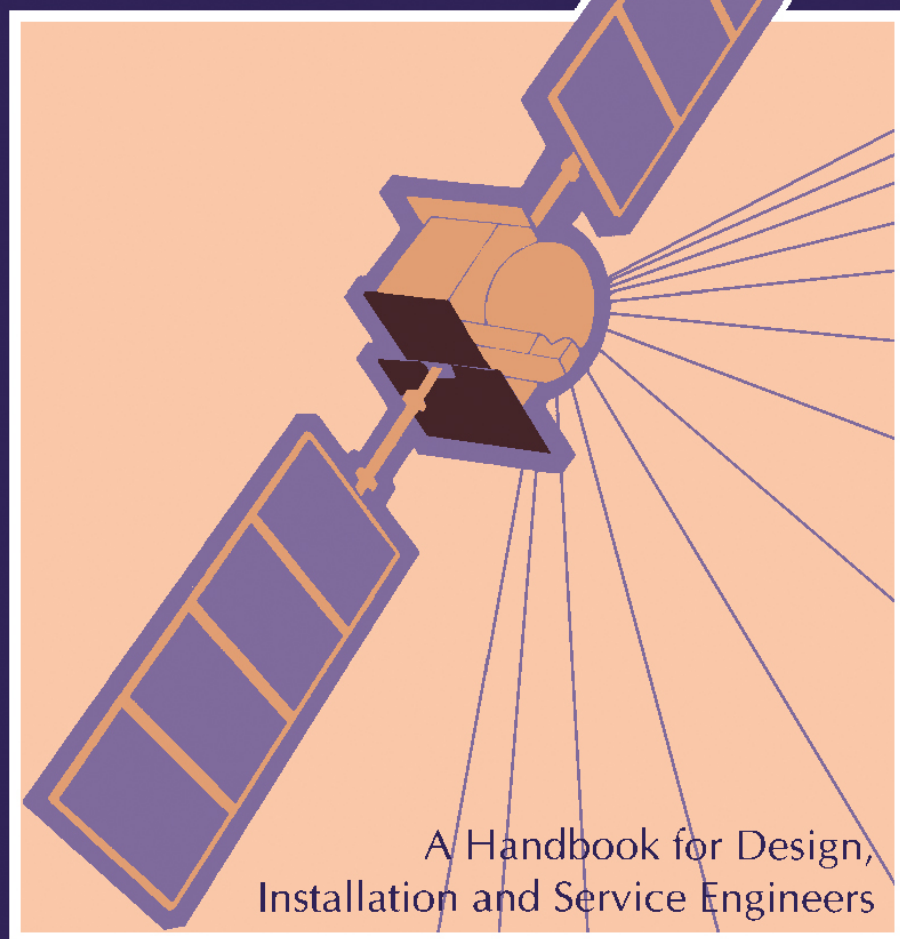


COMMUNICATION SERVICES VIA SATELLITE

G.E. Lewis



A Handbook for Design,
Installation and Service Engineers

SECOND EDITION

DISH 1010

Communication Services via Satellite

This book belongs to Quinn Ford
(qford@perkinscoie.com)

Copyright Elsevier 2022

This second edition is dedicated to the memory of Derek Henry in recognition of his encouragement and long friendship.

Communication Services via Satellite

A Handbook for Design, Installation
and Service Engineers

Second edition

Geoffrey E. Lewis

BA, MSc, MRTS, MIEIE

BUTTERWORTH
HEINEMANN

Butterworth-Heinemann Ltd
Linacre House, Jordan Hill, Oxford OX2 8DP



PART OF REED INTERNATIONAL BOOKS

OXFORD LONDON BOSTON
MUNICH NEW DELHI SINGAPORE SYDNEY
TOKYO TORONTO WELLINGTON

First edition published by BSP Professional Books Ltd 1988
Second edition published by Butterworth-Heinemann Ltd 1992

© Geoffrey E. Lewis 1988, 1992

All rights reserved. No part of this publication may be reproduced in any material form (including photocopying or storing in any medium by electronic means and whether or not transiently or incidentally to some other use of this publication) without the written permission of the copyright holder except in accordance with the provisions of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London, England W1P 9HE. Applications for the copyright holder's written permission to reproduce any part of this publication should be addressed to the publishers

British Library Cataloguing in Publication Data

Lewis, Geoffrey E.

Communication Services Via Satellite:
Handbook for Design, Installation and
Service Engineers. – 2 Rev. ed

I. Title
621.3825

ISBN 0 7506 0437 9

Printed and bound in Great Britain

Contents

<i>Preface to the second edition</i>	ix
<i>Preface to the first edition</i>	x
<i>Acknowledgements</i>	xii
1 The Basic Concepts	1
1.1 Advantages of satellite communications	1
1.2 Orbits in use	1
1.3 Frequencies used	3
1.4 Maximising the availability of the frequency spectrum	5
1.5 System constraints imposed by transponder characteristics	8
1.6 Characteristics of modulation systems	9
1.7 Companding	13
1.8 Energy dispersal	14
1.9 Digital modulation systems	15
1.10 Up and down-link frequencies	19
2 Noise and Interference	20
2.1 Noise sources	20
2.2 System noise factor	20
2.3 Equivalent noise temperature	22
2.4 Comparison of noise factor and noise temperature	23
2.5 Atmospheric effects	24
2.6 Interference with analogue signals	25
2.7 Signal to noise ratios	27
2.8 Interference with digital channels	28
3 Electromagnetic Waves, Propagation and Antennas	30
3.1 Energy in free space	30
3.2 Polarisation and the two components of the E field	30
3.3 Primary properties of antennas	33
3.4 Secondary antenna properties	37
3.5 Choice of antenna type	44
3.6 Helical antennas	44

3.7	Reflection, refraction, depolarisation and feeds	47
3.8	Frequency selective surfaces and planar antennas	54
3.9	Fresnel zone plate antennas	57
3.10	Signal strength contour maps or footprints	58
3.11	Antenna pointing and mounting	61
4	Microwave Circuit Elements	66
4.1	Transmission lines	66
4.2	Waveguides	68
4.3	Resonant cavities	71
4.4	Coupling to cavities and waveguides	72
4.5	Circulators and isolators	73
4.6	Dielectric resonators	75
4.7	Microwave circuit substrates and transmission lines	76
4.8	Some microwave circuit elements	78
4.9	Active devices	84
5	Digital Signal Processing	88
5.1	Analogue and digital signalling compared	88
5.2	Sampling and quantisation	89
5.3	Non-linear quantisation	91
5.4	Quantising ac signals	92
5.5	Pulse spectra	93
5.6	Baseband code formats	93
5.7	Data error control	96
5.8	Pseudo random binary sequences	105
5.9	Bit rate reduction/bandwidth conservation	108
5.10	Quadrature mirror filters and sub-band coding	111
5.11	Transform coding	113
6	Digital and Digital/Analogue Communications Systems	116
6.1	Systems overview	116
6.2	Facsimile system	119
6.3	Satellite weather and environmental services	129
6.4	Mobile systems	160
6.5	Satellite navigation systems	180
6.6	Satellite news gathering (SNG)	193
6.7	Videoconferencing	198
6.8	Bandwidth compression video codec	201
6.9	Small system satellite services (S ⁴)	212
6.10	Global business services	219
7	Television Systems	222
7.1	Review of current systems	222
7.2	Improved processing of current systems	227

7.3	Digit 2000 system	231
7.4	Multiplexed analogue components systems	237
7.5	High definition television developments	249
8	Television Receivers and Distribution Systems	267
8.1	Outdoor electronics unit	267
8.2	TV receivers SAT-TEL models PRK-2 and PRK-2000	273
8.3	Frequency synthesis tuning	275
8.4	TV receivers DRAKE Series II models (1024, 1224 and 1424)	276
8.5	Threshold extension demodulators	278
8.6	Television distribution techniques	280
8.7	An alternative to cable delivery	288
9	Television and Radio Audio Channels	290
9.1	Towards stereo TV	290
9.2	The Leaming System	290
9.3	Broadcast Television System Committee/Multichannel Television Sound System	291
9.4	Japanese FM-FM system	291
9.5	Multi-carrier channel operation	292
9.6	Wegener 1600 stereo system	292
9.7	Warner Amex system	293
9.8	West German dual channel stereo system	293
9.9	Gorizont sound channels	295
9.10	Dolby adaptive delta modulation digital audio system	296
9.11	C-MAC (Packet) sound channels	298
9.12	D-MAC/D2-MAC sound channels	300
9.13	Vertical interval multiple channel audio system (VIM-CAS)	302
9.14	Near instantaneous companded audio multiplex (NICAM-728)	304
9.15	Wegener SDM 2000 digital audio transmission system	309
9.16	Sound radio broadcasting via satellite	311
10	Information Security and Conditional Access	315
10.1	General introduction	315
10.2	Spread spectrum techniques	315
10.3	Scrambling and Encryption	316
10.4	Conditional access and television services	321
10.5	Controlled access in operation	325
11	Installation and Servicing	329
11.1	Initial survey	329
11.2	Mechanical considerations	330

11.3	Antenna/LNB combination	330
11.4	Alignment of the antenna assembly	336
11.5	Surface Mounted Devices	344
11.6	Control of static in the servicing environment	348
11.7	Servicing the LNB	348
11.8	Servicing the main receivers	352
Appendix 1		
A1.1	Geostationary orbits	358
A1.2	Elliptical orbits	359
A1.3	FM/AM noise advantages	360
A1.4	De-emphasis	361
A1.5	Signals in the presence of noise	362
Appendix 2		
A2.1	Noise performance of cascaded stages	366
Appendix 3		
A3.1	The Universal Antenna Constant	368
A3.2	Approximate beamwidth for parabolic reflector antennas	369
A3.3	Azimuth/elevation angles by calculator	370
A3.4	Calculation of angle of elevation α for simple polar mounts	370
A3.5	Modified polar mount	371
Appendix 4		
A4.1	Quantisation noise	375
Appendix 5		
A5.1	Microwave frequency bands classification	377
Index		378

Preface to the second edition

During the space of three short years since the first edition was completed, satellite systems alongside optical fibres, appear to have become the catalysts that were needed to speed up the integration process of global communications services. Now in the early 1990s, trans-Atlantic telephone calls are equally likely to be routed via satellite or submarine cable. Many of the established systems have now adopted Application Specific Integrated Circuit (ASIC) technology, with the advantages of improved reliability and reduced system size and cost, leading to a greater popularity. In spite of this development, discrete component circuits have been retained in a number of cases, because it is felt that with these it is easier to explain the principles involved.

In preparing this second edition, the opportunity has been taken not only to bring the previous work up to date, but also to extend it into the realms of some of the services that might be introduced before the end of the 1990s. Some of these proposals look like changing the concept of personal communications so completely, that saying 'Open Channel 9' to a pocket pen may no longer be science fiction.

Preface to the first edition

Until the introduction of satellites, communications by electrical and electronic means developed in a progressive and evolutionary manner. Now, in a relatively short time span, a new concept of signals delivery has become available – one that is not only capable of carrying all previous services, but also provides the means of distributing many new services, some that could not be supported by previous terrestrial means.

The success of a technological system usually depends upon its financial and social viability. The system costs ultimately being borne by the end user, he or she becomes the final arbiter on the success or failure of the project. It therefore falls upon the design engineer to provision an efficient and effective service. It is then the responsibility of the installation and service engineer to maintain this good quality of service.

The communications engineering fraternity has a history of providing for such developments, achieving this by continually updating skills and knowledge as the technology has developed. Satellite communications provide a whole new series of problems that will have to be solved by those responsible for the system support services.

To these personnel the book is offered, either as an aid for the self-training of practising engineers, or for communications engineering students following a more formalised training programme such as first degree, HNC/D or City and Guilds of London courses. Where it is anticipated that the reader will have some prior knowledge, reference is made to several standard works for added background information. The general philosophy has been to provide the reader with a description of how the systems signal are coded, modulated and demodulated, and processed, in order to create a good understanding of the way in which the system functions.

The book contains five chapters of related communications principles, provided to explain how and why certain system parameters are important. Also covered are as many actual systems as practicable, the approach in these chapters being largely based on system/block diagrams. There are two main reasons for this. Firstly, circuit diagrams for such rapidly developing technology are soon dated; and secondly, the expansion in the use of integrated circuits. These are closely related to sub-system blocks and their

internal circuitry is often irrelevant, particularly from the service aspect. Engineers called upon to service such systems will be equipped with manufacturers' circuit diagrams. This book is intended to give engineers a thorough understanding of the system operation.

The systems design engineer needs to know how to assemble a cost effective and efficient system from the various standard units available, and the advanced student needs to understand similar concepts. For these people, a number of analyses and design rules have been included.

The design and construction of communications satellites and their methods of launch have been purposely omitted. These aspects are already well documented (1), (2).

REFERENCES

- (1) Slater, J.N. and Trinogga, L.A. (1986) *Satellite Broadcasting Systems*. Chichester: Ellis Horwood Ltd.
- (2) Rainger, P., Gregory, D., Harvey, R. and Jennings, A. (1985) *Satellite Broadcasting*. London: John Wiley & Sons.

Acknowledgements

During the research and writing period for this book, I corresponded with and met very many enthusiastic engineers. All provided me with much encouragement and many useful ideas, and provided valuable guidance. To all of those who read this work, I offer my grateful thanks in appreciation of their help. Such is the enthusiasm found among the practitioners of satellite communications that the task of writing this book has been made much easier and most stimulating.

Whilst I have acknowledged all sources of help in the References and Bibliography, I feel that I must express particular thanks to the following organisations and companies, and their personnel:

BICC Research and Engineering, London, U.K.: R. Grigsby, J.S. Buck.
British Telecom Research Laboratories, Martlesham Heath, U.K.: Dr
N.D. Kenyon, M.D. Clark
Communications Systems Ltd, London, U.K.: J.L. Rose.
Electronic and Wireless World (Business Press International): G. Shorter,
M. Eccles.
Eutelsat, Paris, France: P. Binet, G. De La Villetanet.
European Space Agency, Toulouse, France: G. Ferrand.
European Space Agency (Meteosat Operations), Paris, France:
C. Honvault.
European Space Operations Centre, Darmstadt, West Germany:
A. Robson.
European Space Research and Technology Centre, Noordwijk, Holland:
N. Longdon.
Feedback Instruments Ltd, Crowborough, U.K.: P. Rowland-Smith,
M.L. Christieson.
Independent Television Authority, U.K.: B.T. Rhodes, J.N. Slater.
INMARSAT, London, U.K.: D.M. Kennedy.
Marconi Communication Systems Ltd, Chelmsford, U.K.: P.A.T. Turrall.
Marconi International Marine Ltd, Chelmsford, U.K.: C. Riches.
Meteorological Office, Bracknell, U.K.: D.R. Maine, J. Turner.
Muirhead Office Systems Ltd, Beckenham, U.K.: K.A. Knowles.

Mullard Ltd, London, U.K.: M. Rope, B. Whale.
Multipoint Communications Ltd, Witham, U.K.: J.K. Player.
NEC Ltd, London, U.K.: S. Orme.
Plessey Semiconductors Ltd, Swindon, U.K.: J. Salter.
Polytechnic Electronics Ltd, Daventry, U.K.: M. Warman, E.L. York.
Rediffusion Engineering Ltd, Kingston-upon-Thames, U.K.:
K.C. Quinton.
Satellite TV Antenna Systems Ltd, Staines, U.K.: S.J. Birkill, B.G. Taylor.
Scientific Atlanta Inc., Atlanta, U.S.A.: P.C. Bohana, K. Lucas.
Space Communications (Sat-Tel) Ltd, Northampton, U.K.: R. Crossley,
J.P. Knowles.
Television (IPC Magazines Ltd), Contributors: H. Cocks, N. Harrold,
H. Peters.
University of Kent at Canterbury, U.K.: Dr R. Collier, Dr G. McDonald,
Dr E.A. Parker.
Video Systems (SVT) Ltd, Maldon, U.K.: G. Steele.
Wegener Communications Inc., Atlanta, U.S.A.: N.L. Mountain.

This page intentionally left blank

Chapter 5

Digital Signal Processing

5.1 ANALOGUE AND DIGITAL SIGNALLING COMPARED

Wideband communications applications such as television traditionally use analogue signal processing, primarily for reasons of bandwidth conservation and the fact that such systems are well understood. However, each application tends to become unique in certain ways. When the systems become concentrated into integrated circuits, these devices are dedicated to a particular application; relatively few are produced and so their cost is higher than would be the case for mass-produced devices. When the analogue signals are converted into digital form for processing and then back again for such purposes as display, the only dedicated chips are those associated with the interface between the two types of signal. The digital signal processing (DSP) area of the system then uses standard digital ICs that are very much more cost effective.

The increased transmission bandwidth for the digital signal is available on satellite communications links and this overhead is in any case offset by the considerable advantages gained by changing to the digital concept. The systems become more flexible, systems integration can be achieved and computer control gives rise to the concept of *integrated services digital network* (ISDN), where many services can be accommodated with equal performance.

The principal benefits of digital signal processing can be summarised as follows:

- (1) More appropriate for linking devices that operate in the digital mode.
- (2) Provides transmission speeds that are significantly higher than those commonly achieved with analogue processing.
- (3) Provides for improved transmission quality in noisy environments. The noise effects can be reduced using signal regenerators and by using error detection/correction techniques.
- (4) More compatible with digital switching techniques used for distribution, and is a natural technique to apply on systems that are linked by optical fibres.
- (5) Encryption/decryption can easily be adopted for security of information.

- (6) Signal compression techniques (bit rate reduction) can be used to minimise the transmission bandwidth requirements.
- (7) For many applications, time division multiple access (TDMA) can be used over the satellite link and this is more efficient in the use of the service than frequency division multiple access (FDMA).
- (8) The use of onboard microprocessor-controlled satellite switching becomes feasible and the signal can be regenerated before retransmission to improve the received signal quality. If this improvement is not needed, then regeneration allows smaller antennas to be used.

5.2 SAMPLING AND QUANTISATION

There are several ways of converting an analogue signal into digital form and all are effectively based on a sampling technique. It is shown by the Nyquist theorem that provided a complex analogue signal is sampled at a rate at least twice that of its highest frequency component, then the original signal can be reconstructed from these samples without error.

Figure 5.1(a) depicts the general principle for generating what is commonly known as pulse code modulation (PCM). The analogue signal is sampled at very precise intervals of time, to measure its amplitudes. Since only discrete levels are allowed in a digital signal, each of the amplitudes is quantised or allocated a value which is the integer part of the sampled value (i.e. the integer value of the level that the amplitude is just greater than). These values are then converted into a corresponding binary sequence for digital processing. For the waveform shown in Fig. 5.1(a), the quantised and binary codes series would be:

4, 5, 5, 5, 5, 4, 3, 2, 2, 3, 4 ...etc.
 100, 101, 101, 101, 101, 100, 011, 010, 010, 011, 100 ...etc.

It will be noted that each of the eight levels shown can be coded using three bits ($2^3 = 8$).

Any analogue signal that is reconstructed from such a series will obviously be an approximation of the original, the difference or error being referred to as *quantisation noise*. It is shown in Appendix A4.1 that the signal-to-quantisation-noise (SQNR) for a dc signal such as video luminance is given by:

$$\text{SQNR} = (10.8 + 20 \log M) \text{ dB.} \quad (5.1)$$

or:

$$= (10.8 + 6n) \text{ dB} \quad (5.2)$$

where M = number of sampling levels, and n = number of bits per code sample. For the example shown in Fig. 5.1(a);
 $\text{SQNR} = (10.8 + 20 \log 8) \text{ dB} = 29 \text{ dB.}$

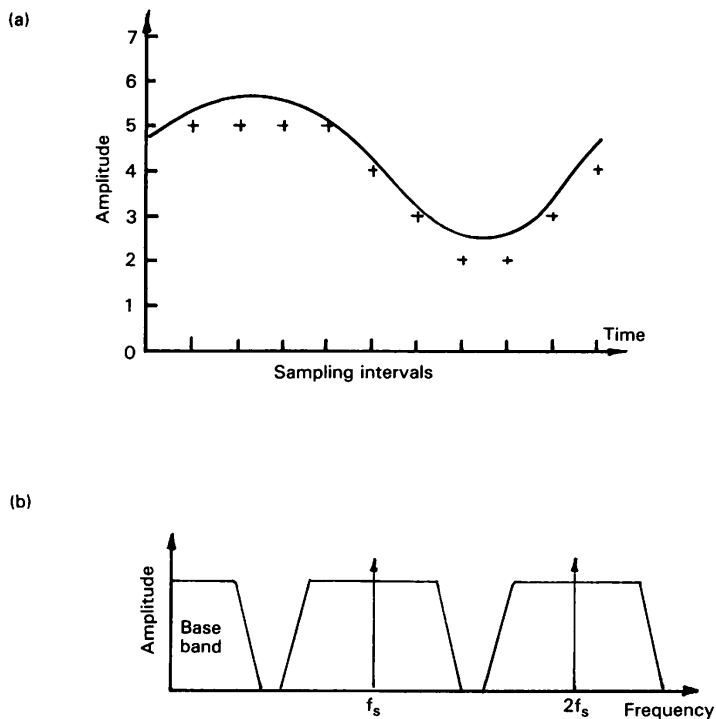


Fig. 5.1 (a) Sampling an analogue signal. **(b)** Spectrum of sampled signal.

Equations 5.1 and 5.2 show that quantisation noise can be reduced to any low level desired by increasing the number of sampling levels and hence the length of the binary code series. The overhead paid for this is the bandwidth of the sampled signal, which can be calculated as follows:

$$2nf_m \quad (5.3)$$

where f_m is the maximum frequency component in the signal, and n is the number of bits per sample.

The process of sampling produces a frequency spectrum similar to that of amplitude modulation but with an infinite range of harmonics, as depicted by Fig. 5.1(b). The reconstruction or demodulation circuit must contain a low-pass filter to separate the baseband component from the harmonics. If the sampling frequency is not high enough, or the filter cut-off not sharp enough, interference from the first lower sideband will result. Figure 5.2 shows the effect of using a sampling frequency less than $2f_m$, where it is impossible to distinguish between some of the components of the baseband and the lower sideband. Such interference is known as *aliasing*.

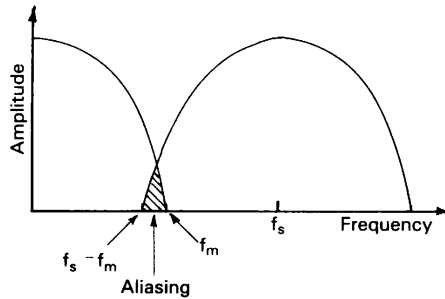


Fig. 5.2 Introduction of aliasing.

5.3 NON-LINEAR QUANTISATION

Reference to Fig. 5.1(a) will show that linear quantisation produces proportionately more noise with smaller amplitude signals. In addition, large amplitude signals are better able to mask the effects of noise. This imbalance can be improved by using a non-linear form of quantisation that behaves in the same way as companding. Such a characteristic, which is relatively easy to implement, is shown in Fig. 5.3 and in the European Standard is referred to as an A-law compander, 'A' being a constant chosen to suite the amplitude distribution on the signal to be processed and typically equal to 87.6.

The curve has a linear region to handle low-level signals and a logarithmic region for the larger amplitudes. Using normalised values, the law for this curve is specified by:

$$y = Ax / (1 + \ln A) \text{ for } 0 \leq x \leq 1/A \quad (5.4)$$

and:

$$y = (1 + \ln Ax) / (1 + \ln A) \text{ for } 1/A \leq x \leq 1 \quad (5.5)$$

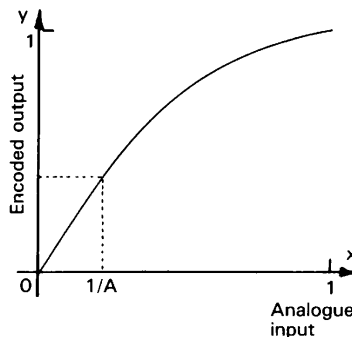


Fig. 5.3 A-law companding.

An alternative compander that finds favour in the North American standard is the μ law device. This is defined by:

$$y = (\log(1 + \mu x)) / (\log(1 + \mu)). \quad (0 < x < 1) \quad (5.6)$$

Where μ is typically equal to 255.

It is neither truly linear nor logarithmic, but is a good approximation at the extremes. Whichever device is used at the transmitter, the complementary device must be used at the receiver to ensure an overall linear response.

In modern systems the compander stage is often incorporated within the same integrated circuit (IC) as the analogue to digital (A/D) and digital to analogue (D/A) convertors. These devices are then often referred to as application specific ICs (ASICs). Since the companding is now carried out in the digital domain, the two laws become modified as:

$$\begin{aligned} \text{A Law: } y &= 0.18(1 + \ln(A|x|))\text{sgn}(x), \text{ for } 1/A < |x| < 1 \\ \text{and } y &= 0.18(A|x|)\text{sgn}(x), \text{ for } 0 < |x| < 1/A \\ \text{u Law: } y &= 0.181\ln(1 + \mu|x|)\text{sgn}(x) \end{aligned} \quad (5.7) \quad (5.8)$$

where

x = normalised analogue signal and encoder input or decoder output.

y = normalised digital signal and encoder output or decoder input.

$A = 87.6$

$u = 255.$

5.4 QUANTISING AC SIGNALS

The methods of sampling and quantising previously described work well for signals that have a large dc component (such as the luminance signal of television). However, signals with both positive and negative excursions, such as audio, need an alternative approach. One possible way involves using an *offset binary* technique, where a constant is added to each sampled value. But in certain cases, such as with audio mixers, where it is necessary to add signals from different sources, the sum can overflow or exceed allowable peak values.

The commonly adopted solution involves using the two's complement method of representing a binary number. By convention, a leading zero indicates a positive number, whilst a leading one indicates a negative value. It will be recalled that the two's complement of a binary number is formed by inverting each bit in turn and adding 1. Thus the two's complement of 01010101 = 10101010 + 1 = 10101011. When reconstructing the analogue signal from a two's complement sample, the excess 1 should be removed before inversion. However, in practice, failure to do so causes such a minute error that it is often neglected.

Typically, then, an 8 bit two's complement code allows for $\pm 2^7 = \pm 128$ sampling levels.

For ac signals, the RMS signal to quantisation ratio is more applicable. For this case, Appendix A4.1 shows that the SQNR is given by:

$$(1.76 + 6.02n) \text{ dB.} \quad (5.9)$$

5.5 PULSE SPECTRA

Figure 5.4(a) shows a single rectangular pulse of amplitude V and time duration t . Such a pulse can be analysed by transforming it into a mathematical model in the frequency domain. In the time domain, the pulse is defined by:

$$\begin{aligned} f(t) &= V \text{ for } |t| < t/2 \text{ and} \\ f(t) &= 0 \text{ for } |t| > t/2. \end{aligned}$$

Transformation to the frequency domain by the Fourier transform:

$$F(\omega) = (Vt \sin(\omega t/2))/\omega t/2 \quad (5.10)$$

where $\omega = 2\pi f$ the angular velocity.

A section of the spectrum for equation 5.10 is shown in Fig. 5.4(b), and this is described generally as a $\sin x/x$ function. Figure 5.4(c) shows a series of such pulses which is typical of the situation at a receiver. As the zero crossings of the tails occur at the bit cell centres, the energy in the tails tends to be self-cancelling. This is particularly true if the receiver clock circuit is accurately timed. Weak synchronism between the receiver clock and the bit stream will lead to decision errors by the detector circuit, a situation which is often made worse by a relatively long string of zeros.

If the pulse stream were filtered before processing, some of the HF energy would have been removed and the tails of the spectrum would have less significance.

5.6 BASEBAND CODE FORMATS

The errors affecting data services via satellite communication links are chiefly characterised by added white noise and an overall propagation delay that is in the order of 250 mS. The noise leads to the concept of a *bit error rate* (BER), whilst delay variation creates timing and synchronism problems, which in turn lead to further bit errors. As satellite transponders tend to be power rather than bandwidth limited, Shannon's channel capacity rule becomes important.

$$C = B \log_2(1 + S/N) \text{ bits/sec} \quad (5.11)$$

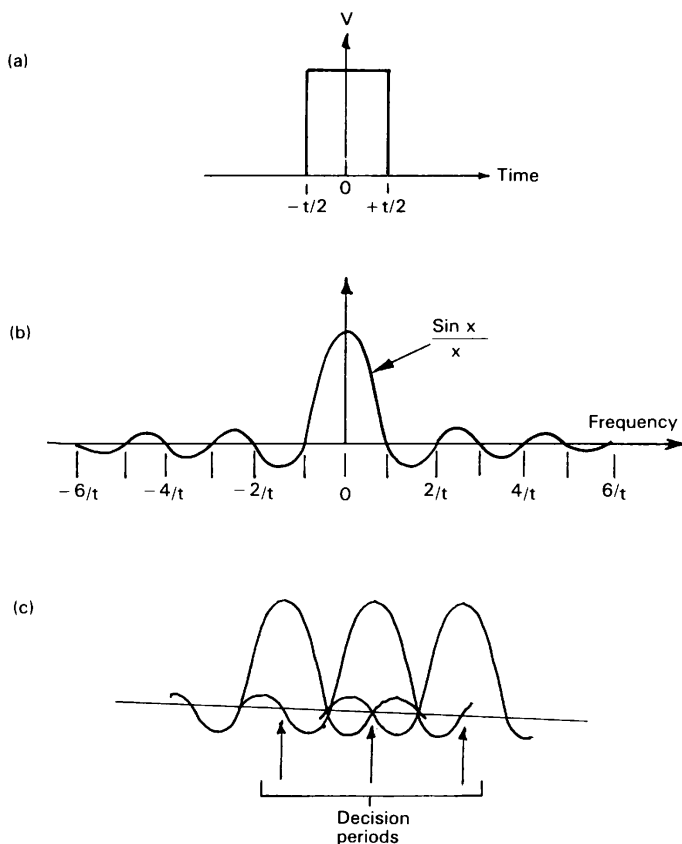


Fig. 5.4 A rectangular pulse and its spectrum.

where C = channel capacity, B = bandwidth, and S/N = S/N ratio.

The balance between bandwidth and S/N ratio can be used to good effect to maximise the channel capacity for an acceptable bit error rate suitable for the particular service.

Binary code formats are designed to insert extra bits into the data stream on a regular basis, in order to take advantage of this trade-off. A few of the many ways of achieving this are shown in Fig. 5.5, the aim being to minimise the number of consecutive similar bits in the transmitted data stream. The receiver detector clock circuit can then be synchronised to a greater number of signal transitions and so improve its timing. An extra advantage may accrue because some formats have no dc component in their power spectrum, a feature that reduces the low frequency response requirement of the receiver and allows ac coupling circuits to be used.

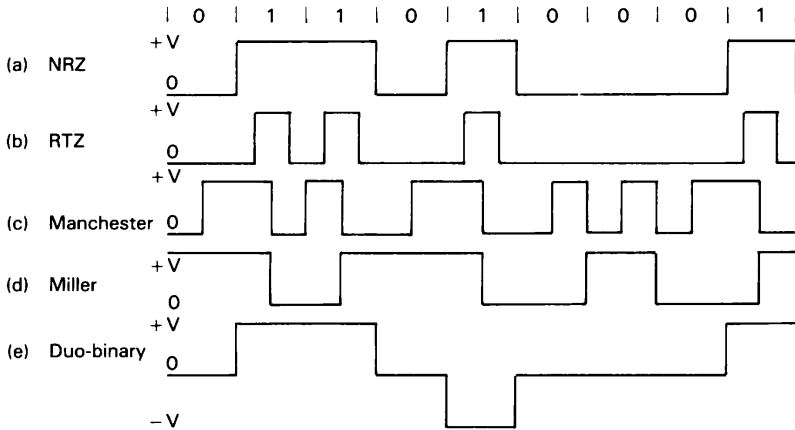


Fig. 5.5 Some baseband code formats.

All the commonly adopted code formats can be generated and decoded using dedicated integrated circuits (ICs), the codes generally being produced from non-return-to-zero (NRZ) basic codes. The return-to-zero (RTZ) format is little used, as its half-width pulses represent an energy/bit penalty.

Codes of the nBmB type are often used to convert n bits/symbol into m bits/symbol ($m > n$), ensuring that, on average the number of ones and zeros being transmitted are equal.

The bi-phase series of formats for which the basic Manchester code shown in Fig. 5.5(c) is representative has the following features. A signal transition occurs at each bit cell centre, so that a zero is represented by 01 and a one by 10. This ensures that there are never more than two identical bits in series. Another variant of this is the *code mark inversion* (CMI) format, where 0 = 01 and 1 = 00, or 11 alternatively. Although bi-phase codes have a 50% redundancy and double the transmission bandwidth, there is no dc component in the power spectrum as depicted by Fig. 5.6.

The Miller format, shown for comparison purposes in Fig. 5.5(d), is not usually used for transmission purposes. It finds favour in the magnetic storage media. A one is represented by a transition at mid-symbol and a zero by no transition. Except for two consecutive zeros when an extra transition is introduced at the end of the first zero, the Miller code has similar properties to the NRZ code (see section 1.9).

The *duo-binary* code is a bi-polar, full-pulse-width code, as shown by Fig. 5.5(e). It has a bit repetition rate that is half the frequency of the original signal. Zero is represented by 0 volts and one by $\pm V$. The V polarity is unchanged if the one follows an even number of zeros and reversed if it follows an odd number.

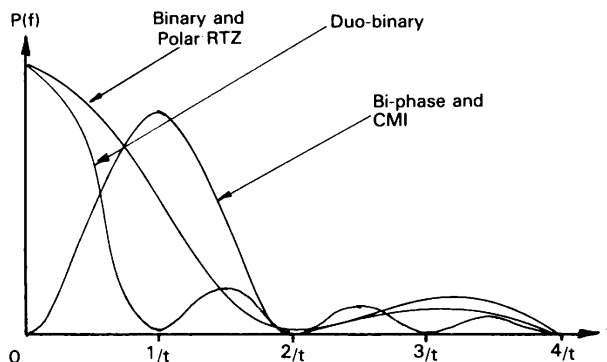


Fig. 5.6 Power spectrum of some code formats.

The more efficient codes use the least redundancy and there is, in general, a trade-off between the complexity of balancing the number of consecutive similar bits and the added redundancy.

The increase in bit rate that these codes produce represents a power penalty. For example, using a 3B4B format involves a bit rate increase by a factor of 4/3 which in turn requires a power increase of about 1.25 dB.

A simple form of error monitoring/detection takes the form of counting the running average of the number of ones and zeros. Any variation outside some predetermined bounds can be used to signal an error situation. With added redundancy, error detection/correction can be provided using *parity checks* or *Hamming codes*.

5.7 DATA ERROR CONTROL (FORWARD ERROR CORRECTION)

When data errors are detected in a bit stream, there are various corrective measures that can be applied. The actual method used often depends upon the source of the original data signal. For instance, had the source been analogue in form, then one of the error concealment techniques might be applicable. This possibility includes:

- (1) Ignore the error and treat it as a zero level signal.
- (2) Repeat the last known correct value.
- (3) Interpolate between two known correct values.

This latter method is really only suitable where a significant amount of data storage is available at the receiver, to allow time to regenerate the analogue signal after processing. In other cases there is the possibility of requesting a repeat transmission when errors have been detected. However, in general this is wasteful of both time and frequency spectrum.

The ASCII code (American Standard Code for Information Interchange) is a commonly used method of representing alphanumeric characters in a digital system. This 7 bit code allows for $2^7 = 128$ different alphabetic, numeric and control characters. The commonly used word length is 8 bits (1 byte), so that space is available for one extra redundant bit.

Even and Odd Parity

A single-error detection code of n binary digits is produced by placing $n - 1$ information or message bits in the first $n - 1$ positions. The n th position is then filled with a 0 or a 1 (the *parity bit*), so that the entire code word (or code vector) contains an even number of ones. If such a code word is received over a noisy link and found to contain an odd number of ones, then an obvious error has occurred. Alternatively, a system might use *odd parity*, where the n th bit is such that the code word will contain an odd number of ones. In either case, a parity check at the receiver will detect when an odd number of errors has occurred. The effects of all even numbers of errors is self-cancelling, so these will pass undetected. The even or odd parity bits can be generated or checked, using Exclusive OR or Exclusive NOR logic respectively.

Such an arrangement of bits is described as an (n, k) code, n bits long and containing k bits of information. It thus follows that there are $n - k = c$ parity or protection bits in the code word. The set of 2^k possible code words is described as a *block code*. It may also be described as a linear code if the set of code words form a sub-space (part) of the vector-space (entirety) of all possible code words.

Hamming Codes

Error correcting codes have been devised and are named after R.W. Hamming (1), the originator of much of the early work on error control. For a code word of length n bits, it is possible to identify n ways in which a single error can occur. Including the possibility of no error, there must be $n + 1$ different code patterns to be recognised.

Hamming showed that the number of parity bits c required for a single error correcting code was given by:

$$c \geq \log_2 (n + 1) \text{ or } 2^c \geq n + 1 \quad (5.12)$$

(Hamming also defined a 'perfect' code for the case where $2^c = (n + 1)$.)

Figure 5.7(a) shows how the message is expanded with redundant check bits. These are usually interleaved with the message bits and placed in positions $2^0, 2^1, 2^3$, etc. in the encoded bit pattern. The mechanics of the encoding and decoding process is explained in Table 5.1 for a (7,4) block

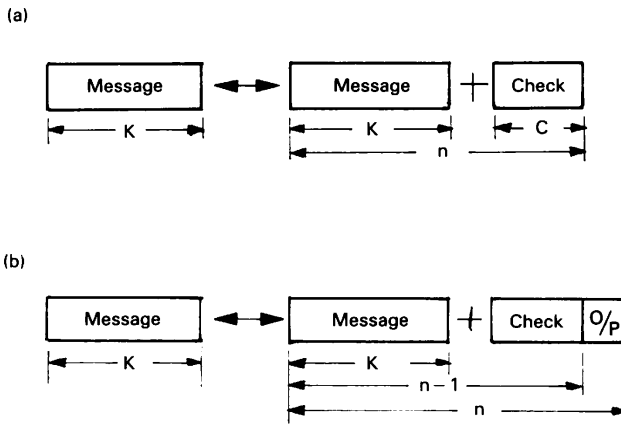


Fig. 5.7 Hamming code formats (a) SEC; (b) SEC/DED.

code (block length $n = 7$, message length $K = 4$, parity bits $c = 3$). The message bits to be transmitted (0011) are placed in the 3rd, 5th, 6th and 7th positions respectively. The three parity checks are carried out, to determine the values (0 or 1) to be placed in positions 1, 2 and 4. In general, the x th parity check bit is given by the sum modulo 2 of all the information bits, where there is a 1 in the x th binary position. The code word to be transmitted is thus 1000011. If this is now received over a noisy link as 1000001, there is an error in the 6th position. The receiver decoder again carries out the three parity checks and generates the following results:

1st check 0
 2nd check 1
 3rd check 1

The reverse of the series is called the *syndrome* (syndrome being a medical term for the symptoms of a disease) and this points to the bit in error ($110 = 6$). Correction is simply achieved by inverting the bit in error. An all correct transmission would have yielded an all-zero syndrome.

By adding an overall parity (O/P) check bit as shown in Fig. 5.7(b), the single-error correction capability is extended to double-error detection, the error patterns being indicated by the following conditions:

- (1) No errors – zero syndrome and overall parity satisfied.
- (2) Single error (correctable) – non-zero syndrome and overall parity fails.
- (3) Double errors (non-correctable) – non-zero syndrome and overall parity is satisfied.

Table 5.1

(Binary)	001	010	011	100	101	110	111
Position	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	P	P	M	P	M	M	M
Check 1	•		•		•		•
Check 2		•	•			•	•
Check 3				•	•	•	•
Message			0		0	1	1
Parity bits	1	0		0			
Transmitted							
Code word	1	0	0	0	0	1	1
Received							
Code word	1	0	0	0	0	0	1
Recheck of parity	0	1		1	(reverse bit order)		
Syndrome = 110 = 6							
Error in position 6. Invert bit 6 to correct error.							

Maximum likelihood Decoding

The *distance* between any two n symbol code words is defined as the number of positions of bit difference. The *minimum distance* d between any valid code words is the minimum number of bit changes necessary to convert one code word into another. For the correction of all combinations of t or fewer errors, the minimum distance must be at least $2t + 1$, thus ensuring that if t errors occur then the received code vector will be nearer to the one transmitted than any other (i.e. $d \geq 2t + 1$).

This concept leads to a simple method of error correction. If a code word is detected as being in error, the decoder checks the distance between the received vector and all valid code words and selects the one exhibiting the minimum number of errors.

Cyclic Codes

In certain transmission systems, the channel noise causes errors to occur not just randomly but in clusters or bursts. A sub-class of the linear block codes valuable in combatting this problem is the group of *cyclic codes*. These have a form such that if a code vector such as 0110 is a valid code word, then so are all its cyclic translates such as 1100, 1001, 0011, etc. which are obtained by shifting the binary sequence one bit at a time to left or right. These codes can easily be encoded or decoded using ICs based on feed-back shift registers.

The *cyclic redundancy checking* (CRC) error detection concept is most effectively described in algebraic form using modulo 2 arithmetic. The basic

rules are that $1 + 1 = 0$ and $1 = -1$, so that addition does not involve a carry and produces the same result as subtraction.

Message and redundant check bits are expressed by the use of polynomials in terms of a dummy variable X , the lowest-order term X^0 , representing the least significant bit (LSB) and the highest-order term X^n the most significant bit (MSB). Since the terms range from X^0 to X^n , there will be $n + 1$ bits in the message. The coefficients of the terms of the polynomial indicate whether a particular bit is 0 or 1. For example, the 5 bit message stream 11010 would be represented by:

$$1.X^4 + 1.X^3 + 0.X^2 + 1.X^1 + 0.X^0 \text{ or } X^4 + X^3 + X$$

The data stream is written with the MSB on the left when this is transmitted first. The degree of a polynomial is the power of the highest-order term, which in this example is 4.

To generate the code for transmission, three polynomials are used: the message polynomial $k(P)$, a generator polynomial $G(P)$ which is selected from a group of *primitive polynomials* to produce the desired characteristics of block length and error detection/correction capability, and a parity check polynomial $c(P)$. The coded word length $n = k + c$, as for block codes.

During encoding, $k(P)$ is first loaded into a shift register and then multiplied by X^c to move the message c bits to the left, thus making room for c parity bits. The shifted polynomial is then divided by the generator polynomial to produce a remainder that forms the parity check polynomial $c(P)$.

$$(k(P).X^{n-k})/G(P) = Q(P) + c(P) \quad (5.13)$$

where $Q(P)$ is a quotient polynomial and $c(P)$ the remainder which is then loaded into the remaining shift register cells.

The transmitted code vectors for a (7,4) cyclic code ($k = 4, c = 3$) using the generator polynomial of $X^3 + X + 1$ are formed as shown by the following example:

Message code = 1101, $k(P) = X^3 + X^2 + 1$,

$$k(P).X^c = (X^3 + X^2 + 1)X^3, \\ = X^6 + X^5 + X^3. \text{ Now divide by } G(P).$$

$$\begin{array}{r} X^3 + X + 1 \overline{) X^6 + X^5 + X^3} \\ \underline{X^6 + X^4 + X^3} \\ X^5 + X^4 \\ \underline{X^5 + X^3 + X^2} \\ X^4 + X^3 + X^2 \\ \underline{X^4 + X^2 + X} \\ X^3 + X \\ \underline{X^3 + X + 1} \\ 1 = \text{REMAINDER} \end{array}$$

$X^6 + X^5 + X^3 + 1$ therefore divides exactly by $G(P)$, and so forms the transmitted polynomial 1101001. The first four bits 1101 are the original $k(P)$ whilst the remaining three bits 001 are $c(P)$, the parity polynomial. The full list of the $2^k = 16$ code words are shown in Table 5.2.

If a code word $T(P)$ is transmitted, and received without error, $T(P)$ divides exactly by $G(P)$, leaving a zero remainder. The last c bits are then stripped off to leave the original message code. If, however, an error occurs, then division leaves a remainder polynomial that forms the syndrome. There is a one-to-one relationship between this and the error pattern, so that correctable errors can be inverted by logic circuits within the decoder.

The effectiveness of cyclic codes depends largely upon the generator polynomial. For a polynomial of degree n , the decoder is generally capable of detecting/correcting error burst of n or less bits (some errors outside of this bound may be detectable), and odd numbers of random errors. A further advantage in cases where speed is not important is that CRC can operate with microprocessor-based coding, where the system characteristics can be reprogrammed.

Table 5.2

Message vector	Parity vector
0000	000
0001	011
0010	110
0011	101
0100	111
0101	100
0110	001
0111	010
1000	101
1001	110
1010	011
1011	000
1100	010
1101	001
1110	100
1111	111

Golay Codes

These are a sub-set of cyclic codes. The (23,12) versions using 11 parity bits are capable of correcting any combination of three random errors (including a burst of three) in a block of 23 bits. The codes are based on the generator polynomials:

$$G_1(P) = X^{11} + X^{10} + X^6 + X^5 + X^4 + X^2 + 1 \text{ and}$$

$$G_2(P) = X^{11} + X^9 + X^7 + X^6 + X^5 + X + 1$$

both of which are factors of $X^{23} + 1$. Encoding and decoding can be accomplished using ICs based on 11 bit feedback shift registers.

Bose-Chaudhuri-Hocquenghem (BCH) Codes

This very important sub-set of cyclic codes is widely used for the control of random errors. For any positive integers m and t , such that $t \leq 2^m - 1$, codes exist with the following parameters:

$$\begin{aligned} \text{Block length } n &= 2^m - 1 \\ \text{Parity bits } c &= n - k \leq mt \\ \text{minimum distance } d &\geq 2t + 1 \end{aligned}$$

which are capable of correcting t or less errors in a block $n = 2^m - 1$ bits. The generator polynomials used must be at least of degree mt . In general, the number of check bits c is almost mt , but if t is less than 5 then $c = mt$.

Values for n , k and t have been tabulated (2) for various polynomials. For example, when $m = 5$, $n = 31$, $k = 26$, $c = 5$, t is equal to 1, but when $m = 5$, $n = 31$, $k = 11$, $c = 20$, t rises to 5.

BCH codes can be processed either by digital logic circuits or by digital computer. The hardware solution is faster but the computer is of course more flexible.

Reed-Solomon Codes

These are a sub-set of BCH codes, which correct any combination of t or fewer errors and require no more than $2t$ parity bits. The parameters are $n - k = c = 2t$ and $d = 2t + 1$. As with all cyclic codes, a generator polynomial is used for encoding, and the decoder again produces a syndrome that identifies the error pattern.

Two Dimensional Reed-Solomon Codes

These have been devised so that an inner code can be used to correct burst errors of a few bytes length, whilst the outer code can correct very long bursts. In one such code, the data is organised into a matrix of 600 bytes $\times$ 30 rows. The first or outer code is then used to add 2 check bytes to each of the columns to create 32 rows. Each row is then divided into 10 blocks of 60 bytes and the second or inner code is used to add 4 check bytes to each block of 60 bytes to give an array of 640 bytes $\times$ 32 rows. The inner code is able to correct all single byte errors and tag all errors that are longer than this. The outer code can correct 2 inner code blocks of 60 bytes each, provided that they are tagged when passed to the decoder. It can therefore

correct 120 bytes. However, since the inner code is interleaved to a depth of 10 bytes, the outer code, when using the inner code tags, can correct an error up to 1200 bytes long.

Interleaved or Interlaced Codes

This is a simple but powerful way of dealing with either random or burst errors. Any cyclically coded (n,k) set of vectors can be used to produce a new code $(\alpha n, \alpha k)$, by loading coded vectors from the original code into a matrix of n columns and α rows and then transmitting the bits column by column. The transmission sequence is then an interleaved or interlaced code, with an interleaving factor of α . If a burst of errors α or less in length occurs, there will only be one error in each affected word of the original code. As a lower bound, if the original code corrects t or fewer errors, the interleaved code will correct any combination of t bursts each of length α or less.

Because the original code was cyclic, so will be the interleaved one. If the original generator polynomial had been $G(X)$, the generator of the new code would be $G(X^\alpha)$. Therefore the interleaved code can also be encoded and decoded in a similar manner to cyclic codes.

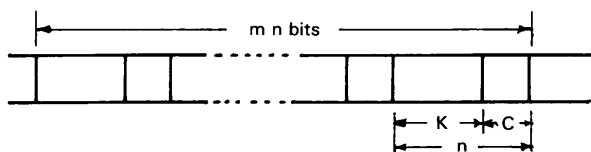
An extension of the interleaving technique is used in certain special cases. This involves the generation of two Reed-Solomon codes from the data, and then cross-interleaving the coded bit patterns before transmission. Provided that the encoder and decoder are synchronised, relatively very long burst errors become correctable as well as single random errors.

Convolution Codes

These codes, used for the control of burst errors, are produced by the complex process of convolution of the message stream with a generator matrix. Unlike cyclic codes, where the parity bits for any given block are contained within that block, the parity bits within a block of convolution code check the message bits in previous blocks as well. The range of blocks over which the check bits are operative is defined as the *constraint length m* for the code (a block code might be described as a convolution code for $m=1$), the parity bits formed on the current block being dependent upon the message bits in the previous $m-1$ blocks. The code structure as transmitted is represented by Fig. 5.8(a) and has the following parameters:

Block length = n
 Message bits per block = k
 Parity bits per block = c
 Constraint length n_0 = m blocks = mn bits
 Message bits within n_0 = k_0 = mk bits

(a)



(b)

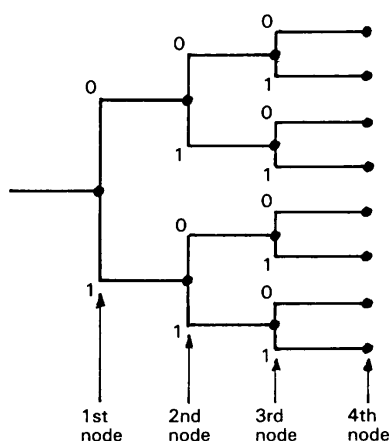


Fig. 5.8 Convolution codes (a) Code structure; (b) decoding search tree,

Parity bits within $n_0 = c_0 = mc$ bits.

For convolution codes, n and k are normally small integers, often less than 5.

The encoder consists essentially of feedback shift registers and EX OR logic. For decoding, there are two strategies that can be used. Logic or algebraic decoding is restricted to the constraint length. This uses an encoder plus logic circuit to recalculate the parity bits and invert any correctable bits in error. The alternative is probabilistic decoding, where the decoder must be able to store more than m blocks. The sequential decoder then makes its decision on one block, based on more than $m - 1$ previous blocks. The decoding follows a predetermined algorithm through a tree structure such as is shown by Fig. 5.8(b). When the decoder detects an error, it backtracks to the previous node and accepts the alternative branch.

Viterbi Decoding Algorithm (5)

The decoder calculates the distance function between each received code word and all the others in the code book. The word that meets the minimum distance requirement is selected as the most likely one to have been transmitted and is therefore often described as the *survivor*. The decoder stores the survivors and the associated distances to enable a tree search to proceed in the manner of Fig. 5.8(b). Since decoding represents an iterative process, the decoder can readily be fabricated on an ASIC to handle most of the standard convolution code formats.

Punctured Codes (8)

At higher data rates, Viterbi/Trellis decoding becomes more complex and hence more costly. *Punctured* convolution codes can achieve higher data rates without this penalty. The technique operates by selectively and periodically removing bits from the encoder output effectively to raise the code rate.

In practice, the output from an adaptive convolutional encoder is passed through an analogue gate circuit in blocks of parallel bits. The switching action of the gates is controlled by the *puncture matrix*, which defines the characteristics of the specific error protection for the bits in an information block.

The puncture matrix has to be available at the receiver decoder either as a hardware element or in software. The decoder then treats the punctured code in the same way as erased bits so that these do not affect the decoder decision making.

5.8 PSEUDO-RANDOM BINARY SEQUENCES (PRBS) (PSEUDO NOISE (PN))

For a series of binary digits to be in random order, each symbol must occur by chance and not be dependent upon any previous symbol. Over a long period, the number of occurrences of ones (n_1) and zeros (n_0) should be the same. In a similar way, runs of 2, 3 or more of each symbol should be equiprobable. Such sequences can be generated using shift registers, as indicated by Fig. 5.9, where the logic state of the switches controls the feedback paths, through modulo-2 adders placed between the serial shift register input and output, the state of the switches being set according to a *characteristic polynomial*. Assuming a 4 bit register with the switches set, $S_1 = S_n = 1$, and $S_2 = S_{n-1} = 0$ (1001), then irrespective of the initial states of the shift register cells, the binary pattern shown in Table 5.3 will be

produced. The bit pattern b_n is the required sequence (010,110,010,001,111), generated on a periodic and cyclic basis. Selecting a new characteristic polynomial will create a new sequence.

Table 5.3

	r_1	r_2	r_3	r_4	b_n
State 1	1	1	1	1	0
2	0	1	1	1	1
3	1	0	1	1	0
4	0	1	0	1	1
5	1	0	1	0	1
6	1	1	0	1	0
7	0	1	1	0	0
8	0	0	1	1	1
9	1	0	0	1	0
10	0	1	0	0	0
11	0	0	1	0	0
12	0	0	0	1	1
13	1	0	0	0	1
14	1	1	0	0	1
15	1	1	1	0	1

There are 2^n possible states for the shift register cells, but the all-zero combination is not valid as this would bring the generator to a halt. Therefore the length of a sequence is $2^n - 1$, the period of repetition being independent of the initial conditions.

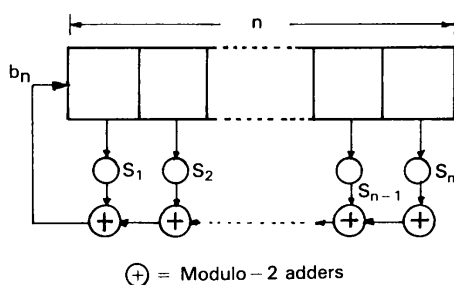


Fig. 5.9 Pseudo-random binary sequence generator.

It can be shown for these sequences that $n_1 = 2^{n-1}$ and $n_0 = 2^{n-1} - 1$, so that if n is large then the sequence has near-random properties. A maximal length sequence is usually described as an m-sequence and finds many uses in communications systems.

Because of the pseudo-random properties, PRSBs can be used:

- (1) As repeatable noise sources for testing digital systems.
- (2) To add redundancy to a transmitted data stream, by coding logic 1 as the m-sequence and its inverse as logic 0.
- (3) Added to the data stream as shown by Fig. 5.10, to act as a key to ensure secure data.

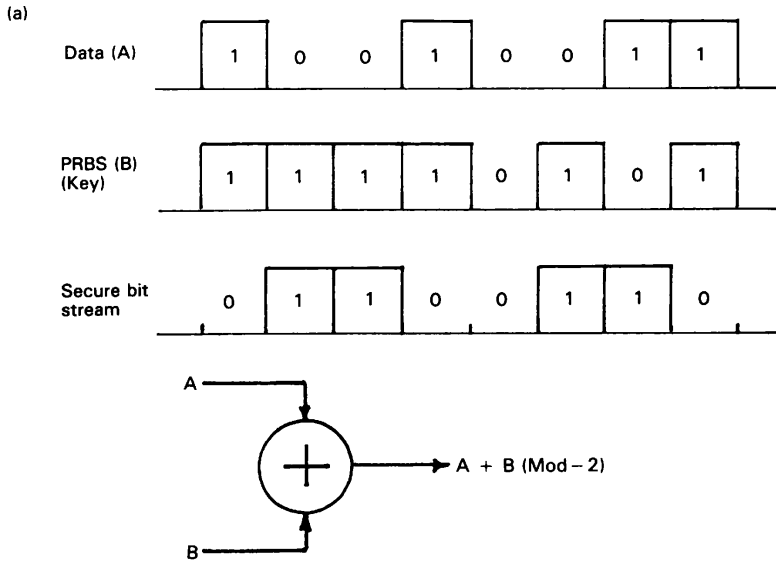


Fig. 5.10 Addition (Modulo-2) of data and m-sequence.

The major difficulties caused by 2 and 3 are synchronism and a reduced data rate.

Gold Code Sequences (6)

These represent a family of pseudo random sequences that were defined by R. Gold in 1967 as having a low cross correlation between each other. As such they can occupy the same signalling channel without producing destructive interference. Because of these properties, Gold codes find applications in code division multiple access (CDMA) or spread spectrum multiple access (SSMA) systems. Each code sequence is produced from a pair of PRBS generators each with n stages by Modulo-2 addition. If all the $2^n + 1$ Gold sequences of period $2^n - 1$ are concatenated, the resulting

sequence period becomes $(2^n + 1)(2^n - 1) = 2^{2n} - 1$, the same as would be generated by a shift register with $2n$ cells.

5.9 BIT RATE REDUCTION/BANDWIDTH CONSERVATION

Pulse Code Modulation

The International Telegraph and Telephone Consultative Committee's (CCITT) standard for pulse code modulation (PCM), as used on telephony channels, operates at a sample frequency of 8 KHz and uses 8 bits per sample (7 bits representing level plus 1 sign bit to represent polarity). The transmitted bit rate is thus $8 \times 8 \text{ KHz} = 64 \text{ Kb/sec}$. In order to minimise the bandwidth required, quaternary phase shift keying (QPSK) is used. Unlike bi-phase PSK, where each phase inversion represents a bit of information, QPSK uses four phase shifts as follows:

$$\begin{aligned} 0^\circ &= 00, \\ 90^\circ &= 01, \\ 180^\circ &= 11, \\ 270^\circ &= 10, \end{aligned}$$

so that each phase now represents 2 bits, thus doubling the information content without the increased bandwidth penalty. However, the phase separation between each code symbol is halved relative to bi-phase PSK, which leads to a 3 dB S/N ratio loss, and so companding is used.

Further bandwidth compression can be achieved by using an extension of QPSK. A system known as 16QAM is in operation in which each of the 8 phases separated by 45° can be amplitude modulated to one of two levels. This provides 16 vectors, each one representing 1 or 16 different code patterns each of 4 bits. The bit rate is thus now increased by a factor of 4 without invoking a bandwidth penalty, but at the expense of S/N ratio.

Telephony by PCM is transmitted over satellite links, either by time division multiple access (TDMA) systems, or over digital data channels using the single channel per carrier (SCPC) concept.

PCM can be used for other services such as high quality music or television, but with these wider bandwidth signals the effects of bit errors can be a problem. With PCM, the effect of a single bit in error depends upon its weighting. Whilst an error in the LSB will probably pass unnoticed, an MSB in error is likely to have a significant effect. Using companded PCM on music channels can cause the noise level to vary audibly as the signal level changes.

Delta Modulation

Alternatively, delta modulation (DM) may be used, where the audio signal is simply coded by 1 bit, positive or negative, based upon whether a signal sample is greater or less than the previous value. Because only 1 bit per sample needs to be transmitted, the sampling rate can be increased significantly. This reduces quantisation noise, simplifies the receiver anti-aliasing filter and at the same time reduces the bandwidth required relative to standard PCM. A 1 bit error simply produces a step of the wrong polarity in the output signal, which then retains its correct general shape as shown in Fig. 5.11. Such errors are then likely to pass unnoticed.

An overload effect can occur (also shown in Fig. 5.11) when the input signal amplitude changes by a greater than quantising step size during the sampling intervals, but this chiefly affects large-amplitude, high-frequency signal components, which in audio occur relatively infrequently.

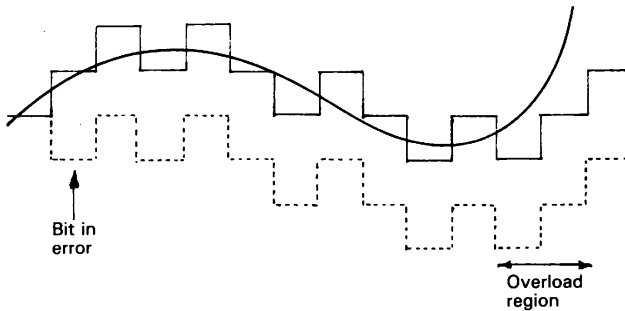


Fig. 5.11 Delta modulation, bit error and overload.

Differential PCM (DPCM)

If the pattern of an analogue signal is known up to some point in time, then because of the correlation between successive values, it is possible to draw certain inferences about future behaviour. This can be achieved by a predictive or extrapolation process.

In predictive coding systems the signal is sampled at regular intervals, but as each sample time approaches the probable value of the next sample is predicted. The difference between the prediction and the actual value is coded as for PCM and transmitted. The receiver must make the same prediction and add the same correction. Figure 5.12(a) shows the basic principles of such a system. However, since the two predictors operate on slightly different signals, this leads to some errors. The system shown in Fig. 5.12(b) is an improvement. Both predictors work on decoded signals and if

no noise were added in the transmission channel, both would yield identical results. In the system illustrated in Fig. 5.12(c), the predictor is replaced by an integrator/LPF, so making better use of the previous sample value because of the removal of the ripple due to sampling.

Although the predictions made under any system must contain errors, the advantages claimed for DPCM can be explained as follows.

‘In standard PCM, no prediction is made; or in other words, the next predicted value is zero. Since an approximate prediction is better than no

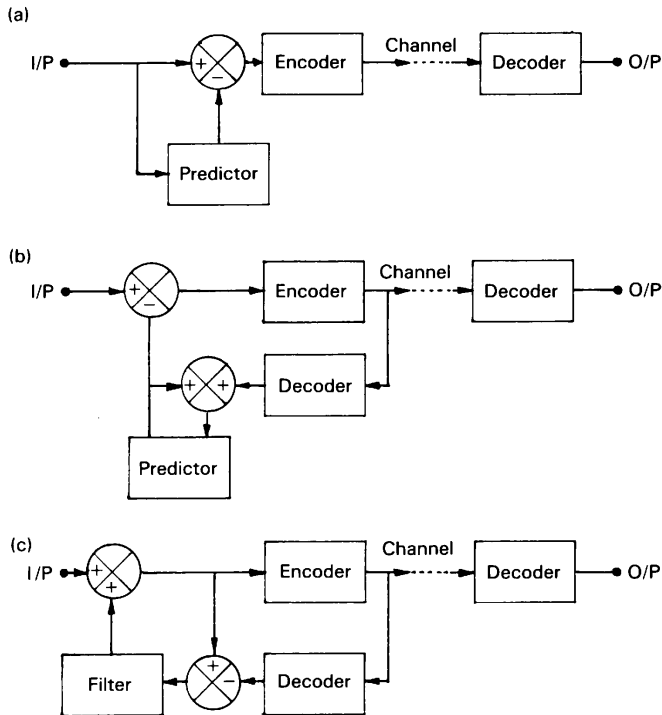


Fig. 5.12 Predictive coding (a) basic principle; (b) improved version; (c) integrator/filter version.

prediction at all and because of the good correlation between successive samples, DPCM must lead to a better S/N ratio as well as a reduced bandwidth.’

Adaptive Systems

Both DM and DPCM systems can give better results if the quantisation step size is made to vary in accordance with the time varying amplitude of the input signal. Such adaptive circuits give a significant improvement in audio quality and behave as companded systems. In addition, pre-emphasis and de-emphasis can be applied to the analogue signals before encoding and after decoding to further improve the S/N ratio.

Sub-Nyquist Sampling

In certain special cases, alternate sampled values are suppressed, so that sampling effectively takes place at half the normal frequency. At the receiver, the missing samples can be replaced by interpolation or predictive coding. The overall effect reduces the bandwidth required by a factor of 2. However, this concept requires the use of a very stable sampling frequency and accurate receiver filtering, plus the extra circuitry for interpolation.

5.10 QUADRATURE MIRROR FILTERS AND SUB-BAND CODING (6 & 7)

It has been shown that to avoid aliasing when digitally processing an analogue signal, the sampling frequency must be at least twice that of the highest frequency component in the complex wave. If aliasing occurs then the original signal cannot normally be extracted (filtered) from the complex spectrum of the sampled signal.

In any wideband signal, many of the frequencies in the band are only rarely present. A good example of this is the television image, where much of the background remains unchanged from frame to frame and with only relatively small areas of movement. By dividing (filtering) the broad spectrum into narrower sub-bands, it will be found that many of these will contain only occasional information. If these sub-bands are sampled, then by using a variable length coding such as Huffman, which allocates the shortest code words to those values that have the highest occurrence probability (the stationary areas of the image), a useful bandwidth compression can be achieved.

Quadrature mirror filters (QMF) are basically multi-rate digital filters so that the sampling frequency is not constant throughout the system. Fig. 5.13 shows the basic principle involved with the system containing *decimators* which *down sample* the signal and *interpolators* that provide complimentary *up sampling*.

If a decimator has the digital sequence $X(n)$ as input, it will produce an output $Y(n) = X(Mn)$, where M is an integer that represents the down

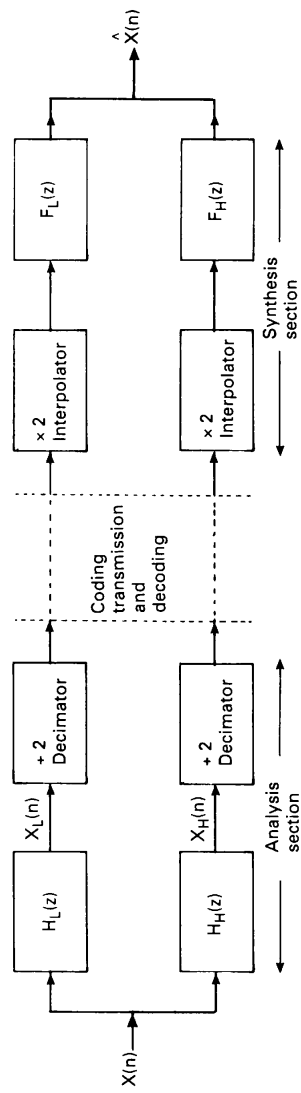


Fig. 5.13 Basic operation of 2 channel QMF system.

sampling ratio. If $M = 2$, then $Y(n)$ will consist of every second sample. The decimator thus produces a compression in the time domain. In a practical system, before the signal is decimated, it will first be bandwidth limited to reduce the effects of aliasing using a *decimation filter*. At the receiver, the M -fold interpolator will replace the missing $M-1$ values between adjacent samples. If $M = 2$, then the interpolator will either average successive values or insert zeros. The effect being a complimentary expansion in the time domain.

The concept is closely related to sub-band coding. As indicated in Fig. 5.13, the input sequence $X(n)$ is divided into separate channels using low and high pass filters with a response of $H_L(z)$ and $H_H(z)$ before down sampling. This signal is then coded for transmission in any suitable manner, before being processed in a complementary manner at the receiver. Higher orders of filter can be produced by further sub-dividing each channel with low and high pass filters. It will be clear from the comments on sub-band coding that some of these channels may contain long strings of zeros, thus allowing significant bit rate reductions to be achieved. The combined response of decimator and interpolator can give rise to aliasing in the frequency domain, but the *synthesis filters* with response of $F_L(z)$ and $F_H(z)$ remove this from the respective channels. In a correctly designed system, the recombined output sequence $\hat{X}(n)$ will be a perfect reconstruction of the input $X(n)$, even under noisy transmission conditions.

The name quadrature mirror filter, derives from the fact that the response of the filter $H_L(z)$ is a mirror image of that of $H_H(z)$ with respect to the normalised frequency $\pi/2$, which is a quarter of the sampling frequency.

5.11 TRANSFORM CODING (See also Chapter 6, H. 261 Codecs)

This technique, which is commonly used for image signal transmission, is a two step process. First a linear transform is performed on the original image by splitting the space into $N \times N$ blocks of *picture elements* or *pixels*. The amplitude coefficient of each pixel is then mapped on to a transform space for coding and transmission. The *discrete cosine transform* (DCT) which is related to the discrete and fast fourier transforms (DFT and FFT) is almost universally adopted. This method is faster because it requires the calculation of fewer coefficients and is more compatible with digital signal processing (DSP) ICs than other transform methods. When used with suitable coding the DCT provides for a high degree of image data compression. This transform technique thus finds applications in video-phones, tele/video conferencing, colour fax systems, interactive video discs, and high-definition colour imaging.

Each block of pixel data is transformed using a two dimensional matrix operation in the following manner:

$[T] = [C].[D].[C]^T$ where
 $[T]$ is the transformed block,
 $[C]$ is the basis of the DCT matrix,
 $[D]$ is the original data block and
 $[C]^T$ is the transpose of $[C]$.

The DCT coefficients are calculated from the relationship

$$F(u,v) = \frac{4C(u)C(v)}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i,j) \cos A \cos B.$$

The inverse transform relationship used at the receiver is given by:

$$f(i,j) = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C(u)C(v)F(u,v) \cos A \cos B$$

$$\text{where } A = \frac{(2i + 1)\mu\pi}{2N} \text{ and } B = \frac{(2j + 1)v\pi}{2N}$$

In both cases:

- i, j and $u, v = 0, 1, 2, \dots, N - 1$ (N being the block size),
- i, j are the spatial coordinates in the original image plane,
- u, v are the corresponding coordinates in the transform plane,
- $C(u) = C(v) = 1/\sqrt{2}$ for $u = v = 0$ and
- $C(u) = C(v) = 1$ for $u = v \neq 0$

Such transforms are readily performed within dedicated ICs. The high degree of image data compression available arises not from the transform itself. But due to the fact that many of the transformed coefficients are very small or zero, indicating little block to block variation. In addition, many of the remaining coefficients can be transmitted with lower precision without significantly affecting the received image quality. Since only a relatively few values need to be transmitted, any form of variable or run length coding can be used to advantage.

Depending upon the system requirements, a standard TV image can be reduced to a bit rate as low 64 kbit/s, but 2 Mbit/s is necessary to provide an acceptable quality of colour image.

REFERENCES

Because of the extensive nature of digital concepts, it has only been possible to give a brief description of the more common techniques in use. For a more expansive explanation the reader is referred to works such as those listed in the references below.

- (1) Hamming, R.W. (1950) 'Error detecting and Error correcting Codes'. *Bell System Technical Journal* Vol XXVI No.2.
- (2) Shu, Lin (1970) *An Introduction to Error-Correcting Codes*. New Jersey: Prentice-Hall Inc.
- (3) Halliwell, B.J. (1974) *Advanced Communication Systems*. London: Newnes-Butterworth Ltd.
- (4) Cattermole, K.W. (1969) *Principles of Pulse Code Modulation*. London: Iliffe Books Ltd.
- (5) Shu, Lin & Costello (1983) *Fundamentals and Applications of Error Control Coding*. New Jersey: Prentice-Hall Inc.
- (6) Spilker, J.J. (1977) *Digital Communications by Satellite*. New Jersey: Prentice-Hall Inc.
- (7) Vaidyanathan, P.P. 'Quadrature Mirror Filter Banks, M-Band Extensions and Perfect-Reconstruction Techniques.' *IEEE ASSP Magazine*, pp. 4–20, July 1987.
- (8) Hagenauer, J. 'Rate compatible punctured convolutional codes'. *Proc ICC '87*, pp. 1032, June 1987.