unit is the bar, but in practice the more engineering forms such as the pascal are used. In many cases, the measurement is assumed to be in conditions of stable P, V, and T. This can be a trap. Placing an instrument inside a sealed insulated chamber to avoid pressure effects can lead to changes as the temperature varies.

Laws of gases relate their pressure, P, temperature, T and volume, V, as

$$PV/T = \text{constant}$$

where T is measured in kelvin.

Changing any one of these will alter others. The need to control this set of interactions in measurements of gaseous forms is obvious.

Units of measurement are discussed in more detail in **Article 9, Nature and Properties of Measurement, Volume 1**.

6 STANDARDS

As has been explained, the chemical measurement field employs numerous means to make measurements, and they are easily corrupted by poor practice and incorrect methodology of setting up and conducting the measurement.

Because of the importance of unwanted species being present, it is necessary to ensure that the test equipment does not introduce other species from foreign sources, such as sample tube contamination, or from residues left over from a previous test where the instrument parts have not been cleaned sufficiently. Purging all feed lines and vessels is usually an essential part of chemical instrument design.

Instrument calibration using samples of known purity is essential to avoid poisoning difficulties by the prior sample. They are also needed to set up the detection sensitivity to the required precision and accuracy. Few instruments are inherently calibrated by features of their design.

In chemical analysis, standards are required for both qualitative and quantitative analysis.

In qualitative analysis, only the purity and identity of the substance are important. Modern fine chemical suppliers of simple molecules usually assure these, although some doubts as to the optical purity (that is, presence of unwanted optical isomers) of samples can occur. These are best resolved by measuring the optical rotation or circular dichroism spectrum of the sample. Indeed, it will often only be in such measurements that optical purity will have significance. Most other chemical transducers are insensitive to optical isomerism.

Real problems and expense arise in quantitative analysis where it is necessary to obtain analyzed samples whose precision and accuracy of analysis is well known. For many common acids and bases, ampoules of carefully standardized solutions are commercially available. Mixtures of known composition are much harder to obtain and the analysis needs to be established by careful collaborative work, involving different laboratories and methods. Such samples, called *standard reference materials* (SRM), can sometimes be obtained commercially or from the appropriate national standards body (such as the National Institute of Science and Technology (NIST), US, and the Physikalisch-Technische Bundesanstalt (PTB), Germany). This applies particularly to standard ores, cements, or alloys, which are essential for the proper standardization of methods such as X-ray fluorescence spectrometry. It will often be necessary to hold a range of such standard materials. As an aside, it must be mentioned that it is possible to buy cylinders of standardized polluted air!

In organic analysis, the problems are much more diverse and it is true to say that rigorous standardization of organic analysis is less developed than it is in inorganic analysis. There is a tendency for industries requiring organic analyses to rely on empirical, often indirect, measures, rather than making use of the concentration of a specific substance. For this and other reasons, the commercial demands for precision and accuracy in quantitative organic analysis are usually not as rigorous as those needed for inorganic analysis.

International bodies, such as the Organization Internationale de Métrologie Légale (OIML), have laid down specific analytical methods and reagent purities for both organic and inorganic analysis. National standards bodies have also defined many standard analytical procedures for local contractual and legal purposes. These procedures are commonly classified according to the product of interest rather than the method, and often rely on the traditional empirical measures referred to above.

More detail on the standards of measurement are found in **Article 43, Units, Volume 1**.

RELATED ARTICLES

Article 174, Typical Signals Arising in Measurement, Volume 3; **Article 186, The Family of Signal Recovery Methods, Volume 3**; **Article 203, Resistance Measurement, Volume 3**; **Article 210, Optical Transducers for Chemical Measurements, Volume 3**; **Article 212, Chromatography, Volume 3**; **Article 213, Electrochemical Measurements, Volume 3**.

REFERENCE

Armutat, A. (2004) Low Level Measurements, Keithley Instruments. Free download on www.keithley.com.

FURTHER READING

Chan, C.C., Lee, Y.C. and Zhang, X.M. (2004) *Analytical Method Validation & Instrument Calibration*, Wiley-Interscience, New York.

Currell, G. (2000) *Analytical Instrumentation: Performance Characteristics and Quality*, Wiley, New York.

Ewing, G.W. (1997) *Analytical Instrumentation Handbook*, Marcel Dekker, New York.

Liptak, B.G. (1994) *Analytical Instrumentation*, CRC Press, Boca Raton.

Monk, M.S. (2001) *Fundamentals of Electro-Analytical Chemistry*, Wiley, New York.

Sherman, R.E. and Rhodes, L. (1996) *Analytical Instrumentation: Practical Guides for Measurement and Control*, ISA, Research Triangle Park.

210: Optical Transducers for Chemical Measurements

Ashutosh Sharma
University of California, Riverside, CA, USA

1 Principles of Optical Chemical Measurements 1427
2 Incident Radiation Methods for Chemical Measurement 1428
3 Intensity Measurement 1428
4 UV–Visible Absorption 1429
5 Atomic Absorption 1429
6 Use of Derivative Spectra 1430
7 Parallel Spectrometers 1431
8 Infrared Absorption 1432
9 Processing Optical Transducer Signals 1433
10 Fiber-optic Transduction 1434
References 1434

1 PRINCIPLES OF OPTICAL CHEMICAL MEASUREMENTS

As explained in **Article 209, Characteristics of Chemical Measurements, Volume 3**, chemical measurements make use of some form of optical interaction with the chemical species of interest. In use, the term *optical* usually includes the Electro Magnetic (EM) spectrum of radiation between the wavelengths 20 μm to 200 nm, which extends from the long Infrared (IR) wavelengths (20 μm and shorter) through the comparatively narrow visible band (red 700 to violet 400 nm) to the outer edge of the ultraviolet wavelengths (below 400 nm and extending out to the X-ray region).

Typically, interaction of an incident optical beam with a chemical species results in changes in its original characteristics. These changes observed in the emerging beam are exploited to identify, characterize, and quantitate a chemical species of interest. Parameters of interest are absorption, emission, wavelength shifts, or associated polarization changes, these being measured using a suitable optical radiation detector. Unlike a radio wave detector, which is capable of coherently measuring both amplitude and phase across the whole radio frequency spectrum, an optical detector is limited to comparatively narrow bands. Excellent photodetectors are now commercially available that detect

- energy within specific bandwidths; examples are photodiodes, photomultipliers, diode arrays, and charged coupled devices;
- total energy integrated as a whole; these are often called *bolometers*.

The performance characteristics of a light source and optical detector deteriorate with their use. It is therefore necessary to calibrate an optical chemical measurement instrument periodically using reference calibration curves, a standard reference material (SRM), (chemical actinometry), secondary reference (such as a standardized optical source or a frequency generator) and an electrochemical half-cell.

Optical systems can be quite sophisticated. A well-drawn optoelectronic element layout is essential to achieve desired optical beam profiles, including shape, energies, power and phase, and so on. Both active and passive elements are used in beam designing, shaping, and processing of

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

optical signals. A photosensitive device (detector) is used to convert an optical output signal into an electronic signal, which is then processed by using electronic circuits to obtain a meaningful output.

Old instruments widely employed photographic plates and later analog circuits. Modern instruments use digital signal processing and microprocessor technologies to achieve the desired output type. In all cases, several energy conversion stages are involved between the original optical signal and the output data – *see* **Article 108, Transducer Fundamentals, Volume 2**.

Instruments methods of optical chemical measurement may be classified into two types of measurement on the basis of

- incident radiation
- emitted radiation.

This article discusses incident radiation–based methods of optical chemical measurement.

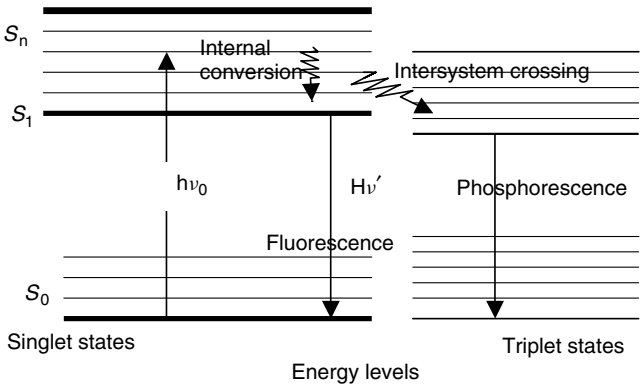


Figure 1. The interaction of a molecule with a photon of frequency ν_0. The molecule initially in a ground equilibrium energy state S_0, is promoted to a higher excited equilibrium state S_n. The molecule deactivates its energy via the *intersystem crossing* and after a time t, it returns to the ground equilibrium state S_0 emitting fluorescence. In favourable conditions the excited molecule may undergo intersystem crossing resulting in phosphorescence emission.

2 INCIDENT RADIATION METHODS FOR CHEMICAL MEASUREMENT

The scientific basis of an optical radiation–based chemical measurement is the interaction of radiation with a molecule.

Typically, a molecule upon interacting with a photon of frequency ν_0 in a ground equilibrium energy state S_0, is promoted to a higher excited equilibrium state S_n. After momentarily residing in this state (known as its *lifetime*), the molecule deactivates its energy via processes termed as the *intersystem crossing*. After a time τ, the molecule returns to the ground equilibrium state S_0.

During this deactivation process, the molecule looses energy resulting in a photon of new frequency ν' (Figure 1). Usually, $\nu' < \nu_0$ and the measurement process involves examination of the population of reemitted photons as compared to absorbed photons. Information such as an emitted photon's frequency, intensity and state of polarization, phase, and delay can then be used to infer the identity or concentration of the molecule.

Given here are some specific examples of optical transducers, classified according to the measurement made on the reemitted photon.

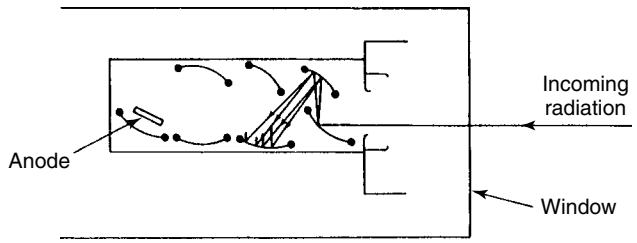


Figure 2. The photomultiplier: Incident radiation strikes an electron emissive surface; secondary electrons collide with successive electrodes in a many-from-one cascade process.

3 INTENSITY MEASUREMENT

The most common type of optical transducer is for the intensity measurement that is made by using a photodetector. The most widely used photodetector is a photomultiplier (Figure 2), which provides measurement capabilities for a varied photon flux. When larger numbers of photons are to be detected, a solid-state detector (for example, photodiode, photovoltaic cell) is sufficient. In the infrared region, the low energy of the photon necessitates careful

Table 1. Infrared detectors.

Name	Type	Detection sensitivity (cm $Hz^{1/2}$ W^{-1})	Response time (ms)
Bolometer	Ambient thermal	10^8	1.0
Golay	Ambient thermal	1.5×10^9	15
Thermocouple	Ambient thermal	1.5×10^9	15
Pyroelectric	Ambient thermal	10^8	2×10^{-3}
Photoconductor	Cooled semiconductor	10^9	10^{-1}–10^{-6}

choice of detector. Several photodetector types, together with their characteristics, are listed in Table 1.

In qualitative analysis, energy dispersion monochromation of the incident photon beam is usually carried out prior to intensity measurement. An exception to this is the X-ray fluorescence spectrometry. Some examples of optical systems are now given.

4 UV–VISIBLE ABSORPTION

The intensity of incident photons I_0 can be related to the transmitted intensity I by the Beer–Lambert law:

$$\log\left(\frac{I_0}{I}\right) = abc \tag{1}$$

where

a = molar absorptivity (a constant for a given species under given conditions) (g^{-1} cm^{-2})
b = cell path length (cm)
c = concentration (g cm^{-3})

This relation is commonly used by chemists in quantitative analysis: solutions of known concentration are used to plot a calibration curve, from which unknown solution concentrations may be read (see Hargis and Howell, 1980).

Most absorption measurement instruments are designed with a 1 × 1-cm cell to allow 1-cm path length of sample, thus simplifying calculations ($b = 1$ in the above relation). There are many variations in instrument design (Lue, 1979), depending on the applications that are available:

1. quantitative analysis, with infrequent wavelength changes and low spectral band-pass requirements (Figure 3);
2. quantitative analysis, with frequent wavelength changes and stringent spectral band-pass requirements (Figure 4);

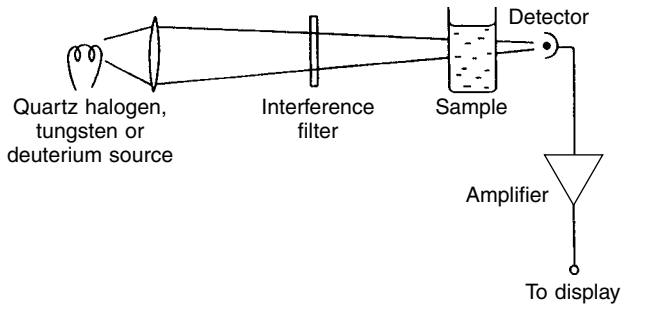


Figure 3. A simple colorimeter, or filter photometer, commonly used in repetitive quantitative analysis in the laboratory or process stream.

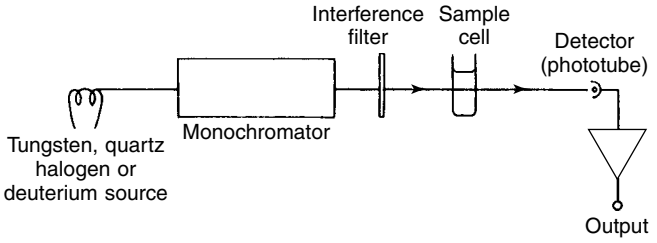


Figure 4. A single-beam spectrophotometer of a design used commonly in routine quantitative analysis.

3. qualitative analysis and structural studies, with wavelength scanning and stringent band-pass requirements (Figure 5).

Many other configurations are possible. Table 2 lists the major instrument components and their characteristics.

5 ATOMIC ABSORPTION

This special case of UV–visible absorption spectrometry is of particular importance because of its wide use in trace metal analysis (Koirtyohann, 1980), and because of the transducer design (Figure 6).

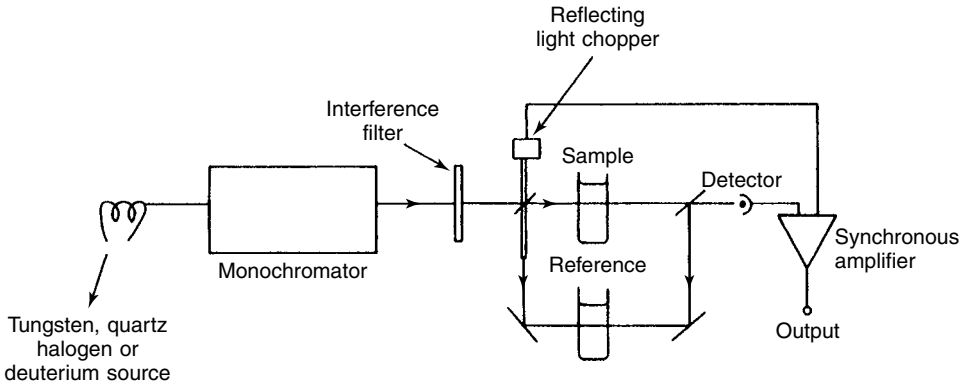


Figure 5. A double-beam UV–visible spectrophotometer of a design used for obtaining absorption spectra for qualitative analysis, structural work, and multicomponent quantitative analysis.

Table 2. Major instrument components in UV–visible spectrometers.

Item	Description
Light source	Visible: tungsten or quartz halogen lamp UV: hydrogen, deuterium, or (more expensively) xenon lamp Usually requires stabilized power supply
Monochromator	Almost always uses a reflection grating – quality of grating determines resolution. Optical design can reduce stray light
Filter(s)	Can be absorption (color) or interference; used to remove unwanted orders of diffraction due to grating
Modulator	May be rotating sector disc (chopper) or modulation achieved by switching the light source power supply
Detector	Solid state in simple instruments, phototube in intermediate, and photomultiplier in most sensitive instruments

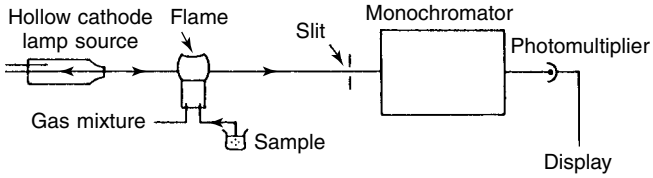


Figure 6. A simple design of atomic absorption spectrometer. Solutions are aspirated, nebulized, and swept into the flame, where neutral analyte atoms are formed. These absorb the characteristic radiation of the source by an amount proportional to the original solution concentration.

The light source is an atomic emission lamp (usually a hollow cathode lamp), in which the cathode is of the same metal as the analyte. Its emission spectrum is, therefore, coincident on the absorption spectrum of the vaporized element in the sample. Since atomic spectra are sharp line spectra, this provides an inherently high degree of freedom from interelement spectral interference. Thus, the monochromator needs only possess resolution sufficient to remove neighboring lines (Figure 7).

Detection is achieved via use of a modulated source being synchronized with this, a lock-in detector amplifier, *see* **Article 184, Photon Counting, Volume 3**. This approach removes much of the ambient light entering the sample or detector and interfering with the amplified signal; photomultiplier saturation must be avoided. The sample is vaporized either by flame, inductively coupled plasma, or in an inductively heated graphite furnace. This heating of the sample serves to remove the solvent and organic sample matrix (if any), but it also renders the analyte into the unionized atomic state, which then absorbs the source radiation.

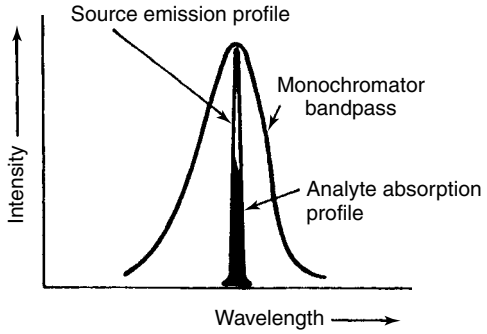


Figure 7. Spectral profiles of atomic spectra and the low-resolution monochromator typically used in atomic absorption spectrometry.

Atomic absorption of photons is primarily a quantitative tool, although the freedom from interelement interference can allow it to be used as a qualitative tool as well. The Beer–Lambert law applies strictly to atomic absorption in a flame; however, self-absorption and the vagaries of the sample vaporization process result in a nonlinear plot. A calibration curve is usually plotted from measurements on solutions of known concentrations.

Some common analytes, limits of detection, and relative precision are listed in Table 3. The major difficulty with the atomic absorption analysis method is interference due to the sample matrix, which can take many forms, including molecular band spectra, failure of the analyte to dissociate, matrix suppressions, and so on.

6 USE OF DERIVATIVE SPECTRA

One disadvantage of instruments designed to relate transmittance ($T = I/I_0$) to concentration using the Beer–Lambert law is the logarithmic dependence between the two. This restricts the sensitivity of simple absorption methods compared with say, fluorescence, where the relationship is approximately linear.

This problem can alternatively be expressed as the limited usable range of the change in concentration with transmittance (or specific response)

$$\frac{\mathrm{d}c/c}{\mathrm{d}T} = \frac{0.4343}{T \log T} \tag{2}$$

In practice, measurements above about 1.2 or below about 0.1 absorbance units will be difficult.

An alternative is to record $\partial^2 T/\partial\lambda^2$, the second derivative of transmittance with respect to wavelength. The relationship between the normalized second derivative of transmittance and concentration is now approximately linear:

$$-\frac{1}{T}\frac{\partial^2 T}{\partial\lambda^2} \simeq bc\frac{\partial^2 a}{\partial\lambda^2}$$

Table 3. Common analytes and limits of detection in atomic absorption spectrometry (AAS). Precision is typically 1 to 4% relative, depending on sample matrix and instrumentation.

Element	Wavelength (nm)	Flame type	Typical limit of detection ($\mu g\,cm^{-3}$)
Aluminum	309.3	N_2O/C_2H_2	0.2
Arsenic	193.7	Air/C_2H_2	0.4
Barium	553.5	N_2O/C_2H_2	0.3
Cadmium	228.8	Air/C_2H_2	0.002
Calcium	422.7	Air/C_2H_2	0.005
Chromium	357.9	Air/C_2H_2	0.008
Cobalt	240.7	Air/C_2H_2	0.02
Copper	324.8	Air/C_2H_2	0.002
Gold	242.8	Air/C_2H_2	0.02
Iron	248.3	Air/C_2H_2	0.02
Lead	217.0	Air/C_2H_2	0.02
Magnesium	285.2	Air/C_2H_2	0.0005
Manganese	279.5	Air/C_2H_2	0.002
Molybdenum	313.3	N_2O/C_2H_2	0.1
Nickel	232.0	Air/C_2H_2	0.02
Potassium	766.5	Ar/H_2	0.004
Rhodium	343.5	Air/C_2H_2	0.01
Selenium	196.0	Air/C_2H_2	0.2
Silicon	251.6	N_2O/C_2H_2	0.5
Silver	328.1	Air/C_2H_2	0.006
Sodium	589.0	Air/C_2H_2	0.001
Strontium	460.7	N_2O/C_2H_2	0.006
Tin	224.0	N_2O/C_2H_2	0.2
Titanium	264.3	N_2O/C_2H_2	0.4
Zinc	213.9	Air/C_2H_2	0.001

Note: N_2O = nitrous oxide, Ar = argon, C_2H_2 = acetylene, H_2 = hydrogen.

The problem of recording second derivatives is one of avoiding noise amplification. For gaseous pollutant studies, this has been done optically (Figure 8) (Kroon, 1978). Many modern recording spectrometers, however, now perform numerical differentiation on the digitized absorption spectrum. This is useful not only for extending the limit of detection in certain cases but also for removing baseline drift due to broad interfering absorption.

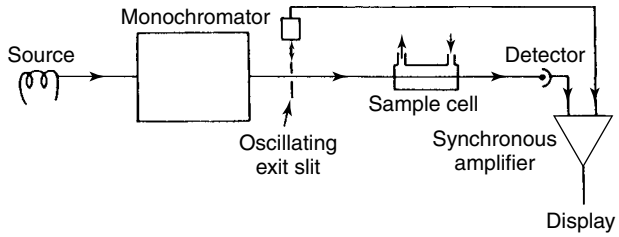


Figure 8. A second derivative spectrometer with optical differentiation. The wavelength of the source is modulated at frequency 2ω and is synchronously amplified. This is proportional to the second derivative of transmittance with respect to wavelength.

The effect of taking the second derivative with respect to wavelength is to emphasize those regions of the spectrum that have a rapid change in transmittance with wavelength that is sharp peaks. This increases the ease with which species that have interfering absorption spectra may be resolved.

7 PARALLEL SPECTROMETERS

The conventional UV–visible recording spectrometer, using a continuum light source, monochromator, and scanning mechanism is still widely used. However, the availability of photodiode array detectors with the requisite sensitivity has now made possible parallel spectrometers, in which a large region of the spectrum is recorded simultaneously by allowing the dispersed radiation to fall on a high resolution detector array after it has passed through the sample (Figure 9).

This design has many advantages which are

1. accumulation is possible of the entire spectra in short time intervals, thus facilitating, for example, the measurement of spectra on the chromatograph eluent;

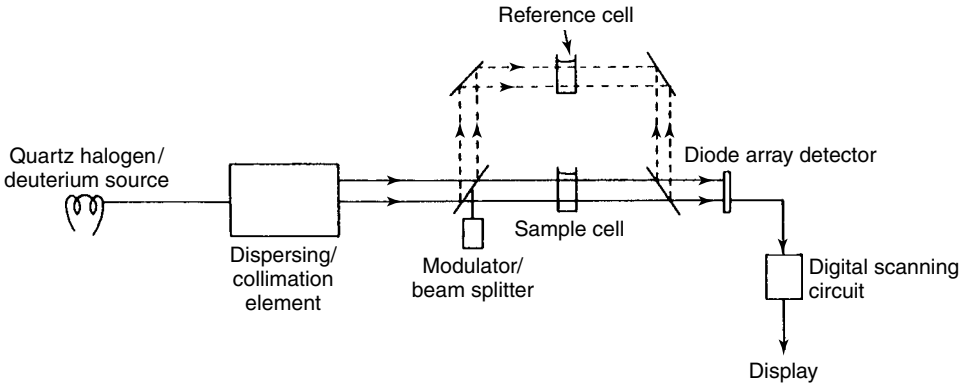


Figure 9. A parallel UV–visible spectrometer. The dispersed radiation is not resolved by slits into small wavelength regions, but is allowed to fall on a photodiode array. This allows simultaneous measurement of intensity at all wavelengths.

2. reduction of instrument noise and complexity through greater use of solid-state components;
3. absence of a monochromator and slits, with a resultant increase in light throughput and relative simplicity of optical design.

Advances in the technology of photodiode arrays have also meant better signal-to-noise ratios in light measurement, and a reduction in blooming (the effect where one light-sensitive element scatters light into a neighboring element, thus decreasing the effective resolution of the spectral information).

Conventional scanning instruments can measure with an absorbance precision of typically no better than 10^{-3} units. Careful design in a parallel spectrometer can reduce this to 10^{-4} units. This has important implications for multicomponent analysis of compounds with overlapping spectra. The least squares analysis of such spectra can be extended to species with almost-coincident spectral profiles, and to mixtures of 6 to 8 components.

Such advances have restored UV–visible spectrometry to its former preeminence as a general method of trace inorganic analysis.

8 INFRARED ABSORPTION

Initially, infrared spectrometry was used mainly in qualitative and structural chemistry. Several instrumental developments changed this situation radically. They were

1. introduction of low cost, large bandpass, single-beam instruments (Figure 10);
2. development of differential reflectance instruments for solids analysis (Figure 11);
3. introduction of better to use, and more affordable, infrared detectors, and sources (notably, for the latter, the use of diode lasers).

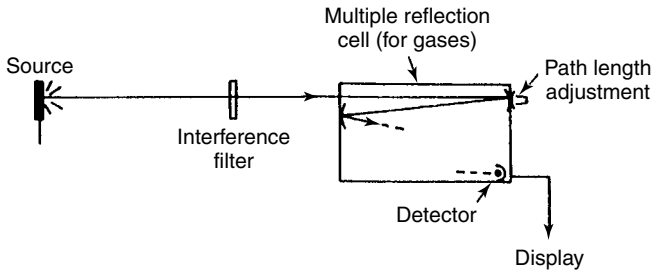


Figure 10. A simple single-beam infrared spectrometer suitable for quantitative analysis. The filter may be fixed or variable; in the latter case, a crude spectrum can be obtained, but the chief use is in quantitative work, where the high energy throughput gives a low limit of detection.

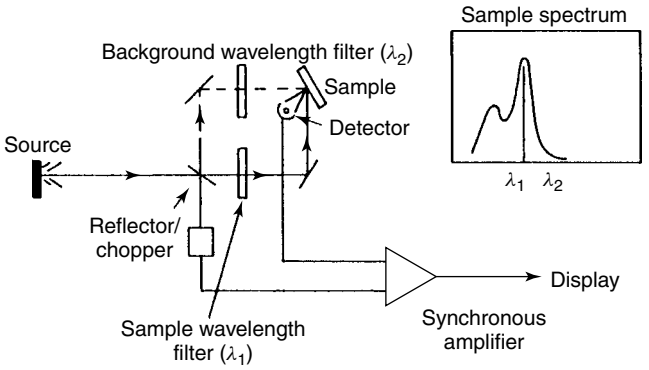


Figure 11. A differential reflectance infrared spectrometer suitable for rapid quantitative analysis in solid samples. The diffuse scattering caused by the uneven solid sample requires that the absorbance at the analyte wavelength be corrected by a simultaneous baseline absorbance (see inset).

Instrumental design for the infrared has always been plagued by the low energy of the infrared photon. This limits the sensitivity of semiconductor detectors, which rely on the incident photon promoting an electron into the conduction band.

Efforts to use thermal detectors meet with another problem: their slow response limits source modulation frequencies to regions where ambient infrared contributes significantly to instrument noise. The lower the frequency of a noise signal, the more difficult it is to filter out a reasonable averaging time. Nevertheless, the conventional double-beam scanning design has found wide application in chemistry. Table 4 lists some of the common components of this design and their characteristics.

Relatively recent developments of the instrument, concomitant with digital recording and control, include an online spectral library search database, and multicomponent mixture analysis with derivative spectra facilities.

Table 4. Major components of scanning infrared spectrometers.

Item	Description
Source	Electrically heated element, for example, Globar, Nernst glower
Monochromator	Reflective grating
Modulator	Rotating sector mechanical disc (low rotation speeds used because of large time constant of some detectors). Switching of source possible only with lasers
Sample cell	Usually sodium chloride (fragile, water soluble); can use silica for some spectra, also calcium fluoride. Sample may be pressed into potassium bromide disc
Detector	See Table 1

Table 5. Common diode lasers[a].

Type	Available range (cm^{-1})	Typical tuning range (cm^{-1})	Continuous power (mW)
$Pb_{1-x}CD_xS$	5000–2600	50	0.25
$PbS_{1-x}Se_x$	2600–1200	200	0.50
$Pb_{1-x}Sn_xTe$	1500–300	50	0.10
$Pb_{1-x}Sn_xSe$	1200–300	50	0.10

[a]Changing the doping of a given semiconductor type alters the range that the diode will cover. For a specific diode, tuning is achieved by varying the temperature within a cryogenic range.

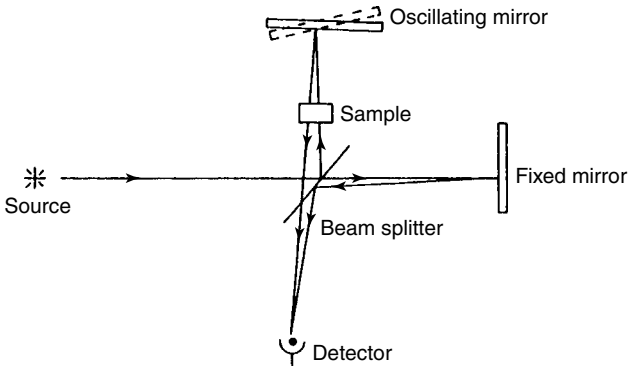


Figure 12. A Fourier transform infrared spectrometer. The sample spectrum, as the Fourier transform, is obtained in a time of the order of a mirror oscillation. Resolution of wavelength is very high (of the order of 0.01 cm^{-1}).

The use of tunable diode lasers, while bringing higher routine resolution to infrared spectrometry, also brings instrumental complexities, such as the need for cryogenic tuning and multiple sources. Table 5 lists some of the more common diode lasers, their wavelength range, and typical power output.

Fourier transform infrared spectrometers (Figure 12) offer resolution higher than conventional spectrometers. The interferometric principle also gives high optical throughput Fellgett's advantage, and a very short scan time, which makes them suitable for chromatograph eluent monitoring (*see* **Article 212, Chromatography, Volume 3**).

9 PROCESSING OPTICAL TRANSDUCER SIGNALS

Setting up both optical and electronic aspects of electro-optical systems is definitely more of an art than science, unless considerable design support is available to allow use of sophisticated computer modeling.

Computer tools have now used to perform otherwise laborious ray tracing work that is needed to establish how much energy exists in positions of interest, and to decide the level of optical aberration due to less than perfect optical elements and types of elements used. Today, a 10-element zoom lens system design is done in hours using computer tools, when by long hand methods it used to take tens of man-years of effort.

Electrical regime modeling is also easier now at least once the model is operating entirely in that regime.

The difficult area is the interface wherein the optical energy is converted. This is because the material properties, optical signal levels and stray radiation all can rarely be modeled to better than 10 to 30% uncertainties.

The development of these systems requires good modeling tools, a costly experimental optical laboratory and a large number of optical high-grade 'optical quality' optical elements. Educational grade parts (those with greater than 1 wavelength defects) are rarely usable for they can introduce significant errors. For example, an uncoated lens can act as weak Fabry-Perot interferometer introducing fringes in the final image, by virtue of their own two surfaces being an interferometer. Lack of flatness of windows can cause distortion in images that give incorrect levels in array detectors.

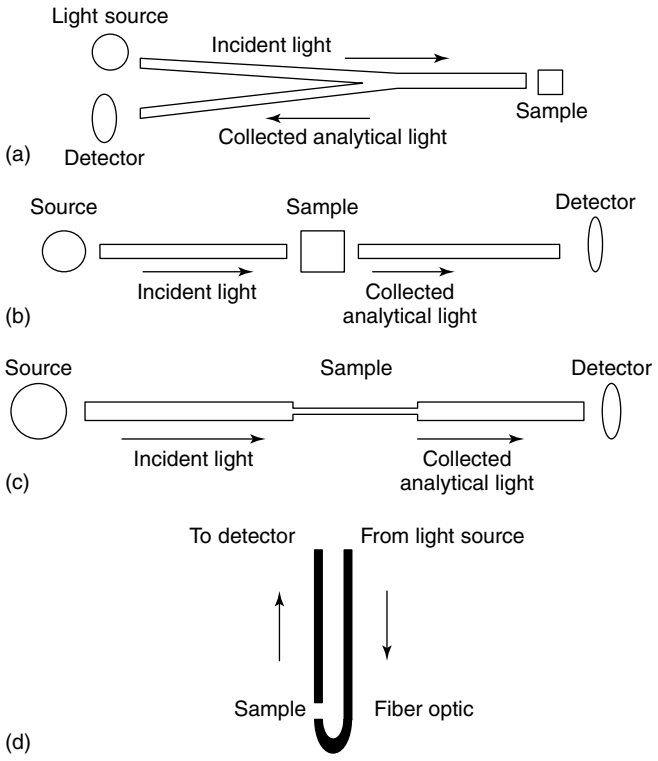


Figure 13. Various fiber-optic light guide configuration for incident light interaction measurement (a) bifurcated fiber optic, (b) two fiber split configuration, (c) evanescence field interrogated fiber-optic configuration and (d) U-shape fiber-optic cell configuration.

The difficulties of designing such systems have been covered in Hobbs (2000); see also Smith (2000), and Weber (2002).

10 FIBER-OPTIC TRANSDUCTION

With the advances of fiber-optic technology, a number of molecular measurement applications have chosen to use this outstanding and flexible means to transport photons to and from a sample. Increasingly, UV–absorption spectrometers are now available that offer fiber-optic attachments or have a fully integrated fiber-optic channel for delivery and collection of light energy, both types allowing analyses of remote chemical samples. Usually, one fiber carries the incident light beam from the light source to the sample, and the other carries the analytical beam from the sample to the detector optics. Some times these fiber designs are referred to as *sensors*. Various fiber-optic sensor configurations for analytical applications are explained in Sharma and Schulman (1999), (Figure 13).

REFERENCES

Hargis, L.G. and Howell, J.A. (1980) Ultraviolet and Light Absorption Spectrometry. *Analytical Chemistry*, **52**, 306R–23R.

Hobbs, P.C.D. (2000) *Building Electro-Optical Systems: Making It All Work*, Wiley-InterScience, New York.

Koirtyohann, S.R. (1980) Current Status and Future Needs in Atomic Absorption Instrumentation. *Analytical Chemistry*, **52**, 736A–8A.

Kroon, D.J. (1978) The Analysis of Ambient Air. *Journal of Physics E: Scientific Instruments*, **11**, 497.

Lue, J.T. (1979) Wavelength Modulated Spectrometer Implemented with Photodiodes. *Journal of Physics E: Scientific Instruments*, **12**, 833.

Sharma, A. and Schulman, S.G. (1999) *Introduction to Fluorescence Spectroscopy*, Wiley-InterScience, New York.

Smith, W.J. (2000) *Modern Optical Engineering*, McGraw-Hill Professional, New York.

Weber, M.J. (2002) *Handbook of Optical Materials*, CRC Press, Boca Raton, FL.

This article is based in part on a chapter by R. Watts originally published in Handbook of Measurement Science Vol 2, P.H. Sydenham, Wiley, 1983.

211: Mass Spectrometry

Peter H. Sydenham
GSEC Pty Ltd, Adelaide, South Australia, Australia

1 Principles of the Mass Spectrometer 1435
2 Theory of the Mass Spectrometer 1435
3 Mass Spectrometer Designs 1436
Related Articles 1437
References 1437

1 PRINCIPLES OF THE MASS SPECTROMETER

Mass spectrometry is the process whereby a molecule in the vapor state is broken into charged fragments that are then separated according to their mass/charge ratio, being detected according to their number. For the basics, see www.chem.vt.edu/chem-ed/ms/ms-intro.html.

Mass particles flowing through a magnetic or electric field are acted upon by the electric forces being continuously deflected as long as they are in the presence of the field.

A sample, in solid, liquid, or gaseous state, is processed into the gas-phase ion state. This is then beamed as emitted protons towards a target. As the mass particles pass through a uniform electric or magnetic field separator region, the various mass particles present will be deflected apart according to their atomic mass/electron charge (m/e) ratio.

The physics of this process is well understood in terms of engineering dimensions, field type, and projection method. Many variations on the basic theme exist. Over time they have become very sophisticated (Nier, 1998).

2 THEORY OF THE MASS SPECTROMETER

The resolution R of the ion separation power of a mass spectrometer is defined as

$$R = \frac{m}{\Delta m} \tag{1}$$

where m is the ion mass and Δm is the difference in mass between two resolvable peaks in a mass spectrum. For example, a mass spectrometer with a resolution of 1000 can resolve an ion with an m/e ratio of 100.0 from an ion with an m/e ratio of 100.1.

A mass spectrometer is therefore a complex set of processing modules for fragmenting, resolving, and detecting. It can, nevertheless, be viewed, as are optical spectrometers, as a transducer for converting a unique chemical structure into a highly individual fingerprint of response versus mass/charge ratio of its daughter fragments.

Figure 1 illustrates the structure of a mass spectrometer in simplified block form. Table 1 lists some common types of fragmentation systems. It may be added that the task of bringing the molecule into the vapor state prior to fragmentation is usually accomplished by using a gas chromatograph – *see* **Article 212, Chromatography, Volume 3** – as a sampling and/or separating device. Samples may also be introduced by direct injection.

The fragments are accelerated by a series of electrostatic slits to pass into the mass–charge separator. The various beam products emerging at different angles are detected by setting the beam or detector in the various locations, by scanning the detector across the region using multiple detectors, or by beam manipulation to scan it across the

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

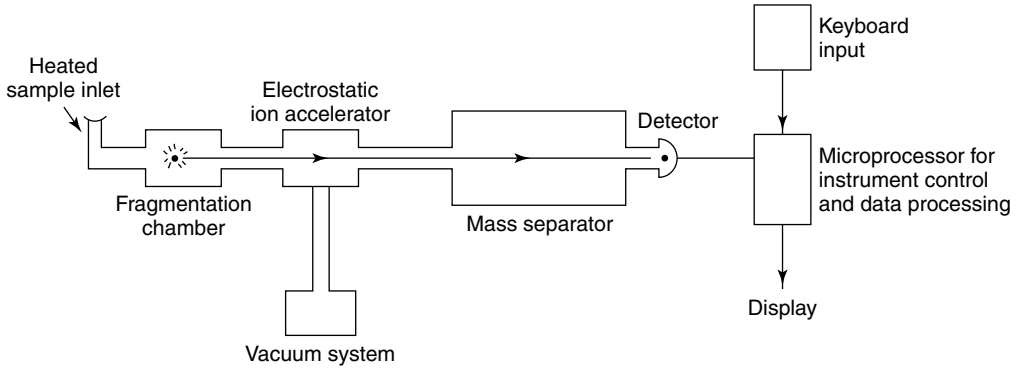


Figure 1. The principal components of a mass spectrometer.

Table 1. Fragmentation systems used in mass spectrometry.

Type	Description
Electron impact	Molecules bombarded with 70 eV (usually) electrons; fragmentation is violent, and molecular ion may not be observed
Chemical ionization	An ionized reactant gas (e.g. methane) forms reactive intermediates with itself; these react with sample to produce a simple spectrum. Quasi-molecular ion (at molecular weight plus one) often formed
Spark source ionization	RF spark (e.g. 800 kHz, 100 kV) applied to high melting point solids; equal sensitivity for all elements; high resolution spectrometer required because of wide energy spread of ions; useful for minor elements in semiconductors
Laser microprobe	Focused pulsed laser (e.g. 10^8 W cm^{-2} on 50 μm^2) used to vaporize sample; electron impact used to fragment; time of flight spectrometer used; can use small samples and plot element distributions

detector. A detector array can read several beam products at once.

3 MASS SPECTROMETER DESIGNS

The most common types of separators are listed in Table 2. Resolution here is defined as

$$R = \frac{m}{\Delta m} \tag{2}$$

where Δm is the mass difference between peaks at m and $(m + \Delta m)$ when the peaks overlap at 10% of their mean height (this is referred to as 10% valley definition).

Detection is accomplished using one of the devices listed in Table 3.

A mass spectrum can be used for the following:

Table 2. Mass separators used in mass spectrometry.

Type	Resolution (see text)	Comment
Magnetic deflection:		
(a) Single focusing	200–600	Available with permanent magnet
(b) Double focusing	to 35 000	
Quadrupole	400–1200	Rapid, linear scan; simple relative to magnetic types
Time of flight	400–600	50-μs scan time; suitable fast kinetic studies

Table 3. Detectors used in mass spectrometers.

Type	Description
Faraday cup	Ions impinge on insulated cup and absorbs an electron – results in ion current, for example, 10^{-12}–10^{-15} A
Electron multiplier	Beryllium–copper dynode chain used to amplify primary emitted electron caused by impact of fragment ions; for example, 10^{-18} A limiting current
Channeltron	Secondary electron multiplier in the form of fine tubes coated with internal layers of secondary emissive material–typical limiting current 10^{-19} A

- Qualitative analysis
- Quantitative analysis
- Structural work.

There are two common ways in which qualitative analysis can be performed:

1. Using the known abundances of isotopes, together with rules derived from the valence (binding power) of atoms, to fix the numbers of atoms of a given type in a molecule whose spectrum is given. This method depends on having available accurate mass/charge ratio measurement and high resolution.

2. Using the observed mass spectrum as an empirical fingerprint of the molecule, which may then be compared with a library of spectra using a statistical matching criterion. Such is the individuality of the information about a molecule contained in a mass spectrum that this method is most demanding of instrument reproducibility rather than resolution. Libraries of the spectra are held in the computer associated with these instruments.

As with many other spectral methods, quantitative analysis is achieved by comparison of peak heights or areas per unit of standards against those of unknowns.

The inherent high resolution of mass spectra allows multicomponent analysis, particularly in gas mixtures, this being done rather more easily than in common forms of optical spectroscopy.

Structural work is of limited interest in analysis and will not be discussed further here.

A useful process monitor based on mass spectrometric design is the quadrupole gas analyzer. This is a simple quadruple mass spectrometer designed for quantifying low mass molecules (up to say, molecular mass 100). Samples are pumped directly from the process stream through a restrictor into the fragmentation chamber; the mass separator is merely required to separate whole number masses. A resolution of 150 is often adequate. Quantification can be obtained by monitoring the parent ion current (that is, the detector current as the molecular mass of the compound).

Separation chamber designs have been made in numerous forms. The challenge of making the separation large for small-length separator units has resulted in multiple-folded separator systems.

Ardrey (2003), Hoffmann and Stroobrant (2002), Siuzdak (1996), and Watson (1997) give detailed information on mass spectrometers.

RELATED ARTICLES

Article 209, Characteristics of Chemical Measurements, Volume 3; **Article 210, Optical Transducers for Chemical Measurements, Volume 3**; **Article 212, Chromatography, Volume 3**; **Article 213, Electrochemical Measurements, Volume 3**.

REFERENCES

Ardrey, R.E. (2003) *Liquid Chromatography – Mass Spectrometry: An Introduction*, John Wiley & Sons, New York.

Hoffmann, E.De. and Stroobrant, V. (2002) *Mass Spectrometry: Principles and Applications*, John Wiley & Sons, New York.

Nier K.A. (1998) Spectrometer – Mass, Entry, in *Instruments of Science: An Historical Encyclopedia*, (eds R. Budd and D.J. Warner), Garland, Taylor and Francis, London (pp. 552–556).

Siuzdak, G. (1996) *Mass Spectrometry for Biotechnology*, Academic Press, London.

Watson, J.T. (1997) *Introduction to Mass Spectrometry*, Lippincott Williams & Wilkins, London.

This article is based in part on a chapter by R. Watts originally published in Handbook of Measurement Science Vol 2, P.H. Sydenham, Wiley, 1983.

212: Chromatography

Brett Paull
Dublin City University, Dublin, Ireland

1 Principles of Chromatography 1438
2 Liquid Chromatographs 1438
3 Gas Chromatographs 1440
4 Column Characteristics 1441
5 Detectors Used with Chromatographs 1441
6 Processing Chromatographic Signals 1443
7 Performance and Limitations of Chromatographs 1443
Related Articles 1444
Further Reading 1444

1 PRINCIPLES OF CHROMATOGRAPHY

The chromatographic separation process occurs because of repeated sorption/desorption of sample components onto and off a stationary phase surface as they are carried over or through that stationary phase by a mobile phase. The relative affinities of these sample components for both the stationary phase and the mobile phase govern how much time is spent in each phase and, therefore, how long it takes for each component to travel over or through the chromatographic system. The higher the affinity of a particular component for the stationary phase, the more time it will spend being stationary. Equally, those components with a low affinity for the stationary phase will travel through the chromatographic system at a rate close to that of the mobile phase.

The stationary phase itself can be either a solid (adsorption chromatography) or a liquid (partition chromatography) coated or chemically bonded to a solid. Planar Chromatography is the name given to the mode of chromatography where the stationary phase is immobilized on a flat planar surface. An example of planar chromatography is (TLC) Thin Layer Chromatography, where the stationary phase, usually alumina or silica gel, is immobilized on a glass plate. Alternatively, the stationary phase can be immobilized or packed within a column. Techniques involving such columns are grouped within the umbrella term *Column Chromatography*. The common column chromatographic techniques form the focus of the following sections.

The mobile phase in column chromatographic methods is usually either a gas or a liquid. Hence, we have either (GC) Gas Chromatography or (LC) Liquid Chromatography. The former is often further characterized as either (GSC) Gas–Solid Chromatography or (GLC) Gas–Liquid Chromatography depending on whether a solid or liquid stationary phase is used.

An alternative to the above two types of mobile phases exists in the form of supercritical fluids. These represent an intermediate phase between a liquid and a gas and exist at certain elevated temperatures and pressures. The technique of (SFC) Supercritical Fluid Chromatography has some advantages over both GC and LC but has not found such widespread application due to cost and complexity of the required instrumentation. In LC, the liquid mobile phase is also called the *eluent*, prior to passage through the column, and the *eluate*, after it has emerged from the end of the column. The emergence of the liquid mobile phase from the column, with the sample components contained therein, is called *elution*.

2 LIQUID CHROMATOGRAPHS

High-performance Liquid Chromatography (HPLC) is the name given to the form of LC carried out using

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

high-performance grade stationary phases (generally, 5–10-μm stationary phase particles) packed into stainless steel or polymer columns, with mobile phases delivered through the stationary phase under high pressures (generally between 500 and 3000 psi). As the solutes emerge from the column, they are carried by the mobile phase through a suitable detector equipped with a flow cell (predominantly, either an (UV–Vis) ultraviolet–visible absorbance spectrophotometer or an electrochemical detector being monitored to produce a chromatogram).

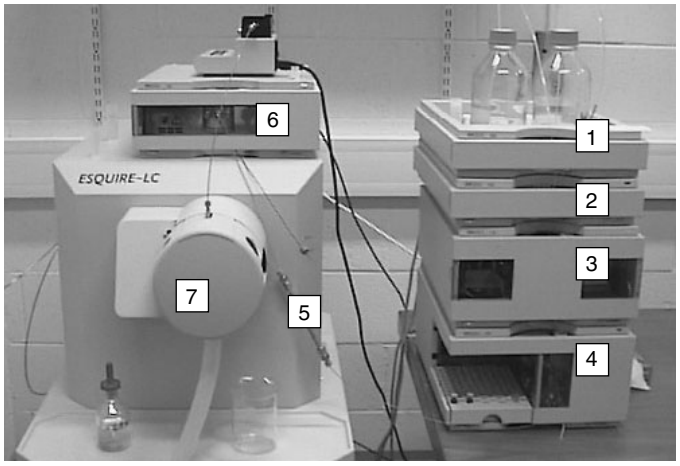
The instrumentation used to perform HPLC is generally modular in type and collectively termed a *high-performance liquid chromatograph*. The complete system generally consists of a mobile phase reservoir and delivery module, a sample/standard injector module, a guard column and the analytical column, a detector module, and, finally, some form of data recorder to produce the resultant chromatograms. A schematic diagram of a typical HPLC system is shown alongside a photograph of a commercial modular HPLC system in Figure 1 below.

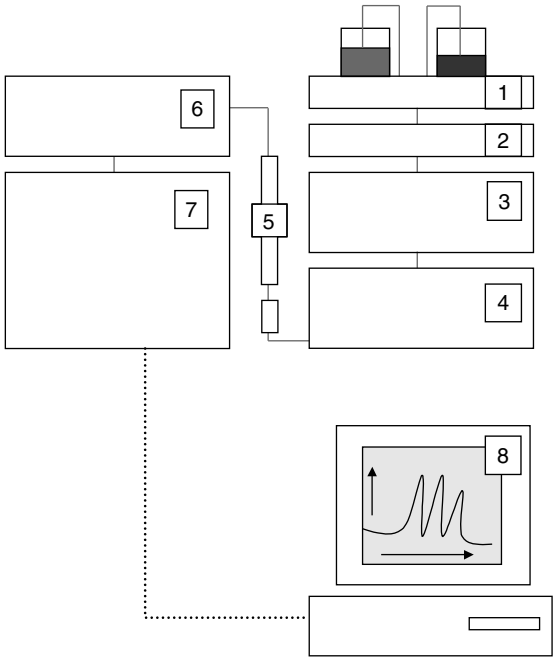
In almost all modern HPLC systems, the mobile phase is delivered to the analytical column via the injector module by one or more high-pressure HPLC pumps. Modern pumps can operate at variable flow rates, generally between 0.1 and 10 ml min^{-1}, with most mobile phase flow rates being between 0.5 and 2.5 ml min^{-1} when using standard HPLC columns.

The majority of modern HPLC pumps are based upon either single or dual alternating pistons. With dual piston pumps, the pistons are operated 180° out of phase to ensure a pulse free flow. With single piston pumps, a pulse dampener device is often included to reduce any such mobile phase pulsations. HPLC pumps can be further classified as either being isocratic or gradient pumps, meaning they are capable of delivering either a single mobile phase of constant composition or a mobile phase whose composition can be changed automatically over time.

Injection of samples and standards onto the column in HPLC is problematic as the mobile phase into which the sample is to be injected is flowing under the high pressure and cannot be stopped during sample injection. In addition, the injection procedure used must be capable of reproducibly injecting microliter quantities of sample. Most modern instruments circumvent this problem with the use of six-port switching valves, which contain a sample loop of fixed volume (generally 5–100 μL) that is filled with sample using a simple syringe. These valves can be switched manually or automatically to quickly insert the filled sample loop into and out of the path of the mobile phase without affecting the flow or pressure of the mobile phase.

The actual HPLC columns and various types of flow-through detectors used in HPLC, which together with the





1 Mobile phase reservoirs
2 Mobile phase degasser module
3 Mobile phase pump module
4 Auto-sampler and injector module
5 Guard column and analytical column
6 UV/Visible absorbance detector module
7 Mass spectrometer detector
8 PC with chromatography data capture software

Figure 1. Schematic diagram of a typical HPLC system alongside a photograph of a commercial modular system with spectrophotometric and mass spectrometric detector.

above make up the complete HPLC instrument, will be discussed in Sections 4 and 5 below.

3 GAS CHROMATOGRAPHS

As mentioned in Section 1, the mobile phase in GC is an inert gas, generally nitrogen, helium, or hydrogen. Just like the liquid mobile phase in HPLC, within a gas chromatograph, the mobile phase (more commonly termed the *carrier gas*) passes through a sample injection port; then through the GC column, which is held within a thermostated oven; and finally, out through a flow-through detector. The instrumentation required to carryout GC could be viewed as being simpler than that required for HPLC as there is no requirement for high-pressure pumping systems. Instead, the gas is delivered from a pressurized source, commonly a pressurized cylinder supply, via a moisture trap to remove any traces of water, directly into the gas chromatograph. This is possible because of the much lower viscosity and increased difussivity of a gas compared to a liquid, which means the pressure drop across the GC column is much less than that found in HPLC.

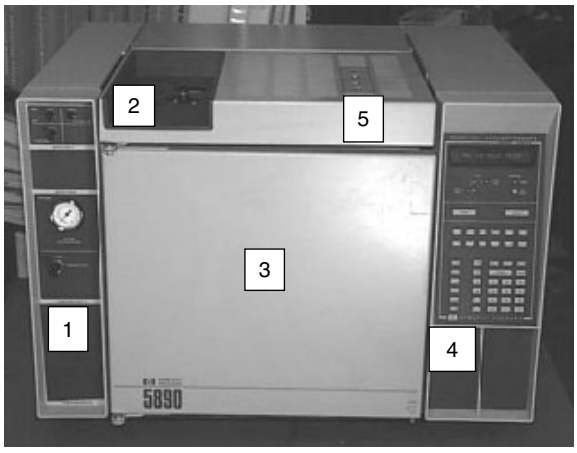
This also means the length of the GC column is less restricted by this pressure drop than a standard HPLC column, allowing open tubular columns typically of 50-m length to be used (for more details on gas and liquid chromatography columns, see Section 4).

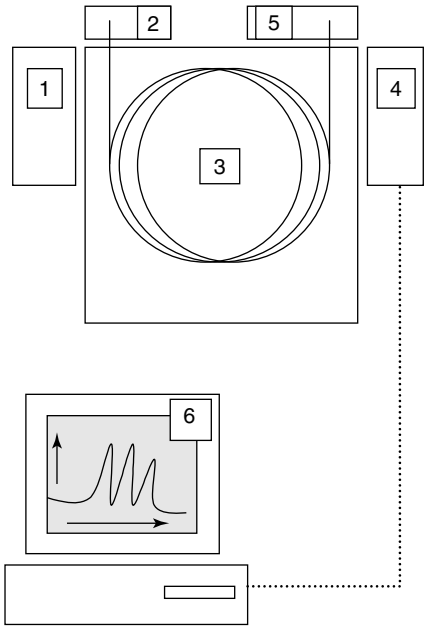
Unlike HPLC, most gas chromatographs are nonmodular single unit designs, although modern instruments will have optional modules, such as autosamplers.

Figure 2 shows a schematic diagram of a typical gas chromatograph alongside a photograph of a commercial system.

Sample injection in GC is perhaps the most complex aspect of the instrumentation. In most common applications of GC, the sample has to be in a gaseous or vapor state prior to entering the chromatographic column. This limits GC itself to the separation of volatile solutes. Therefore, the injector in GC must act not only as an inlet for the sample itself but also to vaporize and mix the sample with the carrier gas before the sample enters the head of the analytical column. To achieve this, the injector is subject to temperature control and is generally operated at 50 °C above the boiling point of the least volatile solute.

Most modern applications of GC utilize capillary columns (discussed in more detail in Section 4), which are only capable of accepting minute quantities ($<1\,\mu L$) of sample before becoming saturated. Therefore, injectors capable of reproducibly and accurately injecting such small volumes have been developed. Split/splitless injection is a technique that allows the analyst to inject larger volumes of sample while controlling by means of a vent valve the proportion of the injected sample that actually reaches the head of the column. Split ratios are used to describe this diverted proportion of sample, with split ratios varying from 1 : 20 to 1 : 500, with the latter indicating that





1 Carrier gas inlet, pressure regulator and flow meter
2 Thermostated sample inlet and injector
3 Analytical column within thermostated column oven
4 Injector, oven, and detector temperature control panel
5 Thermostated detector and carrier gas\sample outlet
6 PC with chromatography data capture software

Figure 2. Schematic diagram of a typical gas chromatograph alongside a photograph of a commercial system.

only 0.2% of the volume of sample injected actually enters the column. The remaining 99.8% is carried to waste by vent gas.

As shown in Figure 2, the column is housed in a temperature-controlled oven, which can be rapidly heated or cooled, from 40 to 450 °C, with most modern chromatographs stabilized to within 0.5 °C. The temperature of the column can be ramped up or down over time, allowing temperature gradients to be used to speed up or slow down elution of solutes.

Upon exit from the column, the carrier gas transports the solutes to the detector where their relative concentrations in the carrier gas are determined. The various types of detectors commonly used in GC are discussed in Section 5.

4 COLUMN CHARACTERISTICS

4.1 HPLC columns

All HPLC columns are straight tubes made of either stainless steel or, less commonly, inert polymers such as (PEEK) poly ether ether ketone. Typical dimensions range from 3.0 to 25.0 cm in length and from 0.2 to 0.5 cm in (ID) internal diameter. The stationary phase substrate is pressure packed within the column and held in place using porous frits (0.2–0.4-μm pores) and pressure-tight end fittings.

The majority of separations using HPLC are achieved using silica gel–based stationary phases. These may be bare silica, which is generally applied to the separation of polar solutes using a nonpolar mobile phase, a mode of HPLC termed *normal phase HPLC*, or bonded silica, where organic functional groups are covalently bonded to reduce the polarity of the stationary phase. Commonly, these organic groups have linear alkyl chains from 2 to 18 carbon atoms, with most HPLC applications using C_{18} bonded silica gels, otherwise known as (ODS) *octadecyl-silica*. These ODS columns are applied to the separation of nonpolar solutes, using polar mobile phases, in what is termed *reversed phase HPLC*. Table 1 shows the range and description of some of the large number of HPLC bonded phases available.

Table 1. HPLC bonded phases.

Type (bonded group)	Description	Typical solutes
Silica gel	Classic normal phase material	Polar nonionic organic compounds
Nitro (NO_2)	Normal phase material	Aromatic compounds, compounds with double bonds
Phenyl (C_6H_5)	Reversed phase material	Large proteins, macromolecules
Butyl (C_4)	Reversed phase material	Aromatic compounds
Hexyl (C_6)	Reversed phase material	Nonpolar/moderately polar solutes
Dodecyl (C_{12})	Reversed phase material	Nonpolar/moderately polar solutes
Octadecyl (C_{18})	Classic reversed phase material	Wide application – nonpolar organic compounds

4.2 GC columns

There are two main types of column used in GC, these being

1. columns that are pressure packed with a stationary phase or stationary phase support material in a similar fashion to HPLC columns; and
2. columns that are not packed and thus open tubular in nature, with the inner walls acting as the support for the stationary phase material.

The former are the older type of GC column and rarely used in most modern GC applications today. The latter open tubular columns are more widespread today and in most instances are prepared from fused silica capillaries, the I.D. of which varies from 0.1 to 0.35 mm and length from anywhere between 5 and 50 m.

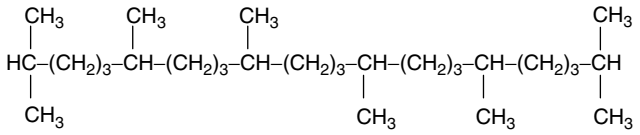
As mentioned above, these open tubular capillary columns have the actual stationary phase material coated or, in some cases, chemically bonded onto their internal surface. These are termed *wall-coated open tubular* columns, with the thickness of the stationary phase layer varying between 0.05 and 5 μm.

A reasonably large number of stationary phase materials have been developed for use in capillary GC columns; however, the majority are viscous organic liquids of varying polarity and generally fall within the classification of polysiloxanes, also known as *silicone oils*, or polyethylene glycols. Figure 3 shows some of the common liquid stationary phases used in capillary column GC.

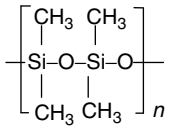
5 DETECTORS USED WITH CHROMATOGRAPHS

Detectors for HPLC and GC can be classified as either universal, meaning they register some degree of response to the presence of the majority of solutes, or as selective, meaning they only detect the presence of a particular class or group of solutes or, in some highly selective detectors, only individual solutes. For most detectors used in HPLC

Squalane: Solutes strongly retained: Sat. hydrocarbons, aromatic hydrocarbons, low polarity solutes.



SE-30 (Silicone gum): Strongly retained solutes: Low polarity as above.



Carbowax 20 M (Polyethylene glycol): Strongly retained solutes: (very polar) Hydroxyacids, polyhydroxyalcohols , polyprotic acids, aminoalcohols.

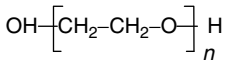


Figure 3. Some common liquid stationary phases used in capillary column GC.

and GC, the main requirement is that the detector does not register a large response for the mobile phase itself, which elutes the solute from the column.

5.1 Detectors for HPLC

The types of detectors used in HPLC can in most cases be classified as either photometric or electrochemical detectors. The former includes UV/Vis absorbance detectors, fluorescence and chemiluminescence detectors, the latter includes amperometric, potentiometric, and conductometric detectors. Alternative detectors that do not fit into the above classifications include refractive index and light scattering detectors, although these have only limited application. The commonality between all the above types of HPLC detectors is they all possess a flow-through cell connected to the end of the analytical column by HPLC grade tubing (<0.3 mm I.D.). These flow-though cells must have small internal volumes to ensure that sharp chromatographic bands eluting from the analytical column are maintained, although this is balanced with the need for sensitive detection, which in the case of spectrophotometric detectors is dependent upon cell length. Figure 4 shows a schematic diagram of a simple single wavelength UV absorbance detector as would be used in basic HPLC systems.

5.2 Detectors for GC

Two detectors common to most modern gas chromatographs are the (TCD) thermal conductivity detector and the (FID) flame ionization detector. Both are considered universal detectors, although the latter only responds to organic compounds.

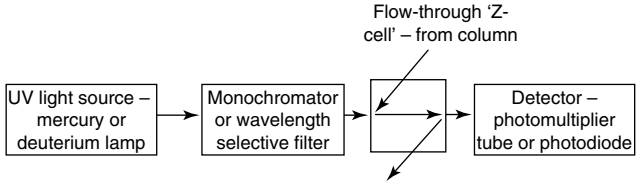


Figure 4. Schematic diagram of a simple single wavelength UV absorbance detector found in basic HPLC systems.

The TCD measures the thermal conductivity of the carrier gas and detects a change in this baseline conductivity when solute bands pass through the detector. Most organic solutes have a vapor phase thermal conductivity significantly lower than the carrier gas, particularly if this is H_2. Therefore, the detector records a decrease in conductivity as these solute bands pass through the detector.

Unlike the TCD, FID is a destructive detector, meaning the sample is actually destroyed as it passes through the detector. This is achieved by burning the solutes as they elute from the GC column in an air–H_2 flame. The burning of organic solutes in this way produces positive ions (CHO^+) and electrons within the flame, which itself is situated between an anode and cathode. The application of a voltage (+400 V) across these electrodes allows a current to be measured, the magnitude of which is dependent upon the concentration of organic solutes entering the flame. Figure 5 below shows a schematic diagram of an FID.

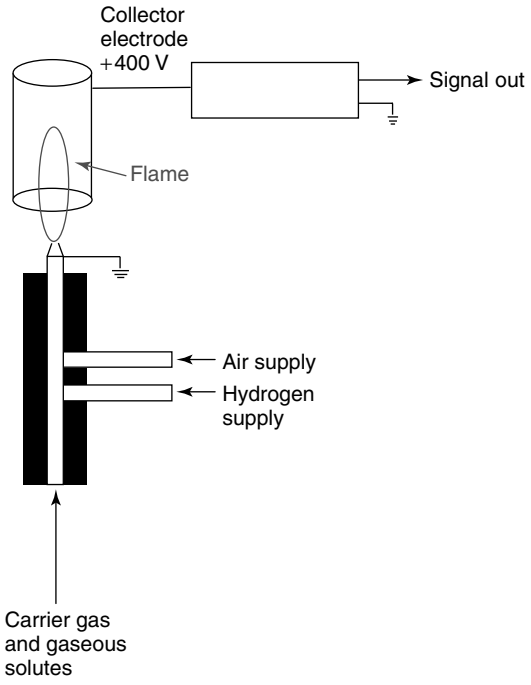


Figure 5. Schematic diagram of a flame ionization detection (FID) as used in standard gas chromatographs.

6 PROCESSING CHROMATOGRAPHIC SIGNALS

The chromatogram is the graphic result of a chromatographic separation. It represents the signal produced from the chromatographic detector plotted against time, beginning at the time of sample injection and ending upon elution of the last solute from the column.

Each solute has a characteristic t_R, retention time, when chromatographed under set stationary and mobile phase conditions. This is measured from injection to the point of maximum signal for each solute peak.

However, the mobile phase itself also takes a set amount of time to pass through the chromatographic column and this is termed t_M, *void time*. The positive difference between these two values represents the time the solute spends in or on the stationary phase and is represented as t'_R, adjusted retention time.

A more useful parameter used to describe solute retention behavior is the k', capacity factor. This is determined by t'_R/t_M. Capacity factors should conveniently be in the range 1 to 5. If k' is less than 1, solutes elute too quickly and so are not separated; if k' is much greater than 5, solutes are retained for long periods, making the chromatography impractical.

The potential of a chromatographic column to separate two solutes is described by α, the separation or selectivity factor. This is obtained by the division of respective capacity factors for the two solutes, $\alpha = k'_2/k'_1$. In general, α must be at least 1.02 to 1.05 for acceptable separations in GC and 1.05 to 1.20 for good separations in HPLC. However, the above parameter is based upon retention times for solutes only and does not include the width of the eluting chromatographic peaks. This, of course, will affect the ability of the system to produce an acceptable separation as broad peaks will tend to overlap if differences in retention are too small. Resolution, R_s, is a quantitative measure determined from a chromatogram, which, as it includes information on relative peak widths, more precisely describes the separation of two solutes. Resolution is obtained by (1), where W_n represents individual peak widths at their base:

$$R_s = \frac{t_{R2} - t_{R1}}{0.5(W_1 + W_2)} \tag{1}$$

The efficiency of a chromatographic separation is basically a measure of the degree of solute band broadening that occurs as the solute band passes through the column.

Ideally, the solute band should remain as narrow and concentrated as possible, as is the case with the most efficient systems. The term used to describe column efficiency is the theoretical plate number, N. The higher the number of theoretical plates, the higher the efficiency. Theoretical plate number can be determined using (2). The height equivalent to a theoretical plate, $HETP$, relates N to column length, L, and is determined using (3).

$$N = \frac{16t_R^2}{W^2} \tag{2}$$

$$HETP = \frac{L}{N} \tag{3}$$

Figure 6 shows a typical chromatogram for a six component sample obtained using HPLC.

7 PERFORMANCE AND LIMITATIONS OF CHROMATOGRAPHS

7.1 Liquid chromatographs

HPLC is an extremely versatile and established analytical technique. Modern liquid chromatographs suffer few of the problems of early instruments and are reliable and robust, with high quality pulsation-free pumps and sensitive low noise detectors. Because of this, liquid chromatographs can be found in a wide variety of laboratories belonging to the chemical industry, the food production and testing

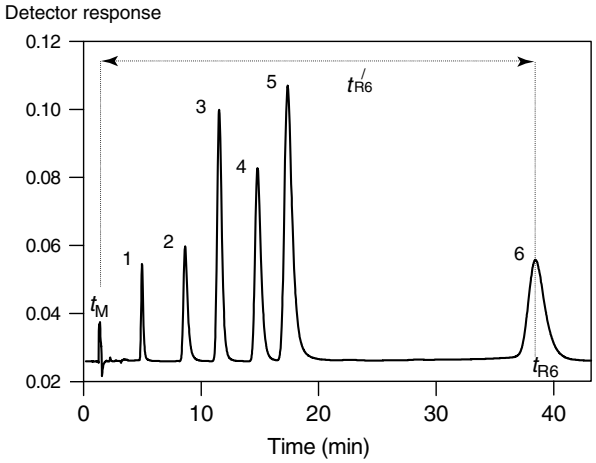


Figure 6. Typical HPLC chromatogram for a six component sample.

industries, environmental monitoring agencies, and pharmaceutical and biochemical plants.

The performance of the technique is to a large extent dependent upon the correct choice of mobile and stationary phases. Obviously, these have to be compatible with each other and with the actual solutes within the sample. Most problems encountered with HPLC are due to problems of this nature. A few simple rules should be followed to help avoid many such problems.

1. Always filter (<0.45-μm filter) and degas the mobile phase before use.
2. Always test the solubility of sample components in the mobile phase before injection.
3. Avoid the use of highly acidic pH < 2 or basic pH > 8 mobile phases with silica-based stationary phases.
4. Be careful with the use of organic solvents as mobile phases when using polymer-based stationary phases, which may dissolve.
5. Always use a short guard column before the analytical column to avoid fouling of the analytical column with strongly retained material.
6. When beginning or ending work with the chromatograph, slowly increase or decrease the mobile phase flow rate to that desired. This avoids sudden increases in back pressure, which can cause unwanted voids within the analytical column.

7.2 Gas chromatographs

In terms of efficiency and separation power, GC outperforms HPLC. Peaks are generally much sharper and resolution of complex mixtures often easier to obtain. However, as with HPLC, limitations of the technique are less to do with instrumentation and more to do with the chemistry involved.

Firstly, all solutes need to be volatile and readily vaporized, yet thermally stable, so that they do not degrade at the elevated temperatures within the injector, column, or detector.

Also, as it is generally not advisable to inject water into capillary GC columns, it is often the case that the solutes of interest first have to be extracted into a more volatile organic solvent prior to injection, which obviously adds unwanted complexity to the analytical method. The stability of the stationary phase itself is also often a cause of problems with GC. Coated capillary columns are often subject to the stationary phase bleeding off the column at elevated temperature, a problem that can show as high background detector noise levels or as reductions in retention times.

RELATED ARTICLES

Article 211, Mass Spectrometry, Volume 3; **Article 213, Electrochemical Measurements, Volume 3**.

FURTHER READING

Beesley, T.E., Buglio, B. and Scott, R.P.W. (2001) *Quantitative Chromatographic Analysis*, Chromatographic Science Series Vol. 85, Marcel Dekker Inc., NY, USA.

Jennings, W., Mittlefehld, E. and Stremple, P. (1997) *Analytical Gas Chromatography*, Academic Press, FL, USA.

Jonsson, J.A. (ed.) (1987) *Chromatographic Theory and Basic Principles*, Chromatographic Science Series Vol. 38, Marcel Dekker Inc., NY, USA.

Wilson, I. (ed.) (2000) *Encyclopedia of Separation Science*, Academic Press, FL, USA.

213: Electrochemical Measurements

David Davey
University of South Australia, Adelaide, South Australia, Australia

1 Potentiometry 1445
2 Ion Transport and Ion Exchange 1445
3 Reference Electrodes and Measurement Stability 1447
4 Further ISE Design Improvements 1448
5 Ion-selective Electrodes and Miniaturization 1448
6 Ion-selective Field Effect Transistor (ISFET) 1448
7 Voltammetry 1449
8 Polarography 1449
9 Improved DME Designs 1450
10 Biosensors 1451
References 1452
Further Reading 1453

1 POTENTIOMETRY

Potentiometry is where the electrical potential (voltage) of an electrochemical cell is measured under conditions of zero current flow.

The potential of a galvanic cell measured in volts correlates with the activity (or approximately, concentration) of the species of interest. This measurement is based on the Nernst equation

$$E = E^s + \frac{0.059}{n} \log \left\{ \frac{[A]^a [B]^b}{[C]^c [D]^d} \right\} \quad (1)$$

for an electrochemical reaction

$$aA + bB + n\mathrm{e} \longrightarrow cC + dD \quad (2)$$

where e denotes an electron and the square brackets denote activity of the species (approximately concentrations). E is the measured potential, and E^s is a constant potential measured at defined conditions and activities. The term '0.059 V' is derived from thermodynamic considerations and includes the Gas constant, the Faraday, a conversion factor for log base 10, and absolute temperature 298 K. It represents the ideal or Nernstian electrode response for a one-electron transfer. Note the indirect relationship between potential and the base 10 logarithm of activity (approximately concentration) for a single chemical species, the analyte.

The potential can be related to the analyte concentration once a comparison or reference electrode is provided. From the cell potential, the electrode potential for the process of interest may in turn be derived. Three reference electrodes are given in Table 1. The first is the primary standard reference and the other two are secondary practical references used in everyday work.

Direct potentiometry is the most used technique in electroanalysis. Here, the potential of a selective electrode relative to a standard is measured at a very low, or zero, current.

2 ION TRANSPORT AND ION EXCHANGE

The performance of an electrode is determined by the control of the charge transfer process and any component voltages that result from that charge transfer. Let us simplify these processes to two major examples:

$$\text{Ion transport, } \mathrm{Ion}^{+(-)}/\mathrm{H}^+ \textit{ interface } \mathrm{Ion}^{+(-)}/\mathrm{H}^+/\mathrm{e}^- \quad (3)$$

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

Table 1. Electroanalytical reference electrodes.

Type	Description
Standard Hydrogen, SHE. The Primary reference electrode for electrochemistry.	A glass jacket surrounds platinum, so hydrogen gas can bubble at 1 atm pressure over the Pt surface immersed in 1 M HCl held at 25 °C.
The half-cell reaction is: $2H^+ + 2e \rightarrow H_2$ (gas)	E^0 value set as 0.00 V
Calomel Secondary Reference	Two concentric glass tubes, the inner one containing mercury, mercury (I) chloride, and potassium chloride paste. The outer tube contains potassium chloride solution and connects to the sample solution via a porous glass frit.
The half-cell reaction is: $Hg_2Cl_2 + 2e \rightarrow 2Cl^- + 2Hg$	E^0 (KCl saturated) +0.241 V against SHE
Silver/silver chloride Secondary Reference	A silver wire coated with silver chloride and immersed in a solution of potassium chloride saturated with silver chloride. Contact with the sample solution is via a porous glass frit.
The half-cell reaction is: $AgCl + e \rightarrow Cl + Ag$	E^0 (KCl saturated) +0.197 V against SHE

and Ion binding then transfer via an intermediate molecule called an *ionophore*

$$\text{Ion}^{+(-)}/\text{ionophore} \longrightarrow \text{complex } \textit{interface } \text{Ion}^{+(-)}/\text{H}^+/\text{e}^- \quad (4)$$

An important element of electrode control occurs at the internal membrane solution *interface*, where processes described in (3) and (4) by the all-embracing term, $\text{Ion}^{+(-)}/\text{H}^+/\text{e}^-$, occur, that is, ion transfer (+ or − analyte ion or via the hydrogen ion, H^+) or electron transfer, e^-, to a conducting surface.

An understanding of these electrodic processes has led to improved measurement performance for potentiometric sensors. Consider the most important ion-selective electrode (ISE), the glass electrode, and pH measurement.

The electrode has a thin-walled membrane of carefully doped glass and an internal reference (generally, silver/silver chloride with an electrolyte of KCl and HCl). The electrode is immersed in the test solution and the potential is measured against an external reference. The electrode falls under the general descriptive $\text{Ion}^{+(-)}/\text{H}^+$ *interface* $\text{Ion}^{+(-)}/\text{H}^+/\text{e}^-$, which may be simplified to hydrogen ion transport across the glass membrane, H^+ *interface* H^+, since other factors are constant.

With both the internal and external reference potentials constant, changes in the cell potential are decided at the membrane. For the glass electrode,

$$E_{\text{membrane}} = E_{\text{external surface}} - E_{\text{internal surface}} \quad (5)$$

or

$$E_{\text{membrane}} = E^s + \frac{0.059}{1}\log a_{\text{H}^+\ \text{external}} - \left(E^s + \frac{0.059}{1}\log a_{\text{H}^+\ \text{internal}}\right) \quad (6)$$

Since E^s, 0.059, and $\log a_{\text{H}^+\ \text{internal}}$ are constant, they can be replaced by an overall system constant, L, and so finally results in

$$E_{\text{membrane}} = L - 0.059\text{pH}_{\text{external}} \quad (7)$$

The transport process through the membrane confers selectivity on the electrode and has led to the modern definition of pH being

$$\text{pH} = -\log a_{\text{H}^+} \quad (8)$$

where a_{H^+} is the activity (approximately concentration) of the hydrated hydrogen ion in solution.

The impedance (~100 MΩ) of the circuit provided measurement difficulties in the past, but today, very high input impedance microelectronic amplifiers have simplified matters considerably.

The principle behind other ISEs (Table 2) is similar to that shown for the glass system, although the membranes used depend upon diverse transport processes for their selectivity. The principle transport mechanisms involve either ion flux through a crystalline lattice (as shown for the glass electrode), where the lattice's stability limits the size and charge of an acceptable ion, or diffusion facilitated through complexation by an ionophore, a specific binding agent for the selected species, held in a plastic membrane.

Essentially, ion transfer creates a concentration cell across the membrane and similar equations to that given for the pH electrode apply. The term 0.059 V or 59 mV represents the ideal or Nernstian electrode response for a one-electron transfer, with a factor of 10 for change in activity (approximately concentration) for the analyte. As the charge, n, on the ion rises, the sensitivity of the measurement declines as $0.059/n$ decreases.

Table 2. Examples of crystalline, pressed salt, and polymer-based ion-selective electrodes and their Limit of Detection (LOD). All listed electrodes are Nernstian.

Analyte	Membrane type	LOD ($mol\,dm^{-3}$)	Reference
H_3O^+ (pH)	doped alkali glass/silicate	10^{-11}	–
F^-	LaF_3 crystal	10^{-7}	Frant and Ross (1966)
Cl^-	Pressed $AgCl/Ag_2S$	10^{-6}	Midgley (1981)
Cl^-, Br^-, I^-	Ag/AgX lattice (X = Cl^-, Br^-, I^-)	10^{-6}	Papeschi, Bordi and Carla (1978)
SO_4^{2-}	$Pb/PbSO_4$	5×10^{-4}	Gorina and Ryvkina (1978)
Ag^+	Ag/Ag_2S	10^{-7}	–
Cu^{2+}	Cu_xSe	10^{-6}	
Ca^{2+}	CaF_2	10^{-6}	Ghosh *et al*. (1978)
K^+	Polymer/valinomycin	10^{-5}	Umezawa (1990)
NO_3^-	Polymer/4,4′-diphenyl 2,2′-bipyridine nickel	10^{-6}	Umezawa (1990)
S^{2-}	Ag_2S	10^{-7}; 10^{-14a}	Midgley (1981)
Cu^{2+}	copolymer/CuS	10^{-6}	Achmatowicz *et al*. (2000)
Ca^{2+}	PVC/organophosphate	10^{-7}	Moody, Saad and Thomas (1988)
Pb^{2+}	DOS/PVC/ETH 5428	10^{-6}	Ceresa and Pretsch (1999)
Al^{3+}	PVC/Furil	10^{-6}	Mousavi, Arvand-Barmchi and Zanjanchi (2001)
Cl^-	Polymer/Mercuracarborand	10^{-7}	Badr *et al*. (1999)
I^-	Polyflourinated phosphazene/AgI, Ag_2S	10^{-7}	Janata and Huber (1979)
Acetate	PVC/Urea-functionalized porphyrin	10^{-5}	Amemiya *et al*. (1999)
Nitrate	PVC/Tetraalkylammonium nitrate	10^{-5}	Midgley (1981)

Note: [a]LOD values are highly dependent on control of the medium. This is done using acid–base and metal ion buffers, and total salt (ionic strength) control (Midgley, 1981).

There are hundreds of different ISEs in use, with a small number having sufficient selectivity for routine field, laboratory, or factory measurements (Umezawa, 1990).

Perhaps the best known after the glass electrode is the fluoride electrode (Figure 1), which has a europium-doped lanthanum fluoride crystal membrane with strong selectivity for fluoride, via the ion transfer F^- *interface* F^-, which leads to

$$E_{\text{membrane}} = L' - 0.059 \log[F^-]_{\text{external}} \qquad (9)$$

Measurements require use of a buffered media, (the OH^- ion is able to compete with fluoride in the LaF_3 lattice, and is thus a major problem). Further solution controls include salt to control the solution resistance and activity factors and a metal complexing agent to 'release' the free fluoride ion, which otherwise binds to the same metal ions. Analyte matrix control, via an ionic strength and buffer solution, such as that for fluoride, is essential for successful application of ISEs. Control of temperature and measurement time is also important; for the fluoride ion the response is often Nernstian to $10^{-7}\,mol\,dm^{-3}$ (Frant and Ross, 1966).

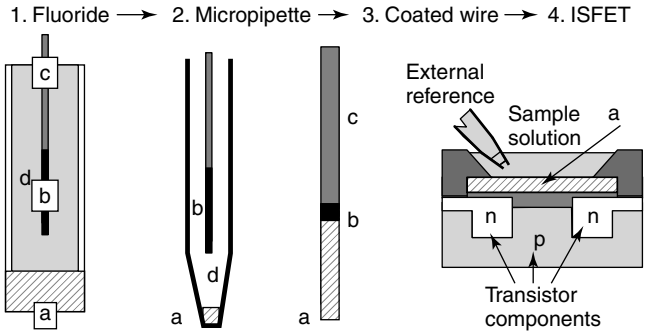


Figure 1. ISE construction and design trends, 1. the fluoride electrode; 2. micropipette; 3. a coated wire electrode; 4. the ISFET.

The antibiotic valinomycin gave important insights into the role of an ionophore in electrode selectivity; valinomycin provides an ion channel for potassium through the bacterial cell wall into the bacterium, the channel consisting of a ring of oxygen atoms of appropriate size for potassium ions. A selective electrode has been marketed using this design. Nonactin has also been found to be selective for ammonium ion for similar reasons. In a recent work by Amemiya *et al*. (1999), acetate ion has been selectively determined through a neutral hydrogen-bonding ionophore, (Table 2).

3 REFERENCE ELECTRODES AND MEASUREMENT STABILITY

Readings are found to drift with time for many potentiometric sensors, and a need to frequently check calibration is

commonly seen as essential practice. Finding the seat of the drift problem is often difficult, but examining the external reference electrode may be productive (Table 1). The reference electrode makes solution contact through a porous plug and the electrolyte flow must remain constant to avoid a so-called *junction potential shift*. Silver/silver chloride reference electrodes are subject to precipitation of silver chloride in the plug, this blocking process changing the overall potential. The problem can be reduced by trapping silver ions in a cation-exchange resin within the internal reference, with AgCl no longer precipitating. Ito *et al*. (1995) found that pH readings were consistent against freshly prepared pH sensors over many days with this modification.

4 FURTHER ISE DESIGN IMPROVEMENTS

Insight into ion fluxes in both the sensor and accompanying reference has led to important design improvements. Sokalski *et al*. (1997) dramatically improved the detection limit of a polymer membrane electrode for lead, using an internal complexing agent to ensure ion flux was directed inwards, avoiding contamination of the external analyte medium by the internal reference solution.

5 ION-SELECTIVE ELECTRODES AND MINIATURIZATION

Ion-selective electrodes have a strong advantage in the simplicity of their design, which has led to their miniaturization and application in many different clinical and industrial areas. As illustrated in Figure 1, micropipette (Gasser, 1988), coated wire (Cattrall and Hamilton, 1984), and transistor (ISFET) designs (Nam *et al*., 2003) are successful, with dimensions in the submillimeter range.

Figure 1 emphasizes the key features of these electrodes.

The Fluoride electrode (1) uses a lanthanum fluoride crystal (a), which promotes F^- ion transfer, the ion flux generating a potential in accord with (9). The internal reference (b) is $Ag/AgCl/Cl^-$ (Table 1) with some KF in solution(c).

The micropipette design (2) has led to clinical applications because of its small size (less than 1 mm). The design is otherwise similar to the fluoride example and excellent performance has been noted with detection limits at 10^{-9} M for calcium (Gasser, 1988)).

A coated wire electrode (3) dispenses with the internal reference and relies on electrical contact at (b) – see (4) and comments below.

The Ion-selective Field Effect Transistor ISFET (4) places the sensing membrane (a) over the gate of an n-p-n transistor. Potential again follows the Nernst expectation for pH or fluoride, as illustrated by (8) and (9).

Micropipette multibarrel potentiometric sensors, developed by clinical chemists and biologists, led to many novel applications, as discussed by Toth and coworkers (Toth *et al*., 1995). A zinc microelectrode provides potentiometric microscopy to nanometer resolution. More importantly, micropipette sensors inspired others to think of unique ways to miniaturize sensors.

The coated wire sensors are intriguing in that the internal reference is replaced by direct contact of the sensor coating with the metal contact (Figure 1, 3(b) and (3)). It is generally thought that a layer of metal salt forms and provides a pseudoreference potential for the sensor membrane (Cattrall and Hamilton, 1984). Despite the lack of understanding, the good performance of many coated wire systems gave the impetus for a similar arrangement in the ISFET where the membrane covers a silicon nitride layer.

6 ION-SELECTIVE FIELD EFFECT TRANSISTOR (ISFET)

The transistor and control of semiconducting properties of silicon has led to the ISFET. In the most common design, the ion-selective membrane is encapsulated in the device as the gate, the section that changes potential and allows a current to flow between regions of n-type silicon separated by a p-silicon section, Figure 1 (Nam *et al*., 2003). An active species, transported through the ISE membrane, can be measured via the potential needed to maintain an unchanged gate current. In the last situation, potential is found to follow the predictions of the Nernst equation (see (1), (7) or (9)).

6.1 Microfabrication of thick-film electrodes and electrode arrays

Electrodes can be made using screen-printing techniques on ceramic substrates with conducting pastes, which are then hardened by firing at high temperatures or by UV curing. Both heat and UV-cured approaches are described by Achmatowicz *et al*. (2000). Screen-printed methyl and butyl methacrylate copolymer-based copper sulfide sensors were found to give excellent Nernstian performance. An application of a multielectrode device to clinical analysis is illustrated by the use of coated silicon sensors by Scheipers *et al*. (2001). The authors discuss the on-line monitoring of blood electrolytes using potassium, sodium, and pH sensors deposited on silicon.

7 VOLTAMMETRY

Voltammetry is a process using applied voltages, which leads to electrolysis, loss of sample, and a current flow in the cell. By the use of a microelectrode, only a small part of the whole sample is used up in this electrochemical reaction. In electrical terms, however, this means the matching of the cell and the measuring instrument is not optimal as it is in potentiometry. Any problem is overcome by appropriate measurement operations designed mainly to protect the reference electrode.

It is a controlled potential method. The applied potential transfers electrons between the electrode surface and solution ions or molecules able to undergo reduction or oxidation at that potential. In this way, voltammetry provides information on what is present in the solution (the applied potential closely correlates with the standard potential of a reacting species). The current magnitude relates to how much is present, and the current direction tells whether oxidation or reduction is occurring.

8 POLAROGRAPHY

Voltammetric techniques are dominated by polarography and its variants. This is the measurement of the diffusion-limited current between a dropping mercury electrode (DME) (or working electrode) and a second or counter electrode. The applied potential is controlled against a reference (see Table 1), with no current drawn through the reference. The dropping mercury electrode (Figure 2) has major advantages over alternatives despite efforts to replace it as the working electrode because of environmental concerns. The mercury microdrop is renewed at a set time (a value within the range 0.5–2 s being commonly selected) at the end of a 100-μm glass capillary.

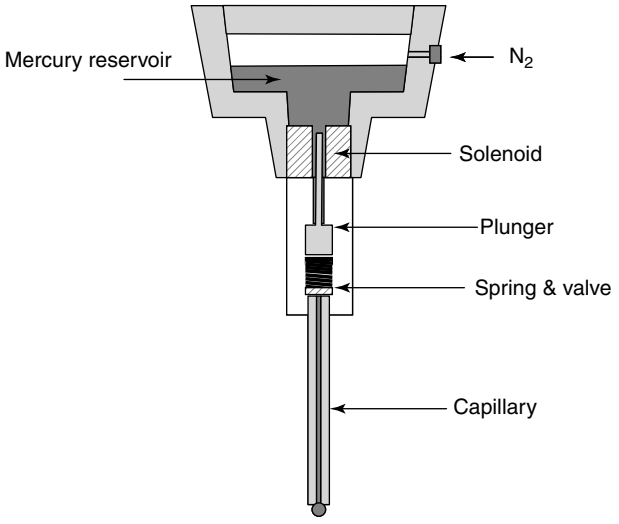


Figure 2. A controlled mercury capillary electrode. Mercury is held under nitrogen in the reservoir and drops are released at strictly controlled times as required. Skoog, Holler and Nieman (1998). (Reproduced by permission of Harcourt Brace College Publishers.)

Both depleted solution and the 'contaminated' electrode surface are renewed by the microdrop process. Secondly, mercury has a large hydrogen overpotential, which means that a wide range of cathode potentials is available without the danger of solution electrolysis.

The three electrode approach has come to dominate all voltammetric applications including polarography and its variants. The potential is scanned relative to a known reference electrode potential, while the diffusion-limited current between a DME and a counter platinum electrode is measured. Information on both the identity and the concentration of many electroactive compounds in a selected potential scan window are obtained. The potential observed is close, in most cases, to the standard potential seen in potentiometry and is typified by the Nernst equation. A small shift due to the influence of diffusion is observed, so the characteristic potential is called the *half wave potential*, $E_{1/2}$.

$$E_{1/2} = E^0 + K \log \left\{ \frac{D_{ox}}{D_{red}} \right\}^{1/2} \tag{10}$$

where E^0 is the standard reduction potential earlier defined for the Nernst equation, measured against a standard hydrogen electrode at $1\,\text{mol}\,\text{dm}^{-3}$ and 25 °C.

K is the Nernst factor, that is, $0.059/n$ V for log base 10, with n being the number of electrons transferred in the redox process. D_{ox} and D_{red} are the diffusion coefficients for the reduced and oxidized ions or molecules characterized by the process at that E^0.

The $E_{1/2}$ value is found half way up the polarographic wave – see later in Figure 4(a). It should perhaps be no surprise that $E_{1/2}$ will shift in value as does E in potentiometry if the conditions for E^0 to stay constant, for example, pH control, are not adhered to. Conversely, this means that pH influences on reactions are neatly studied by polarographic means.

An alternative form of the Nernst equation can be derived, which describes the polarographic waveform given in Figure 4.

$$E = E_{1/2} + K \log \left\{ \frac{i_d - i}{i} \right\} \tag{11}$$

where$\{(i_d - i)/i\}$replaces the concentration ratio in (1).

Measurement of the limiting diffusion current i_d and current, potential, i, $E_{1/2}$, data pairs over the polarographic

wave gives an excellent means of finding n values from calculated K values for reactions under study.

The additional theoretical basis of polarography is the Ilkovic equation:

$$i_d = knD^{1/2}Cm^{2/3}t^{1/6} \tag{12}$$

where

i_d = (measured) diffusion-limiting current
k = a constant dependent on both fundamental and system constants
C = concentration of electroactive species
m = mass flow rate of mercury
t = drop time

D and n are defined as given above.

The current flow in a practical measurement is limited by the rate of diffusion of the analyte molecule and its concentration. For given solution and electrode conditions, a plot of i_d versus C will provide a calibration graph from which unknown concentrations may be read. A useful account of the electrical side of the low-level signal need is Armutat (2004).

9 IMPROVED DME DESIGNS

Polarographic instruments use either a solenoid- or pressure-operated valve to release mercury from a 5-mL reservoir held under nitrogen, and generally a tapping device, to dislodge the mercury microdrop from its capillary tip to give exact time control of the microelectrode renewal. In the DC approach, the 'linear' potential scan is a series of small potential steps in unison with the drop release, a so-called *staircase sweep*, the current being read for say 20 ms late in the drop time (Figure 3a). The latter approach avoids problems in charging the drop and provides a cleaner and more easily processed signal.

The excellent reproducibility of the DME has been exploited in pulse techniques, which generally dominate voltammetry. Additional potential pulses are applied to the DC sweep to provide differential pulse polarography DPP and other variants (Garcia-Armada, Losada and de Vicente-Perez, 1996).

In Figure 3, a linear sweep (a) has been replaced by a staircase sweep in the DC method (b), while an additional potential pulse leads to a change in electrolysis current in the DP method (c). The current is read twice in each drop's life, before and after a potential pulse (e.g. 20–100 mV) is added to each 10-mV step in the staircase.

Comparison of the DC and DP polarographic approaches is further emphasized in Figure 4. The DC method reads a current, i, late in each drop time and displays i versus the applied potential E in a sigmoidal format where a redox process occurs. The DP approach, on the other hand, reads two current values, i_1 and i_2, and plots Δi against the applied potential E for each potential step.

The current change per potential pulse gives a peak data form or differential and has superior detection limits to the DC method. The potential at the DP peak differs slightly from the DC half wave potential because of the pulse size, that is,

$$E_{peak} = E_{1/2} - \Delta E_{pulse} \tag{13}$$

A differential pulse polarogram for two elements is shown Figure 5.

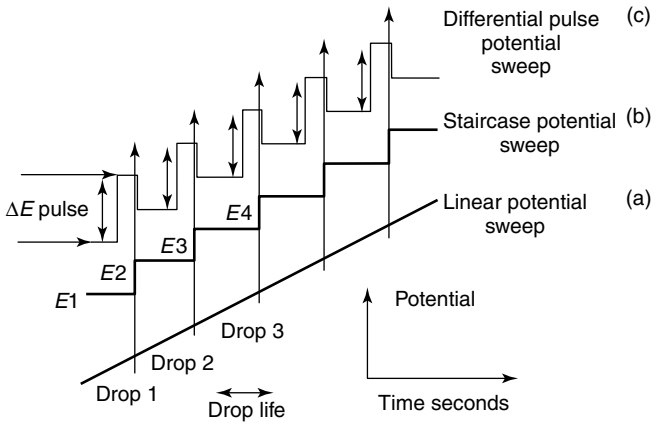


Figure 3. Potential–time sweep profiles used in voltammetric measurement methods. Pulse techniques replace a linear potential sweep with staircase potential time profiles.

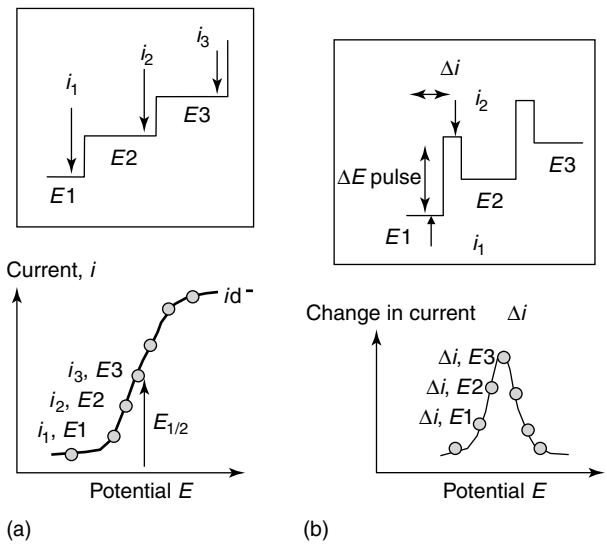


Figure 4. Potential and current relationships in DC and DP Polarography. The diffusion current i_d is proportional to concentration. $E_{1/2}$ is approximately equal to E^0.

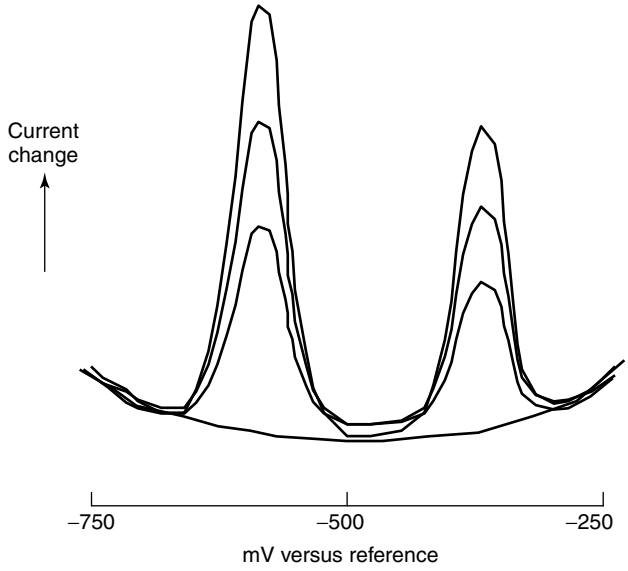


Figure 5. Standard addition traces for lead and cadmium in water, with applied potential read against the silver/silver chloride reference. Current is in nA for ppb concentrations.

9.1 Stripping voltammetry

A further refinement is found in stripping voltammetry (SV). The analyte is first plated on the working electrode at a fixed potential and then oxidized or reduced by a reversed potential sweep. Selected working electrodes include a hanging mercury drop and mercury films on a gold or graphite surface.

In anodic stripping voltammetry (ASV), analytes (importantly, heavy metal ions) are first reduced at a fixed electrode (often a mercury film on a platinum or graphite surface) and then oxidized by a rapidly sweeping reversed potential.

The advantage is that the limit of detection is improved by effectively integrating the measured current over a larger proportion (perhaps all) of the analyte in the sample, the analyte being preconcentrated on the electrode.

A disadvantage of ASV is that only species that undergo reversible electrode processes may be quantitated (Kipton, Powell and Florence, 1990). A summary of the limits of detection for the selected polarographic methods and stripping voltammetry is given in Table 3.

Table 3. Concentration ranges in polarography and voltammetry[a].

Mode	Typical limit of detection (mol dm^{-3})
DC sweep	10^{-4}–10^{-6}
DP sweep	10^{-4}–10^{-7}
Anodic stripping	10^{-6}–10^{-10}

[a]The limit of detection LOD is highly dependent on analyte and experimental conditions.

An interesting alternative to ASV is potentiometric stripping analysis. Plating follows the ASV approach, but stripping uses either a chemical oxidant or small oxidative current and follows the potential/time trend for each analyte's oxidation at its characteristic potential. The method is more tolerant of interferences in real solutions than ASV, and oxygen can be exploited as the oxidant (Jagner, 1983)).

Cathodic stripping voltammetry extends the range of species accessible by stripping procedures by adsorbing analytes held in complex form, the latter undergoing accessible reduction or oxidation. The technique has led to an improved understanding of environmental situations where metal bioavailability is controlled by naturally formed complexes and their equilibria with labile species (Herbert, 1999).

Voltammetric methods are widely used in environmental and industrial applications and are relatively inexpensive. Barek *et al*. (2001) discuss the polarography of carcinogens, pesticides, herbicides, dyes, and pharmaceuticals. A number of clinical and industrial measurements have been automated (Arnold and Meyerhoff, 1988; Hart and Wring, 1997).

9.2 Microvoltammetric electrodes

The application of microelectrodes to monitoring generally has immense promise, with electrode arrays being manufactured with hundreds of individual sensing elements and low cost per array (Montenegro, Queiros and Daschbach, 1991).

Arrays can be tuned to particular elements in a mixture, giving improved selectivity over conventional single electrodes. Saban and Darling (1999) measured heavy metals in aqueous solution with arrays having working electrodes such as Hg, C, Au, Ag, and Pt. Stripping voltammetry of Cu, As, Cd, Pb, Cu, Zn, Sn, and Hg was successfully illustrated to trace levels. Sullivan *et al*. (1999) provide an excellent description of both manufacture and application of simple arrays, in this case using 64 electrodes.

10 BIOSENSORS

An outstanding sensor based on voltammetric principles is the glucose sensor. The concept combines the discovery of a specific enzyme that oxidizes glucose to gluconic acid when oxygen forms hydrogen peroxide. Monitoring of the enzymatic process can follow oxygen depletion, peroxide formation, or changes in mediators.

An applied potential, tuned to the selected product, produces a current, and a glucose reading can be displayed.

This device has dramatically changed the lives of thousands of people suffering from diabetes.

The fundamental principles of the enzymatic electrode were laid down by Clark in the 1960s, and commercial success followed research at Oxford on immobilized mediator chemistry (Turner, Karube and Wilson, 1987).

A multichannel analysis of enzyme reactions is nicely illustrated by the work of Kasai, Matsue and Uchida (1996). A flow-through cell containing sixteen 100-μm Au electrodes probed the kinetics of horseradish oxidase and xanthine oxidase reactions with various electron donors. Each electrode could be scanned independently of the other. The continuing search for implantable electrodes and automated monitoring of health measures is discussed by Henry (1998).

REFERENCES

Achmatowicz, S., Jakubowska, M., Zwierkowska, E., Koncki, R. and Tymecki, L. (2000) Polymer Paste for Ises Potentiometric Measurements of Cu-ions Concentration, *Proceedings of the XXII IMAPS Poland Conference*, Koszalin, http://www.cyf-kr.edu.pl/academic/OBRMHiR/imaps/data/txt/txt13.htm.

Amemiya, S., Buhlmann, P., Umezawa, Y., Jagessar, R.C. and Burns, D.H. (1999) An Ion-Selective Electrode for Acetate Based on a Urea-functionalized Porphyrin as a Hydrogen-bonding Ionophore. *Analytical Chemistry*, **71**, 1049–1054.

Armutat, A. (2004) *Low Level Measurements*, Keithley Instruments. Free download on www.keithley.com.

Arnold, M.A. and Meyerhoff, M.E. (1988) Recent Advances in Development and Analytical Applications of Biosensing Probes. *Critical Reviews in Analytical Chemistry*, **20**, 149–196.

Badr, I.H.A., Diaz, M., Hawthorne, M.F. and Bachas, L.G. (1999) Mercuracarborand "Anti-Crown Ether"-Based Chloride-Sensitive Liquid/Polymeric Membrane Electrodes. *Analytical Chemistry*, **71**, 1371–1377.

Barek, J., Fogg, A.G., Muck, A. and Zima, J. (2001) Polarography and Voltammetry at Mercury Electrodes. *Critical Reviews in Analytical Chemistry*, **31**, 291–309.

Cattrall, R.W. and Hamilton, I.C. (1984) Coated-Wire Ion-Selective Electrodes. *Ion-Selective Electrode Reviews*, **6**, 125–172.

Ceresa, A. and Pretsch, E. (1999) Determination of Formal Complex Formation Constants of Various Pb^{++} Ionophores in the Sensor Membrane Phase. *Analytica Chimica Acta*, **395**, 41–52.

Frant, M.S. and Ross, J.W. (1966) Electrode for Sensing Fluoride Ion Activity in Solution. *Science*, **154**, 1553.

Garcia-Armada, P., Losada, J. and de Vicente-Perez, S. (1996) Cation Analysis Scheme by Differential Pulse Polarography. *Journal of Chemical Education*, **73**, 544–547.

Gasser, R. (1988) Advances in Construction and Methods of Application of Calcium Ion-Selective Microelectrodes. *Selective Electrode Reviews*, **10**, 49–70.

Ghosh, M., Dhaneshwar, M.R., Dhaneshwar, R.G. and Ghosh, B. (1978) Response to calcium ions of the calcium fluoride crystal membrane electrode. *Analyst (London)*, **103**, 768.

Gorina, M.Y. and Ryvkina, L.E. (1978) Ion selective electrode for determining sulfate ion. *Zhurnal Analiticheskoi Khimii*, **33**, 2269.

Hart, J.P. and Wring, S.A. (1997) Recent developments in the design and application of screen-printed electrochemical sensors for biomedical, environmental and industrial analyses. *Trends in Analytical Chemistry*, **16**, 89–103.

Henry, C. (1998) Getting under the Skin. *Analytical Chemistry News and Features*, **70**, 594A–598A.

Herbert, R.B. (1999) MiMi-Ferric iron complexation by organic ligands:Voltammetric studies. Report MiMi 1999:2. *The MISTRA-programme MiMi. Mitigation of the environmental impact from mining waste*, MiMi Print.

Ito, S., Hachiya, H., Baba, K., Asano, Y. and Wada, H. (1995) Improvement of the Silver/silver chloride reference and application to pH measurement. *Talanta*, **42**, 1685–1690.

Jagner, D. (1983) Computerised flow potentiometric stripping analysis. *Trends in Analytical Chemistry*, **2**, 53–56.

Janata, J. and Huber, R.J. (1979) Ion-Sensitive Field Effect Transistors. *Ion-selective Electrode Reviews*, **1**, 31–79.

Kasai, N., Matsue, T. and Uchida, I. (1996) Multichannel Electrochemical Analyses of Enzyme Reactions. *Electroanalysis*, **8**, 748–752.

Kipton, H., Powell, J. and Florence, T.M. (1990) Effect of mercury (II) on metal complex labilities determined by DC ASV on a thin mercury film electrode. *Analytica Chimica Acta*, **228**, 327–331.

Midgley, D. (1981) Detection Limits of Ion-Selective Electrodes. *Ion-Selective Electrode Reviews*, **3**, 43–104.

Montenegro, M.I., Queiros, M.A. and Daschbach, J.L., (eds) (1991) *Microelectrodes, Theory and Applications*, NATO ASI Series, E, Kluwer Academic Publishers, Dordrecht.

Moody, G.J., Saad, B.B. and Thomas, J.D.R. (1988) The Development of Polymer Matrix Membranes for Ion-Selective Electrodes. *Selective Electrode Reviews*, **10**, 71.

Mousavi, M.F., Arvand-Barmchi, M. and Zanjanchi, M.A. (2001) Al(III)-Selective Electrode Based on Furil as Neutral Carrier. *Electroanalysis*, **13**, 1125–1128.

Nam, H., Cha, G.S., Strong, T.D., Ha, J., Sim, J.H., Hower, R.W., Martin, S.M. and Brown, R.B. (2003) Micropotentiometric Sensors. *Proceedings of the IEEE*, **91**, 870–880.

Papeschi, G., Bordi, S. and Carla, M. (1978) Solid state chloride, bromide, or iodide selective electrodes. *Journal of Electrochemical Society*, **125**, 1807.

Saban, S.B. and Darling, R.B. (1999) Multi-element heavy metal ion sensors for aqueous solutions. *Sensors and Actuators B*, **61**, 128–137.

Scheipers, A., Wassmus, O., Sundermeier, C., Eshold, J., Weiss, Th., Gitter, M., Ross, B. and Knoll, M. (2001) Potentiometric ion-selective silicon sensors for the on-line monitoring of blood electrolytes. *Analytica Chimica Acta*, **439**, 29–38.

Skoog, D.A., Holler, F.J. and Nieman, T.A. (1998) *Principles of Instrumental Analysis*, 5th edition, Harcourt Brace College Publishers, Philadelphia.

Sokalski, T., Ceresa, A., Zwickl, T. and Pretsch, E. (1997) Large Improvement of the Lower Detection Limit of Ion-Selective Polymer Membrane Electrodes. *Journal of the American Chemical Society*, **119**, 11347–11348.

Sullivan, M.G., Utomo, H., Fagan, P.J. and Ward, M.D. (1999) Automated Electrochemical Analysis with Combinatorial Electrode Arrays. *Analytical Chemistry*, **71**, 4369–4375.

Toth, K., Nagy, G., Wei, C. and Bard, A.J. (1995) Novel Application of Potentiometric Microelectrodes: Scanning Potentiometric Microscopy. *Electroanalysis*, **7**, 801–811.

Turner, A.P.F., Karube, I. and Wilson, G.S. (1987) *Biosensors: Fundamentals and Applications*, Oxford University Press, Oxford.

Umezawa, Y. (ed.) (1990) *CRC Handbook of Ion Selective Electrodes: Selectivity Coefficients*, CRC Press, Boca Raton, FL.

FURTHER READING

Bakker, E. and Pretsch, E. (2001) Potentiometry at Trace Levels. *Trends in Analytical Chemistry*, **20**, 11–19.

Kalvoda, R. (1978) Modern polarographic and voltammetric techniques. *Euroanalysis*, **3**, 197–208.

Meyerhoff, M.E. (1993) In vivo blood-gas and electrolyte sensors-progress and challenges. *Trends in Analytical Chemistry*, **12**, 257–266. Clinical measurements.

Sim, J.H., Lee, K.M., Lee, J.S., Cho, D.H., Nam, H. and Cha, G.S. (2001) Lithium Ion-Selective Electrode with Improved Lifetime. *Bulletin of the Korean Chemical Society*, **22**, 765–768.

Tani, Y., Swoma, M., Harsanyi, E.G. and Umezawa, Y. (1999) Effect of Dissolved Oxygen on the Response of Cu(II) Ion-selective Electrodes in Metal Buffer Solutions. *Analytica Chimica Acta*, **395**, 53–63.

This article is based in part on a chapter by R. Watts originally published in Handbook of Measurement Science Vol 2, P.H. Sydenham, Wiley, 1983.

214: Accelerated Testing

Patrick D.T. O'Connor
Consultant, UK

1 Definition 1459
2 Role of Accelerated Testing 1459
3 Highly Accelerated Life Testing (HALT) 1461
4 Test Approach for Accelerated Test 1461
5 Examples of Accelerated Testing 1462
Related Articles 1463
References 1463

1 DEFINITION

Accelerated testing is the application of stresses in excess of those that would be expected to be experienced by a product in service, in order to

1. create failures during development, to provide information on how the design might be improved;
2. cause faulty production items to fail without damaging good ones, so that delivered items will be reliable.

2 ROLE OF ACCELERATED TESTING

Stresses, in the widest sense, can lead to failures. Variations of strength, stress, and other conditions can influence the likelihood of failure or the duration (time, distance, cycles, etc.) to failure. (Stress–strength variation and interactions are discussed in **Article 98, Reliability and Maintainability, Volume 2**.) In this module, we will describe how tests should be designed and conducted to provide assurance that designs and products are reliable and durable in service.

Testing for reliability and durability is fundamentally different from testing to measure other performance parameters. The reasons are that we never know what is the 'uncertainty gap' between the theoretical and real reliability and durability of the design and the products made from it, for the whole population, over their operating lives and environments. The effects of these uncertainties can seldom be evaluated with confidence by design analysis methods. Therefore, we must perform tests that will reduce the uncertainty gap to an extent that we can be assured of reliability and durability, whilst taking due account of practical constraints such as cost and time. *To do this we must not test to demonstrate successful achievement, but test to cause failures.*

Then the following questions must be asked, for each failure:

1. Could this failure occur in use (on other items, after longer times, at other stresses, etc.)?
2. Could we prevent it from happening in use?

The stresses that were applied are relevant only in so far as they were the tools to provide the evidence that an opportunity exists to improve the design. We have obtained information on how to reduce the uncertainty gap. Whether the opportunity is taken is a management issue, in which other aspects such as cost, weight, time, and so on must be considered.

The uncertainty described above is shown in Figure 1. If we consider only one stress and the failures it might cause, the stress-to-failure distribution of production items might be as shown. As discussed in **Article 98, Reliability and Maintainability, Volume 2**, the exact nature of this distribution is almost always uncertain, particularly in the

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

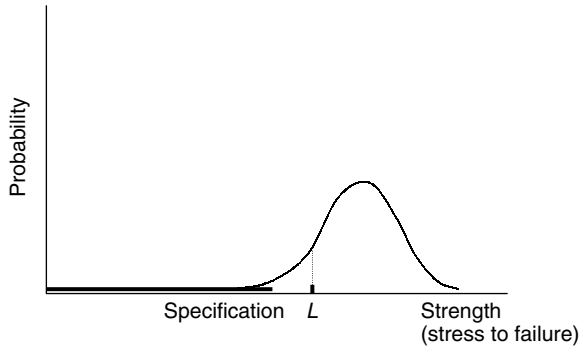


Figure 1. Stress, strength, and test failures (1).

tails, which are the most important areas as far as reliability and durability are concerned.

Suppose the first test failure occurs at stress level L. At this stage, we might have only a few items to test – maybe only one. We can state that the strength of this item represents a point on the distribution, but we cannot say whether it was an average strength item, a strong one, or a weak one. The only way to find out the nature of the strength distribution is to test more items to failure, and to plot and analyze the results. Inevitably, most of the items tested will be near the average, because that is where most of the population will lie. Therefore, it is unlikely that any item tested will represent the weakest in the future population. Also, testing large numbers is usually too expensive to be practicable.

However, if we analyze the actual cause of the failure, by whatever means is appropriate, and take action to prevent its recurrence, then in effect we will move the strength distribution to the right. *We can use high stresses in place of large samples*. The design will therefore be made more reliable.

For many items subject to stress, particularly electronic systems, two types of failures can occur: transient or permanent failures. With transient failures correct operation is recovered when the stress is reduced, but there is no recovery from permanent failures. The general case is illustrated in Figure 2.

If the failure is due to a wear-out mechanism, such as wear or fatigue, the horizontal axis of the distribution will represent time (or cycles), for any particular stress value. In addition to the uncertainty regarding the stress and strength, we now have the further uncertainty of time. Failure on test after time t at stress level L will represent one point on an unknown 3-dimensional distribution (Figure 3). As Figure 3 illustrates, an important feature of wear-out mechanisms is that the resulting distributions of times to failure become wider as damage accumulates, thus further increasing the uncertainty.

In many engineering situations, failures are caused by combinations of stresses and strength values, not just by one stress and one strength variable. Some might be time-dependent, others not. Between any two or more variables there might also be interactions, as described in **Article 98, Reliability and Maintainability, Volume 2**. Therefore, even for relatively simple and common failure situations, there is not just one distribution that is important, but a number of possible distributions and interactions.

Accelerated stress testing can provide quantitative cause-and-effect information when the mechanisms of failure are

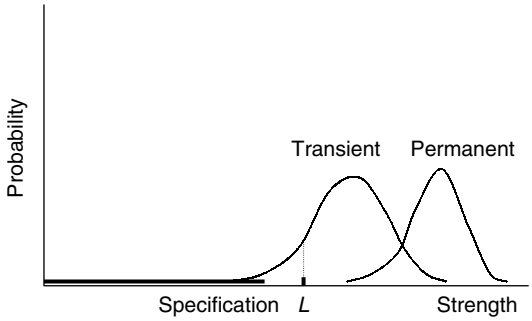


Figure 2. Stress, strength, and test failures (2).

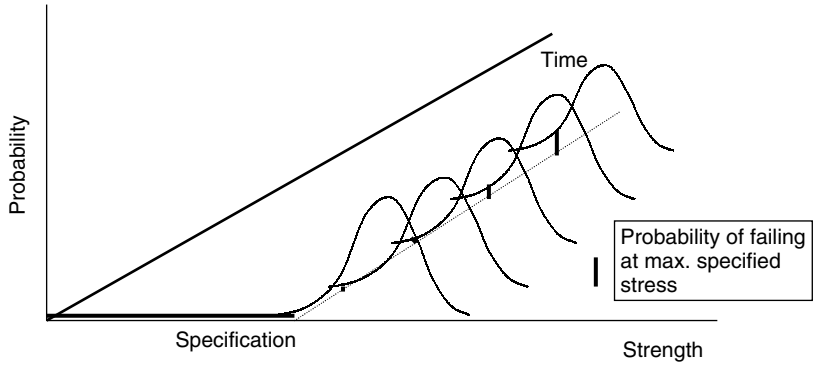


Figure 3. Stress, strength, and test failures (3): wear-out failures.

already understood (e.g. material fatigue), and when the tests are planned specifically to provide such information. We must decide whether we need the extra information that more 'scientific' or statistical tests can provide.

Accelerated tests can also be applied to production and to repair.

3 HIGHLY ACCELERATED LIFE TESTING (HALT)

Nearly all standardized approaches to stress testing (O'Connor, 2001) demand the use of typical or maximum specified stresses. However, the highly accelerated life testing (HALT) approach to accelerated testing is a logical extension of the preceding discussion. With HALT, we increase stresses (individual or combined) so that we cause failures to occur, then use the information to improve reliability and durability. The logic that justifies the use of very high stresses, not necessarily representative of expected application conditions, is based upon four aspects of engineering reality:

1. The causes of failures that will occur in the future are often very uncertain.
2. The probabilities and durations of failures are also highly uncertain.
3. Time spent on testing is expensive, so the more quickly we can reduce the uncertainty gap the better.
4. Finding causes of failure during development and preventing recurrence is far less expensive than finding new failure causes in use.

It cannot be emphasized too strongly: testing at 'representative' stresses, in the hope that failures will not occur, is very expensive in time and money and is mostly a waste of resources.

The HALT approach, and variations of it, are also referred to as (highly) accelerated stress testing ((H)AST), stress-induced failure environmental (STRIFE) testing, failure mode verification test (FMVT), and other terms.

Clearly, there might be practical limits to the stresses applied. These could be set by

- the fundamental limits of the technology, for example, there is no point in testing an electronic system at temperatures above the melting point of the solder used;
- the limits of the test capability, such as the maximum temperature of the test chamber.

4 TEST APPROACH FOR ACCELERATED TEST

4.1 Development

The approach that should be applied to any accelerated test programme during the development phase to assure reliability/durability should be as follows:

1. Try to determine, as far as practicable, what failures might occur in service. This should have been performed during design analysis and review, particularly during the failure modes, effects, (and criticality) analysis (FME(C)A) as described in **Article 98, Reliability and Maintainability, Volume 2**.
2. List the application and environmental stresses that might cause failures.
3. Plan how the stresses that might stimulate foreseeable and unforeseen failures can most effectively be applied in test. Set up the item (or items) to be tested in the test chamber or other facility so that it can be operated and monitored.
4. Apply a single stress, at or near the design maximum, and increase the level stepwise until the first failure is detected.
5. Determine the cause and take action to strengthen the design so that it will survive higher stresses. This action might be a permanent improvement, or a temporary measure to enable testing to be continued.
6. Continue increasing the stress/stresses to discover further failure causes (or the same cause at a higher stress), and take action as above. This approach is called *step-stress testing*.
7. Continue until all of the transient and permanent failure modes for the applied stress are discovered and, as far as technologically and economically practicable, designed out. Repeat for other single stresses.
8. Decide when to stop (fundamental technology limit, limit of stress that can be applied, cost or weight limit).
9. Repeat the process using combined stresses, as appropriate and within the equipment capabilities (temperature, vibration, power supply voltage, etc.).

4.2 Production

The basic difference between the objectives of accelerated test in development and in manufacturing is that, whilst we try to cause all development test items to fail in order to learn how to improve the design (of the product and of the production processes), we must try to avoid damaging good manufactured items, whilst causing weak or defective

ones to fail so that they can be corrected or segregated. Therefore, the stress regime applied is less severe. This is called *environmental stress screening* (ESS). When applied to electronic components and assemblies, it is also referred to as *burn-in*.

Applying accelerated testing during development provides evidence not only on how to make designs more robust but it also provides the information necessary to optimize ESS for manufacturing. The knowledge that we gain by applying the full HALT philosophy, including the design ruggedization, can be used to design a stress test regime that is optimized for the product, and that is far more effective than conventional ESS. This is called *highly accelerated stress screening* (HASS). HASS provides the same benefits in manufacturing as HALT does in development, in greatly increasing the effectiveness of manufacturing screens whilst reducing test cost and time.

4.3 In service

ESS can be applied to items that have been repaired. The HASS approach can be particularly beneficial in confirming that repairs have been successful, and in identifying the causes of intermittent or transient failures that often lead to 'no fault found' (NFF) (or 're-test OK' (RTOK)) diagnoses and repeat failures in service.

4.4 Accelerated test facilities and services

Special test chambers have been developed for HALT/HASS. These enable higher stresses to be applied (primarily temperature rates of change and vibration) than is possible with conventional environmental test chambers. Figure 4 shows a chamber. Training and consulting on HALT/HASS is provided by Hobbs Engineering (URL: http://www.hobbsengr.com) and by some equipment and service providers. The method and the equipment are fully described in Hobbs (2000).

Other facilities are available for the provision of accelerated tests, with the emphasis on specific stresses, for example, vibration in the FMVT approach. Equipment and service suppliers are listed in the homepage of O'Connor (2001) (URL: http://www.pat-oconnor.co.uk/testengineering.htm). More information is provided in **Article 216, Test Facilities, Volume 3**.

5 EXAMPLES OF ACCELERATED TESTING

Accelerated tests, and in particular the HALT/HASS approach, can be applied in principle to any kind of product or technology. For example:

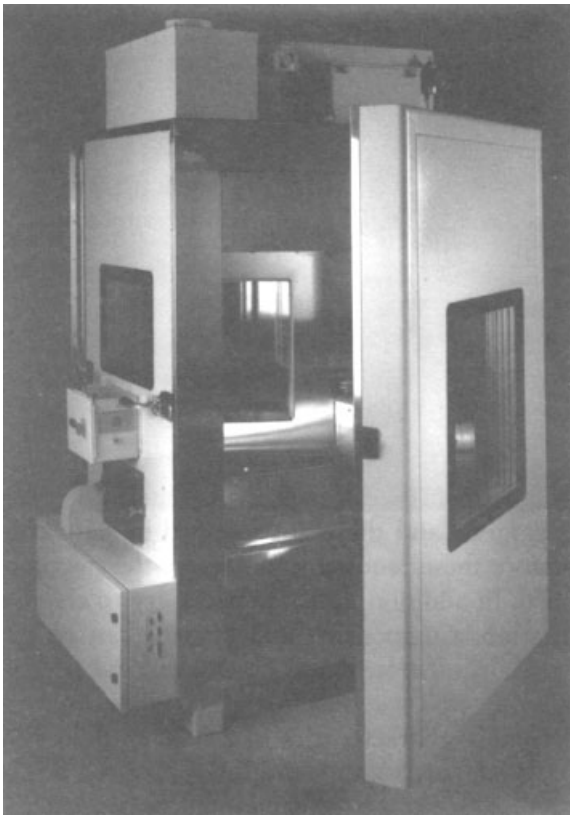


Figure 4. HALT/HASS chamber. (Copyright QualMark Corporation.)

Engines, pumps, power transmission units such as gearboxes, and so on.

- Start tests with old lubricants or other fluids (coolants, hydraulics, etc.), rather than new.
- Run at low fluid levels.
- Use fluids that are heated, cooled, or contaminated.
- Use old filters.
- Misalign shafts, bearings, and so on.
- Apply out-of-balance to rotating components.

Electro-mechanical assemblies such as printers, document, material or component handlers, and so on.

- Apply high/low temperatures, vibration, humidity/damp, and so on.
- Use components with out-of-tolerance dimensions.
- Misalign shafts, bearings, and so forth.
- Use papers/documents/materials/components that exceed specifications (thickness, weight, friction, etc.).

Small components or assemblies such as electronic packages, mechanical latches, switches, transducers, and so on.

- Apply high/low temperatures, vibration, humidity/damp, and so forth.

- Apply high frequency vibration by fixing to suitable transducers, such as loudspeaker coils, and driving with an audio amplifier.

RELATED ARTICLES

Article 98, Reliability and Maintainability, Volume 2; **Article 216, Test Facilities, Volume 3**.

REFERENCES

Hobbs, G. (2000) *Accelerated Reliability Engineering: HALT and HASS*, John Wiley & Sons, Chichester.

O'Connor, P.D.T. (2001) *Test Engineering*, John Wiley & Sons, Chichester. URL: http://www.pat-oconnor.co.uk/testengineering.htm.

215: Automatic Test Systems

Patrick D.T. O'Connor
Consultant, UK

1 Definition of an Automatic Test System (ATS) 1464
2 Role of an Automatic Test System 1464
3 Elements of an ATS 1464
4 Setting up an ATS 1465
5 Example of ATS 1466
References 1466

1 DEFINITION OF AN AUTOMATIC TEST SYSTEM (ATS)

An automatic test system, ATS (also called *automatic test equipment* (ATE)), is one in which the tasks of measurement and of categorizing items under test are controlled and performed automatically. An ATS can in some cases also perform other functions including diagnostics, data logging, and so on. Automatic test equipment is the term generally used in the context of testing electronic components, circuits, and system (*see* **Article 79, Testing Electronic Systems, Volume 2**). The term *unit under test* (UUT), or *system under test* (SUT), is often used in the context of ATS to refer to the item or system being tested. For more information, see Sheppard and Simpson (1994).

2 ROLE OF AN AUTOMATIC TEST SYSTEM

The primary role of an ATS is to improve the efficiency (primarily by reducing test time or cost) of the test process, by replacing human actions with equipment that can perform tests more rapidly and also usually more consistently. ATS can be applied during development, in production, and in service. For some modern systems, particularly complex electronic circuits (microcircuits and assembled systems), ATS is the only feasible approach since the test problem (number of test conditions, system operating speed) makes manual testing impracticable or impossible. In many modern production situations, the throughput is so high that ATS is the only practicable approach. In-service equipment is sometimes tested by ATS that is integrated into the system. This is called *built-in test* (BIT), *built-in test equipment* (BITE), or *built-in self-test* (BIST).

BIT can run tests fully automatically, either continuously or, for example, on system start-up or when selected. Automatic calibration is also a type of BIT. For further information, see Stroud (2002).

3 ELEMENTS OF AN ATS

An ATS can consist of individual test equipments controlled by a separate controller, or it can be a fully integrated system. Most electronic ATE is in the latter category. The primary elements of an ATS are as follows:

- *Interfaces*: The ATS must be connected to the UUT so that stimuli can be applied and responses measured. Interfaces can be electrical connectors, special adaptors, transducers, and so on, and can also include functions such as analog to digital (A/D) conversion, signal conditioning, and so on.
- *Data acquisition*: The ATS must provide automatic data acquisition DAQ – *see* **Article 132, Data Acquisition Systems (DAS) in General, Volume 3**. This typically

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

includes capabilities for sampling the parameters or states being measured, capturing the data, and transmitting it to the appropriate instruments and so on. A wide range of systems is available. See the section below.
- *Measurement*: Appropriate instruments must be provided for measuring the relevant parameters.
- *Display and recording*: Input and output data must be displayed and recorded. Displays can include dials, screens, and so on. Data can be recorded on magnetic tape, or on computer storage media.
- *Analysis*: The ATS might include capabilities for data analysis, such as graphical analysis, Fourier analysis, statistical analysis and control, and so on.
- *Software*: The functions listed above are controlled and performed mainly by software. See the sections below.
- *Others*: Depending on the requirements of the system, the ATS might include other capabilities, such as warnings, communications with design or database systems, Internet connectivity, and so on.

3.1 PC-based ATS

Personal computers (PCs) can be used to implement and control ATS. The PC, with appropriate software, can perform the functions of test and instrument control, data acquisition, display, analysis, and so on. The LABVIEW® software (National Instruments Corp. (URL: http://www.ni.com)) enables PC-based ATS to be designed graphically, and all test functions to be controlled from PCs. See Johnson (1997). PCs can also be used in conjunction with special PC cards in lieu of dedicated test instruments, such as oscilloscopes (*see* **Article 79, Testing Electronic Systems, Volume 2**).

3.2 Test Data Acquisition

There is a wide range of available systems for test data acquisition. These convert analog electrical signals to digital and control the data transfer between the UUT and the tester. The important performance characteristics are high speed–multichannel data sampling, high noise rejection, high bandwidth, and data transfer to computer memory.

Modern electronic test systems use standard data bus software. IEEE488 was the first test data bus standard, and most modern instruments include it. It is also called the *general purpose interface bus* (GPIB). More information is available from instrument suppliers, such as Agilent Technologies (URL: http://www.agilent.com), *see also* **Article 167, Asynchronous and Synchronous Interface Protocols, Volume 3**; **Article 168, RS 232 and EIA/TIA 232 Serial Interface, Volume 3**; and **Article 170, IEEE-488 Instrumentation Bus, Volume 3**.

The PC interface (PCI) bus or PCI extensions for instruments (PXI) bus are the versions used for PC-based testing. More information is available from instrument suppliers, particularly from National Instruments Corp. (URL: http://www.ni.com).

Larger functional test ATE systems use the VXI bus standard, which allows higher test speeds and greater accuracy.

DAQ or DAS systems are described in more detail in **Article 132, Data Acquisition Systems (DAS) in General, Volume 3**; **Article 136, Quantizing Theory Relevant to DAS, Volume 3**; and **Article 140, Integrating Type A/D Converters, Volume 3**. Computer bus systems are described in the section on **Distributed and Networked Measuring Systems**.

3.3 Test Software

Test software must be developed for any system that is to be tested on ATS. For some systems, particularly for aerospace electronics (avionics) and military systems, the abbreviated test language for all systems (ATLAS) is used. The LABVIEW® software, mentioned above, includes the appropriate control software so that it is not necessary to write test programs.

3.4 Conclusions

More information on the topics and systems described above is available in O'Connor (2001). The book's homepage (URL: http://www.pat-oconnor.co.uk/testengineering.htm) provides lists of hardware and software suppliers.

4 SETTING UP AN ATS

To set up an effective ATS, it is necessary to consider

- the items to be tested in terms of quantities, technologies involved, test requirements, and so on;
- the test conditions to be applied (e.g. climatic, vibration, loads, electromagnetic, etc.; *see* **Article 214, Accelerated Testing, Volume 3**; **Article 216, Test Facilities, Volume 3**);
- the costs of the facilities and options (manual vs automatic) in terms of fixed and variable costs;
- the costs of alternative options for setting up the facility (use of other facilities, reliance on previous tests).

The management aspects of ATS are discussed in more detail in O'Connor (2001).

5 EXAMPLE OF ATS

Modern vehicle development testing relies heavily on ATS. PC-based systems control the testing and data acquisition of a number of parameters, while the vehicle is driven on test tracks or subjected to other environmental conditions. Vehicle subsystems such as power trains, suspension, and so on are also tested automatically on special facilities. The PC communicates with the various test instruments via a PXI bus. The design of the tests (parameters, limits, sampling rates, data acquisition and analysis, etc.) are programmed using appropriate software such as LABVIEW (mentioned above), *see* **Article 105, Modeling with LabVIEW™, Volume 2**.

REFERENCES

Johnson, G.W. (1997) *LabVIEW Graphical Programming*, 2nd edn, McGraw-Hill.

O'Connor, P.D.T. (2001) *Test Engineering*, John Wiley.

Sheppard, J.W. and Simpson, W.R. (1994) *System Test and Diagnosis*, Kluwer Academic Publishers.

Stroud, C.E. (2002) *A Designer's Guide to Built-In Self-Test*, Kluwer Academic Publishers.

216: Test Facilities

Patrick D.T. O'Connor
Consultant, UK

1 Nature and Scope of Test Facilities 1467
2 Materials, Components, and Structures 1467
3 Systems 1468
4 Special Test Facilities 1470
5 List of Key Test Facilities 1471
Related Articles 1471
References 1472

1 NATURE AND SCOPE OF TEST FACILITIES

In this module, the term test facility is used to include any location or organization in which test capabilities (product test, instrumentation, etc.) are provided. For convenience, test facilities are classified as follows:

- materials, components, and structures;
- systems;
- specialist facilities, including those for electromagnetic testing and calibration.

These are described in turn, including the main types of equipment used.

Facilities that are used wholly or mainly for manufacturing test are not included.

There are many independent test facilities, offering capabilities for most of the tests described below. Some of them specialize in particular industry or technology areas. Many of the industry-based facilities also offer independent services.

Major national and international test facilities are listed in Section 5.

2 MATERIALS, COMPONENTS, AND STRUCTURES

This article covers the methods used to test materials, components, and structures to determine their strength and other characteristics. These tests can be applied to samples of the materials used in which case the data is applicable to any component made of the material. However, sometimes the tests are performed on finished components and structures as well, particularly if the manufacturing processes can alter the material properties.

The main tests, with links to **Part 8, Elements: B – Signal Conditioning, Volume 2**, are as follows:

Static strength: Static tests are performed to measure strength: tensile, compressive, shear, buckling, and so on. Figure 1 shows a typical test machine.

Fatigue strength: Fatigue tests are performed on material specimens by applying controlled cyclic stresses, for example, using a tensile test machine, and recording the cycles until fracture occurs. For large components and structures, such as vehicle bodies, aircraft wings, and train bogies, special test rigs must be constructed, and the stresses are applied by hydraulic actuators. Figure 2 shows a rail passenger vehicle being tested using a hydraulically excited undersupport system. Such systems exist for testing several hundred tonne equipments such as battle tanks.

Hardness: Hardness is determined by measuring the indentation in material specimens when a defined shape is forced on to the specimen with a defined load. Two commonly applied hardness tests are the Brinell test and the Rockwell test. The Vickers and the Knoop hardness

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

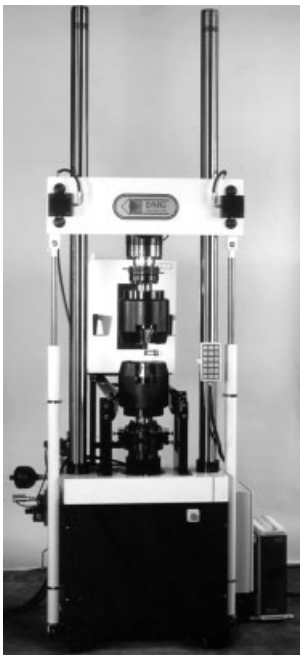


Figure 1. Tensile test machine. (Reproduced by permission of Denison-Mayes.)



Figure 2. Strength/fatigue test on rail passenger vehicle. (Reproduced by permission of TuV Product Service.)

tests are similar. This test is not developed from formal scientific description; in each case, the hardness number is an empirical value, with units of force/length^2. Its use is a comparative test.

Toughness: Toughness is measured by impact tests in which the energy level that causes a prepared specimen to fracture indicates the material toughness. The Charpy V-notch test uses a specimen with a defined V-shaped notch machined across the centre.

Wear resistance: Resistance to wear is measured by subjecting the materials to wear conditions (adhesive, fretting, abrasive, etc.) in suitable test rigs. The amount and nature of the wear is measured by examination of the surfaces and by weight comparisons.

Corrosion and chemical attack: Tests for the ability to withstand corrosion or chemical attack are performed by appropriate methods, such as salt spray, immersion, and so on.

3 SYSTEMS

3.1 Climatic, thermal

Climatic and temperature testing is performed using chambers in which the temperature can be varied. These can range in size from small bench-top units of less than 1 m^3, to very large facilities in which complete items such as cars, rail vehicles, aircraft, medium size houses, rocket motors under simulated space conditions, and major military systems can be tested. High temperatures are generated using electrical heaters. Radiant heating may also be applied. Low temperatures are generated by refrigeration units, or by using liquid nitrogen. Humidity, condensation, freezing, and precipitation conditions can be applied with appropriate chambers.

Typical chambers provide temperature ranges between $+60\,^{\circ}\mathrm{C}$ and $-40\,^{\circ}\mathrm{C}$. An important further performance characteristic is the rate of change of temperature that can be achieved, particularly for tests that require rapid or frequent temperature excursions. Powerful electrical heaters and liquid nitrogen cooling systems can generate rates of change of $20\,^{\circ}\mathrm{C\,min^{-1}}$, and up to $100\,^{\circ}\mathrm{C\,min^{-1}}$ in special chambers such as highly accelerated life test (HALT) chambers (**Article 214, Accelerated Testing, Volume 3**). Another way of generating high rates of change, or temperature shock, is to use a two-chamber system, with one hot chamber and one cold one, and a rapid transfer of the item under test between the two chambers.

3.2 Vibration

Items can be subjected to a vibration test by being secured to platforms that are vibrated. The vibration systems are variously called *shakers, vibrators, exciters*, or *thrusters*,

though vibrator is the ISO preferred term. The VIBKIT98™ (http://www.lds.com) software is an effective tool for determining vibration and shock test parameters. Harris and Crede (1976) cover the subject of vibration and shock testing in detail. *See also* **Article 205, Theory of Vibration Measurement, Volume 3**.

3.2.1 Vibration sources

The main vibration sources for test are as follows:

- Eccentric rotating weights. These are simple, low-cost systems, and being strictly harmonic systems, lack frequency spectrum modification.
- Electrodynamic (ED) (moving element with armature coil, suspended in a static magnetic field, and driven by alternating electric current). The frequency and intensity can be varied by the control system. Such systems can range in size and power from small vibrators with frequency ranges up to 10 kHz, to very large, water-cooled units with frequencies up to 2 kHz and thrust values up to 300 kN. Figure 3 shows an ED vibrator, being used for evaluating squeak and rattle in automotive assemblies.
- Pneumatic and hydraulic-actuator systems are used when larger displacements are needed, and for heavy items such as vehicles.
- Multiaxis systems, such as the following:
  - multiaxis shake tables (MAST). These are large systems that use multiple hydraulic or pneumatic actuators to drive the platform;
  - Multiaxis repetitive shock systems, such as HALT vibrator tables, which are powered by a number of pneumatic hammers mounted on the underside. These input completely random, 6 degree of freedom (3 axial, 3 rotational) vibration up to 0.1 g $(\text{rms})^2$/Hz over a wide frequency range (10 Hz–10 kHz). *See* **Article 214, Accelerated Testing, Volume 3**.
  - The Cube™ of TEAM Corp. (http://www.team.com) is a freely suspended hollow rigid cube, with pneumatic actuators mounted onto the inside faces (upper and four sides). Items to be tested are fitted to the outside faces.

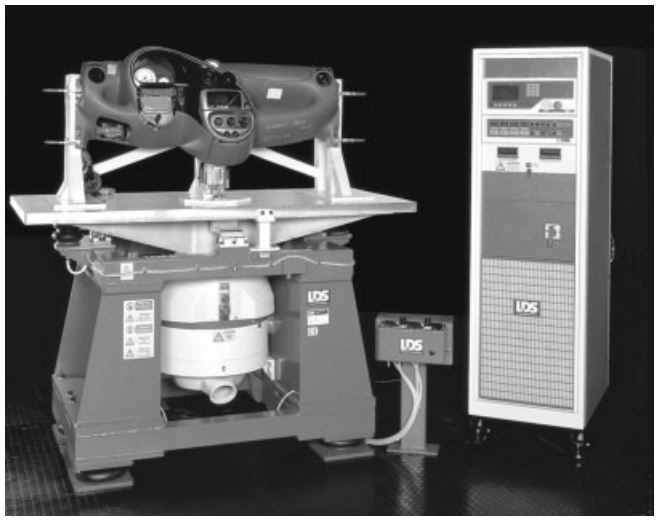


Figure 3. Electrodynamic vibrator (LDS). (Copyright © LDS.)

3.2.2 Vibration characteristics

Vibration may be applied at a simple fixed frequency, over a swept range of frequencies, or simultaneously over a selected range of frequencies. Random vibration is vibration that is applied over a range of frequencies and amplitudes that vary randomly over time. Pseudorandom vibration is what is actually applied, because generation of true random vibration is not practicable. By stimulating a range of vibrations simultaneously, the existence of simultaneous resonances can be determined.

Power spectral density (PSD) is the term used to quantify vibration power at a frequency within the range. The unit of PSD is g_n^2/Hz (or, more correctly, g_n rms^2/Hz or the SI units $\text{m/s}^2/\text{Hz}$). The total force rating of a vibration system, over its frequency range, is expressed as g_n rms, kgf rms, N rms or lbf rms. It is the area under the curve of PSD versus frequency, such as in Figure 4.

Vibration control systems permit the vibration pattern and intensity to be planned and controlled. The vibration parameters measured on the unit under test (usually from suitably positioned accelerometers) are compared with the defined inputs, and the controller provides the required outputs to the drive amplifier of the vibrator. A typical random vibration test specification is shown in Figure 4.

Other vibration and shock inputs can also be applied, such as sine superimposed on random, specific shock patterns, and so on.

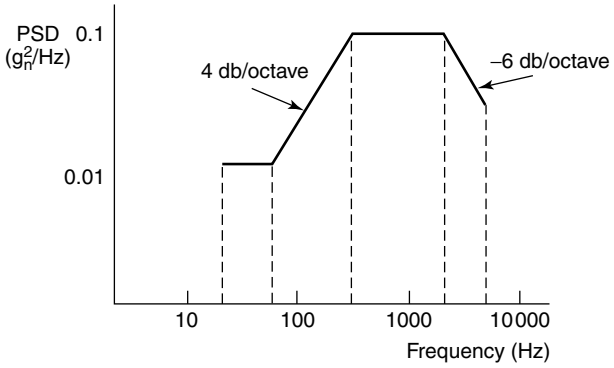


Figure 4. Typical random vibration test profile. (Reproduced from *Test Engineering*, O'Connor, 2001 © John Wiley & Sons Ltd.)



Figure 5. Combined environment test facility. (Repoduced by permission of Thermotron.)

3.3 Shock testing

Shocks are single events of high energy, or which might result in short-duration high frequency vibration. Mechanical shock or impact can cause fracture, or repeated shocks can cause fatigue leading to fracture. Shock tests are performed by dropping the item on to a rigid base, by using shock test machines, which allow defined accelerations to be applied, or by using more dramatic methods such as crash tests on vehicles. The term often used for shock impact is *jerk*, that being the derivative of acceleration.

3.4 Combined environments

To be realistic, environmental stresses are often applied in combination. This is sometimes referred to as *combined environmental reliability testing* (CERT). Typical combinations used are temperature and vibration, as well as other stresses such as electrical power, humidity, solar radiation, and so on. Figure 5 shows a CERT chamber along with its computer control and data acquisition and visual display recording system.

4 SPECIAL TEST FACILITIES

4.1 EMI and EMC

Electromagnetic interference (EMI) is the conducted or radiated emissions ('noise') from an electrical or electronic equipment that could interfere with the correct operation of other equipment.

Electromagnetic compatibility (EMC) is the capability of equipment to operate correctly in the presence of known levels of electromagnetic radiation or conducted electrical or magnetic disturbance. This is sometimes called electromagnetic immunity.

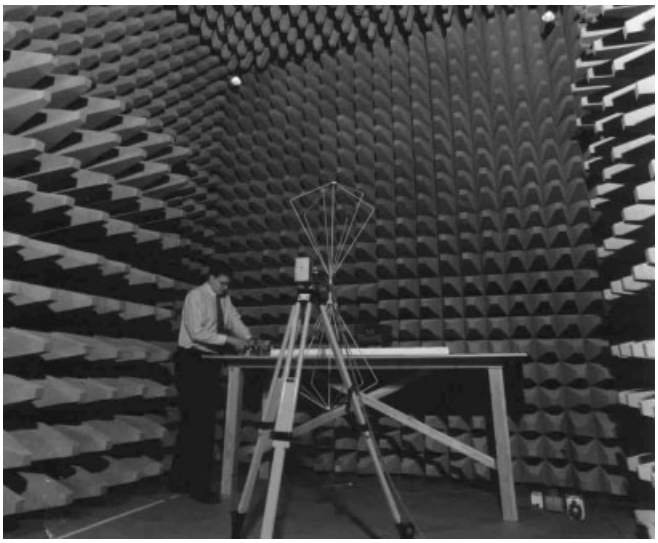


Figure 6. EMI/EMC test chamber. (Reproduced by permission of TuV Product Service.)

Testing for EMI/EMC consists of operating the equipment and measuring the noise outputs, and submitting the equipment to specified conditions. Special Radio Frequency (RF) anechoic chambers, equipped with appropriate measuring and transmitting equipment, are used (Figure 6). The cones on the walls are foam laden with RF absorbing materials to attenuate RF reflections inside the chamber. Doors are sealed, using silver finger contacts, to prevent RF emissions. Walls are provided with RF screening to make the enclosure an effective Faraday Cage.

Williams (1992) describes EMI/EMC requirements and test methods. *See also* **Article 17, EMC and EMI, Volume 1**.

4.2 Industry-specific facilities

Many large companies operate test facilities for their own product development. The major automotive companies operate facilities that include most of the special test equipment described in this module, as well as test tracks, dynamometers, and so on. Military and aerospace companies, as well as other industry sectors (telecommunications, railways, etc.), also operate appropriate facilities. Industry-specific facilities are also operated by consortia and by government and nongovernment organizations.

4.3 Calibration

Nationally accredited facilities are operated in most countries to provide calibration with reference to primary

metrology reference standards, to which all lower-level calibrations must be traceable. Calibration requirements and standards are described in **Article 46, Calibration Interval, Volume 1**; and **Article 47, Internet Calibration, Volume 1**, *see also* Morris (1997).

There is an increasing trend toward international and transnational operations for test and calibration as this avoids double testing and reduces duplication of test plant.

The international standard for the accreditation of measurement and test laboratories is ISO17025.

A test facility may be offered for third-party testing, in which case it is likely to be offered under a certified accreditation scheme. Many test equipments, however, are proprietary and kept for company use only.

4.4 Other specialist facilities

Other specialist test facilities exist to provide services as required for other conditions and susceptibilities, such as humidity, mold growth, atmospheric pressure, space vacuum, ultraviolet radiation, and so on. Information on these can be obtained from the lists in Section 5, and from appropriate specialist publications and standards.

5 LIST OF KEY TEST FACILITIES

This list presents information on other test facility lists that are available. It also provides a selection of the major national and international test facilities.

5.1 Test facility listings

1. Homepage to O'Connor (2001) (http://www.patoconnor.co.uk/testengineering.htm). This includes lists of facilities, information sources, and so on.
2. Institute of Environmental Sciences and Technology (USA) (IEST) (http://www.iest.org). The IEST publishes lists of test facilities and equipment suppliers, mainly in the USA.
3. Society of Environmental Engineers (SEE) (UK) (http://www.environmental.org.uk). The SEE publishes a directory of test facilities and equipment suppliers, mainly in the United Kingdom but also in the rest of Europe and the United States.
4. The relevant national accreditation agencies maintain lists of accredited test facilities. UK Accreditation Service or UKAS (http://www.ukas.org.uk); USA: Underwriters' Laboratory UL (http://www.ul.org).

5.2 Selected UK facilities

1. AEA Technology (http://www.aeat.co.uk). Test facilities, including nondestructive test NDT (UK National NDT Centre).
2. British Standards Institution (BSI) (http://www.bsi.org.uk). BSI operates test and measurement laboratories.
3. Electrical Research Association (ERA) (http://www.era.co.uk). Wide range of materials and component test and research, EMI/EMC test.
4. Motor Industry Research Association (MIRA) (http://www.mira.co.uk). Test facilities, primarily to support automotive applications.
5. National Engineering Laboratory (NEL) (http://www.nel.co.uk). Structural testing and flowmeter evaluation.
6. National Physical Laboratory (NPL) (http://www.npl.co.uk). NPL maintains the primary UK reference standards for calibration and measurement. It undertakes measurement research and development, calibration, and EMI/EMC testing.
7. QinetiQ (was Defence Evaluation and Research Agency (DERA)). (http://www.qinetiq.co.uk). Range of test facilities, primarily for military equipment and systems test.

5.3 Selected USA facilities

1. Underwriters' Laboratory (UL) (http://www.ul.com). In addition to maintaining the primary US national reference standards for calibration and measurement and accrediting lower-tier facilities. UL operates test and measurement laboratories.
2. Wyle Laboratories (http://www.wylelabs.com). Wide range of environmental test facilities.

5.4 International

1. TuV Product Service (TUV-PS) (http://www.tuvglobal.com). TuV operates a number of test laboratories in the United States, United Kingdom, and Europe, covering a wide range of environmental conditions, including HALT/HASS (*see also* **Article 214, Accelerated Testing, Volume 3**).

RELATED ARTICLES

Article 17, EMC and EMI, Volume 1; **Article 214, Accelerated Testing, Volume 3**.

REFERENCES

Harris, C.E. and Crede, C.E. (1976) *Shock and Vibration Handbook*, McGraw-Hill, New York.

Morris, A.S. (1997) *Measurement and Calibration Requirements for Quality Assurance to ISO9000*, Wiley, Chichester.

O'Connor, P.D.T. (2001) *Test Engineering*, Wiley. URL: http://www.pat-oconnor.co.uk/testengineering.htm.

Williams, T. (1992) *EMC for Product Designers: Meeting the EMC Directive*, Newnes, Chichester.

217: Instrument Evaluation

Steve Cork
Sira Test and Certification Ltd, Kent, UK

1 Introduction 1473
2 The Need for Instrument Evaluation 1473
3 Cooperative Instrument Evaluation 1474
4 Instrument Evaluation Methodology 1475
5 Requirements of an Instrument Evaluation Laboratory 1478
6 Results of Instrument Evaluations 1479
References 1480

1 INTRODUCTION

A definition of instrument evaluation that is generally accepted among the major cooperative instrument evaluation groups is 'the investigation and testing of the performance of instruments and related equipment in accordance with an agreed program based on the manufacturer's specification, relevant national, and international standards and specific user requirements'. The agreed program will usually include performance evaluation under reference conditions and simulated environmental conditions (often referred to as *influence conditions*) together with an assessment of standards of construction and adequacy of documentation, such as operating instructions and maintenance handbooks.

The process can most succinctly be described as a comprehensive technical audit. The concept of instrument evaluation has been developed over a 40-year period in order to meet the needs of instrument users and manufacturers of industrial instrumentation and process control equipment. It is aimed at producing the maximum possible amount of information on the performance of instruments and related equipment bearing in mind the needs of the user and the rights of the manufacturer to a fair and consistent assessment of his product.

2 THE NEED FOR INSTRUMENT EVALUATION

Back in the 1950s, with the increasing use of sophisticated measurement and control equipment in process plants (which themselves were increasing in size, complexity, and cost) came the realization of the increasing financial consequences of instrument failure. Such consequences do not only apply in the event of a complete failure of an instrument, which, because of a fail-safe system philosophy, could result in a costly plant shutdown. They also, and more commonly, apply in cases in which the performance of the instrument gradually deteriorates in an undetected manner as a result of an influence condition or just over a period of time. Such deterioration can cause a shift in the performance of an automated plant away from a state of optimum performance as defined by a required cost and quality target. It is relatively straightforward, although sometimes costly, to provide instrument redundancy so that a fail-safe plant shutdown only occurs if, say, two out of three instruments fail simultaneously in a catastrophic sense. However, it is less simple to arrange that such redundancy will compensate for the gradual deterioration of the measurement or for the susceptibility of the instrument to the variations of a particular environmental parameter, such as temperature, vibration, or electromagnetic interference. The financial consequence of instrumentation failing to meet its expected performance criteria led to an increasing insistence by user industries of comprehensive evaluation as

Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn.
© 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8.

a prerequisite to purchasing decisions. It is however recognized that in recent years partial control over the purchasing of equipment has been relinquished by some end users. This has come about by the contracting out of the supply of systems or parts of systems to third parties, with these contractors being more likely to rely on manufacturers' published specification rather than on independent comprehensive evaluation reports.

It is noted however that a concept that is becoming increasingly recognized in industry is the one of Life Cycle Costing, which considers the costs incurred from the conception of an application to the point of decommissioning and abandonment. Life Cycle Costing has two main components: CAPEX (CAPital EXpenditure), which includes all expenditure from conception to the end of commissioning, and OPEX (OPerating EXpenditure), which covers the cost from then on. It is clear for nominally identical products from different manufacturers that apart from the ex-factory cost of the item itself, the CAPEX costs (including those associated with design, specification, installation, commissioning, and documentation) are similar. However, OPEX costs can be quite different, primarily due to the ability of the equipment to operate within certain performance limits, over an extended period of time under the prevailing environmental conditions. For this reason, if for none other, the need for comprehensive instrument evaluation remains.

The cost of carrying out a comprehensive evaluation naturally depends on the type of instrument to be investigated and the details of the testing required, but would typically range from £10 000 to £30 000. While this may seem a reasonably large sum of money, it must be balanced against the savings that can be made from making informed decisions on the basis of evaluation findings.

While the needs of manufacturers for instrument evaluation are less obvious, they nevertheless exist. Particularly in relation to new designs, it is important to identify a serious design weakness before a significant number have been delivered. The direct cost penalty of retrieving and modifying such instruments is obvious enough, the penalty in terms of lost customer confidence and future sales is more serious. While the manufacturer may carry out significant testing of his product before its appearance on the market, there are still a surprising proportion of cases in which an independent evaluation reveals weaknesses that have not previously been identified. Data shows that a reasonably high proportion of evaluations lead to necessary design modifications, and it is not unusual for manufacturers to commission an independent evaluation prior to a major launch of a new instrument.

A more detailed account is given later in this article of the different phases of an evaluation.

3 COOPERATIVE INSTRUMENT EVALUATION

By cooperative instrument evaluation is meant the sponsorship of evaluation programs by groups of companies on the basis of cost sharing and results sharing. Instrument evaluation had been practiced individually for many years by the major instrument using companies in the chemical, oil, steel, and other industries as a means of guiding purchasing decisions, before cooperative instrument evaluation began in the late 1950s. In 1958 and 1959, Sira (then the British Scientific Instrument Research Association) commenced instrument evaluation work under contract separately to the Central Electricity Generating Board, as it was then known, and the United Kingdom Atomic Energy Authority. It was known that other major users of instruments had their own programs of instrument evaluation, and it was suggested that such users would obtain benefits from joining forces to sponsor evaluations cooperatively and share results. With a great deal of help from the UKAEA, a proposal for such a cooperative group project was put to a number of instrument user companies, and in 1961 a meeting was held at Sira, at which the cooperative instrument evaluation scheme was formalized. The founder members comprised CEGB, UKAEA, Imperial Chemical Industries (ICI), Distillers Co. (subsequently part of BP), Ministry of Defence, and Shell Research. The group, which eventually gave itself the acronym SIREP (Sira Instrument Evaluation Panel), had its inaugural meeting on 6 December 1961. SIREP, which recently changed its name to Evaluation International (EI) (http://www.evaluationint.com), contracts evaluation projects to Sira Test & Certification Ltd (http://www.siraservices.com) amongst other test houses.

In 1963, a similar group was formed in The Netherlands called WIB (working-party on instrument behavior). WIB (http://www.wib.nl) mainly uses the laboratories of TNO (Toegepast Natuurwetenscappelijk Onderzoek) (http://www.tno.nl) in Holland, although they have evaluations carried out in various laboratories across Europe. Finally in 1974, a French group, Association des Exploitants D'Equipments de Regulation Automatique (EXERA) (http://www.exera.com), was formed.

Although the administrations of EI, WIB, and EXERA differ in detail, the general objectives are the same, and over the years a close collaboration has grown between the three organizations, examples being the sharing of evaluation reports between the three groups and liaison to ensure that equipment by one group is not then evaluated by another. Although the membership of EI, WIB, and EXERA is still predominantly European there are member companies from North America, Japan, and the Middle East. The joint membership is currently over 70, which includes certain

group memberships such as the Swedish consortium SIP (Stiftelsen for Instrument Proving), which itself comprises some 30 members.

In the early years of instrument evaluation, it would be fair to say that the primary focus was on general process instrumentation, such as pressure, temperature, level, and flow equipment. As time moved on, the emphasis has moved onto more complex devices and certainly over the last five to ten years the main area of interest has been Analytical and Environmental Safety equipment. A list of the equipment having been evaluated by the three organizations can be found at their Web sites.

The commercial benefits of joining a cooperative evaluation group are clear for a major instrument user. Taking Evaluation International as an example, the annual subscription for the year 2004 is £9300, which is less than the cost of a single evaluation project. For this subscription, the members receive copies of all the reports issued by EI as well as those produced by WIB and EXERA. While a user may not be interested in all the reports produced, the contents of a single report each year can easily save a member his annual cost many times over.

4 INSTRUMENT EVALUATION METHODOLOGY

A number of steps can be identified in a typical instrument evaluation program, consisting of

- agreement on test program;
- tests under reference conditions;
- tests under simulated environmental (influence) conditions;
- assessment of construction standards, ease of use, and literature quality;
- review of manufacturer's quality assurance system and instrument status;
- draft report;
- final report.

Figure 1 shows how these steps are incorporated into a typical Evaluation International sponsored project.

4.1 The test program

The test program is a key document. It is prepared by the evaluator, who is guided by the manufacturer's specification, standard specifications, and standards relating to methods of evaluating and calibrating equipment. In particular, his experience guides him to seek performance information not revealed by the manufacturer's specification.

The basic reference document for the development of the test program is the manufacturer's specification; however, it is often a very incomplete statement of the performance, particularly in relation to influence conditions that depart significantly from laboratory reference conditions. In such circumstances, it is necessary to look for other guides to performance, which are acceptable both to manufacturer and user. Whether or not the specification is complete, the actual methods used for the evaluation need to be acceptable to manufacturer and user and not subject to significant modification when applied to different instruments of the same kind.

For these reasons, a great deal of attention has been given to specifications and testing procedures by national and international standards bodies. A standard specification is a technical specification or other document that has been evolved by cooperation between various bodies, and represents a consensus of the parties concerned. Such a standard may set out the quantities to be specified, the boundary values of accuracy classes, an explanation of the terminology used, and an outline of methods of measuring the quantities. Therefore, in the absence of a sufficiently detailed manufacturer's specification, a test program can be based on a national or international standard specification, with the manufacturer concerned having the reassurance that his instrument's performance will be compared with a specification evolved by consensus, that comparative instruments will be subject to the same comparison, and that the test methods used for the evaluation will be the same as that used for other instruments. There will of course be occasions where no such standard specification exists and the formulation of the test program relies even more on the experience of the evaluator, although a useful document in the generation of any such program would be IEC 60770-1:1999 'Transmitters for use in industrial-process control systems. Methods for performance evaluation', which largely cross references to the four-part standard IEC 61298 'Process measurement and control devices – General methods and procedures for evaluating performance'.

Once complete the test program is submitted to the sponsor and the manufacturer (if not the sponsor) and the evaluation only proceeds when the three parties, laboratory, manufacturer, and sponsor, have agreed to the program. In many instances, the agreement to a program is almost a formality. In other cases, the methods to be used to establish basic accuracy and environmental effects can be the subject of controversy, and lengthy discussions maybe required before an agreement is reached. For example, evaluation and calibration procedures for temperature or pressure transmitters are widely documented and accepted, but this is not the case for some

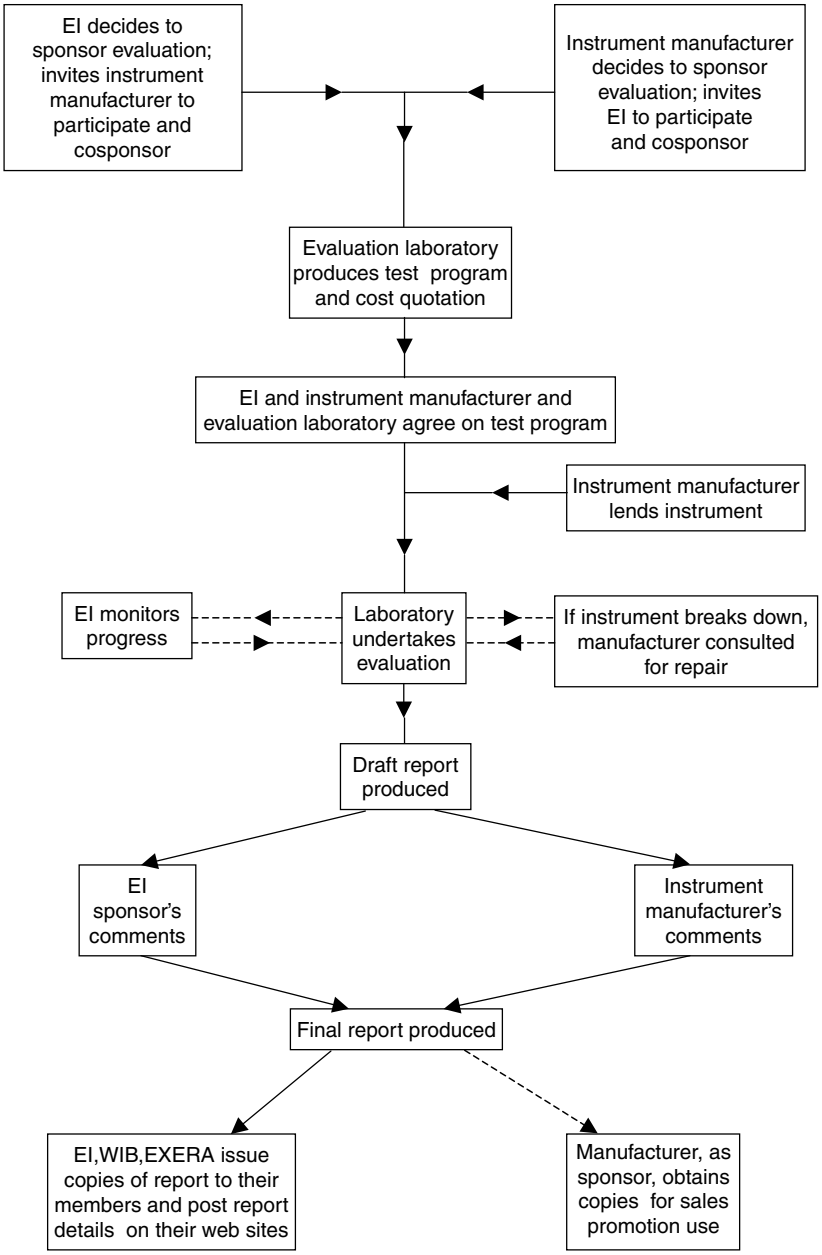


Figure 1. Evaluation International evaluation procedure.

chemical analysis devices. In such cases, evaluation work is often preceded by an amount of research and development in order to arrive at mutually acceptable methods of evaluation.

4.2 Tests under reference conditions

Presented below is a set of standard tests as indicated by IEC 60770-1.

Basic accuracy: Average error, repeatability, hysteresis, and linearity determined from three to five upscale and downscale full-range traverses, measuring at least six points.

Dead band: The smallest change results in a detectable output change at 10, 50, and 90% output.

Frequency response: The frequency response measured with a 20% span peak-to-peak input.

Step response: Record the step response time and the time for the output to reach and remain within 1% of its

final steady value, for varying input steps corresponding to 10 and 80% of the output span.

Start-up drift: Output monitored for 4 h after power is switched on.

Long-term drift: Output monitored for 30 days with an input of 90% span.

4.3 Tests under influence conditions

Presented below is a selection of standard tests from IEC 60770-1, which maybe applied to an electrically powered transmitter. It could be argued that some of the tests are covered by current legislation, such as immunity to radiated electromagnetic interference testing as part of CE marking in accordance with a European Directive. However, it is possible that the requirements of the legislation do not fully meet the needs of the demanding instrument user and indeed the work carried out in obtaining the approval of the legislation may not have been based on actual test results at all.

Ambient temperature: Specified tests to determine effects of exposing the instrument to

1. dry heat (+55 °C maximum or higher if applicable);
2. cold (−25 °C or as applicable)

in comparison to the operation under reference conditions.

Humidity: Specified tests to determine the effects of exposing the instrument to 40 °C, 93% relative humidity in comparison to the operation under reference conditions.

Vibration (sinusoidal): Effect of vibrating the instrument along three mutually perpendicular axes at specified peak displacements, or peak accelerations, over a specified frequency range. The effects being looked at for being both changes on the output and any mechanical resonances that may result in a premature failure.

Mounting position: Effect on output of 10° inclination about two planes at right angles to each other.

Over-range: Measurement of residual changes at 0 and 100% span, which result from over-ranging the input by 50% for a period of 1 min and after allowing a 5-min recovery period.

Accelerated life: Measurement of the change in performance of the instrument after the application of 100 000 cycles of amplitude equal to half the span.

Supply voltage and frequency variations: Effect at 0 and 100% span of changing the supplied voltage (and supply frequency where appropriate) between specified limits.

Short-term supply voltage interruptions: Measurement of the output changes, and recovery time, following a series of interruptions of specified duration.

Common mode interference: The effect of applying extraneous voltages between earth and each input and output terminal. Initially, by applying 250 V rms at mains frequency, followed by positive and negative 50 V DC potentials.

Magnetic field interference: Effect on output of a mains frequency magnetic field of 400 A m^{-1}.

Open and short-circuit of input/output: Effect of interrupting each input connection in turn for 5 min, followed by the effect of shorting the connection. Test repeated for each output connection.

Radiated electromagnetic interference: Effect on output of subjecting instrument to a specified field strength, typically 10 V m^{-1}, over a specified frequency range, typically 80 MHz to 1 GHz.

Output load variations: Effect on the instrument performance of varying the output load within specified limits.

Degree of protection provided by enclosure: Determination of the effectiveness of the enclosure's protection to water and dust ingress by performing appropriate tests as defined in IEC 60529 'Degrees of protection provided by enclosures (IP Code)'.

4.4 Assessment of construction standards, ease of use and literature quality

Standards of construction and finish: Overall design quality, robustness, and quality of components.

Packaging: Adequacy and method of delivery.

Use: Ease of operation in terms of installation and commissioning, general operation, maintenance requirements, and component accessibility.

Safety: Clarity of labeling, accessibility to live terminals, whether certified for use in potentially explosive atmospheres.

Documentation: Quality, comprehensiveness and language(s) of documentation, adequacy of installation, operating and servicing instructions, inclusion of circuit diagrams, troubleshooting routines.

4.5 Review of manufacturer's quality assurance system and instrument status

Quality assurance system: The manufacturer is asked to provide details of any structured quality assurance policy that they may have adopted, if so, on which national/international standard it is based and if it is registered under any particular scheme.

Instrument status: The manufacturer is requested to confirm the expected life expectancy of the instrument, the guarantee period for the equipment and for how long after manufacture the product would be supported.

4.6 The evaluation report

A standardized format is used for reporting and IEC 61298 – 4 gives guidance as to the content of the report. The principle followed is to present findings in a concise and easy-to-refer-to manner. Major findings are presented immediately after a short introduction and contain enough information to enable a reader who does not wish to immerse himself in details to assimilate quickly the salient facts. A very important aspect of the report is a section devoted to the manufacturer's comments. This contains verbatim comments, additional information, and criticism (occasionally) by the manufacturer on the findings.

This section is followed by tabular and graphical results, and the main part of the report is concluded with a description of the test methods and references.

The report may have appendices, when appropriate, for additional theoretical discussion, descriptive material, quality assurance and instrument status information (provided by the manufacturer), and a copy of the manufacturer's relevant specification sheet.

5 REQUIREMENTS OF AN INSTRUMENT EVALUATION LABORATORY

In order for a laboratory to be competent to carry out comprehensive instrument evaluation projects, they must have appropriate staff, facilities, and procedures, and the adequacy of these can be independently confirmed by means of accreditation to an appropriate international standard.

5.1 Staff

As discussed above, a key document in the process of an evaluation is the test program. It is therefore important that in addition to knowledge of the current status of relevant specification standards, staff involved in the formulation of the program should have a fundamental understanding of the operation of the device, as well as the likely applications in which it would be used. To cope with the range of instruments being evaluated, staff experience with a mix of basic disciplines – physics, chemistry, electronics, and mechanics – is essential.

Once underway, in order that an evaluation is carried out efficiently, it is important that staff involved in the technical aspects of the work have a good understanding of the operation of the device being tested. In addition, it is beneficial for them to have an appreciation of the likely performance of the equipment under certain conditions in order that any spurious readings can be recognized as such, allowing work to be repeated at the time of testing rather than having the costly exercise of setting up a repeat test at a later date.

5.2 Facilities

The facilities that are used during an evaluation fall into three categories. Those that are used to supply an input to the equipment under test, those that are used to measure the instrument's response, and finally, there are the facilities that are used to generate an influence on the equipment.

5.2.1 Input generating devices

For the vast majority of evaluations, these would be well-proven devices, such as a pressure balance for the evaluation of pressure transmitter or a temperature bath and reference thermometers combination for the evaluation of a temperature device. It would be normal to expect the input device to be traceable to a national standard and its uncertainty to be significantly better than the quoted performance of the instrument under test. However, in occasional circumstances, for example, where a reference standard does not exist, it maybe acceptable to have an input that has been shown to be repeatable but the absolute performance is unknown.

5.2.2 Output measurement devices

A growing number of instruments tested either have a digital display or a digital output, which can be read using some form of communicator or communications software

usually supplied with the instrument. In these cases, the laboratory may at most need to provide a personal computer for taking output measurements.

However, many instruments still provide what would be regarded as standard outputs, whether they be voltage, current, frequency, or even pneumatic. In all these cases, the laboratory would be expected to have suitable measurement equipment, which would be traceable to national standards.

5.2.3 Influence generating facilities

From the list of possible tests given previously, it is clear that a wide range of facilities are required to perform a comprehensive evaluation program. Again, traceability of the measurement of the influence is important in order to demonstrate that any test criteria have actually been met. In addition, by ensuring traceability it is possible at some point in the future to replicate the test conditions either on the same piece of equipment, possibly in the case of a dispute, or on a subsequent device for comparative reasons.

5.3 Procedures

Having competent staff and appropriate equipment may in itself not be sufficient to ensure that the work is performed correctly, an important third component is written procedures. These may already be part of the agreed test program but are commonly taken from national or international standards or in-house laboratory procedures. The purpose of the procedures is to ensure that the work is carried out in a reproducible fashion to allow comparisons to be made between results obtained from different instruments. The content of the procedure may be more or less detailed depending on the nature of the work.

5.4 Accreditation

For a customer of an evaluation project, it is important for them to know that the laboratory they have chosen for the work has the above three components – competent staff, suitable facilities, and the appropriate procedures. Clearly an option open to them is to carry out an audit of the three aspects; however, there is an alternative and that is to choose a laboratory that has been independently accredited. It may not be reasonable to expect a laboratory to be formally accredited for all the tests they are to perform, primarily due to financial constraints. However, it would be reasonable to expect them to be accredited for the primary measurement, whether it be pressure, temperature, flow or suchlike, as well as for many of the influences that the device will see. The most appropriate standard for laboratories carrying out comprehensive instrument evaluation to be accredited against is ISO 17025:1999 'General requirements for the competence of testing and calibration laboratories'.

In the United Kingdom, the sole national accreditation body recognized by government to assess, against internationally agreed standards, organizations that provide calibration and testing services is the United Kingdom Accreditation Service (http://www.ukas.com). UKAS has a number of mutual recognition agreements with other national accreditation bodies.

6 RESULTS OF INSTRUMENT EVALUATIONS

Periodically, EI, WIB, and EXERA commission an analysis of the results of evaluation reports that they have commissioned. The last such analysis carried out in 2002, covered 107 reports on equipment supplied from 12 different countries, with the sample comprising the following instrument types.

Analytical and environmental safety	42%
Pressure	15%
Flow	10%
Temperature	10%
Control valves and ancillaries	12%
Level	10%
Others	1%

One might think that the results of an evaluation lose some validity in that they are based on tests carried out on a model that has been hand picked by the manufacturer and could therefore be unrepresentative. However, the figures would suggest that such hand picking does not occur, and the author's knowledge of instrument manufacturing companies suggests that it is unlikely to occur.

The analysis carried out shows the proportion of equipment deemed unsatisfactory in various categories. The categories and the relevant proportions are as follows:

1.	Unsatisfactory as received	20%
2.	Outside specification under reference conditions	39%
3.	Outside specification under simulated environmental (influence) conditions	74%
4.	Breakdown during evaluation	20%
5.	Inadequate documentation	89%

In addition, the analysis showed that 21% of the devices met all the required specifications under both reference and influence conditions, and 29% of the equipment evaluated in the sample were subject to modification in design or manufacturing method as a result of the evaluation.

Category 1 records the proportion of instruments, which, in some way had a fault on receipt, sometimes trivial such as a missing fuse, sometimes serious such as completely uncalibrated or missing one or more components.

In categories 2 and 3, the manufacturer's specification was used wherever possible. However, as can be seen below from the comments on category 5, this was more often not the case. Where performance was not specified by the manufacturer, it was judged in relation to a specification based on current practice in instruments of a similar kind.

Category 4 records the proportion of instruments that broke down during the evaluation. It was found that the breakdowns were more likely to be due to a mechanical fault than an electrical one.

Category 5 records the results of an assessment of the documentation accompanying the equipment. The main documents are the operating handbook, covering installation, operating principle, and operating instructions, the maintenance handbook, and a performance specification sheet. The evaluator forms an opinion of the adequacy of these as a guide to the potential purchaser.

It was indicated above that about one-quarter of the devices in the sample were subjected to modification as a consequence of the evaluation. The criterion for inclusion in this figure was that the manufacturer had made a statement in his published comments to the effect that such a change would be made.

While a number of conclusions can be drawn from the results, category 1 results probably cause more concern to users than the other categories. Other than failure occurring during finished goods storage or transit, the conclusion is that some instruments do not receive sufficient final testing for an obvious malfunction to be detected. There was no significant evidence of physical damage, which could be correlated with the instruments found to be faulty.

While in categories 2 and 3 no attempt has been made to assess the proportions of instruments that could be regarded as serious or less serious departures from specified performance, the figures suggest a tendency to optimism in specifications. Some manufacturers have indicated that this is so with regard to certain instruments and state that it is often due to the commercial pressures caused by competitors making performance claims known to be exaggerated. The problem is made more difficult by variations in the way in which performance specifications of similar kinds of instruments are described. For example, it is often unclear whether a specified performance relates only to reference conditions or takes some account of influence conditions. A significant improvement could be effected in this situation if the methods of specifying performance of instruments of a similar kind could be standardized.

Category 4, dealing with failures leading to instrument breakdown, cannot be used to draw conclusions about component reliability. The number and types used in instruments vary widely, and deductions of component reliability are better made from controlled tests on particular categories of components.

Category 5 results indicate that insufficient attention is being given to the preparation of information, which ensures that users get maximum effectiveness from the instrument after delivery. Further analysis shows that 80% of the instruments tested had inadequate performance specifications, and only 60% of the instruments had a manufacturer's specification for any influence condition. The subject of documentation for instrumentation is covered by British Standard 4462:1969 'Guide for the preparation of technical sales literature for measuring instruments and process control equipment'. It is hoped that manufacturers would produce their documentation in line with this standard in order that end users could benefit from the information available and allow them to make better comparisons between like equipment. However, it is recognized that such a hope is probably a forlorn one.

A reasonably high proportion of instruments are modified as a direct result of evaluation. The figure of 29% emphasizes both the benefit of and the need for instrument evaluation. It is a direct action to improve performance resulting from evaluation, and it indicates the positive way in which instrument manufacturers respond to the results of impartial evaluation.

Some of the books that may be of value are Anonymous (1989), Coggan (1990), Driscoll, Gudzinowicz and Martin (1984) and Sydenham (1983). There appears to be no more recent statements on the subject.

REFERENCES

Anonymous (1989) *Evaluation Verification and Maintenance: Laboratory Instrument Manual*, 4th edn, College of American Pathologists.

Coggan, D.A. (1990) *Preparing for Instrumentation Technician Evaluation, Self-Study System/Book and Disk*, ISA – The Instrumentation, Systems, and Automation Society.

Driscoll, J.L., Gudzinowicz, B.J., and Martin, H.F. (1984) *Instrument Evaluation in Biomedical Sciences*, *Clinical and Biochemical Analysis: A Series of Monographs and Textbooks*, Vol. 15, Marcel Dekker.

Sydenham, P.H. (1983) *Handbook of Measurement Science*, Chapter 31 part Laskaris E.K. Evaluation, Vol. 2, Wiley, New York.