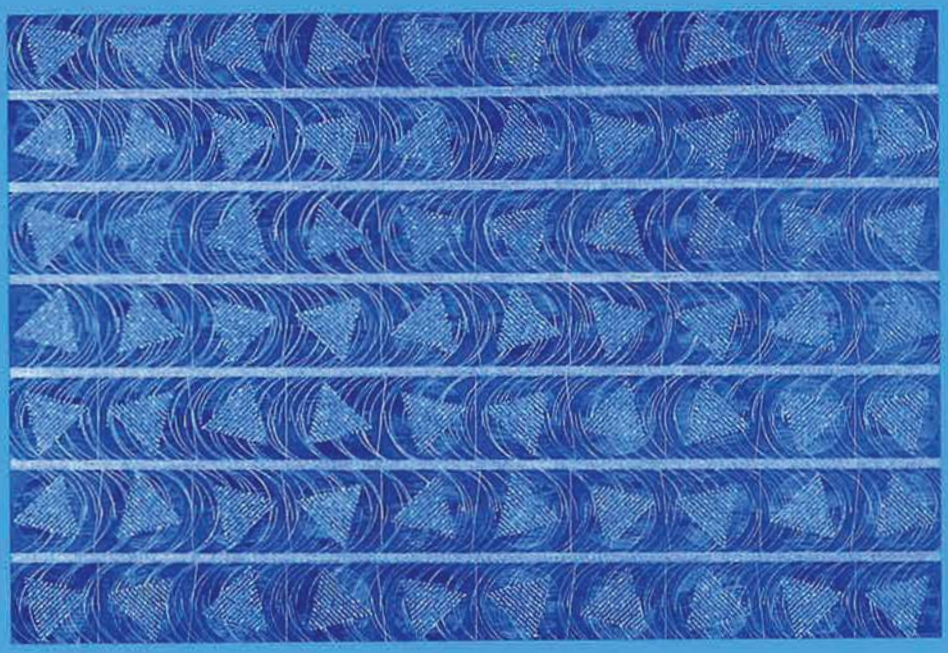
Ziemer
Tranter

Principles of Communications



Fourth Edition

Systems,

Modulation,

and Noise

MAKE LEKC

PRINCIPLES OF COMMUNICATIONS

Systems, Modulation, and Noise

Fourth Edition

R. E. ZIEMER
University of Colorado at Colorado Springs

W. H. TRANTER
University of Missouri–Rolla



JOHN WILEY & SONS, INC.
New York • Chichester • Brisbane • Toronto • Singapore

Cover designer: DesignHeads, Boston

Cover image: Andrea Way, Tidepool, 1988.
Ink and pencil on paper, 36 x 51½″
Private collection
Courtesy Brody's Gallery, Washington DC

Copyright © 1995 by John Wiley & Sons, Inc.
Previously published by Houghton Mifflin Company.

Reproduction or translation of any part of this work beyond that permitted by Sections107 and 108 of the 1976 United States Copyright Act without the permission of the copyright owner is unlawful. Request for permission or further information should be addressed to the Permission Department, John Wiley & Sons, Inc.

ISBN: 0 471 12496 6

Printed in the Unites States of America
10 9 8 7 6 5 4 3

CONTENTS

PREFACE ix

1 INTRODUCTION 1

1.1 The Block Diagram of a Communication System 2
1.2 Channel Characteristics 5
1.3 Summary of Systems-Analysis Techniques 15
1.4 Probabilistic Approaches to System Optimization 16
1.5 Preview of This Book 18
Further Reading 19

2 SIGNAL AND LINEAR SYSTEM ANALYSIS 20

2.1 Signal Models 20
2.2 Signal Classifications 27
2.3 Generalized Fourier Series 29
2.4 Fourier Series 33
2.5 The Fourier Transform 43
2.6 Power Spectral Density and Correlation 62
2.7 Signals and Linear Systems 65
2.8 Sampling Theory 90
2.9 The Hilbert Transform 94
2.10 The Discrete Fourier Transform and Fast Fourier Transform 104
Summary 107
Further Reading 114
Problems 114
Computer Exercises 131

3 BASIC MODULATION TECHNIQUES 134

3.1 Linear Modulation 135
3.2 Angle Modulation 166

3.3 Interference 191
3.4 Feedback Demodulators 201
3.5 Pulse Modulation 216
3.6 Multiplexing 231
3.7 Comparison of Modulation Systems 239
Summary 241
Further Reading 247
Problems 247
Computer Exercises 259

4 PROBABILITY AND RANDOM VARIABLES 261

4.1 What Is Probability? 261
4.2 Random Variables, Distribution Functions, and Density Functions 275
4.3 Statistical Averages 292
4.4 Some Useful pdf's 306
Summary 317
Further Reading 320
Problems 322
Computer Exercises 332

5 RANDOM SIGNALS AND NOISE 334

5.1 A Relative-Frequency Description of Random Processes 334
5.2 Some Terminology of Random Processes 336
5.3 Correlation Functions and Power Spectra 344
5.4 Linear Systems and Random Processes 354
5.5 Narrowband Noise 362
5.6 Distributions Encountered in FM Demodulation of Signals in Gaussian Noise 370
Summary 377
Further Reading 380
Problems 380
Computer Exercises 393

6 NOISE IN MODULATION SYSTEMS 394

6.1 Signal-to-Noise Ratios 395
6.2 Noise and Phase Errors in Coherent Systems 408
6.3 Noise in Angle Modulation Systems 414
6.4 Thresholds and Threshold Extension in FM 424
6.5 Noise in Pulse Modulation 432
Summary 441
Further Reading 443
Problems 443
Computer Exercises 451

7 BINARY DATA TRANSMISSION 453

7.1 Baseband Data Transmission in White Gaussian Noise 455
7.2 Binary Synchronous Data Transmission with Arbitrary Signal Shapes 462
7.3 Error Probabilities for Coherent Binary Signaling Schemes 475
7.4 Modulation Schemes Not Requiring Coherent References 482
7.5 Digital Signaling Through Bandlimited Channels 491
7.6 Multipath Interference 497
7.7 Flat Fading Channels 504
Summary 511
Further Reading 515
Problems 515
Computer Exercises 526

8 ADVANCED DATA COMMUNICATIONS TOPICS 529

8.1 M-ary Data Communications Systems 529
8.2 Bandwidth Efficiencies of Digital Modulation Formats 556
8.3 Synchronization 562
8.4 Spread-Spectrum Communication Systems 570
8.5 Satellite Communications 580
Summary 594
Further Reading 596
Problems 596
Computer Exercises 601

9 OPTIMUM RECEIVERS AND SIGNAL-SPACE CONCEPTS 603

9.1 Bayes Optimization 603
9.2 Vector Space Representation of Signals 615
9.3 Map Receivers for Digital Data Transmission 625
9.4 Estimation Theory 640
9.5 Applications of Estimation Theory to Communications 649
Summary 654
Further Reading 656
Problems 656
Computer Exercises 665

10 INFORMATION THEORY AND CODING 667

10.1 Basic Concepts 668
10.2 Source Encoding 681
10.3 Reliable Communication in the Presence of Noise 690
10.4 Modulation Systems 723
10.5 Bandwidth and Power-Efficient Modulation 726
Summary 732
Further Reading 735
Problems 736
Computer Exercises 745

APPENDIX A: PHYSICAL NOISE SOURCES AND NOISE CALCULATIONS IN COMMUNICATION SYSTEMS 747

A.1 Physical Noise Sources 747
A.2 Characterization of Noise in Systems—Noise Figure and Noise Temperature 754
A.3 Free-Space Propagation Example 764
Further Reading 768
Problems 769

APPENDIX B: JOINTLY GAUSSIAN RANDOM VARIABLES 773

B.1 Joint pdf and Characteristic Function 773
B.2 Linear Transformations of Gaussian Random Variables 774

APPENDIX C: MATHEMATICAL AND NUMERICAL TABLES 775

C.1 Fourier Transform Pairs 776
C.2 Fourier Transform Theorems 777
C.3 Trigonometric Identities 778
C.4 Series Expansions 779
C.5 Definite Integrals 780
C.6 The Q–Function 781

BIBLIOGRAPHY 783

AUTHOR INDEX 787

SUBJECT INDEX 789

PREFACE

In the fourth edition of this book we have maintained the objectives of the previous three editions. For the past 20 years we have attempted to provide in a single book a relatively complete treatment on the principles of communications suitable for seniors and beginning graduate students in electrical engineering. In addition to topics in communications, sections have been included on signals, systems, probability, and random processes, so that all of the necessary background is included within the pages of a single book. This has worked well for undergraduates, for new graduate students who may have forgotten some of the fundamentals, and for the working engineer who may use the book as a reference or who may be taking a course after hours.

A significant feature of the fourth edition of *Principles of Communications* is the inclusion of a number of "computer exercises" at the end of each chapter. These exercises are designed to make use of a computer in order to illustrate the principles involved. They follow the end-of-chapter problems in a section marked by a computer symbol. There are a number of motivations for the inclusion of these problems. The computer is now accepted as a valuable tool in engineering education and is becoming integrated into engineering curricula. Computer usage allows students to work more complex problems and to gain important insights through the working of these problems. While these problems can all be worked by writing programs in FORTRAN, Pascal, or C, we strongly recommend the use of scientific problem-solving programs such as MATLAB®, Mathcad® and Mathematica® to name only a few. Through the use of these tools, the programming time for the student is significantly minimized. In addition, the graphing support provided by these tools allows the student to plot waveforms and performance characteristics with ease. We also feel that it is important for students to be familiar with these tools and to use them in their day-to-day problem solving efforts.

The past twenty years have seen many new applications for communication systems and a number of new technologies have been developed for the implementation of these systems. While there is always a strong desire to include a wide variety of new applications and technologies in a new edition of a book, we strongly feel that a first course in communications serves the student best if the emphasis is placed on fundamentals. We feel that appli-

cations and technologies, which often have short lifetimes, are best treated in subsequent courses after students have mastered the basic theory and analysis techniques. We have, however, been sensitive to new techniques that are fundamental in nature and have added material as appropriate. A new section on trellis-coded modulation is an example. Other new sections are discussed in later paragraphs. In addition, a number of new problems and in-chapter examples have been added. Several less-useful sections have been deleted, resulting in a fourth edition that is actually a few pages shorter than the previous edition.

In order to compare the fourth edition of this book with the previous edition, we briefly consider the changes chapter by chapter. In Chapter 1, the tables have been updated and a brief discussion on atmospheric characteristics has been added. Chapter 2, which is essentially a review of signal and system theory, has been tightened by deleting several examples that seemed superfluous and a brief discussion of the decimation-in time FFT algorithm has been added.

Chapter 3, which is devoted to basic modulation methods, has been significantly expanded. A number of examples have been added which illustrate the effect of modulation in the time domain. Students usually find it much easier to visualize the effect of modulation in the frequency domain than in the time domain and therefore a number of waveforms corresponding to different modulation techniques have been added. In addition, the effect of interference in linear and nonlinear modulation has been combined into a single section. This section has also been expanded to include a discussion on spike noise resulting from origin encirclement in FM systems due to signal-tone interference. An understanding of this simple deterministic process will significantly help the student understand the origin encirclement phenomenon in the presence of additive Gaussian noise in Chapter 5 and thresholding in FM systems, studied in Chapter 6.

Chapters 4 and 5, which deal with probability and stochastic processes, have changed little from previous editions of this book. The most significant change in Chapter 4 is the use of the probability Q-function rather than the error function, erf(x), to represent probabilities of events in a Gaussian environment. While this really represents only a notational change, it recognizes that the probability Q-function is now used extensively in communications engineering. In edition four, a table has been added that gives the most common probability density functions along with the means and variances of the random variables described by the density functions. Chapter 5 has been shortened slightly, and a short section has been added illustrating the determination of the system equivalent-noise bandwidth from the impulse response.

In Chapter 6, which treats the noise performance of various modulation

schemes, the section on threshold extension has been rewritten so that the results are more general. The thresholding characteristics of an FM discriminator with a sinusoidal message signal are presented as a special case in an example problem.

Binary digital data transmission is the subject of Chapter 7. A new section has been added to this chapter dealing with fading channels. This is an important subject in today's world because of the rapidly expanding use of mobile communication systems. Chapter 8 treats more advanced topics in data communication systems including M-ary systems, bandwidth efficiency, synchronization, spread-spectrum systems and satellite links.

Chapter 9, which deals with optimum receivers and signal-space concepts is little changed from the previous edition. The chapter has been tightened, however, by deleting the topics of Weiner filtering and nonlinear estimation.

Chapter 10 provides the student with a brief introduction to the subjects of information theory and coding. As mentioned previously, an addition to this chapter is a brief discussion of the important subject of trellis-coded modulation, a very useful technique for the implementation of bandwidth and power efficient communication systems.

We have used this text for a two-semester course sequence for a number of years. The first course is a first-semester senior-level one covering Chapters 1, 2, 3, and Appendix A. The second course is a one-semester senior-level one covering Chapters 4 and 5 and selected topics from the remainder of the book. Some good students who have had a previous course on signal and system theory have successfully taken the second course without the first. A previous course on signal and system theory also allows the material in Chapter 2 to be reviewed rather than covered in detail. Hence, more material from Chapter 3 can be covered; for example, the coverage of the phase-lock loop can be expanded. Similarly, a previous course on probability allows the material in Chapters 4 and 5 to be reviewed, thereby allowing more topics on communications to be covered.

Although the text was designed primarily for a two-semester sequence, we included sufficient material to allow considerable flexibility in structuring many different course sequences based on the text. A solutions manual is available from the publisher. This manual contains complete solutions to all odd-numbered problems and answers to all even-numbered problems. Partial solutions are given to even-numbered problems in those cases in which we felt that some extra guidance would be helpful. In addition, complete solutions to all computer exercises are given in the solutions manual.

We wish to thank the many persons who have contributed to the development of this textbook and who have suggested improvements for the fourth edition. We especially thank our colleagues and students at the University of

Colorado at Colorado Springs and at the University of Missouri-Rolla for their comments and suggestions. We also express our thanks to the many colleagues who have offered suggestions to us by correspondence or verbally at technical conferences. The industries and agencies that have supported our research deserve special mention since, by working on these projects, we have expanded our knowledge and insight significantly. These include the National Aeronautics and Space Administration, the Office of Naval Research, the National Science Foundation, GE Aerospace, Motorola, Inc., Emerson Electric Company, Battelle Memorial Institute, and the McDonnell Douglas Corporation.

We also thank the reviewers of all editions of this book. The reviewers for the fourth edition deserve special mention. They were:

Derald Cummings, Penn State
Bruce Black, Rose-Hulman Institute of Technology
Steven Bibyk, Ohio State
Michael Fitz, Purdue
Laurence Hassebrook, Kentucky
Donley Winger, California Polytechnic State University (San Luis Obispo)
John Kieffer, Minnesota
Toby Berger, Cornell
B.V.K. Vijaya Kumar, Carnegie Mellon
John Komo, Clemson
Dan Schonfeld, Illinois (Chicago)
H.K. Hwang, California State Polytechnic University (Pomona)
Rao Yarlagadda, Oklahoma State

Dr. Ronald Iltis, from the University of California at Santa Barbara, deserves special mention. He provided a very helpful checkout review by reading the entire manuscript in page-proof form. All reviewers have contributed significantly to the fourth edition of this book. All remaining shortcomings are solely the responsibility of the authors.

Our students have provided continuing encouragement for our writing projects, and it is to them that we dedicate this book.

Finally, our wives, Sandy and Judy, deserve much more than a simple thanks for the patience that they have shown throughout our seemingly endless writing projects.

R. E. Z.
W. H. T.

PRINCIPLES OF COMMUNICATIONS

INTRODUCTION 1

As we approach the close of the twentieth century, we live in a world in which electrical communication is so commonplace that we pick up our cordless telephones without a second thought. Yet the importance of such communication in today's world is so crucial that we cannot imagine modern society without it. We are in an era of change, which some people refer to as the "information age," much like the era—more than 100 years ago—when the world underwent drastic changes because of the industrial revolution. The prosperity and continued development of modern nations will depend more on the originating and disseminating of *information,* rather than of manufactured goods. Examples illustrating that the value of information is often more than that of goods and services are provided by the fact that many portable computers are equipped with modems so that persons traveling on business are linked with their offices; by the number of motorists using cellular telephones in metropolitan areas; and by the fact that more and more mail transactions are now accomplished through computer, or electronic, mail.*

This book is concerned with the theory of systems for the conveyance of information. A *system* is a combination of circuits and devices put together to accomplish a desired result, such as the transmission of intelligence from one point to another. A characteristic of communication systems is the presence of uncertainty. This uncertainty is due in part to the inevitable presence in any system of unwanted signal perturbations, broadly referred to as *noise,* and in part to the unpredictable nature of information itself. Systems analysis in the presence of such uncertainty requires the use of probabilistic techniques.

Noise has been an ever-present problem since the early days of electrical communication, but it was not until the 1940s that probabilistic systems-analysis procedures were used to analyze and optimize communication sys-

* For more on this fascinating subject, see the *IEEE Communications Magazine,* Vol. 30, Oct. 1992 (special issue on "Realizing Global Communications").

tems operating in its presence [Wiener 1949; Rice 1944, 1945].* It is also somewhat surprising that the unpredictable nature of information was not widely recognized until the publication of Claude Shannon's mathematical theory of communications [Shannon 1948] in the late 1940s. This work was the beginning of the science of information theory, a topic that will be considered in some detail later.

A better appreciation of the accelerating pace at which electrical communication is developing can perhaps be gained by the historical outline of selected communications-related events given in Table 1.1.

With this brief introduction and history, we now look in more detail at the various components that make up a typical communication system.

1.1 THE BLOCK DIAGRAM OF A COMMUNICATION SYSTEM

Figure 1.1 shows a commonly used model for a communication system. Although it suggests a system for communication between two remotely located points, this block diagram is also applicable to remote sensing systems, such as radar or sonar, in which the system input and output may be located at the same site. Regardless of the particular application and configuration, all information transmission systems invariably involve three major subsystems—a transmitter, the channel, and a receiver. In this book we will usually be thinking in terms of systems for transfer of information between remotely located points. It is emphasized, however, that the techniques of systems analysis developed are not limited to such systems.

We will now discuss in more detail each functional element shown in Figure 1.1.

Input Transducer The wide variety of possible sources of information results in many different forms for messages. Regardless of their exact form, however, messages may be categorized as *analog* or *digital*. The former may be modeled as functions of a continuous-time variable (for example, pressure, temperature, speech, music), whereas the latter consist of discrete symbols (for example, written text). Almost invariably, the message produced by a source must be converted by a transducer to a form suitable for the particular type of communication system employed. For example, in electrical communications, speech waves are converted by a microphone to voltage variations. Such a converted message is referred to as the *message signal*. In this book, therefore, a *signal* can be interpreted as the variation of a quantity, often a voltage or current, with time.

* References in brackets [ ] refer to Historical References in the Bibliography.

TABLE 1.1 A Brief History of Communications

Time Period	Year	Event
	1826	Ohm's law
	1838	Samuel F. B. Morse demonstrates telegraph
	1864	James C. Maxwell predicts electromagnetic radiation
80 years	1876	Alexander Graham Bell patents the telephone
	1887	Heinrich Hertz verifies Maxwell's theory
	1897	Marconi patents a complete wireless telegraph system
	1904	Fleming invents the diode
	1906	Lee De Forest invents the triode amplifier
	1915	Bell System completes a transcontinental telephone line
	1918	B. H. Armstrong perfects the superheterodyne radio
40 years		receiver
	1920	J. R. Carson applies sampling to communications
	1937	Alec Reeves conceives pulse-code modulation (PCM)
	1938	Television broadcasting begins
	World	Radar and microwave systems are developed; statistical
	War II	methods are applied to signal extraction problems
	1948	The transistor is invented; Claude Shannon's "A
20 years		Mathematical Theory of Communications" is published
	1950	Time-division multiplexing is applied to telephony
	1956	First transoceanic telephone cable
	1960	Laser demonstrated by Maiman (1960)
		First communication satellite, Telstar I, launched (1962)
		Live television coverage of moon exploration;
10 years	to	experimental PCM systems; experimental laser communications; integrated circuits; digital signal
	1970	processing; color TV; time-shared computing
	1970	Commercial relay satellite communications (voice and digital); gigabit signaling rates; large-scale integration; integrated circuit realization of communications
	to	circuits; intercontinental computer communication nets; low-loss light fibers; off-the-shelf optical
10 years		communication systems; packet-switched digital data systems; interplanetary grand tour launched (1977)—communications accomplished from Jupiter, Saturn, Uranus, and Neptune (encounter in August, 1989); microprocessor, computed tomography, and
	1980	supercomputers developed
	1980	Satellite "switchboards in the sky"; mobile, cellular telephone systems; multifunction digital displays; 2 gigasample/s digital oscilloscopes; desktop publishing systems; programmable digital signal processors; digitally tuned receivers with autoscan;

(*continued*)

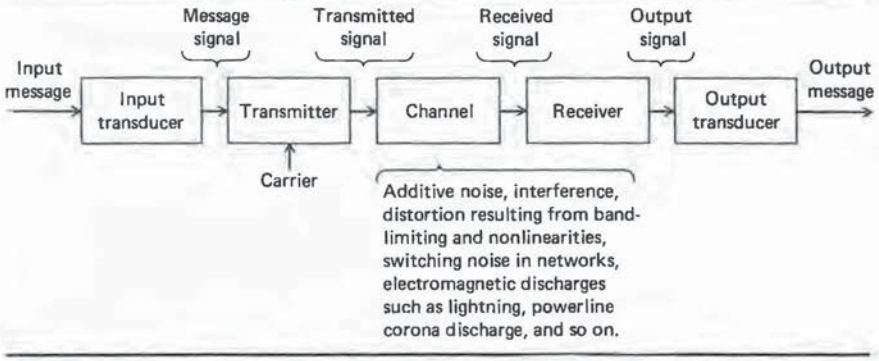
TABLE 1.1 (*continued*)

Time Period	Year	Event
10 years	to	cryptography on a chip; single-chip digital encoders and decoders; infrared data/control links; compact disk audio players; 200,000-word optical storage media; Ethernet developed; Bell system disbands, allowing competing long-haul telephone services;
	1990	digital signal processors developed
	1990's	Global positioning system (GPS) completed; high-definition television, very small antenna aperture satellites (VSATs), global satellite–based cellular telephones and integrated services digital network (ISDN) developed; personal communications systems come of age

Transmitter The purpose of the transmitter is to couple the message to the channel. Although it is not uncommon to find the input transducer directly coupled to the transmission medium, as for example in some intercom systems, it is often necessary to *modulate* a carrier wave with the signal from the input transducer. *Modulation* is the systematic variation of some attribute of the carrier, such as amplitude, phase, or frequency, in accordance with a function of the message signal. There are several reasons for using a carrier and modulating it. Important ones are (1) for ease of radiation, (2) to reduce noise and interference, (3) for channel assignment, (4) for multiplexing or transmission of several messages over a single channel, and (5) to overcome equipment limitations. Several of these reasons are self-explanatory; others, such as the second, will become more meaningful later.

In addition to modulation, other primary functions performed by the

FIGURE 1.1 Block diagram of a communication system



transmitter are filtering, amplification, and coupling the modulated signal to the channel (for example, through an antenna or other appropriate device).

Channel The channel can have many different forms; the most familiar, perhaps, is the channel that exists between the transmitting antenna of a commercial radio station and the receiving antenna of a radio. In this channel, the transmitted signal propagates through the atmosphere, or free space, to the receiving antenna. However, it is not uncommon to find the transmitter hard-wired to the receiver, as in most local telephone systems. This channel is vastly different from the radio example. However, all channels have one thing in common: the signal undergoes degradation from transmitter to receiver. Although this degradation may occur at any point of the communication system block diagram, it is customarily associated with the channel alone. This degradation often results from noise and other undesired signals or interference but also may include other distortion effects as well, such as fading signal levels, multiple transmission paths, and filtering. More about these unwanted perturbations will be presented shortly.

Receiver The receiver's function is to extract the desired message from the received signal at the channel output and to convert it to a form suitable for the output transducer. Although amplification may be one of the first operations performed by the receiver, especially in radio communications, where the received signal may be extremely weak, the main function of the receiver is to *demodulate* the received signal. Often it is desired that the receiver output be a scaled, possibly delayed, version of the message signal at the modulator input, although in some cases a more general function of the input message is desired. However, as a result of the presence of noise and distortion, this operation is less than ideal. Ways of approaching the ideal case of perfect recovery will be discussed as we proceed.

Output Transducer The output transducer completes the communication system. This device converts the electric signal at its input into the form desired by the system user. Perhaps the most common output transducer is a loudspeaker. However, there are many other examples, such as tape recorders, personal computers, meters, and cathode-ray tubes, to name only a few.

1.2 CHANNEL CHARACTERISTICS

Noise Sources

Noise in a communication system can be classified into two broad categories, depending on its source. Noise generated by components within a commu-

nication system, such as resistors, electron tubes, and solid-state active devices is referred to as *internal noise*. The second category, *external noise*, results from sources outside a communication system, including atmospheric, man-made, and extraterrestrial sources.

Atmospheric noise results primarily from spurious radio waves generated by the natural electrical discharges within the atmosphere associated with thunderstorms. It is commonly referred to as *static* or *spherics*. Below about 100 MHz, the field strength of such radio waves is inversely proportional to frequency. Atmospheric noise is characterized in the time domain by large-amplitude, short-duration bursts and is one of the prime examples of noise referred to as *impulsive*. Because of its inverse dependence on frequency, atmospheric noise affects commercial AM broadcast radio, which occupies the frequency range from 540 kHz to 1.6 MHz, more than it affects television and FM radio, which operate in frequency bands above 50 MHz.

Man-made noise sources include high-voltage powerline corona discharge, commutator-generated noise in electrical motors, automobile and aircraft ignition noise, and switching-gear noise. Ignition noise and switching noise, like atmospheric noise, are impulsive in character. Impulse noise is the predominant type of noise in switched wireline channels, such as telephone channels. For applications such as voice transmission, impulse noise is only an irritation factor; however, it can be a serious source of error in applications involving transmission of digital data.

Yet another important source of man-made noise is radio-frequency transmitters other than the one of interest. Noise due to interfering transmitters is commonly referred to as *radio-frequency interference* (RFI). RFI is particularly troublesome in situations in which a receiving antenna is subject to a high-density transmitter environment, as in mobile communications in a large city.

Extraterrestrial noise sources include our sun and other hot heavenly bodies, such as stars. Owing to its high temperature (6000°C) and relatively close proximity to the earth, the sun is an intense, but fortunately localized source of radio energy that extends over a broad frequency spectrum. Similarly, the stars are sources of wideband radio energy. Although much more distant and hence less intense than the sun, nevertheless they are collectively an important source of noise because of their vast numbers. Radio stars such as quasars and pulsars are also intense sources of radio energy. Considered a signal source by radio astronomers, such stars are viewed as another noise source by communications engineers. The frequency range of solar and cosmic noise extends from a few megahertz to a few gigahertz.

Another source of interference in communication systems is multiple transmission paths. These can result from reflection off buildings, the earth,

airplanes, and ships or from refraction by stratifications in the transmission medium. If the scattering mechanism results in numerous reflected components, the received multipath signal is noiselike and is termed *diffuse*. If the multipath signal component is composed of only one or two strong reflected rays, it is termed *specular*. Finally, signal degradation in a communication system can occur because of random changes in attenuation within the transmission medium. Such signal perturbations are referred to as *fading*, although it should be noted that specular multipath also results in fading due to the constructive and destructive interference of the received multiple signals.

Internal noise results from the random motion of charge carriers in electronic components. It can be of three general types: the first is referred to as *thermal noise*, which is caused by the random motion of free electrons in a conductor or semiconductor excited by thermal agitation; the second is called *shot noise* and is caused by the random arrival of discrete charge carriers in such devices as thermionic tubes or semiconductor junction devices; the third, known as *flicker noise*, is produced in semiconductors by a mechanism not well understood and is more severe the lower the frequency. The first two types of noise sources are modeled analytically in Appendix A, and examples of system characterization using these models are given there.

Types of Transmission Channels

There are many types of transmission channels. We will discuss the characteristics, advantages, and disadvantages of three common types: electromagnetic-wave propagation channels, guided electromagnetic-wave channels, and optical channels. The characteristics of all three may be explained on the basis of electromagnetic-wave propagation phenomena. However, the characteristics and applications of each are different enough to warrant our considering them separately.

Electromagnetic-Wave Propagation Channels The possibility of the propagation of electromagnetic waves was predicted in 1864 by James Clerk Maxwell (1831–1879), a Scottish mathematician who based his theory on the experimental work of Michael Faraday. Heinrich Hertz (1857–1894), a German physicist, carried out experiments between 1886 and 1888 using a rapidly oscillating spark to produce electromagnetic waves, thereby experimentally proving Maxwell's predictions. Therefore, by the latter part of the nineteenth century, the physical basis for many modern inventions utilizing electromagnetic-wave propagation—such as radio, television, and radar—was already established.

The basic physical principle involved is the coupling of electromagnetic

energy into a propagation medium, which can be free space or the atmosphere, by means of a radiation element referred to as an *antenna*. Many different propagation modes are possible, depending on the physical configuration of the antenna and the characteristics of the propagation medium. The simplest case—which never occurs in practice—is propagation from a point source in a medium that is infinite in extent. The propagating wave fronts (surfaces of constant phase) in this case would be concentric spheres. Such a model might be used for the propagation of electromagnetic energy from a distant spacecraft to earth. Another idealized model, which approximates the propagation of radio waves from a commercial broadcast antenna, is that of a conducting line perpendicular to an infinite conducting plane. These and other idealized cases have been analyzed in books on electromagnetic theory. Our purpose is not to summarize all the idealized models, but to point out basic aspects of propagation phenomena in practical channels.

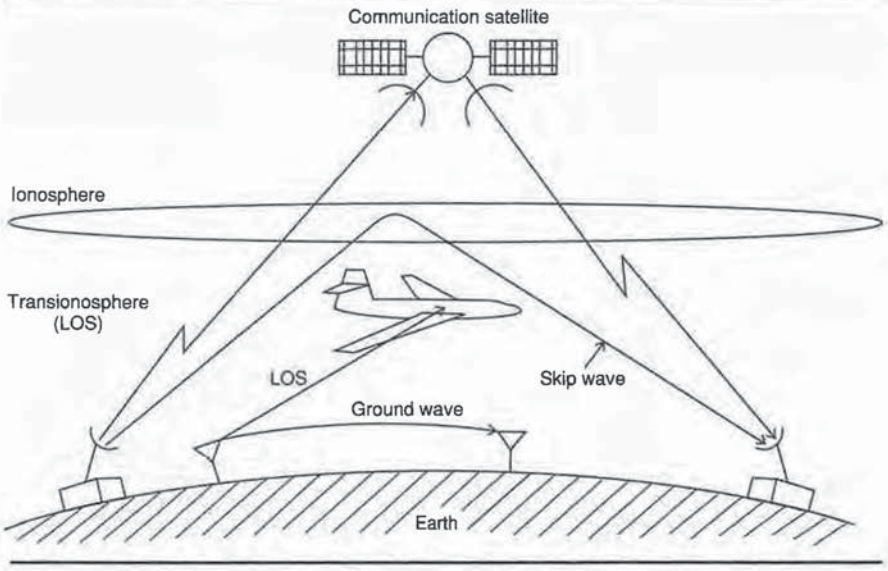
Except for the case of propagation between two spacecraft in outer space, the intermediate medium between transmitter and receiver is never well approximated by free space. Depending on the distance involved and the frequency of the radiated waveform, a terrestrial communication link may depend on line-of-sight, ground-wave, or ionospheric skip-wave propagation (see Fig. 1.2). Table 1.2 lists frequency bands from 3 kHz to 10^7 GHz, along with typical uses in each band. Note that the frequency bands are given in decades; the VHF band has 10 times as much frequency space as the HF band. Table 1.3 shows some bands of particular interest.

General application allocations are arrived at by international agreement. The present system of frequency allocations is administered by the International Telecommunications Union (ITU), which is responsible for the periodic convening of Administrative Radio Conferences on a regional (RARC) or a worldwide (WARC) basis.* In the United States, the Federal Communications Commission (FCC) awards specific applications within a band as well as licenses for their use. For example, in 1974, the FCC decided to alleviate congestion of land mobile radio in the 150 and 450 MHz bands by allocating 115 MHz of bandwidth in the 806 to 947 MHz band to mobile radio. The reasons for the move to UHF were twofold: (1) frequency space was available; and (2) a new concept in mobile radio, called *cellular mobile radio,* in which the geographic area of coverage is divided into cells with a base station in each cell and handoff between cells, had emerged.†

* See A. F. Inglis, *Electronic Communications Handbook,* New York: McGraw-Hill (1988), Chapter 3.

† See John Oetting, "Cellular Mobile Radio—An Emerging Technology." *IEEE Communications Magazine,* Vol. 21, pp. 10–15, Nov. 1983. Also see *IEEE Communications Magazine,* Vol. 24, Feb. 1986 (the whole issue is devoted to mobile radio communications).

FIGURE 1.2 The various propagation modes for electromagnetic waves (LOS stands for line of sight)



WARC-92, the most recent World Administrative Radio Conference, was attended by 127 out of 166 member countries. It made the following changes:

1. Added 790 kHz of extra spectrum to the shortwave bands—200 kHz below 10 MHz and 590 kHz between 11 and 19 MHz
2. Made an allocation between 1.70 and 2.69 GHz for future public land-mobile telecommunication systems
3. Allocated the 1.605 to 1.675 and 1.800 to 1.805 GHz bands worldwide for public telephone service to and from commercial aircraft
4. Allocated additional spectrum for low-earth-orbit satellite communications, mostly below 1 GHz

At lower frequencies, or long wavelengths, propagating radio waves tend to follow the earth's surface. At higher frequencies, or short wavelengths, radio waves propagate in straight lines. Another phenomenon that occurs at lower frequencies is reflection (or refraction) of radio waves by the ionosphere (a series of layers of charged particles at altitudes between 30 and 250 miles above the earth's surface). Thus, for frequencies below about 100 MHz, it is possible to have skip-wave propagation. At night, when lower ionospheric layers disappear due to less ionization from the sun (the E, F_1, and F_2 layers coalesce into one layer—the F layer), longer skip-wave prop-

TABLE 1.2 Frequency Bands with Typical Uses

Frequency Band[a]	Name	Microwave Band (GHz)	Letter Designations	Typical Uses
3–30 kHz	Very low frequency (VLF)			Long-range navigation; sonar
30–300 kHz	Low frequency (LF)			Navigational aids; radio beacons
300–3000 kHz	Medium frequency (MF)			Maritime radio; direction finding; distress and calling; Coast Guard comm.; commercial AM radio
3–30 MHz	High frequency (HF)			Search and rescue; aircraft comm. with ships; telegraph, telephone, and facsimile; ship-to-coast
30–300 MHz	Very high frequency (VHF)		P R E V I O U S C U R R E N T	VHF television channels; FM radio; land transportation; private aircraft; air traffic control; taxi cab; police; navigational aids
0.3–3 GHz	Ultrahigh frequency (UHF)	0.5–1.0	VHF C	UHF television channels; radiosonde; nav. aids; surveillance radar; satellite comm.; radio altimeters; microwave links; airborne radar; approach radar; weather radar; common carrier land mobile
		1.0–2.0	L D	
		2.0–3.0	S E	
		3.0–4.0	S F	
		4.0–6.0	C G	
3–30 GHz	Superhigh frequency (SHF)	6.0–8.0	C H	
		8.0–10.0	X I	
		10.0–12.4	X J	
		12.4–18.0	Ku J	
		18.0–20.0	K J	
		20.0–26.5	K K	
		26.5–40.0	Ka K	
30–300 GHz	Extremely high frequency (EHF)			Railroad service; radar landing systems; experimental

(*continued*)

TABLE 1.2 (*continued*)

Frequency Band[a]	Name	Microwave Band (GHz)	Letter Designations	Typical Uses
300 GHz–3 THz	(0.1–1 mm)			Experimental
43–430 THz	Infrared (7–0.7 μm)			Optical comm. systems
430–750 THz	Visible light (0.7–0.4 μm)			Optical comm. systems
750–3000 THz	Ultraviolet light (0.4–0.1 μm)			Optical comm. systems

[a] Abbreviations: kHz = kilohertz = $\times 10^3$ hertz; MHz = megahertz = $\times 10^6$ hertz; GHz = gigahertz = $\times 10^9$ hertz; THz = terahertz = $\times 10^{12}$ hertz; μm = micrometers = 10^{-6} meters.

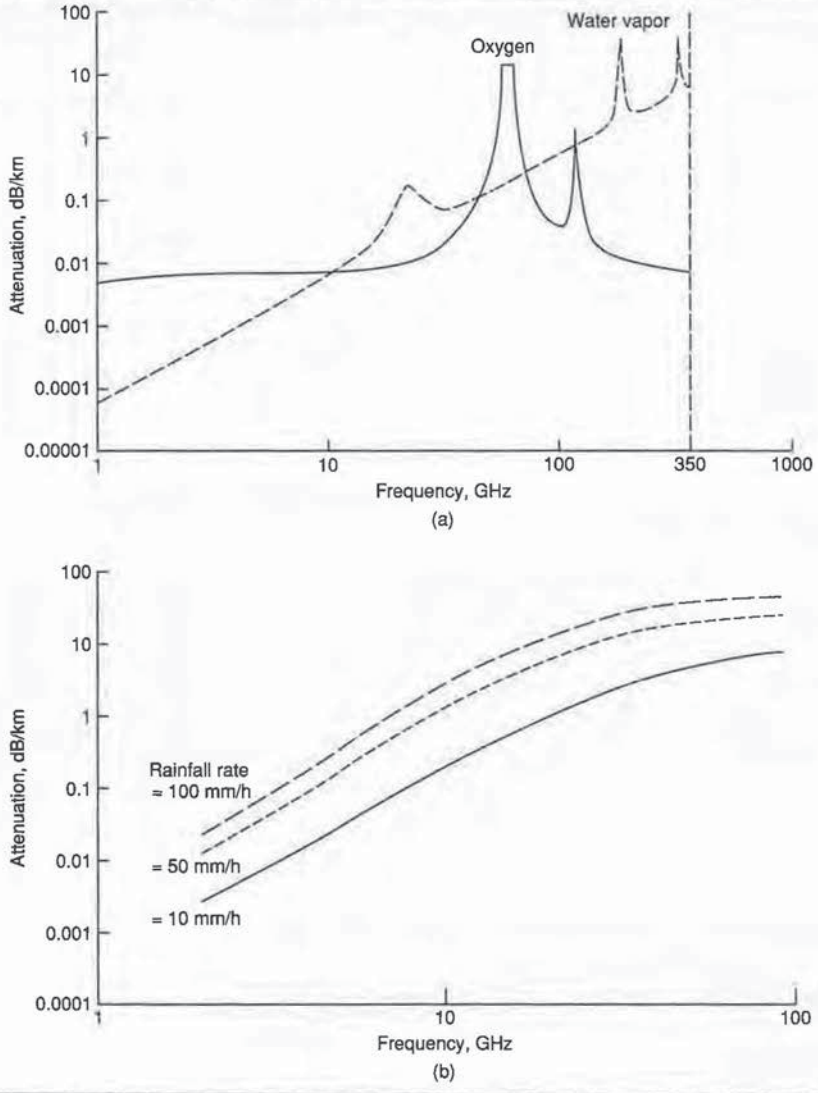
agation occurs as a result of reflection from the higher, single reflecting layer of the ionosphere.

Above about 300 MHz, propagation of radio waves is by line of sight, because the ionosphere will not bend radio waves in this frequency region sufficiently to reflect them back to the earth. At still higher frequencies, say above 1 or 2 GHz, atmospheric gases (mainly oxygen), water vapor, and precipitation absorb and scatter radio waves. This phenomenon manifests itself as attenuation of the received signal, with the attenuation generally being more severe the higher the frequency (there are resonance regions for absorption by gases that peak at certain frequencies). Figure 1.3 shows

TABLE 1.3 Selected Frequency Bands for Public Use and Military Communications

Use	Frequency
Omega navigation	10–14 kHz
Worldwide submarine communication	30 kHz
Loran C navigation	100 kHz
Standard (AM) broadcast	540–1600 kHz
FM broadcast	88–108 MHz
Television: Channels 2–4	54–72 MHz
Channels 5–6	76–88 MHz
Channels 7–13	174–216 MHz
Channels 14–83	420–890 MHz

FIGURE 1.3 Specific attenuation for atmospheric gases and rain. (a) Specific attenuation due to oxygen and water vapor (concentration of 7.5 g/m^3). (b) Specific attenuation due to rainfall at rates of 10, 50, and 100 mm/h.



specific attenuation curves as a function of frequency* for oxygen, water vapor, and rain [recall that 1 decibel (dB) is ten times the logarithm to the base 10 of a power ratio]. One must account for the possible attenuation by

* Data from Louis J. Ippolito, Jr., *Radiowave Propagation in Satellite Communications*. New York, Van Nostrand Reinhold, 1986, Chapters 3 and 4.

such atmospheric constituents in the design of microwave links, which are used in transcontinental telephone links and ground-to-satellite communications links.

At about 23 GHz, the first absorption resonance due to water vapor occurs, and at about 62 GHz a second one occurs due to oxygen absorption. These frequencies should be avoided in transmission of desired signals through the atmosphere, or undue power will be expended (one might, for example, use 62 GHz as a signal for cross-linking between two satellites, where atmospheric absorption is no problem, and thereby prevent an enemy on the ground from listening in). Another absorption frequency for oxygen occurs at 120 GHz, and two other absorption frequencies for water vapor occur at 180 and 350 GHz.

Communication at millimeter-wave frequencies (that is, at 30 GHz and higher) is becoming more important now that there is so much congestion at lower frequencies (the Advanced Technology Satellite, to be launched in the mid-1990s, employs an uplink frequency band around 20 GHz and a downlink frequency band at about 30 GHz). Communication at millimeter-wave frequencies is becoming more feasible because of technological advances in components and systems.

Somewhere above 1 THz (1000 GHz), the propagation of radio waves becomes optical in character. At a wavelength of 10 μm (0.00001 m), the carbon dioxide laser provides a source of coherent radiation, and visible-light lasers (for example, helium-neon) radiate in the wavelength region of 1 μm and shorter. Terrestrial communications systems employing such frequencies experience considerable attenuation on cloudy days, and laser communications over terrestrial links are restricted to optical fibers for the most part. Analyses have been carried out for the employment of laser communications cross-links between satellites, but there are as yet no optical satellite communications links actually flying.

Guided Electromagnetic-Wave Channels Perhaps the most extensive example of guided electromagnetic-wave channels is the part of the long-distance telephone network that uses wire lines.* Communication between persons a continent apart was first achieved by means of voice frequency transmission (below 10,000 Hz) over open wire. Quality of transmission was rather poor. By 1952, use of the types of modulation known as *double-sideband* and *single-sideband* on high-frequency carriers was established. Communication over predominantly multipair and coaxial-cable lines produced transmission of

* For a summary of guided transmission systems as applied to telephone systems, see F. T. Andrews, Jr., "Communications Technology: 25 Years in Retrospect. Part III, Guided Transmission Systems, 1952–1973." *IEEE Communications Society Magazine,* Vol. 16, pp. 4–10, Jan. 1978.

much better quality. With the completion of the first trans-Atlantic cable in 1956, intercontinental telephone communication was no longer dependent on high-frequency radio, and the quality of intercontinental telephone service improved significantly.

Bandwidths on coaxial-cable links are a few megahertz. The need for greater bandwidth initiated the development of millimeter-wave waveguide transmission systems. However, with the development of low-loss optical fibers, efforts to improve millimeter-wave systems to achieve greater bandwidth have ceased. The development of optical fibers, in fact, has made the concept of a wired city—wherein digital data and video can be piped to any residence or business within a city—a reality. Modern coaxial-cable systems can carry only 13,000 voice channels per cable, but optical links are capable of carrying several times this number (the limiting factor being the current driver for the light source).

Optical Links The use of optical links was, until recently, limited to short and intermediate distances. With the installation of trans-Pacific and trans-Atlantic optical cables in 1988 and early 1989, this is no longer true.* The technological breakthroughs that have made possible the use of light waves for communication are the development of small coherent light sources (semiconductor lasers), low-loss (approximately 1 dB/km) optical fibers or waveguides, and low-noise detectors.†

A typical fiber-optic communication system has a light source, which may be either a light-emitting diode or a semiconductor laser, in which the intensity of the light is varied by the message source. The output of this modulator is the input to a light-conducting fiber (analogous to the coaxial cable of a standard telephone channel). The receiver, or light sensor, typically consists of a photodiode. In a photodiode, an average current flows that is proportional to the optical power of the incident light. However, the exact number of charge carriers (that is, electrons) is random. The output of the detector is the sum of the average current—which is proportional to the

* See Inglis, op. cit., Chapter 8.

† For an overview of optical communication systems, see S. D. Personick, "Optical Fibers: A New Transmission Medium," *IEEE Communications Society Magazine,* Vol. 13, pp. 20–24, Jan. 1975. Also see S. D. Personick, "Fiber Optic Communications: A Technology Coming of Age," *IEEE Communications Society Magazine,* Vol. 16, pp. 12–20, Mar. 1978. For an overview of future trends in terrestrial optical fiber communications systems see P. Kaiser, J. Midwinter, and S. Shimada, "Status and Future Trends in Terrestrial Optical Fiber Systems in North America," *IEEE Communications Magazine,* Vol. 25, pp. 8–21, Oct. 1987. For an overview on the use of signal-processing methods to improve optical communications, see J. H. Winters, R. D. Gitlin, and S. Kasturia, "Reducing the Effects of Transmission Impairments in Digital Fiber Optic Systems," *IEEE Communications Magazine,* Vol. 31, pp. 68–76, June 1993.

modulation—and a noise component. This noise component differs from the thermal noise generated by the receiver electronics in that it is "bursty" in character. It is referred to as *shot noise,* in analogy to the noise made by shot hitting a metal plate. Another source of degradation is the dispersion of the optical fiber itself. For example, pulse-type signals sent to the fiber are observed as "smeared out" at the receiver. Losses also occur as a result of the connections between cable pieces and between cable and system components.

Finally, it should be mentioned that optical communications can take place through free space.*

1.3 SUMMARY OF SYSTEMS-ANALYSIS TECHNIQUES

Having identified and discussed the main subsystems in a communication system and certain characteristics of transmission media, let us now look at the techniques at our disposal for systems analysis and design.

Time and Frequency-Domain Analyses

From circuits courses or prior courses in linear systems analysis, you are well aware that the electrical engineer lives in the two worlds, so to speak, of time and frequency. Also, you should recall that dual time-frequency analysis techniques are especially valuable for linear systems for which the principle of superposition holds. Although many of the subsystems and operations encountered in communication systems are for the most part linear, many are not. Nevertheless, frequency-domain analysis is an extremely valuable tool to the communications engineer, more so perhaps than to other systems analysts. Since the communications engineer is concerned primarily with signal bandwidths and signal locations in the frequency domain, rather than with transient analysis, the essentially steady-state approach of the Fourier series and transforms is used. Accordingly, we will discuss the Fourier series, the Fourier integral, and their role in systems analysis in Chapter 2. We postpone a more complete discussion of these tools for systems analysis until then.

Modulation and Communication Theories

Modulation theory employs time and frequency-domain analyses to analyze and design systems for modulation and demodulation of information-bearing signals. To be specific, consider the message signal $m(t)$, which is to be trans-

* See Joseph Katz, "Semiconductor Optoelectric Devices for Free-Space Optical Communications," *IEEE Communications Society Magazine,* Vol. 21, pp. 20–27, Sept. 1983.

mitted through a channel using the method of double-sideband modulation. The modulated carrier for double-sideband modulation is of the form $x_c(t) = A_c m(t) \cos \omega_c t$, where ω_c is the carrier frequency in radians per second and A_c is the carrier amplitude. Not only must a modulator be built that can multiply two signals, but amplifiers are required to provide the proper power level of the transmitted signal. The exact design of such amplifiers is not of concern in a systems approach. However, the frequency content of the modulated carrier, for example, is important to their design and therefore must be specified. The dual time-frequency analysis approach is especially helpful in providing such information.

At the other end of the channel, there must be a receiver configuration capable of extracting a replica of $m(t)$ from the modulated signal, and one can again apply time and frequency-domain techniques to good effect.

The analysis of the effect of interfering signals on system performance and the subsequent modifications in design to improve performance in the face of such interfering signals are part of *communication theory*, which, in turn, makes use of modulation theory.

This discussion, although mentioning interfering signals, has not explicitly emphasized the uncertainty aspect of the information-transfer problem. Indeed, much can be done without applying probabilistic methods. However, as pointed out previously, the application of probabilistic methods, coupled with optimization procedures, has been one of the key ingredients of the modern communications era and led to the development during the 1950s and 1960s of new techniques and systems totally different in concept from those which existed before World War II.

We will now survey several approaches to statistical optimization of communication systems.

1.4 PROBABILISTIC APPROACHES TO SYSTEM OPTIMIZATION

The works of Wiener and Shannon, previously cited, were the beginning of modern statistical communication theory. Both these investigators applied probabilistic methods to the problem of extracting information-bearing signals from noisy backgrounds, but they worked from different standpoints. In this section we briefly examine these two approaches to optimum system design.

Statistical Signal Detection and Estimation Theory

Wiener considered the problem of optimally filtering signals from noise, where "optimum" is used in the sense of minimizing the average squared

error between the desired and the actual output. The resulting filter structure is referred to as the *Wiener optimum filter*. This type of approach is most appropriate for analog communication systems in which the demodulated output of the receiver is to be a faithful replica of the message input to the transmitter.

Wiener's approach is reasonable for analog communications. However, in the early 1940s, North [1943] provided a more fruitful approach to the digital communications problem, in which the receiver must distinguish between a number of discrete signals in background noise. Actually, North was concerned with radar, which requires only the detection of the presence or absence of a pulse. Since fidelity of the detected signal at the receiver is of no consequence in such signal-detection problems, North sought the filter that would maximize the peak-signal-to-rms-noise ratio at its output. The resulting optimum filter is called the *matched filter,* for reasons that will become apparent in Chapter 7, where we consider digital data transmission. Later adaptations of the Wiener and matched-filter ideas to time-varying backgrounds resulted in *adaptive filters*. We will consider a subclass of such filters in Chapter 7 when *equalization* of digital data signals is discussed.

The signal-extraction approaches of Wiener and North, formalized in the language of statistics in the early 1950s by several researchers [see Middleton 1960, p. 832, for several references], were the beginnings of what is today called *statistical signal detection* and *estimation theory*. Woodward and Davies [1952], in considering the design of receivers utilizing *all* the information available at the channel output, determined that this so-called ideal receiver computes the probabilities of the received waveform given the possible transmitted messages. These computed probabilities are known as *a posteriori* probabilities. The ideal receiver then makes the decision that the transmitted message was the one corresponding to the largest *a posteriori* probability. Although perhaps somewhat vague at this point, this *maximum a posteriori* (MAP) *principle,* as it is called, is one of the cornerstones of detection and estimation theory. We will examine these ideas in more detail in Chapters 7 through 9.

Information Theory and Coding

The basic problem that Shannon considered is, "Given a message source, how shall the messages produced be represented so as to maximize the information conveyed through a given channel?" Although Shannon formulated his theory for both discrete and analog sources, we will think here in terms of discrete systems. Clearly, a basic consideration in this theory is a measure of information. Once a suitable measure has been defined (and we will do so in Chapter 10), the next step is to define the information-

carrying capacity, or simply capacity, of a channel as the maximum rate at which information can be conveyed through it. The obvious question that now arises is, "Given a channel, how closely can we approach the capacity of the channel, and what is the quality of the received message?" A most surprising, and the singularly most important, result of Shannon's theory is that by suitably restructuring the transmitted signal, we can transmit information through a channel *at any rate less than the channel capacity with arbitrarily small error,* despite the presence of noise, provided we have an arbitrarily long time available for transmission. This is the gist of Shannon's *second theorem.* Limiting our discussion at this point to binary discrete sources, a proof of Shannon's second theorem proceeds by selecting codewords at random from the set of 2^n possible binary sequences n digits long at the channel input. The probability of error in receiving a given n-digit sequence, when averaged over all possible code selections, becomes arbitrarily small as n becomes arbitrarily large. Thus many suitable codes exist, *but we are not told how to find these codes.* Indeed, this has been the dilemma of information theory since its inception and is an area of active research. In recent years, great strides have been made in finding good coding and decoding techniques that are implementable with a reasonable amount of hardware and require only a reasonable amount of time to decode. Several basic coding techniques will be discussed in Chapter 10.

1.5 PREVIEW OF THIS BOOK

From the previous discussion, the importance of probability and noise characterization in analysis of communication systems should be apparent. Accordingly, after presenting basic signal, system, and noiseless modulation theory in Chapters 2 and 3, we shall discuss probability and noise theory in Chapters 4 and 5. Following this, we apply these tools to the noise analysis of analog communications schemes in Chapter 6. In Chapters 7 and 8, we use probabilistic techniques to find optimum receivers when we consider digital data transmission. Various types of digital modulation schemes are analyzed in terms of error probability. In Chapter 9, we approach optimum signal detection and estimation techniques on a generalized basis and use signal-space techniques to provide insight as to why systems that have been analyzed previously perform as they do. As already mentioned, information theory and coding are the subjects of Chapter 10. This provides us with a means of comparing actual communication systems with the ideal. Such comparisons are then considered in Chapter 10 to provide a basis for selection of systems.

In closing, we must note that large areas of communications technology—

such as optical, computer, and military communications—are not touched on in this book. However, one can apply the principles developed in this text in those areas as well.

FURTHER READING

The references for this chapter were chosen to indicate the historical development of modern communications theory and by and large are not easy reading. They are found in the Historical References section of the Bibliography. You also may consult the introductory chapters of the books listed in the Further Reading sections of Chapters 2 and 3. These books appear in the main portion of the Bibliography.

2 SIGNAL AND LINEAR SYSTEM ANALYSIS

The study of information transmission systems is inherently concerned with the transmission of signals through systems. Recall that in Chapter 1 a *signal* was defined as the time history of some quantity, usually a voltage or current. A *system* is a combination of devices and networks (subsystems) chosen to perform a desired function. Because of the sophistication of modern communication systems, a great deal of analysis and experimentation with trial subsystems occurs before actual building of the desired system. Thus the communications engineer's tools are mathematical models for signals and systems.

In this chapter, we review techniques useful for modeling and analysis of signals and systems used in communications engineering.* Of primary concern will be the dual time-frequency viewpoint for signal representation, and models for linear, time-invariant, two-port systems. It is important always to keep in mind that a model is not the signal or the system, but a mathematical idealization of certain characteristics of it that are most important to the problem at hand.

With this brief introduction, we now consider signal classifications and various methods for modeling signals and systems. These include frequency-domain representations for signals via the complex exponential Fourier series and the Fourier transform followed by linear system models and techniques for analyzing the effects of such systems on signals.

2.1 SIGNAL MODELS

Deterministic and Random Signals

In this book we are concerned with two broad classes of signals, referred to as deterministic and random. *Deterministic signals* can be modeled as com-

* More complete treatments of these subjects can be found in texts on linear system theory. (See, for example, Ziemer, Tranter, and Fannin (1993).)

pletely specified functions of time. For example, the signal

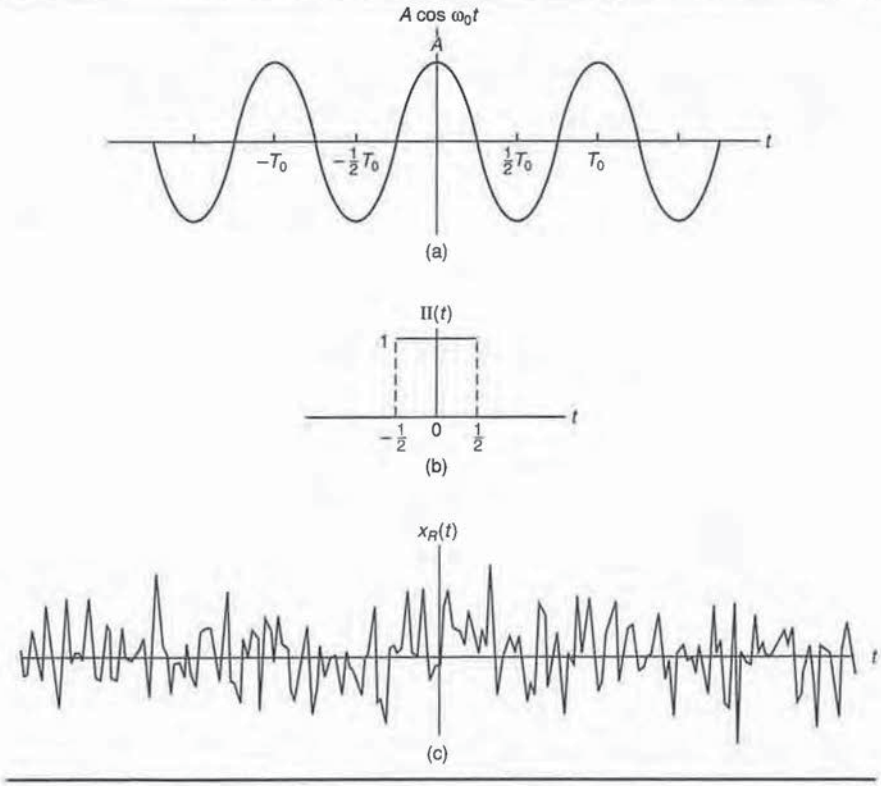
$$x(t) = A \cos \omega_0 t, \qquad -\infty < t < \infty \tag{2.1}$$

where A and ω_0 are constants, is a familiar example of a deterministic signal. Another example of a deterministic signal is the unit rectangular pulse, denoted as $\Pi(t)$ and defined as

$$\Pi(t) = \begin{cases} 1, & |t| < \frac{1}{2} \\ 0, & \text{otherwise} \end{cases} \tag{2.2}$$

Random signals are signals that take on random values at any given time instant and must be modeled probabilistically. They will be considered in Chapter 5. Figure 2.1 illustrates the various types of signals just discussed.

FIGURE 2.1 Examples of various types of signals. (a) Deterministic (sinusoidal) signal. (b) Unit rectangular pulse signal. (c) Random signal.



Periodic and Aperiodic Signals

The signal defined by (2.1) is an example of a *periodic signal*. A signal $x(t)$ is periodic if and only if

$$x(t + T_0) = x(t), \qquad -\infty < t < \infty \tag{2.3}$$

where the constant T_0 is the period. The smallest such number satisfying (2.3) is referred to as the *fundamental period* (the modifier "fundamental" is often excluded). Any signal not satisfying (2.3) is called *aperiodic*.

Phasor Signals and Spectra

A useful periodic signal in system analysis is the signal

$$\tilde{x}(t) = Ae^{j(\omega_0 t+\theta)}, \qquad -\infty < t < \infty \tag{2.4}$$

which is characterized by three parameters: amplitude A, phase θ in radians, and frequency ω_0 in radians per second or $f_0 = \omega_0/2\pi$ hertz. We will refer to $\tilde{x}(t)$ as a *rotating phasor* to distinguish it from the *phasor* $Ae^{j\theta}$, for which $e^{j\omega_0 t}$ is implicit. Using Euler's theorem,* we may readily show that $\tilde{x}(t) = \tilde{x}(t + T_0)$, where $T_0 = 2\pi/\omega_0$. Thus $\tilde{x}(t)$ is a periodic signal with period $2\pi/\omega_0$.

The rotating phasor $Ae^{j(\omega_0 t+\theta)}$ can be related to a real, sinusoidal signal $A \cos(\omega_0 t + \theta)$ in two ways. The first is by taking its real part,

$$\begin{aligned} x(t) = A\cos(\omega_0 t + \theta) &= \operatorname{Re} \tilde{x}(t) \\ &= \operatorname{Re} Ae^{j(\omega_0 t+\theta)} \end{aligned} \tag{2.5}$$

and the second is by taking one-half of the sum of $\tilde{x}(t)$ and its complex conjugate,

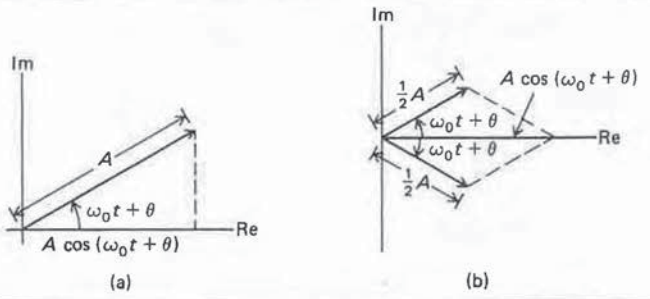
$$\begin{aligned} A\cos(\omega_0 t + \theta) &= \tfrac{1}{2}\tilde{x}(t) + \tfrac{1}{2}\tilde{x}^*(t) \\ &= \tfrac{1}{2}Ae^{j(\omega_0 t+\theta)} + \tfrac{1}{2}Ae^{-j(\omega_0 t+\theta)} \end{aligned} \tag{2.6}$$

Figure 2.2 illustrates these two procedures graphically.

Equations (2.5) and (2.6), which give alternative representations of the sinusoidal signal $x(t) = A\cos(\omega_0 t + \theta)$ in terms of the rotating phasor $\tilde{x}(t) = A\exp[j(\omega_0 t + \theta)]$, are time-domain representations for $x(t)$. Two equivalent representations of $x(t)$ in the frequency domain may be obtained by noting that the rotating phasor signal is completely specified if the parameters A and θ are given for a particular f_0. Thus plots of the magnitude and angle of $Ae^{j\theta}$ versus frequency give sufficient information to characterize $\tilde{x}(t)$

* Euler's theorem, given in Appendix C, is $e^{\pm ju} = \cos u \pm j \sin u$. Also recall that $e^{j2\pi} = 1$.

FIGURE 2.2 Two ways of relating a phasor signal to a sinusoidal signal. (a) Projection of a rotating phasor onto the real axis. (b) Addition of complex conjugate rotating phasors.

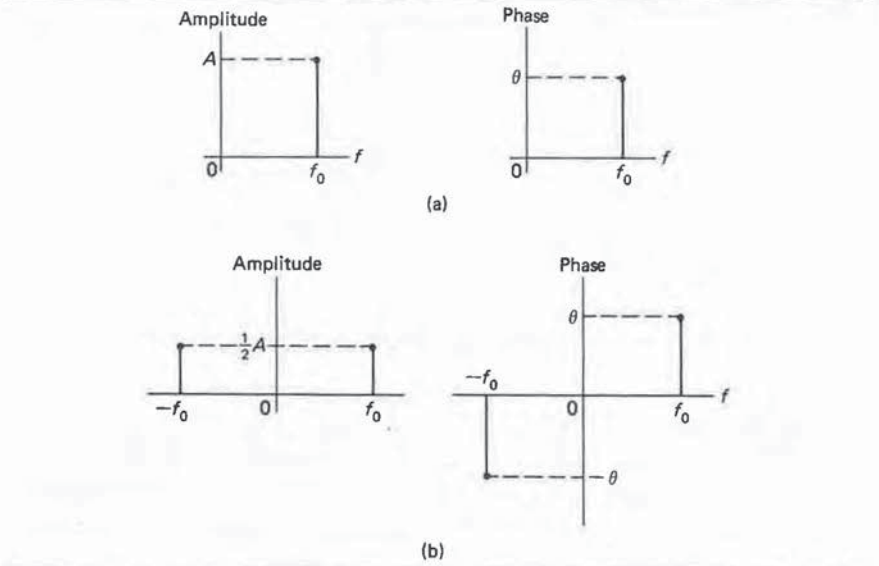


completely. Because $\tilde{x}(t)$ exists only at the single frequency f_0 for this case of a single sinusoidal signal, the resulting plots consist of single lines. Such plots, consisting of discrete lines and referred to as *amplitude and phase line spectra* for $\tilde{x}(t)$, are shown in Figure 2.3(a). These are *frequency-domain* representations not only of $\tilde{x}(t)$ but of $x(t)$ as well, by virtue of (2.5). The plots of Figure 2.3(a) are referred to as the *single-sided amplitude and phase spectra* of $x(t)$ because they exist only for positive frequency. For a signal consisting of a sum of sinusoids of differing frequencies, the single-sided spectra consist of a multiplicity of lines, one for each sinusoidal component of the sum.

By plotting the amplitude and phase of the complex conjugate phasors of (2.6) versus frequency, one obtains another frequency-domain representation for $x(t)$, referred to as the *double-sided amplitude and phase spectra*. This representation is shown in Figure 2.3(b). Two important observations may be made from Figure 2.3(b). First, the lines at the *negative* frequency $f = -f_0$ exist precisely because it is necessary to add complex conjugate (or oppositely rotating) phasor signals to obtain the real signal $A \cos(\omega_0 t + \theta)$. Second, we note that the amplitude spectrum has *even* symmetry about $f = 0$ and that the phase spectrum has *odd* symmetry. This symmetry is again a consequence of $x(t)$ being a real signal. As in the single-sided case, the two-sided spectra for a sum of sinusoids consist of a multiplicity of lines, with one pair for each sinusoidal component.

Figures 2.3(a) and 2.3(b) are therefore equivalent spectral representations for the signal $A \cos(\omega_0 t + \theta)$, consisting of lines at the frequency $f = f_0$ (and its negative). For this simple case, the use of spectral plots seems to be an unnecessary complication, but we will find shortly how the Fourier series and Fourier transform lead to spectral representations for more complex signals.

FIGURE 2.3 Amplitude and phase spectra for the signal $A \cos(\omega_0 t + \theta)$. (a) Single-sided. (b) Double-sided.



▼ **EXAMPLE 2.1** (a) To sketch the single-sided and double-sided spectra of

$$x(t) = 2 \sin(10\pi t - \tfrac{1}{6}\pi) \tag{2.7}$$

we note that $x(t)$ can be written as

$$\begin{aligned} x(t) &= 2 \cos(10\pi t - \tfrac{1}{6}\pi - \tfrac{1}{2}\pi) = 2 \cos(10\pi t - \tfrac{2}{3}\pi) \\ &= \text{Re}\, 2e^{j(10\pi t - 2\pi/3)} = e^{j(10\pi t - 2\pi/3)} + e^{-j(10\pi t - 2\pi/3)} \end{aligned} \tag{2.8}$$

Thus the single-sided and double-sided spectra are as shown in Figure 2.3, with $A = 2$, $\theta = -\frac{2}{3}\pi$ rad, and $f_0 = 5$ Hz.

(b) If more than one sinusoidal component is present in a signal, its spectra consist of multiple lines. For example, the signal

$$y(t) = 2 \sin(10\pi t - \tfrac{1}{6}\pi) + \cos(20\pi t) \tag{2.9}$$

can be rewritten as

$$\begin{aligned} y(t) &= 2 \cos(10\pi t - \tfrac{2}{3}\pi) + \cos(20\pi t) = \text{Re}\,[2e^{j(10\pi t - 2\pi/3)} + e^{j20\pi t}] \\ &= e^{j(10\pi t - 2\pi/3)} + e^{-j(10\pi t - 2\pi/3)} + \tfrac{1}{2}e^{j20\pi t} + \tfrac{1}{2}e^{-j20\pi t} \end{aligned} \tag{2.10}$$

Its single-sided amplitude spectrum consists of a line of amplitude 2 at $f = 5$ Hz and a line of amplitude 1 at $f = 10$ Hz. Its single-sided phase spectrum consists of a single line of amplitude $-2\pi/3$ at $f = 5$ Hz. To get the double-sided amplitude spectrum, one simply *halves* the amplitude of the lines in the single-sided amplitude spectrum and takes the mirror image of this result about $f = 0$. The double-sided phase spectrum is obtained by taking the mirror image of the single-sided phase spectrum about $f = 0$ and inverting the left-hand (negative frequency) portion. ▲

Singularity Functions

An important subclass of aperiodic signals is the singularity functions. In this book we will be concerned with only two: the unit impulse function $\delta(t)$ and the unit step function $u(t)$. The unit impulse function is *defined* in terms of the integral

$$\int_{-\infty}^{\infty} x(t)\,\delta(t)\,dt = x(0) \tag{2.11}$$

where $x(t)$ is any test function that is continuous at $t = 0$. A change of variables and redefinition of $x(t)$ results in the *sifting property*

$$\int_{-\infty}^{\infty} x(t)\,\delta(t - t_0)\,dt = x(t_0) \tag{2.12}$$

where $x(t)$ is continuous at $t = t_0$. We will make considerable use of the sifting property in systems analysis. By considering the special cases $x(t) = 1$ and $x(t) = 0$ for $t < t_1$ and $t > t_2$ the two properties

$$\int_{t_1}^{t_2} \delta(t - t_0)\,dt = 1, \qquad t_1 < t_0 < t_2 \tag{2.13a}$$

and

$$\delta(t - t_0) = 0, \qquad t \neq t_0 \tag{2.13b}$$

are obtained that provide an alternative definition of the unit impulse. Equation (2.13b) allows the integrand in Equation (2.12) to be replaced by $x(t_0)\,\delta(t - t_0)$, and the sifting property then follows from (2.13a).

Other properties of the delta function that can be proved from the definition (2.11) are the following:

1. $\delta(at) = \dfrac{1}{|a|}\,\delta(t)$
2. $\delta(-t) = \delta(t)$

3. $\int_{t_1}^{t_2} x(t)\,\delta(t - t_0)\,dt = \begin{cases} x(t_0), & t_1 < t_0 < t_2 \\ 0, & \text{otherwise} \end{cases}$
4. $x(t)\,\delta(t - t_0) = x(t_0)\,\delta(t - t_0)$
5. $\int_{t_1}^{t_2} x(t)\delta^{(n)}(t - t_0)\,dt = (-1)^n x^{(n)}(t_0), \qquad t_1 < t_0 < t_2$. [In this equation, the superscript n denotes the nth derivative; $x(t)$ and its first n derivatives are assumed continuous at $t = t_0$.]
6. If $f(t) = g(t)$, where $f(t) = a_0\,\delta(t) + a_1\,\delta^{(1)}(t) + \cdots + a_n\,\delta^{(n)}(t)$ and $g(t) = b_0\,\delta(t) + b_1\,\delta^{(1)}(t) + \cdots + b_n\,\delta^{(n)}(t)$, this implies that $a_0 = b_0$, $a_1 = b_1, \ldots, a_n = b_n$.

It is reassuring to note that (2.13a) and (2.13b) correspond to the intuitive notion of a unit impulse as the limit of a suitably chosen conventional function having unity area in an infinitesimally small width. An example is the signal

$$\delta_\epsilon(t) = \frac{1}{2\epsilon}\,\Pi\left(\frac{t}{2\epsilon}\right) = \begin{cases} \dfrac{1}{2\epsilon}, & |t| < \epsilon \\ 0, & \text{otherwise} \end{cases} \tag{2.14}$$

which is shown in Figure 2.4(a). It seems apparent that any signal having unity area and zero width in the limit as some parameter approaches zero is a suitable representation for $\delta(t)$, for example, the signal

$$\delta_{1\epsilon}(t) = \epsilon\left(\frac{1}{\pi t}\sin\frac{\pi t}{\epsilon}\right)^2 \tag{2.15}$$

which is sketched in Figure 2.4(b).

Other singularity functions may be defined as integrals or derivatives of unit impulses. We will need only the unit step $u(t)$, defined to be the integral of the unit impulse. Thus

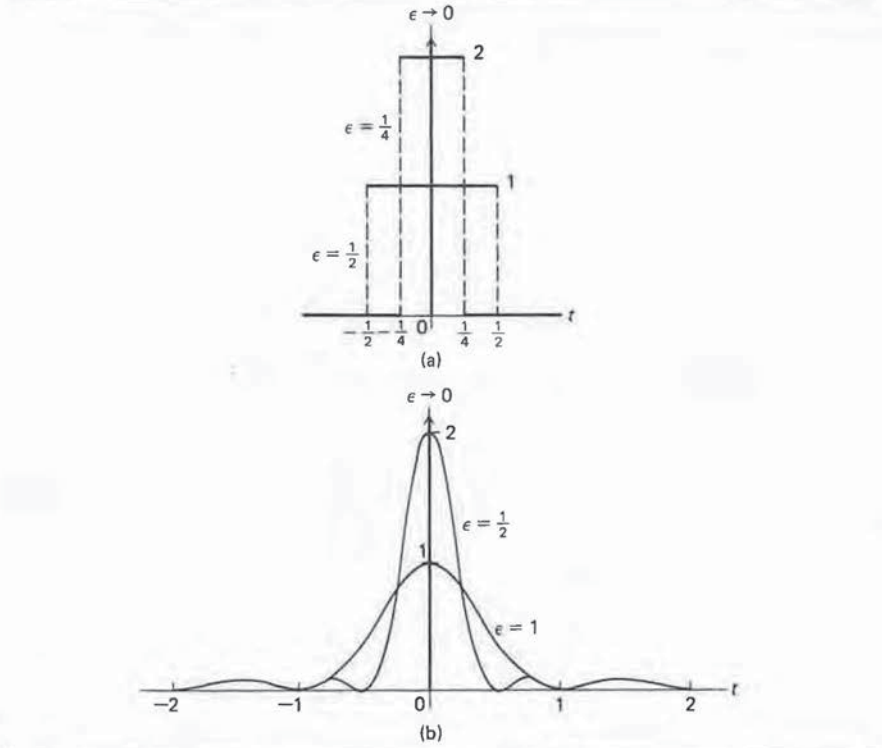
$$u(t) \triangleq \int_{-\infty}^{t} \delta(\lambda)\,d\lambda = \begin{cases} 0, & t < 0 \\ 1, & t > 0 \\ \text{undefined}, & t = 0 \end{cases} \tag{2.16}$$

or

$$\delta(t) = \frac{du(t)}{dt} \tag{2.17}$$

You are no doubt familiar with the usefulness of the unit step for "turning on" signals of doubly infinite duration and for representing signals of the

FIGURE 2.4 Two representations for the unit impulse function in the limit as $\epsilon \to 0$. (a) $(1/2\epsilon)\Pi(t/2\epsilon)$. (b) $\epsilon[(1/\pi t)\sin(\pi t/\epsilon)]^2$.



staircase type. For example, the unit rectangular pulse function defined by (2.2) can be written in terms of unit steps as

$$\Pi(t) = u(t + \tfrac{1}{2}) - u(t - \tfrac{1}{2}) \tag{2.18}$$

We are now ready to consider power and energy signal classifications.

2.2 SIGNAL CLASSIFICATIONS

Because the particular representation used for a signal depends on the type of signal involved, it is useful to pause at this point and introduce signal classifications. In this chapter we will be considering two signal classes, those with finite energy and those with finite power.* As a specific example, sup-

* Signals that are neither energy nor power signals are easily found. For example, $x(t) = t^{-1/4}$, $t \geq t_0 > 0$ and zero otherwise.

pose $e(t)$ is the voltage across a resistance R producing a current $i(t)$. The instantaneous power per ohm is $p(t) = e(t)i(t)/R = i^2(t)$. Integrating over the interval $|t| \leq T$, the total energy and the average power on a per-ohm basis are obtained as the limits

$$E = \lim_{T\to\infty} \int_{-T}^{T} i^2(t)\, dt \tag{2.19a}$$

and

$$P = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} i^2(t)\, dt \tag{2.19b}$$

respectively.

For an arbitrary signal $x(t)$, which may, in general, be complex, we define total (normalized) energy as

$$E \triangleq \lim_{T\to\infty} \int_{-T}^{T} |x(t)|^2\, dt = \int_{-\infty}^{\infty} |x(t)|^2\, dt \tag{2.20a}$$

and (normalized) power as

$$P \triangleq \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |x(t)|^2\, dt \tag{2.20b}$$

Based on the definitions (2.20a) and (2.20b), we can define two distinct classes of signals:

1. We say $x(t)$ is an *energy signal* if and only if $0 < E < \infty$, so that $P = 0$.
2. We classify $x(t)$ as a *power signal* if and only if $0 < P < \infty$, thus implying that $E = \infty$.

▼ **EXAMPLE 2.2** As an example of determining the classification of a signal, consider

$$x_1(t) = Ae^{-\alpha t}u(t), \qquad \alpha > 0 \tag{2.21}$$

where A and α are constants. Using (2.20a), we may readily verify that $x_1(t)$ is an energy signal, since $E = A^2/2\alpha$. Letting $\alpha \to 0$, we obtain the signal $x_2(t) = Au(t)$, which has infinite energy. Applying (2.20b), we find that $P = \frac{1}{2}A^2$, thus verifying that $x_2(t)$ is a power signal. ▲

▼ **EXAMPLE 2.3** Consider the rotating phasor signal given by Equation (2.4). We may verify that $\tilde{x}(t)$ is a power signal, since

$$P = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |\tilde{x}(t)|^2 \, dt = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} A^2 \, dt = A^2 \tag{2.22}$$

▲ is finite.

We note that there is no need to carry out the limiting operation to find P for a periodic signal, since an average carried out over a single period gives the same result as (2.20b); that is, for a periodic signal $x_p(t)$,

$$P = \frac{1}{T_0} \int_{t_0}^{t_0+T_0} |x_p(t)|^2 \, dt \tag{2.23}$$

where T_0 is the period. The proof of (2.23) is left to the problems.

▼ **EXAMPLE 2.4** The sinusoidal signal

$$x_p(t) = A \cos(\omega_0 t + \theta) \tag{2.24}$$

has average power

$$\begin{aligned} P &= \frac{1}{T_0} \int_{t_0}^{t_0+T_0} A^2 \cos^2(\omega_0 t + \theta) \, dt \\ &= \frac{\omega_0}{2\pi} \int_{t_0}^{t_0+(2\pi/\omega_0)} \frac{A^2}{2} \, dt + \frac{\omega_0}{2\pi} \int_{t_0}^{t_0+(2\pi/\omega_0)} \frac{A^2}{2} \cos 2(\omega_0 t + \theta) \, dt \\ &= \frac{A^2}{2} \end{aligned} \tag{2.25}$$

where the identity $\cos^2 u = \frac{1}{2} + \frac{1}{2} \cos 2u$ has been used and the second integral is zero because the integration is over two complete periods of the ▲ integrand.

†2.3 GENERALIZED FOURIER SERIES*

Our discussion of the phasor signal given by (2.4) illustrated the dual time-frequency nature of such signals. Fourier series and transform representations for signals are the key to generalizing this dual nature, since they amount to expressing signals as superpositions of complex exponential functions $e^{j\omega t}$.

* Sections preceded by a dagger can be omitted without loss of continuity.

In anticipation of signal space concepts, to be introduced and applied to communication systems analysis in Chapters 8 and 9, the discussion in this section is concerned with the representation of signals as series of orthogonal functions or, as referred to here, generalized Fourier series. Such generalized Fourier series representations allow signals to be represented as points in a generalized vector space, referred to as *signal space,* thereby allowing information transmission to be viewed in a geometrical context. In the following section, the generalized Fourier series will be specialized to the complex exponential form of the Fourier series.

To begin our consideration of generalized Fourier series, we recall from vector analysis that any vector **A** in a three-dimensional space can be expressed in terms of any three vectors **a**, **b**, and **c** that do not all lie in the same plane and are not collinear:

$$\mathbf{A} = A_1\mathbf{a} + A_2\mathbf{b} + A_3\mathbf{c} \tag{2.26}$$

where A_1, A_2, and A_3 are appropriately chosen constants. The vectors **a**, **b**, and **c** are said to be *linearly independent,* for no one of them can be expressed as a linear combination of the other two. For example, it is impossible to write $\mathbf{a} = \alpha\mathbf{b} + \beta\mathbf{c}$, no matter what choice is made for the constants α and β.

Such a set of linearly independent vectors is said to form a *basis set* for three-dimensional vector space. Such vectors *span* a three-dimensional vector space in the sense that *any* vector **A** can be expressed as a linear combination of them.

We may, in an analogous fashion, consider the problem of representing a time function, or signal, $x(t)$ on a T-second interval $(t_0, t_0 + T)$ as a similar expansion. Thus we consider a set of time functions $\phi_1(t), \phi_2(t), \ldots, \phi_N(t)$, which are specified independently of $x(t)$, and seek a series expansion of the form

$$x_a(t) = \sum_{n=1}^{N} X_n\phi_n(t), \qquad t_0 \leq t \leq t_0 + T \tag{2.27}$$

in which the N coefficients X_n are independent of time and the subscript a indicates that (2.27) is considered an approximation.

We assume that the $\phi_n(t)$'s in (2.27) are *linearly independent;* that is, no one of them can be expressed as a sum of the other $N - 1$. A set of linearly independent $\phi_n(t)$'s will be called a *basis function set.*

We now wish to examine the error in the approximation of $x(t)$ by $x_a(t)$. As in the case of ordinary vectors, the expansion (2.27) is easiest to use if

the $\phi_n(t)$'s are orthogonal on the interval $(t_0, t_0 + T)$. That is,

$$\int_{t_0}^{t_0+T} \phi_m(t)\phi_n^*(t)\,dt = c_n\delta_{mn} \triangleq \begin{cases} c_n, & n = m \\ 0, & n \neq m \end{cases} \qquad \text{(all } m \text{ and } n\text{)} \quad (2.28)$$

where, if $c_n = 1$ for all n, the $\phi_n(t)$'s are said to be *normalized*. A normalized orthogonal set of functions is called an *orthonormal basis set*. The asterisk in (2.28) denotes complex conjugate, since we wish to allow the possibility of complex-valued $\phi_n(t)$'s. The symbol δ_{mn}, called the *Kronecker delta function*, is defined as unity if $m = n$ and zero otherwise.

The error in the approximation of $x(t)$ by the series of (2.27) will be measured in the integral-squared sense:

$$\text{Error} = \epsilon_N = \int_T |x(t) - x_a(t)|^2\,dt \quad (2.29)$$

where $\int_T (\,)\,dt$ denotes integration over t from t_0 to $t_0 + T$. The *integral-squared error* (ISE) is an applicable measure of error only when $x(t)$ is an energy signal or a power signal. If $x(t)$ is an energy signal of infinite duration, the limit as $T \to \infty$ is taken.

We now find the set of coefficients X_n that minimizes the ISE. Substituting (2.27) in (2.29), expressing the magnitude squared of the integrand as the integrand times its complex conjugate and expanding, we obtain

$$\epsilon_N = \int_T |x(t)|^2\,dt - \sum_{n=1}^{N}\left[X_n^* \int_T x(t)\phi_n^*(t)\,dt + X_n \int_T x^*(t)\phi_n(t)\,dt\right] + \sum_{n=1}^{N} c_n|X_n|^2 \quad (2.30)$$

in which the orthogonality of the $\phi_n(t)$'s has been used after interchanging the orders of summation and integration. The details are left to the problems. To find the X_n's that minimize ϵ_N, we add and subtract the quantity

$$\sum_{n=1}^{N} \frac{1}{c_n}\left|\int_T x(t)\phi_n^*(t)\,dt\right|^2$$

which yields, after rearrangement of terms, the following result for ϵ_N:

$$\epsilon_N = \int_T |x(t)|^2\,dt - \sum_{n=1}^{N} \frac{1}{c_n}\left|\int_T x(t)\phi_n^*(t)\,dt\right|^2 + \sum_{n=1}^{N} c_n\left|X_n - \frac{1}{c_n}\int_T x(t)\phi_n^*(t)\,dt\right|^2 \quad (2.31)$$

The first two terms on the right-hand side of (2.31) are independent of the coefficients X_n. Since the last sum on the right-hand side is nonnegative, we will minimize ϵ_N if we choose each X_n such that the corresponding term in the sum is zero. Thus, since $c_n > 0$, the choice of

$$X_n = \frac{1}{c_n} \int_T x(t)\phi_n^*(t)\, dt \tag{2.32}$$

for X_n minimizes the ISE. The resulting minimum-error coefficients will be referred to as the *Fourier coefficients*.

The minimum value for ϵ_N, from (2.31), is obviously

$$\begin{aligned}(\epsilon_N)_{min} &= \int_T |x(t)|^2\, dt - \sum_{n=1}^{N} \frac{1}{c_n} \left| \int^T x(t)\phi_n^*(t)\, dt \right|^2 \\ &= \int_T |x(t)|^2\, dt - \sum_{n=1}^{N} c_n |X_n|^2 \end{aligned} \tag{2.33}$$

If we can find an infinite set of orthonormal functions such that

$$\lim_{N\to\infty} (\epsilon_N)_{min} = 0 \tag{2.34}$$

for any signal that is integrable square,

$$\int_T |x(t)|^2\, dt < \infty \tag{2.35}$$

we say that the $\phi_n(t)$'s are *complete*. In the sense that the ISE is zero, we may then write

$$x(t) = \sum_{n=1}^{\infty} X_n \phi_n(t) \qquad (\text{ISE} = 0) \tag{2.36}$$

although there may be a number of isolated points of discontinuity where actual equality does not hold. For almost all points in the interval $(t_0, t_0 + T)$, Equation (2.34) requires that $x(t)$ be equal to $x_a(t)$ as $N \to \infty$.

Assuming a complete orthogonal set of functions, we obtain from (2.33) the relation

$$\int_T |x(t)|^2\, dt = \sum_{n=1}^{\infty} c_n |X_n|^2 \tag{2.37}$$

This equation is known as *Parseval's theorem*.

▼ **EXAMPLE 2.5** Consider the set of two orthonormal functions shown in Figure 2.5(a). The signal

$$x(t) = \begin{cases} \sin \pi t, & 0 \leq t \leq 2 \\ 0, & \text{otherwise} \end{cases} \tag{2.38}$$

is to be approximated by a two-term generalized Fourier series of the form given by (2.27). The Fourier coefficients, from (2.32), are given by

$$X_1 = \int_0^2 \phi_1(t) \sin \pi t \, dt = \int_0^1 \sin \pi t \, dt = \frac{2}{\pi} \tag{2.39a}$$

$$X_2 = \int_0^2 \phi_2(t) \sin \pi t \, dt = \int_1^2 \sin \pi t \, dt = -\frac{2}{\pi} \tag{2.39b}$$

Thus the generalized two-term Fourier series approximation for this signal is

$$x_a(t) = \frac{2}{\pi} \phi_1(t) - \frac{2}{\pi} \phi_2(t) = \frac{2}{\pi} \left[\Pi\left(t - \frac{1}{2}\right) - \Pi\left(t - \frac{3}{2}\right) \right] \tag{2.40}$$

where $\Pi(t)$ is the unit rectangular pulse defined by (2.1). The signal $x(t)$ and the approximation $x_a(t)$ are compared in Figure 2.5(b). Figure 2.5(c) emphasizes the signal space interpretation of $x_a(t)$ by representing it as the point $(2/\pi, -2/\pi)$ in the two-dimensional space spanned by the orthonormal functions $\phi_1(t)$ and $\phi_2(t)$. Representation of an arbitrary $x(t)$ exactly ($\epsilon_N = 0$) would require an infinite set of properly chosen orthogonal functions (that is, a complete set).

The minimum ISE, from (2.33), is

$$(\epsilon_2)_{\min} = \int_0^2 \sin^2 \pi t \, dt - 2\left(\frac{2}{\pi}\right)^2 = 1 - \frac{8}{\pi^2} \cong 0.189 \tag{2.41}$$

▲

2.4 FOURIER SERIES

Complex Exponential Fourier Series

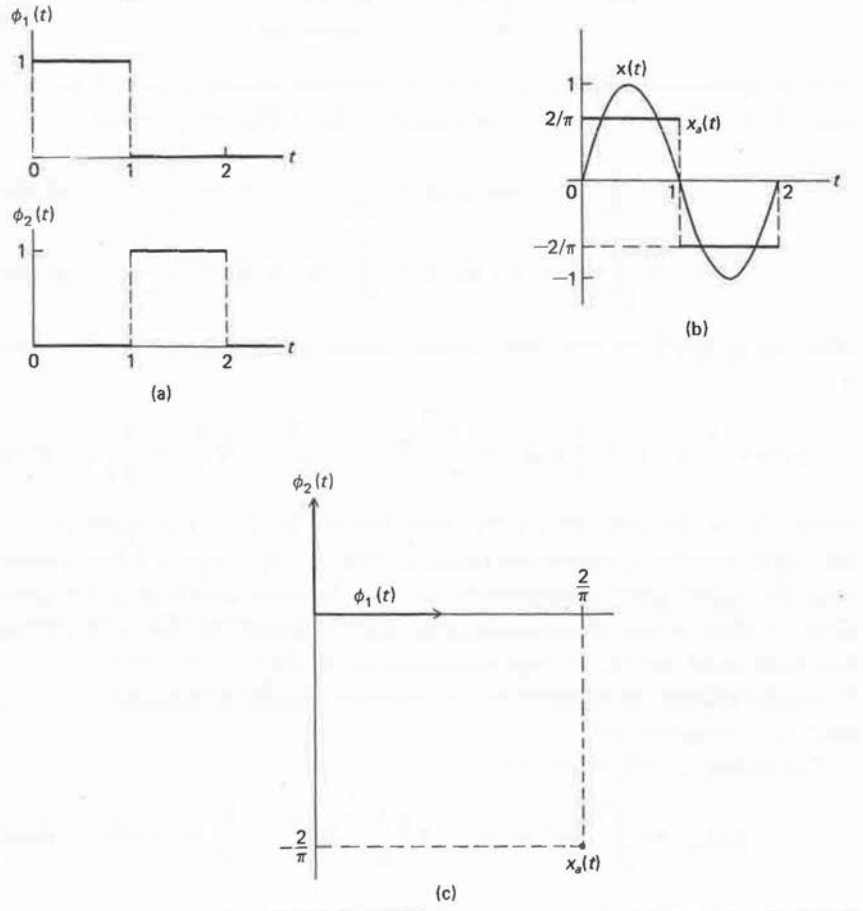
Given a signal $x(t)$ defined over the interval $(t_0, t_0 + T_0)$ with the definition

$$\omega_0 = 2\pi f_0 = \frac{2\pi}{T_0} \tag{2.42}$$

we define the *complex exponential Fourier series* as

$$x(t) = \sum_{n=-\infty}^{\infty} X_n e^{jn\omega_0 t}, \qquad t_0 \leq t \leq t_0 + T_0 \tag{2.43}$$

FIGURE 2.5 Approximation of a sinewave pulse with a generalized Fourier series. (a) Orthonormal functions. (b) Sinewave and approximation. (c) Signal space representation.



where

$$X_n = \frac{1}{T_0}\int_{t_0}^{t_0+T_0} x(t)e^{-jn\omega_0 t}\,dt \tag{2.44}$$

It can be shown to represent the signal $x(t)$ exactly in the interval $(t_0, t_0 + T_0)$, except at a point of jump discontinuity where it converges to the arithmetic mean of the left- and right-hand limits.* Outside the interval $(t_0,$

* Dirichlet's conditions state that sufficient conditions for convergence are that $x(t)$ be defined and bounded on the range $(t_0, t_0 + T_0)$ and have only a finite number of maxima and minima and a finite number of discontinuities on this range.

$t_0 + T_0$), of course, nothing is guaranteed. However, we note that the right-hand side of (2.43) is periodic with period T_0, since it is the sum of periodic rotating phasors with harmonic frequencies. Thus, if $x(t)$ is periodic with period T_0, the Fourier series of (2.43) is an accurate representation for $x(t)$ for *all* t (except at points of discontinuity). The integration of (2.44) can then be taken over any period.

A useful observation about a complete orthonormal-series expansion of a signal is that the series is *unique*. For example, if we somehow find a Fourier expansion for a signal $x(t)$, we know that no other Fourier expansion for that $x(t)$ exists, since $\{e^{jn\omega_0 t}\}$ forms a complete set. The usefulness of this observation is illustrated with the following example.

▼ **EXAMPLE 2.6** Consider the signal

$$x(t) = \cos \omega_0 t + \sin^2 2\omega_0 t \tag{2.45}$$

where $\omega_0 = 2\pi/T_0$. Find the complex exponential Fourier series.

SOLUTION We could compute the Fourier coefficients using (2.44), but by using appropriate trigonometric identities and Euler's theorem, we obtain

$$\begin{aligned} x(t) &= \cos \omega_0 t + \tfrac{1}{2} - \tfrac{1}{2} \cos 4\omega_0 t \\ &= \tfrac{1}{2}e^{j\omega_0 t} + \tfrac{1}{2}e^{-j\omega_0 t} + \tfrac{1}{2} - \tfrac{1}{4}e^{j4\omega_0 t} - \tfrac{1}{4}e^{-j4\omega_0 t} \end{aligned} \tag{2.46}$$

Invoking uniqueness and equating the second line term by term with $\sum_{n=-\infty}^{\infty} X_n e^{jn\omega_0 t}$, we find that

$$\begin{aligned} X_0 &= \tfrac{1}{2} \\ X_1 &= \tfrac{1}{2} = X_{-1} \\ X_4 &= -\tfrac{1}{4} = X_{-4} \end{aligned} \tag{2.47}$$

with all other X_n's equal to zero. Thus considerable labor is saved by noting that the Fourier series of a signal is unique. ▲

The Fourier Coefficients as Time Averages and Their Symmetry Properties

The expression for X_n (2.44) can be interpreted as the time average of $x(t)e^{-jn\omega_0 t}$, where, for future convenience, we define the *time average* of a

signal $v(t)$ as

$$\langle v(t)\rangle = \begin{cases} \lim_{T\to\infty} \dfrac{1}{2T}\displaystyle\int_{-T}^{T} v(t)\,dt & [v(t) \text{ not periodic}] \\ \dfrac{1}{T_0}\displaystyle\int_{T_0} v(t)\,dt & [v(t) \text{ periodic}] \end{cases} \tag{2.48a}$$

(2.48b)

where $\int_{T_0}$ indicates integration over any period. For $n = 0$, we see that $X_0 = \langle x(t)\rangle$ is the average, or dc, value of $x(t)$.

Using Euler's theorem, we can write (2.44) as

$$X_n = \langle x(t)\cos n\omega_0 t\rangle - j\langle x(t)\sin n\omega_0 t\rangle \tag{2.49}$$

and assuming $x(t)$ is real, it follows that

$$X_n^* = X_{-n} \tag{2.50}$$

by replacing n by $-n$ in (2.49). Writing X_n as

$$X_n = |X_n| e^{j\underline{/X_n}} \tag{2.51}$$

we obtain

$$|X_n| = |X_{-n}| \qquad \text{and} \qquad \underline{/X_n} = -\underline{/X_{-n}} \tag{2.52}$$

Thus, for real signals, the magnitude of the Fourier coefficients is an even function of n, and the argument is odd.

Several symmetry properties can be derived for the Fourier coefficients, depending on the symmetry of $x(t)$. For example, suppose $x(t)$ is even; that is, $x(t) = x(-t)$. Then the second term in (2.49) is zero, since $x(t)\sin n\omega_0 t$ is an odd function. Thus X_n is purely real, and furthermore, X_n is an even function of n since $\cos n\omega_0 t$ is an even function of n. These consequences of $x(t)$ being even are illustrated by Example 2.6.

On the other hand, if $x(t) = -x(-t)$ [that is, $x(t)$ is odd], it readily follows that X_n is purely imaginary, since $\langle x(t)\cos n\omega_0 t\rangle = 0$ by virtue of $x(t)\cos n\omega_0 t$ being odd. In addition, X_n is an odd function of n, since $\sin n\omega_0 t$ is an odd function of n.

Another type of symmetry is *halfwave symmetry,* defined as

$$x(t \pm \tfrac{1}{2}T_0) = -x(t) \tag{2.53}$$

where T_0 is the period of $x(t)$. For signals with halfwave symmetry,

$$X_n = 0, \qquad n = 0, \pm 2, \pm 4, \ldots \tag{2.54}$$

which states that the Fourier series for such a signal consists only of odd-indexed terms. The proof of this is left to the problems.

Trigonometric Form of the Fourier Series

Using (2.52) and assuming $x(t)$ real, we can regroup the complex exponential Fourier series by pairs of terms of the form

$$\begin{aligned} X_n e^{jn\omega_0 t} + X_{-n} e^{-jn\omega_0 t} &= |X_n| e^{j(n\omega_0 t + \underline{/X_n})} + |X_n| e^{-j(n\omega_0 t + \underline{/X_n})} \\ &= 2|X_n| \cos(n\omega_0 t + \underline{/X_n}) \end{aligned} \tag{2.55}$$

Hence (2.43) can be written in the equivalent trigonometric form:

$$x(t) = X_0 + \sum_{n=1}^{\infty} 2|X_n| \cos(n\omega_0 t + \underline{/X_n}) \tag{2.56}$$

Expanding the cosine in (2.56), we obtain still another equivalent series of the form

$$x(t) = X_0 + \sum_{n=1}^{\infty} A_n \cos n\omega_0 t + \sum_{n=1}^{\infty} B_n \sin n\omega_0 t \tag{2.57}$$

where

$$\begin{aligned} A_n &= 2|X_n| \cos \underline{/X_n} \\ &= 2\langle x(t) \cos n\omega_0 t \rangle \\ &= \frac{2}{T_0} \int_{t_0}^{t_0+T_0} x(t) \cos(n\omega_0 t)\, dt \end{aligned} \tag{2.58}$$

and

$$\begin{aligned} B_n &= -2|X_n| \sin \underline{/X_n} \\ &= 2\langle x(t) \sin n\omega_0 t \rangle \\ &= \frac{2}{T_0} \int_{t_0}^{t_0+T_0} x(t) \sin(n\omega_0 t)\, dt \end{aligned} \tag{2.59}$$

In either the trigonometric or the exponential form of the Fourier series, X_0 represents the average or dc component of $x(t)$. The term for $n = 1$ is called the *fundamental*, the term for $n = 2$ is called the *second harmonic*, and so on.

Parseval's Theorem

Using (2.23) for average power of a periodic signal, substituting (2.43) for $x(t)$, and interchanging the order of integration and summation, we find Parseval's theorem to be

$$P = \frac{1}{T_0}\int_{T_0} |x(t)|^2\, dt = \sum_{n=-\infty}^{\infty} |X_n|^2 \tag{2.60a}$$

$$= X_0^2 + 2\sum_{n=1}^{\infty} |X_n|^2 \tag{2.60b}$$

which is a specialization of (2.37). In words, (2.60a) simply states that the average power of a periodic signal $x(t)$ is the sum of the powers in the phasor components of its Fourier series.

Examples of Fourier Series—The Sinc Function

Table 2.1 gives Fourier series for several commonly occurring periodic waveforms. The left-hand column specifies the signal over one period. The definition of periodicity,

$$x(t) = x(t + T_0)$$

specifies it for all t. The derivation of the Fourier coefficients given in the right-hand column of Table 2.1 is left to the problems. Note that the full-rectified sinewave actually has the period $\frac{1}{2}T_0$.

For the periodic pulse train, it is convenient to express the coefficients in terms of the *sinc function,* defined as

$$\text{sinc } z = \frac{\sin \pi z}{\pi z} \tag{2.61}$$

The sinc function will arise often in our discussion. It is shown in Figure 2.6 along with $\text{sinc}^2 z$.

▼ **EXAMPLE 2.7** Specialize the results of Table 2.1 to the complex exponential and trigonometric Fourier series of a squarewave with even symmetry and amplitudes zero and A.

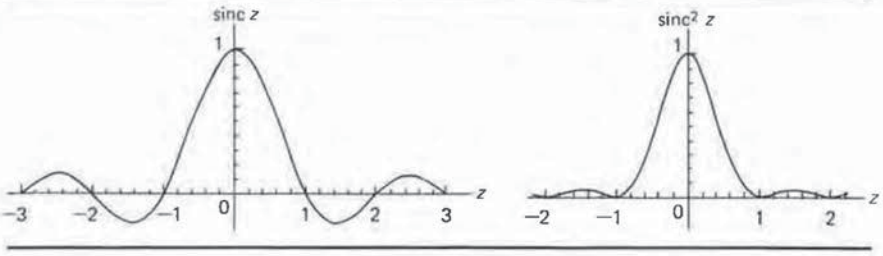
SOLUTION The solution proceeds by letting $t_0 = 0$ and $\tau = \frac{1}{2}T_0$ in item 1 of Table 2.1. Thus

$$X_n = \tfrac{1}{2}A \text{ sinc } (\tfrac{1}{2}n) \tag{2.62}$$

TABLE 2.1 Fourier Series for Several Periodic Signals

Signal: $-\frac{1}{2}T_0 \le t \le \frac{1}{2}T_0$ (One Period)	Coefficients for Exponential Fourier Series
1. Asymmetrical pulse train:	
$x(t) = A\Pi\left(\frac{t - t_0}{\tau}\right), \tau < T_0$ $x(t) = x(t + T_0)$, all t	$X_n = \frac{A\tau}{T_0}\,\text{sinc}\,(nf_0\tau)e^{-j2\pi nf_0t_0}$ $n = 0, \pm1, \pm2, \ldots$
2. Half-rectified sinewave:	
$x(t) = \begin{cases} A \sin \omega_0 t, & 0 \le t \le \frac{1}{2}T_0 \\ 0, & -\frac{1}{2}T_0 \le t \le 0 \end{cases}$ $x(t) = x(t + T_0)$, all t	$X_n = \begin{cases} \frac{A}{\pi(1 - n^2)}, & n = 0, \pm2, \pm4, \ldots \\ 0, & n = \pm3, \pm5, \ldots \\ -\frac{1}{4}jA, & n = 1 \\ \frac{1}{4}jA, & n = -1 \end{cases}$
3. Full-rectified sinewave:	
$x(t) = A\lvert\sin \omega_0 t\rvert$	$X_n = \begin{cases} \frac{2A}{\pi(1 - n^2)}, & n = 0, \pm2, \pm4, \ldots \\ 0, & n = \pm1, \pm3, \ldots \end{cases}$
4. Triangular wave:	
$x(t) = \begin{cases} \frac{4A}{T_0}t + A, & -\frac{1}{2}T_0 \le t < 0 \\ -\frac{4A}{T_0}t + A, & 0 \le t \le \frac{1}{2}T_0 \end{cases}$ $x(t) = x(t + T_0)$, all t	$X_n = \begin{cases} \frac{4A}{\pi^2n^2}, & n \text{ odd} \\ 0, & n \text{ even} \end{cases}$

FIGURE 2.6 Graphs of sinc z and sinc^2 z



But

$$\operatorname{sinc}\left(\tfrac{1}{2}n\right) = \frac{\sin\left(\frac{1}{2}n\pi\right)}{\frac{1}{2}n\pi} \tag{2.63}$$

$$= \begin{cases} 1, & n = 0 \\ 0, & n = \text{even} \\ |2/n\pi|, & n = \pm 1, \pm 5, \pm 9, \ldots \\ -|2/n\pi|, & n = \pm 3, \pm 7, \ldots \end{cases}$$

Thus

$$x(t) = \cdots \frac{A}{5\pi} e^{-j5\omega_0 t} - \frac{A}{3\pi} e^{-j3\omega_0 t} + \frac{A}{\pi} e^{-j\omega_0 t}$$
$$+ \frac{A}{2} + \frac{A}{\pi} e^{j\omega_0 t} - \frac{A}{3\pi} e^{j3\omega_0 t} + \frac{A}{5\pi} e^{j5\omega_0 t} - \cdots$$
$$= \frac{A}{2} + \frac{2A}{\pi}\left(\cos \omega_0 t - \frac{1}{3}\cos 3\omega_0 t + \frac{1}{5}\cos 5\omega_0 t - \cdots\right) \tag{2.64}$$

The first equation is the complex exponential form of the Fourier series, and the second equation is the trigonometric form. The dc component of this squarewave is $X_0 = \frac{1}{2}A$. Setting this term to zero in the preceding Fourier series, we have the Fourier series of a squarewave of amplitudes $\pm\frac{1}{2}A$. Such a squarewave has halfwave symmetry, and this is precisely the reason that no even harmonics are present in its Fourier series. ▲

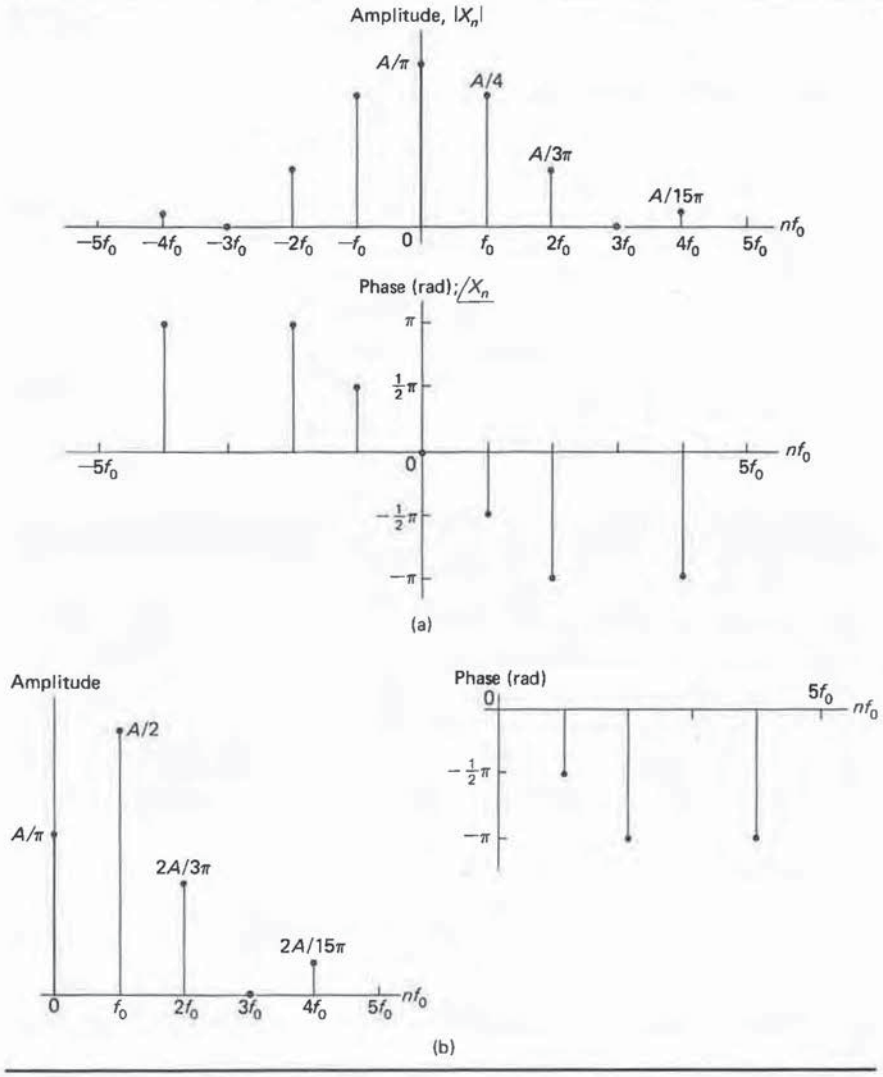
Line Spectra

The complex exponential Fourier series (2.43) of a signal is simply a summation of phasors. In Section 2.1 we showed how a phasor could be characterized in the frequency domain by two plots: one showing its amplitude versus frequency and one showing its phase. Similarly, a periodic signal can be characterized in the frequency domain by making two plots: one showing amplitudes of the separate phasor components versus frequency and the other showing their phases versus frequency. The resulting plots are called the *two-sided amplitude* and phase spectra,* respectively, of the signal. From (2.52) it follows that, for a real signal, the amplitude spectrum is even and the phase spectrum is odd, which is simply a result of the addition of complex conjugate phasors to get a real sinusoidal signal.

Figure 2.7(a) shows the double-sided spectrum for a half-rectified sine-

* *Magnitude spectrum* would be a more accurate term, although *amplitude spectrum* is the customary term.

FIGURE 2.7 Line spectra for half-rectified sinewave. (a) Double-sided. (b) Single-sided.



wave as plotted from the results given in Table 2.1. For $n = 2, 4, \ldots,$ X_n is represented as follows:

$$X_n = -\left|\frac{A}{\pi(1 - n^2)}\right| = \frac{A}{\pi(n^2 - 1)} e^{-j\pi} \qquad (2.65a)$$

For $n = -2, -4, \ldots$, it is represented as

$$X_n = -\left|\frac{A}{\pi(1 - n^2)}\right| = \frac{A}{\pi(n^2 - 1)} e^{j\pi} \tag{2.65b}$$

to ensure that the phase is odd, as it must be. (Note that $e^{\pm j\pi} = -1$.) Thus

$$|X_n| = \begin{cases} \frac{1}{4}A, & n = \pm 1 \\ \left|\dfrac{A}{\pi(1 - n^2)}\right|, & \text{all even } n \end{cases} \tag{2.66a}$$

$$\underline{/X_n} = \begin{cases} -\pi, & n = 2, 4, \ldots \\ -\frac{1}{2}\pi, & n = 1 \\ 0, & n = 0 \\ \frac{1}{2}\pi, & n = -1 \\ \pi, & n = -2, -4, \ldots \end{cases} \tag{2.66b}$$

The *single-sided line spectra* are obtained by plotting the amplitudes and phase angles of the terms in the trigonometric Fourier series (2.56) versus nf_0. Because the series (2.56) has only nonnegative frequency terms, the single-sided spectra exist only for $nf_0 \geq 0$. From (2.56) it is readily apparent that the single-sided phase spectrum of a periodic signal is identical to its double-sided phase spectrum for $nf_0 \geq 0$ and zero for $nf_0 > 0$. The single-sided amplitude spectrum is obtained from the double-sided amplitude spectrum by doubling the amplitude of all lines for $nf_0 > 0$. The line at $nf_0 = 0$ stays the same. The single-sided spectra for the half-rectified sinewave are shown in Figure 2.7(b).

As a second example, consider the pulse train

$$x(t) = \sum_{n=-\infty}^{\infty} A\Pi\left(\frac{t - nT_0 - \frac{1}{2}\tau}{\tau}\right) \tag{2.67}$$

From Table 2.1, with $t_0 = \frac{1}{2}\tau$ substituted in item 1, the Fourier coefficients are

$$X_n = \frac{A\tau}{T_0} \operatorname{sinc}(nf_0\tau) e^{-j\pi nf_0\tau} \tag{2.68}$$

The Fourier coefficients can be put in the form $|X_n| \exp(j\underline{/X_n})$, where

$$|X_n| = \frac{A\tau}{T_0} |\operatorname{sinc}(nf_0\tau)| \tag{2.69a}$$

and

$$\underline{/X_n} = \begin{cases} -\pi n f_0 \tau & \text{if} \quad \text{sinc } (nf_0\tau) > 0 \\ -\pi n f_0 \tau + \pi & \text{if} \quad nf_0 > 0 \quad \text{and} \quad \text{sinc } (nf_0\tau) < 0 \\ -\pi n f_0 \tau - \pi & \text{if} \quad nf_0 < 0 \quad \text{and} \quad \text{sinc } (nf_0\tau) < 0 \end{cases} \tag{2.69b}$$

The $\pm\pi$ on the right-hand side of (2.69b) accounts for $|\text{sinc } nf_0\tau| = -\text{sinc } nf_0\tau$ whenever sinc $nf_0\tau < 0$. Since the phase spectrum must have odd symmetry if $x(t)$ is real, π is subtracted if $nf_0 < 0$ and added if $nf_0 > 0$. The reverse could have been done—the choice is arbitrary. With these considerations, the double-sided amplitude and phase spectra can now be plotted. They are shown in Figure 2.8 for several choices of τ and T_0. Note that appropriate multiples of 2π are subtracted from the lines in the phase spectrum.

Comparing Figures 2.8(a) and 2.8(b), we note that the zeros of the envelope of the amplitude spectrum, which occur at multiples of $1/\tau$ Hz, move out along the frequency axis as the pulse width decreases. That is, *the time duration of a signal and its spectral width are inversely proportional,* a property that will be shown to be true in general later. Second, comparing Figures 2.8(a) and 2.8(c), we note that the separation between lines in the spectra is $1/T_0$. Thus the density of the spectral lines with frequency increases as the period of $x(t)$ increases.

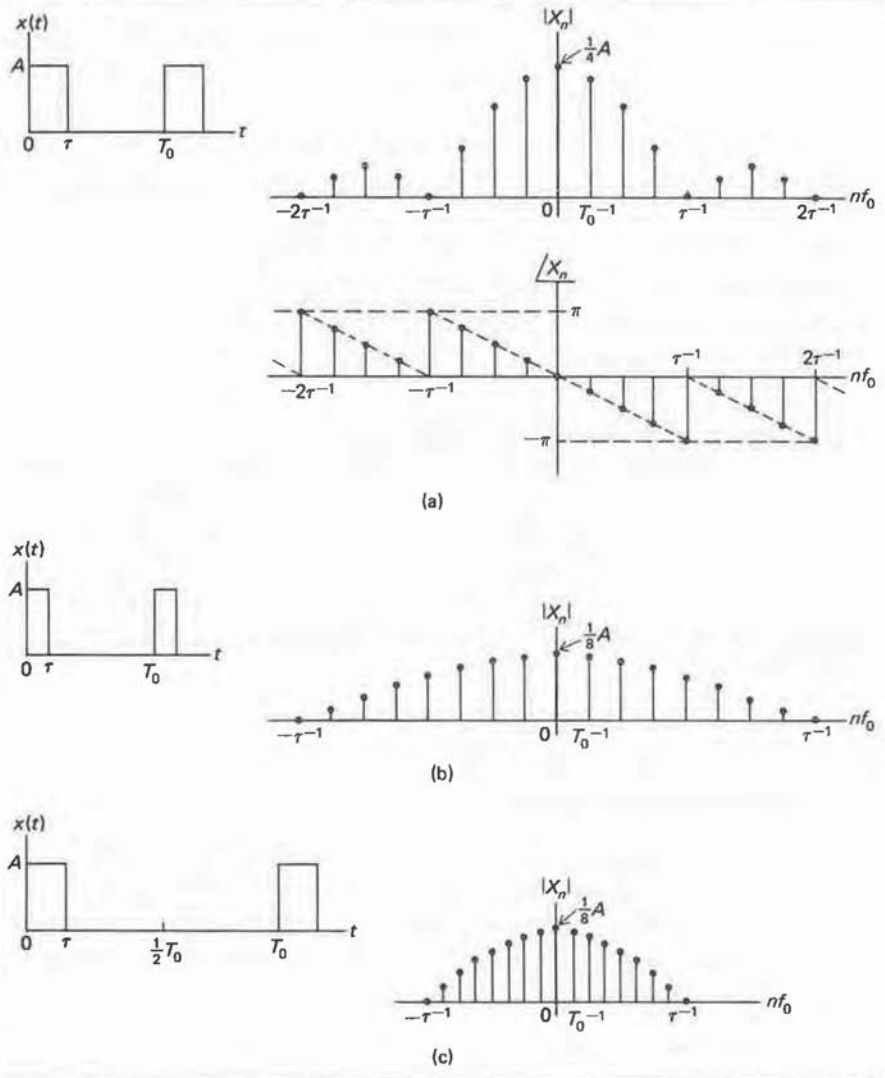
2.5 THE FOURIER TRANSFORM

To generalize the Fourier series representation (2.43) to a representation valid for aperiodic signals, we consider the two basic relationships (2.43) and (2.44). Suppose that $x(t)$ is nonperiodic but is an energy signal, so that it is integrable square in the interval $(-\infty, \infty)$.* In the interval $|t| < \frac{1}{2}T_0$, we can represent $x(t)$ as the Fourier series

$$x(t) = \sum_{n=-\infty}^{\infty} \left[\frac{1}{T_0} \int_{-T_0/2}^{T_0/2} x(\lambda) e^{-j2\pi n f_0 \lambda} \, d\lambda \right] e^{j2\pi n f_0 t}, \qquad |t| < \frac{T_0}{2} \tag{2.70}$$

* This means that $x(t)$ should be an energy signal. Dirichlet's conditions give sufficient conditions for a signal to have a Fourier transform. They are that $x(t)$ be (1) single-valued with a finite number of maxima and minima and a finite number of discontinuities in any finite time interval, and (2) absolutely integrable, that is, $\int_{-\infty}^{\infty} |x(t)| \, dt < \infty$. These conditions include all energy signals.

FIGURE 2.8 Spectra for a periodic pulse train signal. (a) $\tau = \frac{1}{4}T_0$. (b) $\tau = \frac{1}{8}T_0$; T_0 same as in (a). (c) $\tau = \frac{1}{8}T_0$; τ same as in (a).



where $f_0 = 1/T_0$. To represent $x(t)$ for all time, we simply let $T_0 \to \infty$ such that $nf_0 = n/T_0$ becomes the continuous variable f, $1/T_0$ becomes the differential df, and the summation becomes an integral. Thus

$$x(t) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} x(\lambda) e^{-j2\pi f\lambda}\, d\lambda \right] e^{j2\pi ft}\, df \tag{2.71}$$

Defining the inside integral as

$$X(f) = \int_{-\infty}^{\infty} x(\lambda) e^{-j2\pi f\lambda}\, d\lambda \tag{2.72}$$

we can write (2.71) as

$$x(t) = \int_{-\infty}^{\infty} X(f) e^{j2\pi ft}\, df \tag{2.73}$$

The existence of these integrals is assured, since $x(t)$ is an energy signal. We note that

$$X(f) = \lim_{T_0 \to \infty} T_0 X_n \tag{2.74}$$

which avoids the problem that $|X_n| \to 0$ as $T_0 \to \infty$.

The frequency-domain description of $x(t)$ provided by (2.72) is referred to as the *Fourier transform* of $x(t)$, written symbolically as $X(f) = \mathcal{F}[x(t)]$. Conversion back to the time domain is achieved via the *inverse Fourier transform* (2.73), written symbolically as $x(t) = \mathcal{F}^{-1}[X(f)]$.

Expressing (2.72) and (2.73) in terms of $f = \omega/2\pi$ results in easily remembered symmetrical expressions. Integrating (2.73) with respect to the variable ω requires a factor of $(2\pi)^{-1}$.

Amplitude and Phase Spectra

Writing $X(f)$ in terms of amplitude and phase as

$$X(f) = |X(f)| e^{j\theta(f)}, \qquad \theta(f) = \underline{/X(f)} \tag{2.75}$$

we can show, for real $x(t)$, that

$$|X(f)| = |X(-f)| \quad \text{and} \quad \theta(f) = -\theta(-f) \tag{2.76}$$

just as for Fourier series. This is done by using Euler's theorem to write

$$R = \text{Re}\, X(f) = \int_{-\infty}^{\infty} x(t) \cos 2\pi ft\, dt \tag{2.77a}$$

and

$$I = \text{Im}\, X(f) = -\int_{-\infty}^{\infty} x(t) \sin 2\pi ft\, dt \tag{2.77b}$$

Thus the real part of $X(f)$ is even and the imaginary part is odd if $x(t)$ is a real signal. Since $|X(f)|^2 = R^2 + I^2$ and $\tan\theta(f) = +I/R$, the symmetry

properties (2.76) follow. A plot of $|X(f)|$ versus f is referred to as the *amplitude spectrum** of $x(t)$, and a plot of $\underline{/X(f)} = \theta(f)$ versus f is known as the *phase spectrum*.

Symmetry Properties

If $x(t) = x(-t)$, that is, if $x(t)$ is even, then $x(t) \sin 2\pi ft$ is odd and Im $X(f) = 0$. Furthermore, Re $X(f)$ is an even function of f. Thus the Fourier transform of a real, even function is real and even.

On the other hand, if $x(t)$ is odd, $x(t) \cos 2\pi ft$ is odd and Re $X(f) = 0$. Thus the Fourier transform of a real, odd function is imaginary. In addition, Im $X(f)$ is an odd function of frequency because $\sin 2\pi ft$ is an odd function.

▼ **EXAMPLE 2.8** Consider the pulse

$$x(t) = A\Pi\left(\frac{t - t_0}{\tau}\right) \tag{2.78}$$

Its Fourier transform is

$$\begin{aligned} X(f) &= \int_{-\infty}^{\infty} A\Pi\left(\frac{t - t_0}{\tau}\right) e^{-j2\pi ft}\, dt \\ &= A\int_{t_0-\tau/2}^{t_0+\tau/2} e^{-j2\pi ft}\, dt = A\tau\ \text{sinc}\, f\tau\ e^{-j2\pi ft_0} \end{aligned} \tag{2.79}$$

Its amplitude spectrum is

$$|X(f)| = A\tau\ |\text{sinc}\, f\tau| \tag{2.80a}$$

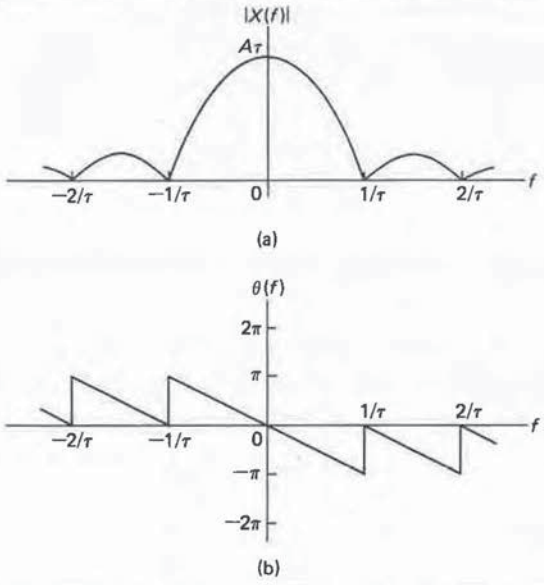
and its phase spectrum is

$$\theta(f) = \begin{cases} -2\pi t_0 f & \text{if} \quad \text{sinc}\, f\tau > 0 \\ -2\pi t_0 f \pm \pi & \text{if} \quad \text{sinc}\, f\tau < 0 \end{cases} \tag{2.80b}$$

The term $\pm\pi$ is used to account for sinc $f\tau$ being negative, and if $+\pi$ is used for $f > 0$, $-\pi$ is used for $f < 0$, or vice versa, to ensure that $\theta(f)$ is odd. When $|\theta(f)|$ exceeds 2π, an appropriate multiple of 2π may be subtracted from $\theta(f)$. Figure 2.9 shows the amplitude and phase spectra for this signal. The similarity to Figure 2.8 is to be noted, especially the *inverse relationship between spectral width and pulse duration.* ▲

* *Amplitude density spectrum* would be more correct, since its dimensions are (amplitude units)(time) = (amplitude units)/(frequency), but we will use the term *amplitude spectrum* for simplicity.

FIGURE 2.9 Amplitude and phase spectra for a pulse signal. (a) Amplitude spectrum. (b) Phase spectrum ($t_0 = \frac{1}{2}\tau$ is assumed).



Energy Spectral Density

The energy of a signal, defined by (2.20a), can be expressed in the frequency domain as follows:

$$E \triangleq \int_{-\infty}^{\infty} |x(t)|^2 \, dt$$

$$= \int_{-\infty}^{\infty} x^*(t) \left[\int_{-\infty}^{\infty} X(f) e^{j2\pi ft} \, df \right] dt \tag{2.81}$$

where $x(t)$ has been written in terms of its Fourier transform. Reversing the order of integration, we obtain

$$E = \int_{-\infty}^{\infty} X(f) \left[\int_{-\infty}^{\infty} x^*(t) e^{j2\pi ft} \, dt \right] df$$

$$= \int_{-\infty}^{\infty} X(f) \left[\int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} \, dt \right]^* df = \int_{-\infty}^{\infty} X(f) X^*(f) \, df$$

or

$$E = \int_{-\infty}^{\infty} |x(t)|^2 \, dt = \int_{-\infty}^{\infty} |X(f)|^2 \, df \tag{2.82}$$

This is referred to as *Rayleigh's energy theorem* or Parseval's theorem for Fourier transforms.

Examining $|X(f)|^2$ and recalling the definition of $X(f)$ given by (2.72), we note that the former has the units of (volts-seconds)^2 or, since we are considering power on a per-ohm basis, (watts-seconds)/hertz = joules/hertz. Thus we see that $|X(f)|^2$ has the units of energy density, and we define the energy spectral density of a signal as

$$G(f) \triangleq |X(f)|^2 \tag{2.83}$$

By integrating $G(f)$ over all frequency, we obtain the total energy.

EXAMPLE 2.9 Rayleigh's energy theorem is convenient for finding the energy in a signal whose square is not easily integrated in the time domain, or vice versa. For example, the signal

$$x(t) = 40 \text{ sinc } (20t) \tag{2.84}$$

has energy density

$$G(f) = |X(f)|^2 = \left[2\Pi\left(\frac{f}{20}\right)\right]^2 = 4\Pi\left(\frac{f}{20}\right) \tag{2.85}$$

where $\Pi(f/20)$ need not be squared because it has unity amplitude. Using Rayleigh's energy theorem, we find that the energy in $x(t)$ is

$$E = \int_{-\infty}^{\infty} G(f)\, df = \int_{-10}^{10} 4\, df = 80 \text{ J} \tag{2.86}$$

This checks with the result that is obtained by integrating $x^2(t)$ over all t using the definite integral $\int_{-\infty}^{\infty} \text{sinc}^2 u\, du = 1$.

The energy contained in the frequency interval $(0, W)$ can be found from the integral

$$E_W = \int_{-W}^{W} G(f)\, df = 2\int_0^W \left[2\Pi\left(\frac{f}{20}\right)\right]^2 df = \begin{cases} 8W, & W \leq 10 \\ 80, & W > 10 \end{cases} \tag{2.87}$$

Convolution

We digress somewhat from our consideration of the Fourier transform to define the convolution operation and illustrate it by example.

The convolution of two signals, $x_1(t)$ and $x_2(t)$, is a new function of time, $x(t)$, written symbolically in terms of x_1 and x_2 as

$$x(t) = x_1(t) * x_2(t) = \int_{-\infty}^{\infty} x_1(\lambda)x_2(t - \lambda)\, d\lambda \tag{2.88}$$

Note that t is a parameter as far as the integration is concerned. The integrand is formed from x_1 and x_2 by three operations: (1) time reversal to obtain $x_2(-\lambda)$, (2) time shifting to obtain $x_2(t - \lambda)$, and (3) multiplication of $x_1(\lambda)$ and $x_2(t - \lambda)$ to form the integrand. An example will illustrate the implementation of these operations to form $x_1 * x_2$. Note that the dependence on time is often suppressed.

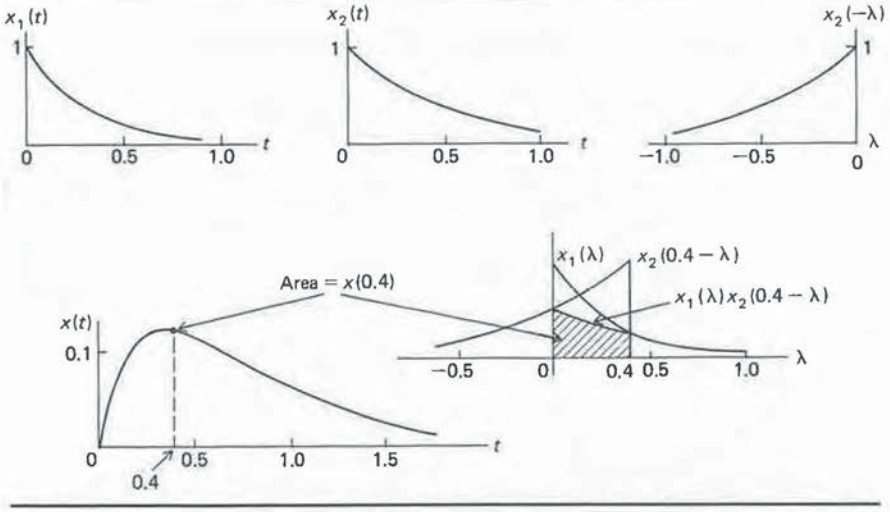
▼ **EXAMPLE 2.10** Find the convolution of the two signals

$$x_1(t) = e^{-\alpha t}u(t) \quad \text{and} \quad x_2(t) = e^{-\beta t}u(t), \qquad \alpha > \beta > 0 \tag{2.89}$$

SOLUTION The steps involved in the convolution are illustrated in Figure 2.10 for $\alpha = 4$ and $\beta = 2$. Mathematically, we can form the integrand by direct substitution:

$$x(t) = x_1 * x_2 = \int_{-\infty}^{\infty} e^{-\alpha\lambda}u(\lambda)e^{-\beta(t-\lambda)}u(t - \lambda)\, d\lambda \tag{2.90}$$

FIGURE 2.10 The operations involved in the convolution of two exponentially decaying signals



But

$$u(\lambda)u(t-\lambda) = \begin{cases} 0, & \lambda < 0 \\ 1, & 0 < \lambda < t \\ 0, & \lambda > t \end{cases} \tag{2.91}$$

Thus

$$x(t) = \begin{cases} 0, & t < 0 \\ \displaystyle\int_0^t e^{-\beta t} e^{-(\alpha-\beta)\lambda}\, d\lambda = \frac{1}{\alpha-\beta}(e^{-\beta t} - e^{-\alpha t}), & t \geq 0 \end{cases} \tag{2.92}$$

This result for $x(t)$ is also shown in Figure 2.10.

*Transform Theorems: Proofs and Applications**

Several useful theorems involving Fourier transforms can be proved. These are useful for deriving Fourier transform pairs as well as deducing general frequency domain relationships. The notation $x(t) \leftrightarrow X(f)$ will be used to denote a Fourier transform pair.

Each theorem will be stated along with a proof in most cases. Several examples giving applications will be given after the statements of all the theorems. In the statement of the theorems, $x(t)$, $x_1(t)$, and $x_2(t)$ denote signals with $X(f)$, $X_1(f)$, and $X_2(f)$ denoting their respective Fourier transforms. Constants are denoted by a, a_1, a_2, t_0, and f_0.

▷ SUPERPOSITION THEOREM

$$a_1x_1(t) + a_2x_2(t) \leftrightarrow a_1X_1(f) + a_2X_2(f) \tag{2.93}$$

PROOF By the defining integral for the Fourier transform,

$$\begin{aligned} \mathcal{F}\{a_1x_1(t) + a_2x_2(t)\} &= \int_{-\infty}^{\infty} [a_1x_1(t) + a_2x_2(t)]e^{-j2\pi ft}\, dt \\ &= a_1 \int_{-\infty}^{\infty} x_1(t)e^{-j2\pi ft}\, dt + a_2 \int_{-\infty}^{\infty} x_2(t)e^{-j2\pi ft}\, dt \\ &= a_1X_1(f) + a_2X_2(f) \end{aligned} \tag{2.94}$$

▷ TIME-DELAY THEOREM

$$x(t - t_0) \leftrightarrow X(f)e^{-j2\pi ft_0} \tag{2.95}$$

* See Tables C.1 and C.2 in Appendix C for a listing of Fourier transform pairs and theorems.

PROOF Using the defining integral for the Fourier transform, we have

$$\begin{aligned}\mathcal{F}\{x(t - t_0)\} &= \int_{-\infty}^{\infty} x(t - t_0)e^{-j2\pi ft}\,dt \\ &= \int_{-\infty}^{\infty} x(\lambda)e^{-j2\pi f(\lambda + t_0)}\,dt \\ &= e^{-j2\pi ft_0}\int_{-\infty}^{\infty} x(\lambda)e^{-j2\pi f\lambda}\,d\lambda \\ &= X(f)e^{-j2\pi ft_0}\end{aligned} \tag{2.96}$$

where the substitution $\lambda = t - t_0$ was used in the first integral.

▷**SCALE-CHANGE THEOREM**

$$x(at) \leftrightarrow \frac{1}{|a|}X\left(\frac{f}{a}\right) \tag{2.97}$$

PROOF First, assume that $a > 0$. Then

$$F\{x(at)\} = \int_{-\infty}^{\infty} x(at)e^{-j2\pi ft}\,dt = \int_{-\infty}^{\infty} x(\lambda)e^{-j2\pi f\lambda/a}\frac{d\lambda}{a} = \frac{1}{a}X\left(\frac{f}{a}\right) \tag{2.98}$$

where the substitution $\lambda = at$ has been used. Next considering $a < 0$, we write

$$\begin{aligned}\mathcal{F}\{x(at)\} &= \int_{-\infty}^{\infty} x(-|a|t)e^{-j2\pi ft}\,dt = \int_{-\infty}^{\infty} x(\lambda)e^{+j2\pi f\lambda/|a|}\frac{d\lambda}{|a|} \\ &= \frac{1}{|a|}X\left(-\frac{f}{|a|}\right) = \frac{1}{|a|}X\left(\frac{f}{a}\right)\end{aligned} \tag{2.99}$$

where use has been made of the relation $-|a| = a$.

▷**DUALITY THEOREM**

$$X(t) \leftrightarrow x(-f) \tag{2.100}$$

That is, if the Fourier transform of $x(t)$ is $X(f)$, then the Fourier transform of $X(f)$ with f replaced by t is the original time-domain signal with t replaced by $-f$.

PROOF The proof of this theorem follows by virtue of the fact that the only difference between the Fourier transform integral and the inverse Fourier transform integral is a minus sign in the exponent of the integrand.

▷FREQUENCY TRANSLATION THEOREM

$$x(t)e^{j2\pi f_0 t} \leftrightarrow X(f - f_0) \tag{2.101}$$

PROOF To prove the frequency translation theorem, note that

$$\int_{-\infty}^{\infty} x(t)e^{j2\pi f_0 t}e^{-j2\pi ft}\,dt = \int_{-\infty}^{\infty} x(t)e^{-j2\pi(f-f_0)t}\,dt = X(f - f_0) \tag{2.102}$$

▷MODULATION THEOREM

$$x(t)\cos(2\pi f_0 t) \leftrightarrow \tfrac{1}{2}X(f - f_0) + \tfrac{1}{2}X(f + f_0) \tag{2.103}$$

PROOF The proof of this theorem follows by writing $\cos(2\pi f_0 t)$ in exponential form as $(e^{j2\pi f_0 t} + e^{-j2\pi f_0 t})/2$ and applying the superposition and frequency translation theorems.

▷DIFFERENTIATION THEOREM

$$\frac{d^n x(t)}{dt^n} \leftrightarrow (j2\pi f)^n X(f) \tag{2.104}$$

PROOF We prove the theorem for $n = 1$ by using integration by parts on the defining Fourier transform integral as follows:

$$\begin{aligned}\mathcal{F}\left\{\frac{dx}{dt}\right\} &= \int_{-\infty}^{\infty} \frac{dx(t)}{dt}\,e^{-j2\pi ft}\,dt \\ &= x(t)e^{-j2\pi ft}\Big|_{-\infty}^{\infty} + j2\pi f\int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}\,dt \\ &= j2\pi f\,X(f)\end{aligned} \tag{2.105}$$

where $u = e^{-j2\pi ft}$ and $dv = (dx/dt)dt$ have been used in the integration-by-parts formula, and the first term of the middle equation vanishes at each end point by virtue of $x(t)$ being an energy signal. The proof for values of $n > 1$ follows by induction.

▷INTEGRATION THEOREM

$$\int_{-\infty}^{t} x(\lambda)\,d\lambda \leftrightarrow (j2\pi f)^{-1}X(f) + \frac{1}{2}X(0)\delta(f) \tag{2.106}$$

PROOF If $X(0) = 0$ the proof of the integration theorem can be carried out by using integration by parts as in the case of the differentiation theorem. We obtain

$$\mathcal{F}\left\{\int_{-\infty}^{t} x(\lambda)\, d\lambda\right\} = \left[\int_{-\infty}^{t} x(\lambda)\, d\lambda\right]\left(-\frac{1}{j2\pi f} e^{-j2\pi ft}\right)\Bigg|_{-\infty}^{\infty} + \frac{1}{j2\pi f}\int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}\, dt \tag{2.107}$$

The first term vanishes if $X(0) = \int_{-\infty}^{\infty} x(t)\, dt = 0$, and the second term is just $X(f)/j2\pi f$. For $X(0) \neq 0$, a limiting argument must be used to account for the Fourier transform of the nonzero average value of $x(t)$.

▷CONVOLUTION THEOREM

$$\int_{-\infty}^{\infty} x_1(\lambda)x_2(t-\lambda)\, d\lambda = \int_{-\infty}^{\infty} x_1(t-\lambda)x_2(\lambda)\, d\lambda \leftrightarrow X_1(f)X_2(f) \tag{2.108}$$

PROOF To prove the convolution theorem of Fourier transforms, we represent $x_2(t-\lambda)$ in terms of the inverse Fourier transform integral as

$$x_2(t-\lambda) = \int_{-\infty}^{\infty} X_2(f)e^{j2\pi f(t-\lambda)}\, df \tag{2.109}$$

Denoting the convolution operation as $x_1(t) * x_2(t)$, we have

$$\begin{aligned} x_1(t) * x_2(t) &= \int_{-\infty}^{\infty} x_1(\lambda)\left[\int_{-\infty}^{\infty} X_2(f)e^{j2\pi f(t-\lambda)}\, df\right] d\lambda \\ &= \int_{-\infty}^{\infty} X_2(f)\left[\int_{-\infty}^{\infty} x_1(\lambda)e^{-j2\pi f\lambda}\, d\lambda\right] e^{j2\pi ft}\, df \end{aligned} \tag{2.110}$$

where the last step results from reversing the orders of integration. The bracketed term inside the integral is $X_1(f)$, the Fourier transform of $x_1(t)$. Thus

$$x_1 * x_2 = \int_{-\infty}^{\infty} X_1(f)X_2(f)e^{j2\pi ft}\, df \tag{2.111}$$

which is the inverse Fourier transform of $X_1(f)X_2(f)$. Taking the Fourier transform of this result yields the desired transform pair.

▷MULTIPLICATION THEOREM

$$x_1(t)x_2(t) \leftrightarrow X_1(f) * X_2(f) = \int_{-\infty}^{\infty} X_1(\lambda)X_2(f-\lambda)\,d\lambda \qquad (2.112)$$

PROOF The proof of the multiplication theorem proceeds in a manner analogous to the proof of the convolution theorem.

▼ EXAMPLE 2.11 Use the duality theorem to show that

$$A \text{ sinc } 2Wt \leftrightarrow \frac{A}{2W}\Pi\left(\frac{f}{2W}\right) \qquad (2.113)$$

SOLUTION From Example 2.8, we know that

$$x(t) = A\Pi\left(\frac{t}{\tau}\right) \leftrightarrow A\tau \text{ sinc } f\tau = X(f)$$

Considering $X(t)$, and using the duality theorem, we obtain

$$X(t) = A\tau \text{ sinc } \tau t \leftrightarrow A\Pi\left(\frac{-f}{\tau}\right) = x(-f) \qquad (2.114)$$

where τ is a parameter with dimension $(s)^{-1}$, which may be somewhat confusing at first sight! By letting $\tau = 2W$, the given relationship follows. ▲

▼ EXAMPLE 2.12 Obtain the following Fourier transform pairs:

1. $A\delta(t) \leftrightarrow A$
2. $A\delta(t - t_0) \leftrightarrow Ae^{-j2\pi ft_0}$
3. $A \leftrightarrow A\delta(f)$
4. $Ae^{j2\pi f_0 t} \leftrightarrow A\delta(f - f_0)$

SOLUTION Even though these signals are not energy signals, we can formally derive the Fourier transform of each by obtaining the Fourier transform of a "proper" energy signal that approaches the given signal in the limit as some parameter approaches zero or infinity. For example, formally,

$$\mathcal{F}[A\delta(t)] = \mathcal{F}\left[\lim_{\tau\to 0}\left(\frac{A}{\tau}\right)\Pi\left(\frac{t}{\tau}\right)\right] = \lim_{\tau\to 0} A \text{ sinc } (f\tau) = A \qquad (2.115)$$

We can use a formal procedure such as this to define Fourier transforms for the other three signals as well. It is easier, however, to use the sifting property of the delta function and the appropriate Fourier transform theorems.

The same results are obtained. For example, we obtain the first transform pair directly by writing down the Fourier transform integral with $x(t) = \delta(t)$ and invoking the sifting property:

$$\mathcal{F}[A\delta(t)] = A\int_{-\infty}^{\infty}\delta(t)e^{-j\omega t}\,dt = A \tag{2.116}$$

Transform pair 2 follows by application of the time-delay theorem to pair 1.

Transform pair 3 can be obtained by using the inverse-transform relationship or the first transform pair and the duality theorem. Using the latter, we obtain

$$X(t) = A \leftrightarrow A\delta(-f) = A\delta(f) = x(-f) \tag{2.117}$$

where the impulse function is assumed to be even.

Transform pair 4 follows by applying the frequency-translation theorem to pair 3. The Fourier transform pairs of Example 2.12 will be used often in the discussion of modulation. ▲

▼ **EXAMPLE 2.13** Use the differentiation theorem to obtain the Fourier transform of the triangular signal, defined as

$$\Lambda\left(\frac{t}{\tau}\right) \triangleq \begin{cases} 1 - |t|/\tau, & |t| < \tau \\ 0, & \text{otherwise} \end{cases} \tag{2.118}$$

SOLUTION Differentiating $\Lambda(t/\tau)$ twice, we obtain, as shown in Figure 2.11,

$$\frac{d^2\Lambda(t/\tau)}{dt^2} = \frac{1}{\tau}\delta(t+\tau) - \frac{2}{\tau}\delta(t) + \frac{1}{\tau}\delta(t-\tau) \tag{2.119}$$

Using the differentiation, superposition, and time-shift theorems and the result of Example 2.12, we obtain

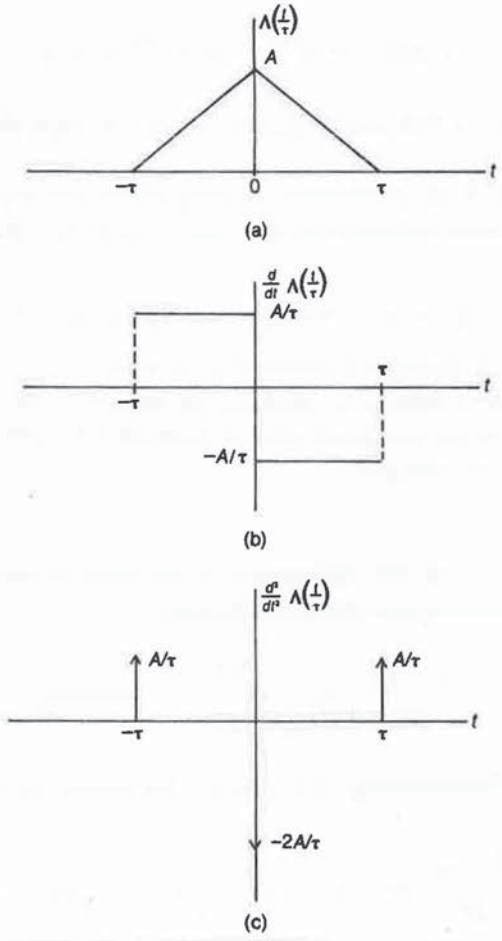
$$\mathcal{F}\left[\frac{d^2\Lambda(t/\tau)}{dt^2}\right] = (j2\pi f)^2\mathcal{F}\left[\Lambda\left(\frac{t}{\tau}\right)\right]$$

$$= \frac{1}{\tau}(e^{j2\pi f\tau} - 2 + e^{-j2\pi f\tau}) \tag{2.120a}$$

or

$$\mathcal{F}\left[\Lambda\left(\frac{t}{\tau}\right)\right] = \frac{2\cos 2\pi f\tau - 2}{\tau(j2\pi f)^2} = \tau\frac{\sin^2\pi f\tau}{(\pi f\tau)^2} \tag{2.120b}$$

FIGURE 2.11 Triangular signal and its first two derivatives. (a) Triangular signal. (b) First derivative of triangular signal. (c) Second derivative of triangular signal.



where we have used $\frac{1}{2}(1 - \cos 2\pi ft) = \sin^2 \pi ft$. Summarizing, we have shown that

$$\Lambda\left(\frac{t}{\tau}\right) \leftrightarrow \tau \operatorname{sinc}^2 f\tau \tag{2.121}$$

▲ where $(\sin \pi f\tau)/(\pi f\tau)$ has been replaced by $\operatorname{sinc} f\tau$.

EXAMPLE 2.14 As another example of obtaining Fourier transforms of signals involving impulses, let us consider the signal

$$y_s(t) = \sum_{m=-\infty}^{\infty} \delta(t - mT_s) \tag{2.122}$$

It is a periodic waveform referred to as the *ideal sampling waveform* and consists of an infinite sequence of impulses spaced by T_s seconds.

SOLUTION To obtain the Fourier transform of $y_s(t)$, we note that it is periodic and, in a formal sense, therefore can be represented by a Fourier series. Thus

$$y_s(t) = \sum_{m=-\infty}^{\infty} \delta(t - mT_s) = \sum_{n=-\infty}^{\infty} Y_n e^{jn\omega_s t}, \qquad \omega_s = \frac{2\pi}{T_s} \tag{2.123}$$

where

$$Y_n = \frac{1}{T_s} \int_{T_s} \delta(t) e^{-jn\omega_s t}\, dt = f_s, \qquad f_s = \frac{\omega_s}{2\pi} \tag{2.124}$$

by the sifting property of the impulse function. Therefore,

$$y_s(t) = f_s \sum_{n=-\infty}^{\infty} e^{jn\omega_s t} \tag{2.125}$$

Fourier-transforming term by term, we obtain

$$Y_s(f) = f_s \sum_{n=-\infty}^{\infty} \mathcal{F}[1 \cdot e^{jn\omega_s t}] = f_s \sum_{n=-\infty}^{\infty} \delta(f - nf_s) \tag{2.126}$$

where we have used the results of Example 2.12. Summarizing, we have shown that

$$\sum_{m=-\infty}^{\infty} \delta(t - mT_s) \leftrightarrow f_s \sum_{n=-\infty}^{\infty} \delta(f - nf_s) \tag{2.127}$$

The transform pair (2.127) is useful in spectral representations of periodic signals by the Fourier transform, which will be considered shortly.

EXAMPLE 2.15 The convolution theorem can be used to obtain the Fourier transform of the triangle $\Lambda(t/\tau)$ defined by (2.118).

SOLUTION We proceed by first showing that the convolution of two pulses is a triangle. The steps in computing

$$y(t) = \int_{-\infty}^{\infty} \Pi\left(\frac{t-\lambda}{\tau}\right) \Pi\left(\frac{\lambda}{\tau}\right) d\lambda \tag{2.128}$$

are carried out in Table 2.2. Summarizing the results, we have

$$\tau\Lambda\left(\frac{t}{\tau}\right) = \Pi\left(\frac{t}{\tau}\right) * \Pi\left(\frac{t}{\tau}\right) = \begin{cases} 0, & t < -\tau \\ \tau - |t|, & |t| \leq \tau \\ 0, & t > \tau \end{cases} \tag{2.129}$$

Using the transform pair

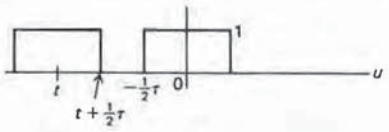
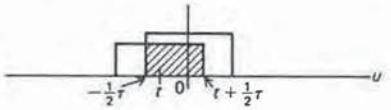
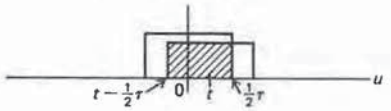
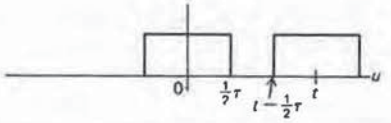
$$\Pi\left(\frac{t}{\tau}\right) \leftrightarrow \tau \operatorname{sinc} ft \tag{2.130}$$

and the convolution theorem, we obtain the transform pair

$$\Lambda\left(\frac{t}{\tau}\right) \leftrightarrow \tau \operatorname{sinc}^2 f\tau \tag{2.131}$$

▲ as in Example 2.13.

able 2.2 Computation of $\Pi(t/\tau) * \Pi(t/\tau)$

Range on t	Integrand	Limits	Area
$o < t < -\tau$			0
$< t < 0$		$-\frac{1}{2}\tau$ to $t+\frac{1}{2}\tau$	$\tau + t$
$: t < \tau$		$t-\frac{1}{2}\tau$ to $\frac{1}{2}\tau$	$\tau - t$
$: t < \infty$			0

A useful result is the convolution of an impulse $\delta(t - t_0)$ with a signal $x(t)$. Carrying out the operation, we obtain

$$\delta(t - t_0) * x(t) = \int_{-\infty}^{\infty} \delta(\lambda - t_0)x(t - \lambda)\, d\lambda$$
$$= x(t - t_0) \tag{2.132}$$

by the sifting property of the delta function. That is, convolution of $x(t)$ with an impulse occurring at time t_0 simply shifts $x(t)$ to t_0.

▼ **EXAMPLE 2.16** Consider the Fourier transform of the cosinusoidal pulse

$$x(t) = A\Pi\left(\frac{t}{\tau}\right)\cos \omega_0 t, \qquad \omega_0 = 2\pi f_0 \tag{2.133}$$

Using the transform pair

$$e^{\pm j2\pi f_0 t} \leftrightarrow \delta(f \mp f_0) \tag{2.134}$$

obtained earlier and Euler's theorem, we find that

$$\cos 2\pi f_0 t \leftrightarrow \tfrac{1}{2}\delta(f - f_0) + \tfrac{1}{2}\delta(f + f_0) \tag{2.135}$$

We have shown also that

$$A\Pi\left(\frac{t}{\tau}\right) \leftrightarrow A\tau \operatorname{sinc} f\tau$$

Therefore, using the multiplication theorem, we obtain

$$X(f) = \tfrac{1}{2}A\tau\,(\operatorname{sinc} f\tau) * [\delta(f - f_0) + \delta(f + f_0)]$$
$$= \tfrac{1}{2}A\tau\,[\operatorname{sinc}(f - f_0)\tau + \operatorname{sinc}(f + f_0)\tau] \tag{2.136}$$

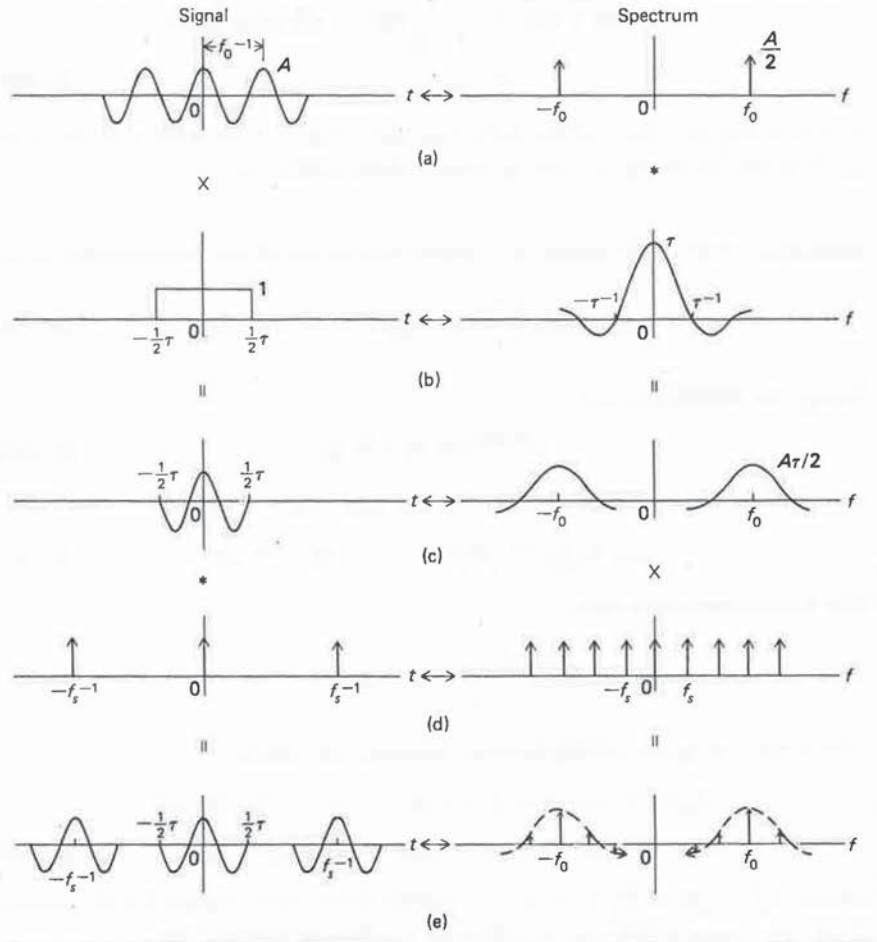
where $\delta(f - f_0) * Z(f) = Z(f - f_0)$ has been used. Figure 2.12(c) shows $X(f)$. The same result can be obtained via the modulation theorem. ▲

Fourier Transforms of Periodic Signals

The Fourier transform of a periodic signal, in a strict mathematical sense, does not exist, since periodic signals are not energy signals. However, using the transform pairs derived in Example 2.12 for a constant and a phasor signal, we could, in a formal sense, write down the Fourier transform of a periodic signal by Fourier-transforming its complex Fourier series term by term.

A somewhat more useful form for the Fourier transform of a periodic signal is obtained by applying the convolution theorem and the transform

FIGURE 2.12 (a)–(c) Application of the multiplication theorem. (c)–(e) Application of the convolution theorem. Note: × denotes multiplication; * denotes convolution, ↔ denotes transform pairs.



pair (2.127) for the ideal sampling wave. To obtain it, consider the result of convolving the ideal sampling waveform with a pulse-type signal $p(t)$ to obtain a new signal $x(t)$, where $x(t)$ is a periodic power signal. This is apparent when one carries out the convolution with the aid of (2.132):

$$x(t) = \left[\sum_{m=-\infty}^{\infty} \delta(t - mT_s)\right] * p(t) = \sum_{m=-\infty}^{\infty} p(t - mT_s) \quad (2.137)$$

Applying the convolution theorem and the Fourier transform pair of (2.127), we find that the Fourier transform of $x(t)$ is

$$X(f) = \mathcal{F}\left\{\sum_{m=-\infty}^{\infty} \delta(t - mT_s)\right\} P(f)$$

$$= f_s P(f) \sum_{n=-\infty}^{\infty} \delta(f - nf_s) = \sum_{n=-\infty}^{\infty} f_s P(nf_s)\, \delta(f - nf_s) \quad (2.138)$$

where $P(f) = \mathcal{F}[p(t)]$. Summarizing, we have obtained the Fourier transform pair

$$\sum_{m=-\infty}^{\infty} p(t - mT_s) \leftrightarrow \sum_{n=-\infty}^{\infty} f_s P(nf_s)\, \delta(f - nf_s) \quad (2.139)$$

The usefulness of (2.139) is illustrated with an example.

▼ **EXAMPLE 2.17** The Fourier transform of a single cosinusoidal pulse was found in Example 2.16 and is shown in Figure 2.12(c). The Fourier transform of a periodic cosinusoidal pulse train, which could represent the output of a radar transmitter, for example, is obtained by writing it as

$$y(t) = \left[\sum_{m=-\infty}^{\infty} \delta(t - mT_s)\right] * \Pi\left(\frac{t}{\tau}\right) \cos \omega_0 t, \qquad \omega_0 \gg 2\pi/\tau$$

$$= \sum_{m=-\infty}^{\infty} \Pi\left(\frac{t - mT_s}{\tau}\right) \cos \omega_0(t - mT_s), \qquad f_s \leq \tau^{-1} \quad (2.140)$$

This signal is illustrated in Figure 2.12(e). According to (2.139), the Fourier transform of $y(t)$ is

$$Y(f) = \sum_{n=-\infty}^{\infty} \frac{Af_s\tau}{2} [\text{sinc}\,(nf_s - f_0)\tau + \text{sinc}\,(nf_s + f_0)\tau]\, \delta(f - nf_s) \quad (2.141)$$

▲ The spectrum is illustrated on the right-hand side of Figure 2.12(e).

Poisson Sum Formula

We can develop the *Poisson sum formula* by taking the inverse Fourier transform of the right-hand side of (2.139). When we use the transform pair $\exp(-j2\pi nf_s t) \leftrightarrow \delta(f - nf_s)$ (see Example 2.12), it follows that

$$\mathcal{F}^{-1}\left\{\sum_{n=-\infty}^{\infty} f_s P(nf_s)\, \delta(f - nf_s)\right\} = f_s \sum_{n=-\infty}^{\infty} P(nf_s) e^{j2\pi nf_s t} \quad (2.142)$$

Equating this to the left-hand side of (2.139), we obtain the Poisson sum formula:

$$\sum_{m=-\infty}^{\infty} p(t - mT_s) = f_s \sum_{n=-\infty}^{\infty} P(nf_s)e^{j2\pi nf_s t} \tag{2.143}$$

The Poisson sum formula is useful when one goes from the Fourier transform to sampled approximations of it. For example, Equation (2.143) says that the sample values $P(nf_s)$ of $P(f) = \mathcal{F}\{p(t)\}$ are the Fourier series coefficients of the periodic function $T_s\sum_{n=-\infty}^{\infty} p(t - mT_s)$.

†2.6 POWER SPECTRAL DENSITY AND CORRELATION

Recalling the definition of energy spectral density, Equation (2.83), we see that it is of use only for energy signals for which the integral of $G(f)$ over all frequencies gives total energy, a finite quantity. For power signals, it is meaningful to speak in terms of *power spectral density*. Analogous to $G(f)$, we define the power spectral density $S(f)$ of a signal $x(t)$ as a real, even, nonnegative function of frequency, which gives total average power per ohm when integrated; that is,

$$P = \int_{-\infty}^{\infty} S(f)\, df = \langle x^2(t) \rangle \tag{2.144}$$

Since $S(f)$ is a function that gives the variation of density of power with frequency, we conclude that it must consist of a series of impulses for the periodic power signals that we have so far considered. Later, in Chapter 5, we will consider random signals, which have power spectral densities that are continuous functions of frequency.

▼ EXAMPLE 2.18 Considering the sinusoidal signal

$$x(t) = A \cos(\omega_0 t + \theta) \tag{2.145}$$

we note that its average power per ohm, $\frac{1}{2}A^2$, is concentrated at the single frequency $f_0 = \omega_0/2\pi$. However, since the power spectral density must be an even function of frequency, we split this power equally between $+f_0$ and $-f_0$. Thus the power spectral density of $x(t)$ is, from intuition, given by

$$S(f) = \tfrac{1}{4}A^2\,\delta(f - f_0) + \tfrac{1}{4}A^2\,\delta(f + f_0) \tag{2.146}$$

Checking this by using (2.144), we see that integration over all frequencies results in the average power per ohm, $\frac{1}{2}A^2$. ▲

The Time-Average Autocorrelation Function

To introduce the time-average autocorrelation function, we return to the energy spectral density of an energy signal. Without any apparent reason, suppose we take the inverse Fourier transform of $G(f)$, letting the independent variable be τ:

$$\begin{aligned}\phi(\tau) &\triangleq \mathcal{F}^{-1}[G(f)] = \mathcal{F}^{-1}[X(f)X^*(f)] \\ &= \mathcal{F}^{-1}[X(f)] * \mathcal{F}^{-1}[X^*(f)]\end{aligned} \tag{2.147}$$

The last step follows by application of the convolution theorem. Applying the time-reversal theorem (item 3b in Table C.2 in Appendix C) to write $\mathcal{F}^{-1}[X^*(f)] = x(-\tau)$ and then the convolution theorem, we obtain

$$\begin{aligned}\phi(\tau) &= x(\tau) * x(-\tau) = \int_{-\infty}^{\infty} x(\lambda)x(\lambda + \tau)\, d\lambda \\ &= \lim_{T\to\infty} \int_{-T}^{T} x(\lambda)x(\lambda + \tau)\, d\lambda \qquad \text{(energy signal)}\end{aligned} \tag{2.148}$$

Equation (2.148) will be referred to as the *time-average autocorrelation function for energy signals*. We see that it gives a measure of the similarity, or coherence, between a signal and a delayed version of the signal. Note that $\phi(0) = E$, the signal energy. The similarity of the correlation operation to convolution also should be noted. The major point of (2.147) is that the autocorrelation function and energy spectral density are Fourier transform pairs. We forgo further discussion of the time-average autocorrelation function for energy signals in favor of analogous results for power signals.

The time-average autocorrelation function $R(\tau)$ of a power signal $x(t)$ is defined as the time average

$$\begin{aligned}R(\tau) &= \langle x(t)x(t + \tau)\rangle \\ &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)x(t + \tau)\, dt \qquad \text{(power signal)}\end{aligned} \tag{2.149}$$

If $x(t)$ is periodic with period T_0, the integrand of (2.149) is periodic, and the time average can be taken over a single period:

$$R(\tau) = \frac{1}{T_0} \int_{T_0} x(t)x(t + \tau)\, dt \qquad [x(t) \text{ periodic}] \tag{2.150}$$

Just like $\phi(\tau)$, $R(\tau)$ gives a measure of the similarity between the signal at time t and at time $t + \tau$; it is a function of the delay variable τ, since time

is the variable of integration. In addition to being a measure of the similarity between a signal and its time displacement, we note that

$$R(0) = \langle x^2(t) \rangle = \int_{-\infty}^{\infty} S(f)\, df \tag{2.151}$$

Thus we suspect that the time-average autocorrelation function and power spectral density of a power signal are closely related, just as they are for energy signals. This relationship is stated formally by the *Wiener-Khinchine theorem,* which says that the time-average autocorrelation function of a signal and its power spectral density are Fourier transform pairs:

$$S(f) = \mathcal{F}[R(\tau)] = \int_{-\infty}^{\infty} R(\tau) e^{-j2\pi f\tau}\, d\tau \tag{2.152a}$$

and

$$R(\tau) = \mathcal{F}^{-1}[S(f)] = \int_{-\infty}^{\infty} S(f) e^{j2\pi f\tau}\, df \tag{2.152b}$$

A formal proof of the Wiener-Khinchine theorem will be given in Chapter 5. We simply take (2.152a) as the definition of power spectral density at this point. We note that (2.151) follows immediately from (2.152b) by setting $\tau = 0$.

Properties of R(τ)

The time-average autocorrelation function has several useful properties, which are listed below:

1. $R(0) = \langle x^2(t) \rangle \geq |R(\tau)|$, for all τ; that is, a relative maximum of $R(\tau)$ exists at $\tau = 0$.
2. $R(-\tau) = \langle x(t)x(t - \tau) \rangle = R(\tau)$; that is, $R(\tau)$ is even.
3. $\lim_{|\tau| \to \infty} R(\tau) = \langle x(t) \rangle^2$ if $x(t)$ does not contain a periodic component.
4. If $x(t)$ is periodic in t with period T_0, then $R(\tau)$ is periodic in τ with period T_0.
5. The time-average autocorrelation function of any signal has a Fourier transform that is nonnegative.

Property 5 results by virtue of the fact that normalized power is a nonnegative quantity. These properties will be proved in Chapter 5.

▼ EXAMPLE 2.19 Returning to the sinusoidal signal of Example 2.18, we see that by the Wiener-Khinchine theorem its autocorrelation function is

$$R(\tau) = \mathcal{F}^{-1}\left[\frac{A^2}{4}\delta(f + f_0) + \frac{A^2}{4}\delta(f - f_0)\right]$$
$$= \frac{A^2}{2}\cos \omega_0 \tau, \quad \omega_0 = 2\pi f_0 \tag{2.153}$$

Checking this result by computing the time average using (2.149), we find that

$$R(\tau) = \frac{1}{T_0}\int_{T_0} A^2 \cos(\omega_0 t + \theta)\cos(\omega_0 t + \omega_0 \tau + \theta)\, dt$$
$$= \frac{1}{T_0}\int_{T_0} \frac{A^2}{2}\cos \omega_0 \tau\, dt + \frac{1}{T_0}\int_{T_0} \frac{A^2}{2}\cos(2\omega_0 t + \omega_0 \tau + 2\theta)\, dt$$
$$= \frac{A^2}{2}\cos \omega_0 t \tag{2.154}$$

where the second integral of the second equation is zero by virtue of the periodicity of the integrand. Note that all the properties listed above for $R(\tau)$ are satisfied by $R(\tau) = \frac{1}{2}A^2 \cos \omega_0 \tau$ except Property 3, which does not apply. ▲

The autocorrelation function and power spectral density are important tools for systems analysis involving random signals.

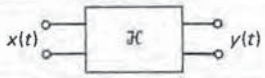
2.7 SIGNALS AND LINEAR SYSTEMS

In this section we are concerned with the characterization of systems and their effect on signals. In system modeling, the actual elements, such as resistors, capacitors, inductors, springs, and masses, that compose a particular system are usually not of concern. Rather, we view a system in terms of the operation it performs on an input to produce an output. Symbolically, this is accomplished, for a single-input, single-output system, by writing

$$y(t) = \mathcal{H}[x(t)] \tag{2.155}$$

where $\mathcal{H}$ is the operator that produces the output $y(t)$ from the input $x(t)$, as illustrated in Figure 2.13. We now consider certain classes of systems, the first of which is linear time-invariant systems.

FIGURE 2.13 Operator representation of a linear system



Definition of a Linear Time-Invariant System

If a system is *linear*, superposition holds. That is, if $x_1(t)$ results in the output $y_1(t)$ and $x_2(t)$ results in the output $y_2(t)$, then the output due to $\alpha_1 x_1(t) + \alpha_2 x_2(t)$, where α_1 and α_2 are constants, is given by

$$\begin{aligned} y(t) &= \mathcal{H}[\alpha_1 x_1(t) + \alpha_2 x_2(t)] = \alpha_1 \mathcal{H}[x_1(t)] + \alpha_2 \mathcal{H}[x_2(t)] \\ &= \alpha_1 y_1(t) + \alpha_2 y_2(t) \end{aligned} \tag{2.156}$$

If the system is *time-invariant*, or *fixed*, the delayed input $x(t - t_0)$ gives the delayed output $y(t - t_0)$; that is,

$$y(t - t_0) = \mathcal{H}[x(t - t_0)] \tag{2.157}$$

With these properties explicitly stated, we are now ready to obtain a more concrete description of a linear time-invariant (LTI) system.

Impulse Response and the Superposition Integral

The *impulse response* $h(t)$ of an LTI system is defined to be the response of the system to an impulse applied at $t = 0$, that is

$$h(t) \triangleq \mathcal{H}[\delta(t)] \tag{2.158}$$

By the time-invariant property of the system, the response to an impulse applied at any time t_0 is $h(t - t_0)$, and the response to the linear combination of impulses $\alpha_1\delta(t - t_1) + \alpha_2\delta(t - t_2)$ is $\alpha_1 h(t - t_1) + \alpha_2 h(t - t_2)$ by the superposition property and time-invariance. Through induction, we may therefore show that the response to the input

$$x(t) = \sum_{n=1}^{N} \alpha_n \delta(t - t_n) \tag{2.159}$$

is

$$y(t) = \sum_{n=1}^{N} \alpha_n h(t - t_n) \tag{2.160}$$

We will use (2.160) to obtain the *superposition integral,* which expresses the response of an LTI system to an arbitrary input (with suitable restrictions) in terms of the impulse response of the system. Considering the arbitrary input signal $x(t)$ of Figure 2.14(a), we can represent it as

$$x(t) = \int_{-\infty}^{\infty} x(\lambda)\delta(t - \lambda)\, d\lambda \tag{2.161}$$

by the sifting property of the unit impulse. Approximating the integral of (2.161) as a sum, we obtain

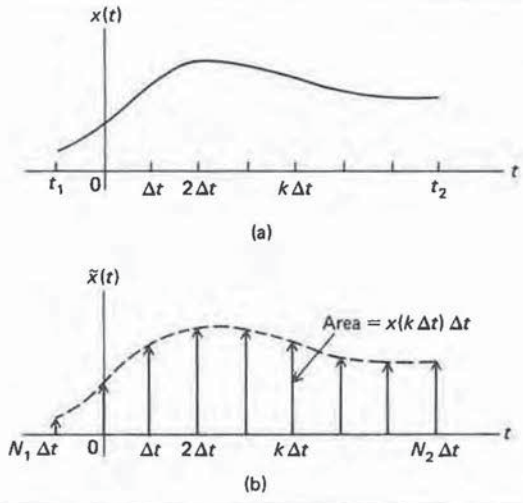
$$x(t) \cong \sum_{n=N_1}^{N_2} x(n\,\Delta t)\,\delta(t - n\,\Delta t)\,\Delta t, \qquad \Delta t \ll 1 \tag{2.162}$$

where $t_1 = N_1\,\Delta t$ is the starting time of the signal and $t_2 = N_2\,\Delta t$ is the ending time. The output, using (2.160) with $\alpha_n = x(n\,\Delta t)\,\Delta t$ and $t_n = n\,\Delta t$, is

$$\tilde{y}(t) = \sum_{n=N_1}^{N_2} x(n\,\Delta t)h(t - n\,\Delta t)\,\Delta t \tag{2.163}$$

where the tilde denotes the output resulting from the approximation to the input given by (2.162). In the limit as Δt approaches zero and $n\,\Delta t$

FIGURE 2.14 A signal and an approximate representation. (a) Signal. (b) Approximation with a sequence of impulses.



approaches the continuous variable λ, the sum becomes an integral, and we obtain

$$y(t) = \int_{-\infty}^{\infty} x(\lambda)h(t - \lambda)\, d\lambda \tag{2.164a}$$

where the limits have been changed to $\pm\infty$ to allow arbitrary starting and ending times for $x(t)$. Making the substitution $\sigma = t - \lambda$, we obtain the equivalent result

$$y(t) = \int_{-\infty}^{\infty} x(t - \sigma)h(\sigma)\, d\sigma \tag{2.164b}$$

Because these equations were obtained by superposition of a number of elementary responses due to each individual impulse, they are referred to as *superposition integrals*. A simplification results if the system under consideration is *causal*, that is, is a system that does not respond before an input is applied. For a causal system, $h(t - \lambda) = 0$ for $t < \lambda$ and the upper limit on (2.164a) can be set equal to t. Furthermore, if $x(t) = 0$ for $t < 0$, the lower limit becomes zero.

Stability

A fixed, linear system is bounded-input, bounded-output (BIBO) stable if every bounded input results in a bounded output. It can be shown* that a system is BIBO stable if and only if

$$\int_{-\infty}^{\infty} |h(t)|\, dt < \infty \tag{2.165}$$

Transfer Function

Applying the convolution theorem of Fourier transforms, item 8 of Table C.2 in Appendix C, to either (2.164a) or (2.164b), we obtain

$$Y(f) = H(f)X(f) \tag{2.166}$$

where $X(f) = \mathcal{F}\{x(t)\}$, $Y(f) = \mathcal{F}\{y(t)\}$, and

$$H(f) = \mathcal{F}\{h(t)\} = \int_{-\infty}^{\infty} h(t)e^{-j2\pi ft}\, dt \tag{2.167a}$$

or

$$h(t) = \mathcal{F}^{-1}\{H(f)\} = \int_{-\infty}^{\infty} H(f)e^{j2\pi ft}\, df \tag{2.167b}$$

* See Ziemer, Tranter, and Fannin (1993), Chapter 2.

$H(f)$ is referred to as the *transfer function* of the system. We see that either $h(t)$ or $H(f)$ is an equally good characterization of the system. By an inverse Fourier transform on (2.166), the output becomes

$$y(t) = \int_{-\infty}^{\infty} X(f)H(f)e^{j2\pi ft}\,df \tag{2.168}$$

Causality

A system is causal if it does not anticipate the input. In terms of the impulse response, it follows that for a time-invariant causal system,

$$h(t) = 0, \qquad t < 0 \tag{2.169}$$

When causality is viewed from the standpoint of the transfer function of the system, a celebrated theorem by Wiener and Paley* states that if

$$\int_{-\infty}^{\infty} |h(t)|^2\,dt = \int_{-\infty}^{\infty} |H(f)|^2\,df < \infty \tag{2.170a}$$

with $h(t) \equiv 0$ for $t < 0$, it is then necessary that

$$\int_{-\infty}^{\infty} \frac{|\ln|H(f)||}{1+f^2}\,df < \infty \tag{2.170b}$$

Conversely, if $|H(f)|$ is square-integrable, and if the integral in (2.170b) is unbounded, then we cannot make $h(t) \equiv 0$, $t < 0$, no matter what we choose for $\underline{/H(f)}$. Consequences of (2.170) are that no filter can have $|H(f)| \equiv 0$ over a finite band of frequencies (i.e., a filter cannot perfectly reject any band of frequencies). In fact, the Paley-Wiener criterion restricts the rate at which $|H(f)|$ for a linear causal time-invariant system can vanish. For example,

$$|H(f)| = e^{-k_1|f|} \tag{2.171a}$$

and

$$|H(f)| = e^{-k_2 f^2} \tag{2.171b}$$

are not allowable amplitude responses for causal filters because (2.170b) does not give a finite result in either case.

The sufficiency statement of the Payley-Wiener criterion is: Given any square-integrable function $|H(f)|$ for which (2.170) is satisfied, there exists an $\underline{/H(f)}$ such that $H(f) = |H(f)|e^{j\underline{/H(f)}}$ is the Fourier transform of $h(t)$ for a causal filter.

* See William Siebert, *Circuits, Signals, and Systems*, McGraw-Hill, New York, 1986, p. 476.

Symmetry Properties of H(f)

The transfer function $H(f)$ is, in general, a complex quantity. We therefore write it in terms of magnitude and argument as

$$H(f) = |H(f)|e^{j\underline{/H(f)}} \tag{2.172}$$

where $|H(f)|$ is called the *amplitude-response function* and $\underline{/H(f)}$ is called the *phase-shift function* of the network. Also, $H(f)$ is the Fourier transform of a real-time function $h(t)$. Therefore, it follows that

$$|H(f)| = |H(-f)| \tag{2.173a}$$

and

$$\underline{/H(f)} = -\underline{/H(-f)} \tag{2.173b}$$

▼ EXAMPLE 2.20 Consider the lowpass RC filter shown in Figure 2.15. We may find its transfer function by a number of methods. First, we may write down the governing differential equation (integral-differential equations, in general) as

$$RC\frac{dy}{dt} + y(t) = x(t) \tag{2.174}$$

and Fourier-transform it, obtaining

$$(j2\pi fRC + 1)Y(f) = X(f)$$

or

$$H(f) = \frac{Y(f)}{X(f)} = \frac{1}{1 + j(f/f_3)}$$

$$= \frac{1}{\sqrt{1 + (f/f_3)^2}} e^{-j \tan^{-1}(f/f_3)} \tag{2.175}$$

where $f_3 = 1/2\pi RC$ is the 3-dB frequency, or half-power frequency. Second, we can use Laplace transform theory with s replaced by $j2\pi f$. Third, we can

FIGURE 2.15 An RC network

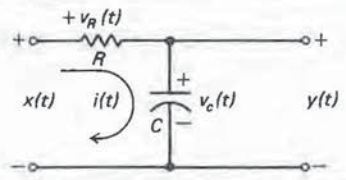
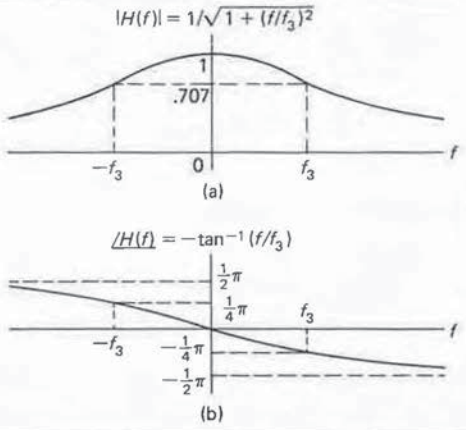


FIGURE 2.16 Amplitude and phase responses of the lowpass RC filter. (a) Amplitude response. (b) Phase shift.



use ac sinusoidal steady-state analysis. The amplitude response and phase response of this system are illustrated in Figures 2.16(a) and (b).

Using the transform pair

$$\alpha e^{-\alpha t}u(t) \leftrightarrow \frac{\alpha}{\alpha + j2\pi f} \tag{2.176}$$

we find the impulse response of the filter to be

$$h(t) = \frac{1}{RC}\, e^{-t/RC}\, u(t) \tag{2.177}$$

Finally, we consider the response of the filter to the pulse

$$x(t) = A\Pi\left(\frac{t - \frac{1}{2}T}{T}\right) \tag{2.178}$$

Using appropriate Fourier transform pairs, we can readily find $Y(f)$, but its inverse Fourier transformation requires some effort. Thus it appears that the superposition integral is the best approach in this case. Choosing the form

$$y(t) = \int_{-\infty}^{\infty} h(t - \sigma)x(\sigma)\, d\sigma \tag{2.179}$$

we find, by direct substitution in $h(t)$, that

$$h(t - \sigma) = \frac{1}{RC}\, e^{-(t-\sigma)/RC}u(t - \sigma) = \begin{cases} \dfrac{1}{RC}\, e^{-(t-\sigma)/RC}, & \sigma < t \\ 0, & \sigma > t \end{cases} \tag{2.180}$$

Since $x(\sigma)$ is zero for $\sigma < 0$ and $\sigma > T$, we find that

$$y(t) = \begin{cases} 0, & t < 0 \\ \displaystyle\int_0^t \frac{A}{RC} e^{-(t-\sigma)/RC}\, d\sigma, & 0 < t < T \\ \displaystyle\int_0^T \frac{A}{RC} e^{-(t-\sigma)/RC}\, d\sigma, & t > T \end{cases} \tag{2.181}$$

Carrying out the integrations, we obtain

$$y(t) = \begin{cases} 0, & t < 0 \\ A(1 - e^{-t/RC}), & 0 < t < T \\ A(e^{-(t-T)/RC} - e^{-t/RC}), & t > T \end{cases} \tag{2.182}$$

This result is plotted in Figure 2.17 for several values of T/RC. Also shown are $|X(f)|$ and $|H(f)|$. Note that $T/RC = 2\pi f_3/T^{-1}$ is proportional to the ratio of the 3-dB frequency of the filter to the spectral width (T^{-1}) of the pulse. When this ratio is large, the spectrum of the input pulse is essentially passed undistorted by the system, and the output looks like the input. On the other hand, for $2\pi f_3/T^{-1} \ll 1$, the system distorts the input signal spectrum, and $y(t)$ looks nothing like the input. These ideas will be put on a firmer basis when signal distortion is discussed. ▲

Input-Output Relationships for Spectra

Consider a fixed linear two-port system with transfer function $H(f)$, input $x(t)$, and output $y(t)$. If $x(t)$ and, therefore, $y(t)$ are energy signals, their energy spectral densities are $G_x(f) = |X(f)|^2$ and $G_y(f) = |Y(f)|^2$, respectively. Since $Y(f) = H(f)X(f)$, it follows that

$$G_y(f) = |H(f)|^2 G_x(f) \tag{2.183}$$

A similar relationship holds for power signals and spectra:

$$S_y(f) = |H(f)|^2 S_x(f) \tag{2.184}$$

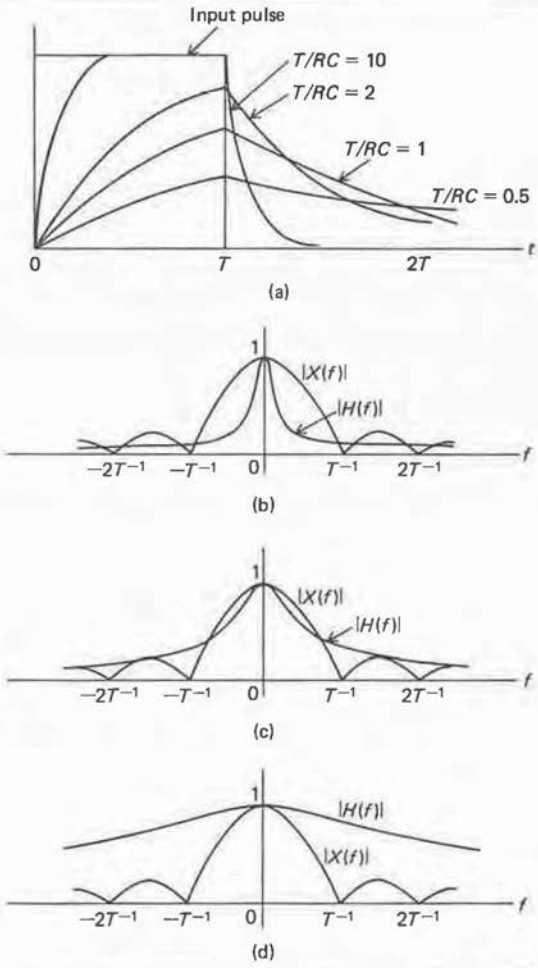
This will be proved in Chapter 5.

Response to Periodic Inputs

Consider the steady-state response of a fixed linear system to the complex exponential input signal $Ae^{j2\pi f_0 t}$. Using the superposition integral, we obtain

$$y_{ss}(t) = \int_{-\infty}^{\infty} h(\lambda) A e^{j2\pi f_0 (t - \lambda)}\, d\lambda$$

FIGURE 2.17 (a) Waveforms and (b)–(d) spectra for a lowpass RC filter with pulse input. (a) Input and output signals. (b) $T/RC = 0.5$. (c) $T/RC = 2$. (d) $T/RC = 10$.



$$= Ae^{j2\pi f_0 t} \int_{-\infty}^{\infty} h(\lambda) e^{-j2\pi f_0 \lambda}\, d\lambda$$

$$= H(f_0) A e^{j2\pi f_0 t} \tag{2.185}$$

That is, the output is a complex exponential signal of the same frequency but with amplitude scaled by $|H(f_0)|$ and phase-shifted by $\angle H(f_0)$ relative to the amplitude and phase of the input. Using superposition, we conclude

that the steady-state output due to an arbitrary periodic input is represented by the complex exponential Fourier series

$$y(t) = \sum_{n=-\infty}^{\infty} X_n H(nf_0) e^{jn2\pi f_0 t} \tag{2.186}$$

or

$$y(t) = \sum_{n=-\infty}^{\infty} |X_n||H(nf_0)| \exp\{j[n2\pi f_0 t + \underline{/X_n} + \underline{/H(nf_0)}]\} \tag{2.187}$$

Thus, for a periodic input, the magnitude of each spectral component of the input is attenuated (or amplified) by the amplitude-response function *at the frequency of the particular spectral component,* and the phase of each spectral component is shifted by the value of the phase-shift function of the system *at the frequency of the spectral component.*

▼ EXAMPLE 2.21 Consider the response of a filter having the transfer function

$$H(f) = 2\Pi\left(\frac{f}{42}\right) e^{-j\pi f/10} \tag{2.188}$$

to a unit-amplitude triangular signal with period 0.1 s. From Table 2.1 and Equation (2.43), the exponential Fourier series of the input signal is

$$\begin{aligned} x(t) &= \cdots \frac{4}{25\pi^2} e^{-j100\pi t} + \frac{4}{9\pi^2} e^{-j60\pi t} + \frac{4}{\pi^2} e^{-j20\pi t} + \frac{4}{\pi^2} e^{j20\pi t} \\ &\quad + \frac{4}{9\pi^2} e^{j60\pi t} + \frac{4}{25\pi^2} e^{j100\pi t} + \cdots \\ &= \frac{8}{\pi^2}\left[\cos(20\pi t) + \frac{1}{9}\cos(60\pi t) + \frac{1}{25}\cos(100\pi t) + \cdots\right] \end{aligned} \tag{2.189}$$

The filter eliminates all harmonics above 21 Hz and passes all those below 21 Hz, imposing an amplitude scale factor of 2 and a phase shift of $-\pi f/10$ rad. The only harmonic of the triangular wave to be passed by the filter is the fundamental, which has a frequency of 10 Hz, giving a phase shift of $-\pi(10)/10 = -\pi$ rad. The output is therefore

$$y(t) = \frac{16}{\pi^2}\cos 20\pi\left(t - \frac{1}{20}\right) \tag{2.190}$$

▲ where the phase shift is seen to be equivalent to a delay of $\frac{1}{20}$ s.

Distortionless Transmission

Equation (2.187) shows that both the amplitudes and phases of the spectral components of a periodic signal will, in general, be altered as the signal is sent through a two-port system. This modification may be desirable in signal *processing* applications, but it amounts to distortion in signal *transmission* applications. While it may appear at first that ideal signal transmission results only if there is *no* attenuation and phase of the spectral components of the input, this requirement is too stringent. A system will be classified as distortionless if it introduces the same attenuation and time delay to all spectral components of the input, for then the output *looks like* the input. In particular, if the output of a system is given in terms of the input as

$$y(t) = H_0 x(t - t_0) \tag{2.191}$$

where H_0 and t_0 are constants, the output is a scaled, delayed replica of the input ($t_0 > 0$ for causality). Employing the time-delay theorem to Fourier-transform (2.191) and using the relation $H(f) = Y(f)/X(f)$, we obtain

$$H(f) = H_0 e^{-j2\pi f t_0} \tag{2.192}$$

as the transfer function of a distortionless system; that is, the amplitude response of a distortionless system is constant and the phase shift is linear with frequency. Of course, these restrictions are necessary only within the frequency ranges where the input has significant spectral content. Figure 2.18 and Example 2.22 will illustrate these comments.

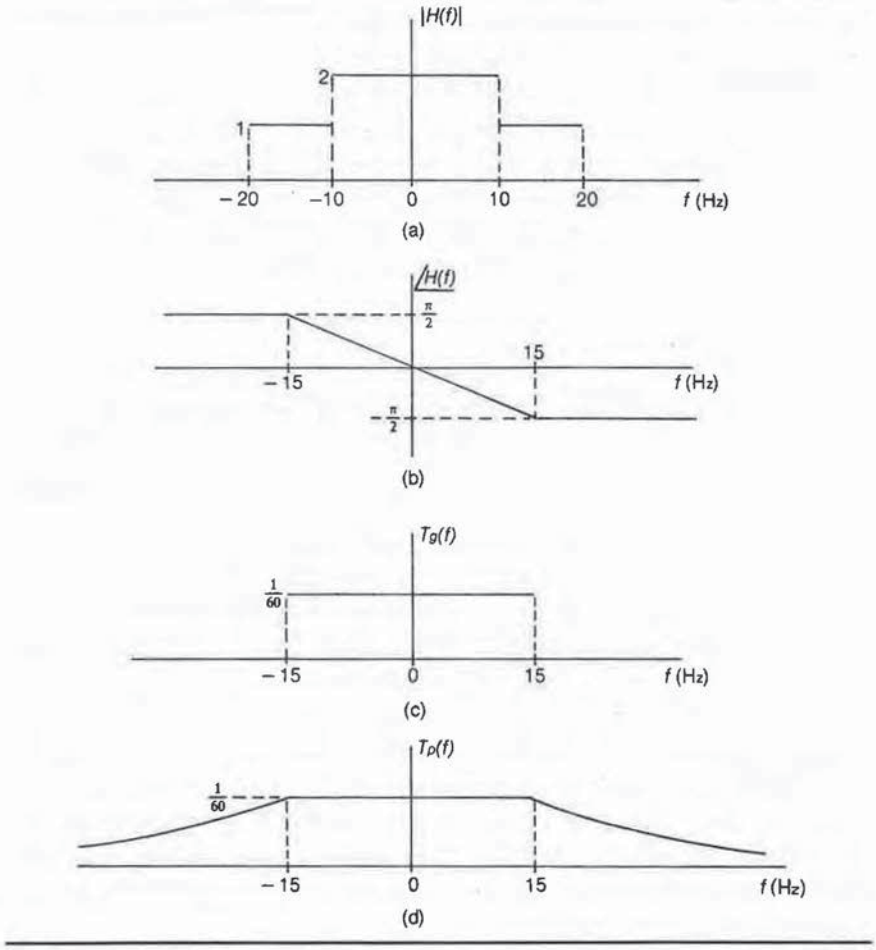
In general, we can isolate three major types of distortion. First, if the system is linear but the amplitude response is not constant with frequency, the system is said to introduce *amplitude distortion*. Second, if the system is linear but the phase shift is not a linear function of frequency, the system introduces *phase,* or *delay, distortion*. Third, if the system is not linear, we have *nonlinear distortion*. Of course, these three types of distortion may occur in combination with one another.

Group and Phase Delay

One can often identify phase distortion in a linear system by considering the derivative of phase with respect to frequency. A distortionless system exhibits a phase response in which phase is directly proportional to frequency. Thus the derivative of phase with respect to frequency of a distortionless system is a constant. The negative of this constant is called the *group delay* of the system. In other words, the *group delay* is defined by the equation

$$T_g(f) = -\frac{1}{2\pi}\frac{d\theta(f)}{df} \tag{2.193}$$

FIGURE 2.18 Amplitude and phase response and group and phase delays of the filter for Example 2.22. (a) Amplitude response. (b) Phase response. (c) Group delay. (d) Phase delay.



in which $\theta(f)$ is the phase response of the system. For a distortionless system, the phase is given by (2.192) as

$$\theta(f) = -2\pi f t_0 \tag{2.194}$$

This yields a group delay of

$$T_g(f) = -\frac{1}{2\pi}\frac{d}{df}(-2\pi f t_0)$$

or

$$T_g(f) = t_0 \tag{2.195}$$

This confirms the preceding observation that the group delay of a distortionless system is a constant.

Group delay is the delay that a group of two or more frequency components undergo in passing through a linear system. If a linear system has a single-frequency component as the input, the system is always distortionless, since the output can be written as an amplitude-scaled and phase-shifted (time-delayed) version of the input. As an example, assume that the input to a linear system is given by

$$x(t) = A \cos 2\pi f_1 t \tag{2.196}$$

It follows from (2.187) that the output can be written as

$$y(t) = A|H(f_1)| \cos [2\pi f_1 t + \theta(f_1)] \tag{2.197}$$

where $\theta(f_1)$ is the phase response of the system evaluated at $f = f_1$. Equation (2.197) can be written as

$$y(t) = A|H(f_1)| \cos 2\pi f_1 \left[t + \frac{\theta(f_1)}{2\pi f_1} \right] \tag{2.198}$$

The delay of the single component is defined as the phase delay:

$$T_p(f) = -\frac{\theta(f)}{2\pi f} \tag{2.199}$$

Thus (2.198) can be written as

$$y(t) = A|H(f_1)| \cos 2\pi f_1 [t - T_p(f_1)] \tag{2.200}$$

Use of (2.194) shows that for a distortionless system, the phase delay is given by

$$T_p(f) = -\frac{1}{2\pi f}(-2\pi f t_0) = t_0 \tag{2.201}$$

Thus we see that distortionless systems have equal group and phase delays. The following example should clarify the preceding definitions.

▼ **EXAMPLE 2.22** Consider a system with amplitude response and phase shift as shown in Figure 2.18 and the following four inputs:

1. $x_1(t) = \cos 10\pi t + \cos 12\pi t$
2. $x_2(t) = \cos 10\pi t + \cos 26\pi t$

3. $x_3(t) = \cos 26\pi t + \cos 34\pi t$
4. $x_4(t) = \cos 32\pi t + \cos 34\pi t$

Although this system is somewhat unrealistic from a practical standpoint, we can use it to illustrate various combinations of amplitude and phase distortion. Using (2.187) and superposition, we obtain the following corresponding outputs:

1. $y_1(t) = 2\cos(10\pi t - \frac{1}{6}\pi) + 2\cos(12\pi t - \frac{1}{5}\pi)$
   $= 2\cos 10\pi(t - \frac{1}{60}) + 2\cos 12\pi(t - \frac{1}{60})$
2. $y_2(t) = 2\cos(10\pi t - \frac{1}{6}\pi) + \cos(26\pi t - \frac{13}{30}\pi)$
   $= 2\cos 10\pi(t - \frac{1}{60}) + \cos 26\pi(t - \frac{1}{60})$
3. $y_3(t) = \cos(26\pi t - \frac{13}{30}\pi) + \cos(34\pi t - \frac{1}{2}\pi)$
   $= \cos 26\pi(t - \frac{1}{60}) + \cos 34\pi(t - \frac{1}{68})$
4. $y_4(t) = \cos(32\pi t - \frac{1}{2}\pi) + \cos(34\pi t - \frac{1}{2}\pi)$
   $= \cos 32\pi(t - \frac{1}{64}) + \cos 34\pi(t - \frac{1}{68})$

Checking these results with (2.191), we see that only the input $x_1(t)$ is passed without distortion by the system. For $x_2(t)$, amplitude distortion results, and for $x_3(t)$ and $x_4(t)$, phase (delay) distortion is introduced.

The group delay and phase delay are also illustrated in Figure 2.18. It can be seen that for $|f| \leq 15$ Hz, the group and phase delays are both equal to $\frac{1}{60}$ s. For $|f| > 15$ Hz, the group delay is zero, and the phase delay is

$$T_p(f) = \frac{1}{4f} \tag{2.202}$$

▲

Nonlinear Distortion

To illustrate the idea of nonlinear distortion, let us consider a nonlinear system with the input-output characteristic

$$y(t) = a_1 x(t) + a_2 x^2(t) \tag{2.203}$$

and with the input

$$x(t) = A_1 \cos \omega_1 t + A_2 \cos \omega_2 t \tag{2.204}$$

The output is therefore

$$y(t) = a_1(A_1 \cos \omega_1 t + A_2 \cos \omega_2 t) + a_2(A_1 \cos \omega_1 t + A_2 \cos \omega_2 t)^2 \tag{2.205}$$

Using trigonometric identities, we can write the output as

$$\begin{aligned} y(t) = {} & a_1(A_1 \cos \omega_1 t + A_2 \cos \omega_2 t) \\ & + \tfrac{1}{2}a_2(A_1^2 + A_2^2) + \tfrac{1}{2}a_2(A_1^2 \cos 2\omega_1 t + A_2^2 \cos 2\omega_2 t) \\ & + a_2 A_1 A_2[\cos(\omega_1 + \omega_2)t + \cos(\omega_1 - \omega_2)t] \end{aligned} \tag{2.206}$$

As can be seen in Figure 2.19, the system has produced frequencies in the output other than the frequencies of the input. In addition to the first term in (2.206), which may be considered the desired output, there are distortion terms at harmonics of the input frequencies as well as distortion terms involving sums and differences of the harmonics of the input frequencies. The former are referred to as *harmonic distortion terms,* and the latter are referred to as *intermodulation distortion terms.* Note that a second-order nonlinearity could be used as a device to *double* the frequency of an input sinusoid. Third-order nonlinearities can be used as *triplers,* and so forth.

A general input signal can be handled by applying the multiplication theorem given in Table C.2 in Appendix C. Thus, for the nonlinear system with the transfer characteristic given by (2.203), the output spectrum is

$$Y(f) = a_1 X(f) + a_2 X(f) * X(f) \tag{2.207}$$

The second term, considered as distortion, is seen to give interference at all frequencies occupied by the desired output (the first term), and it is impos-

FIGURE 2.19 Input and output spectra for nonlinear system with discrete frequency input. (a) Input spectrum. (b) Output spectrum.

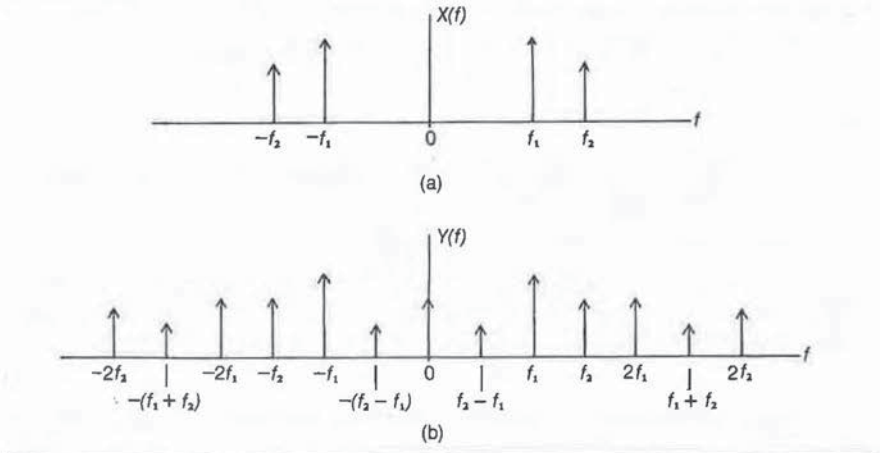
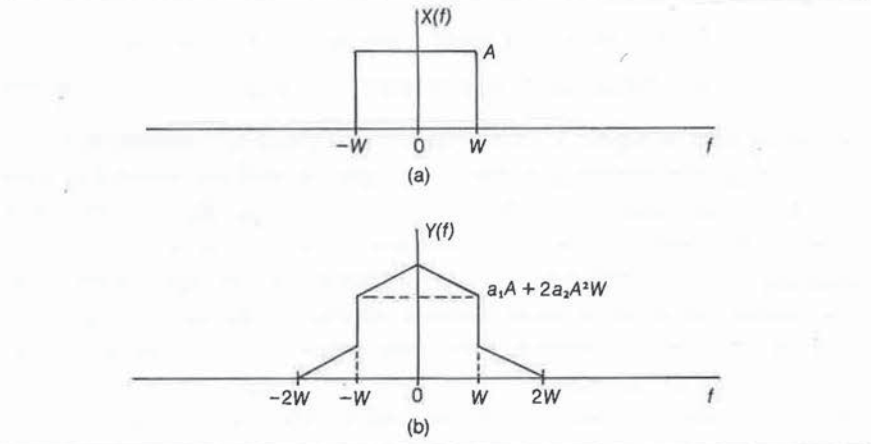


FIGURE 2.20 Input and output spectra for nonlinear system with continuous frequency input. (a) Input spectrum. (b) Output spectrum.



sible to isolate harmonic and intermodulation distortion components as before. For example, if

$$X(f) = A\Pi\left(\frac{f}{2W}\right) \tag{2.208}$$

Then the distortion term is

$$a_2X(f) * X(f) = 2a_2WA^2\Lambda\left(\frac{f}{2W}\right) \tag{2.209}$$

The input and output spectra are shown in Figure 2.20. Note that the spectral width of the distortion term is *double* that of the input.

Ideal Filters

It is often convenient to work with filters having idealized transfer functions with rectangular amplitude-response functions that are constant within the passband and zero elsewhere. We will consider three general types of ideal filters: lowpass, highpass, and bandpass. Within the passband, a linear phase response characteristic is assumed. Thus, if B is the single-sided bandwidth *(width of the stopband* for the highpass filter)* of the filter in ques-

* The *stopband* of a filter will be defined here as the frequency range(s) for which $|H(f)|$ is below 3 dB of its maximum value.

tion, the transfer functions of these three ideal filters can be written as follows:

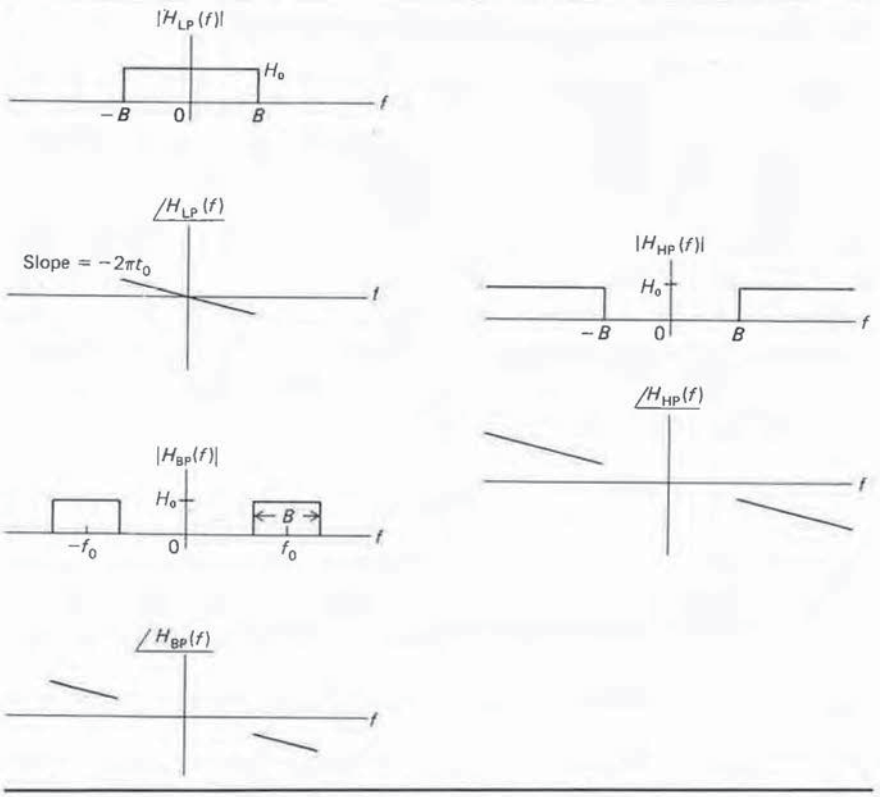
1. $H_{LP}(f) = H_0\Pi(f/2B)e^{-j2\pi ft_0}$ (lowpass)
2. $H_{HP}(f) = H_0[1 - \Pi(f/2B)]e^{-j2\pi ft_0}$ (highpass)
3. $H_{BP}(f) = [H_l(f - f_0) + H_l(f + f_0)]e^{-j2\pi ft_0}$, where $H_l(f) = H_0\Pi(f/B)$ (bandpass)

The amplitude- and phase-response functions for these filters are shown in Figure 2.21.

The corresponding impulse responses are obtained by inverse Fourier transformation. For example, the impulse response of an ideal lowpass filter is, from Example 2.11 and the time-delay theorem, given by

$$h_{LP}(t) = 2BH_0 \text{ sinc } [2B(t - t_0)] \tag{2.210}$$

FIGURE 2.21 Amplitude- and phase-response functions for ideal filters.



Since $h_{LP}(t)$ is not zero for $t < 0$, we see that an ideal lowpass filter is noncausal. Nevertheless, ideal filters are useful concepts because they simplify calculations and can give satisfactory results.

Turning to the ideal bandpass filter, we may use the modulation theorem to write its impulse response as

$$h_{BP}(t) = 2h_l(t - t_0) \cos 2\pi f_0(t - t_0) \tag{2.211}$$

where

$$h_l(t) = \mathcal{F}^{-1}[H_l(f)] = H_0 B \text{ sinc } Bt \tag{2.212}$$

Thus the impulse response of an ideal bandpass filter is the oscillatory signal

$$h_{BP}(t) = 2H_0 B \text{ sinc } B(t - t_0) \cos 2\pi f_0(t - t_0) \tag{2.213}$$

Figure 2.22 illustrates $h_{LP}(t)$ and $h_{BP}(t)$. If $f_0 \gg B$, it is convenient to view $h_{BP}(t)$ as the slowly varying envelope $2H_0$ sinc Bt *modulating* the high-frequency oscillatory signal $\cos 2\pi f_0 t$ and shifted to the right by t_0 seconds.

*Approximation of Ideal Lowpass Filters by Realizable Filters**

Although ideal filters are noncausal and therefore unrealizable devices, there are several practical filter types that may be designed to approximate ideal filter characteristics as closely as desired. In this section we consider three such approximations for the lowpass case. Bandpass and highpass approximations may be obtained through suitable frequency transformation. The three filter types to be considered are (1) Butterworth, (2) Chebyshev, and (3) Bessel.

The Butterworth filter is a filter design chosen to maintain a constant amplitude response in the passband at the cost of less stopband attenuation. An nth-order Butterworth filter is characterized by a transfer function, in terms of the complex frequency s, of the form

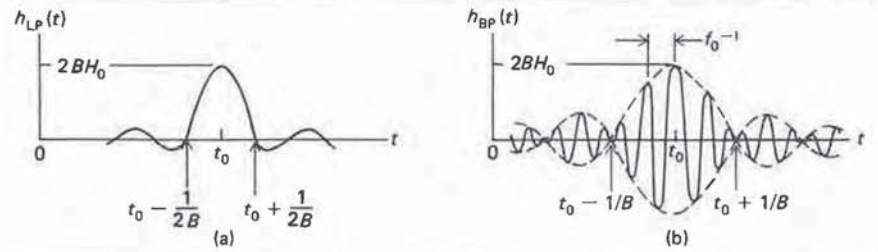
$$H_{BU}(s) = \frac{\omega_3{}^n}{(s - s_1)(s - s_2) \cdots (s - s_n)} \tag{2.214}$$

where the poles $s_1, s_2, \ldots, s_n$ are symmetrical with respect to the real axis and equally spaced about a semicircle of radius ω_3 in the left half s-plane and $f_3 = \omega_3/2\pi$ is the 3-dB cutoff frequency.† Typical pole locations are

* See Williams and Taylor (1988), Chapter 2, for a detailed discussion of classical filter designs.

† From basic circuit theory courses you will recall that the poles and zeros of a rational function of s, $H(s) = N(s)/D(s)$, are those values of complex frequency $s \triangleq \sigma + j\omega$ for which $D(s) = 0$ and $N(s) = 0$, respectively. See, for example, Cunningham and Stuller (1991).

FIGURE 2.22 Impulse responses for ideal lowpass and bandpass filters. (a) $h_{LP}(t)$. (b) $h_{BP}(t)$.



shown in Figure 2.23(a). The amplitude response for an nth-order Butterworth filter is of the form

$$|H_{BU}(f)| = \frac{1}{\sqrt{1 + (f/f_3)^{2n}}} \tag{2.215}$$

Note that as n approaches infinity, $|H_{BU}(f)|$ approaches an ideal lowpass filter characteristic. However, the filter delay also approaches infinity.

The Chebyshev lowpass filter has an amplitude response chosen to maintain a maximum allowable attenuation in the passband while maximizing the attenuation in the stopband. A typical pole-zero diagram is shown in Figure 2.23(b). The amplitude response of a Chebyshev filter is of the form

$$|H_C(f)| = \frac{1}{\sqrt{1 + \epsilon^2 C_n^2(f)}} \tag{2.216}$$

The parameter ϵ is specified by the maximum allowable attenuation in the passband, and $C_n(\ )$, known as a *Chebyshev polynomial,* is given by the recursion relation

$$C_n(f) = 2\left(\frac{f}{f_c}\right) C_{n-1}(f) - C_{n-2}(f), \qquad n = 2, 3, \ldots \tag{2.217}$$

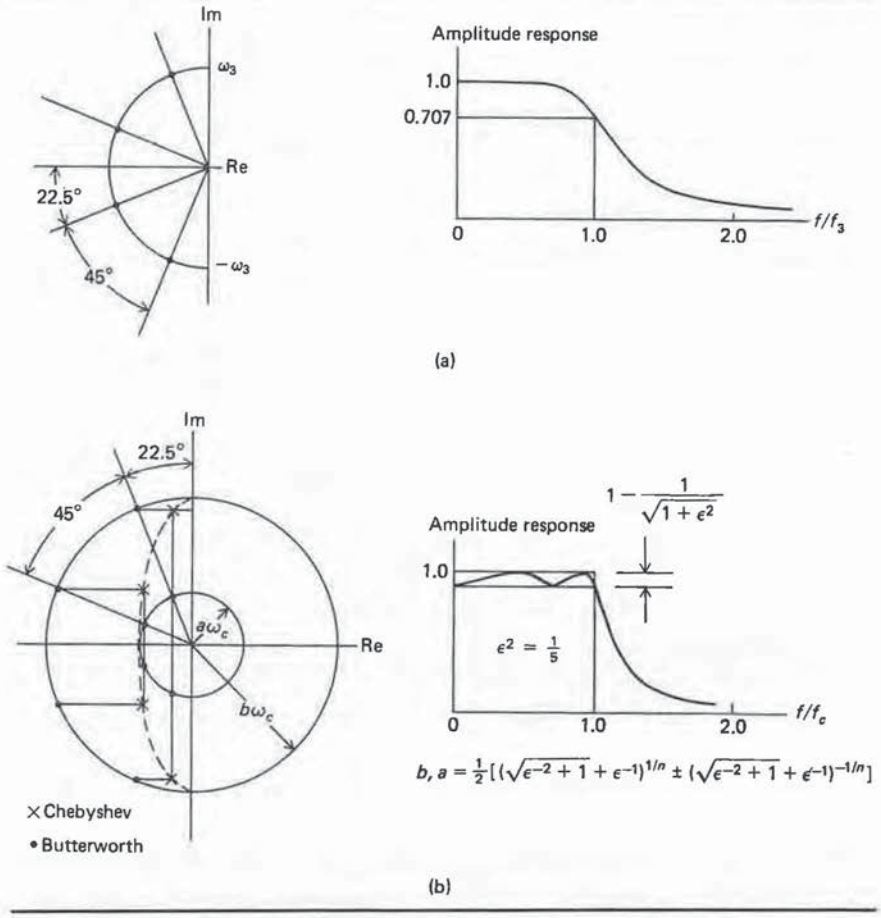
where

$$C_1(f) = \frac{f}{f_c} \qquad \text{and} \qquad C_0(f) = 1 \tag{2.218}$$

Regardless of the value of n, it turns out that $C_n(f_c) = 1$, so that $|H_c(f_c)| = (1 + \epsilon^2)^{-1/2}$. (Note that f_c is not necessarily the 3-dB frequency here.)

The Bessel lowpass filter is a design that attempts to maintain a linear phase response in the passband at the expense of the amplitude response.

FIGURE 2.23 Pole locations and amplitude responses for fourth-order Butterworth and Chebyshev filters. (a) Butterworth filter. (b) Chebyshev filter.



The cutoff frequency of a Bessel filter is defined by

$$f_c = (2\pi t_0)^{-1} = \frac{\omega_c}{2\pi} \tag{2.219}$$

where t_0 is the nominal delay of the filter. The transfer function of an nth-order Bessel filter is given by

$$H_{BE}(f) = \frac{K_n}{B_n(f)} \tag{2.220}$$

where K_n is a constant chosen to yield $H(0) = 1$, and $B_n(f)$ is a Bessel polynomial of order n defined by

$$B_n(f) = (2n - 1)B_{n-1}(f) - \left(\frac{f}{f_c}\right)^2 B_{n-2}(f) \tag{2.221}$$

where

$$B_0(f) = 1 \qquad \text{and} \qquad B_1(f) = 1 + j\left(\frac{f}{f_c}\right) \tag{2.222}$$

Figure 2.24 illustrates the amplitude and group delay characteristics of third-order Butterworth, Bessel, and Chebyshev filters. All four filters are normalized to have 3-dB amplitude attenuation at a frequency of f_c Hz. The amplitude responses show that the Chebyshev filters have more attenuation than the Butterworth and Bessel filters do for frequencies exceeding the 3-dB frequency. Increasing the passband ($f < f_c$) ripple increases the stopband ($f > f_c$) attenuation.

The group delay characteristics shown in Figure 2.24(b) illustrate, as expected, that the Bessel filter has the most constant group delay. Comparison of the Butterworth and the 0.1-dB ripple Chebyshev group delays shows that although the group delay of the Chebyshev filter has a higher peak, it has a more constant group delay for frequencies less than about $0.3f_c$.

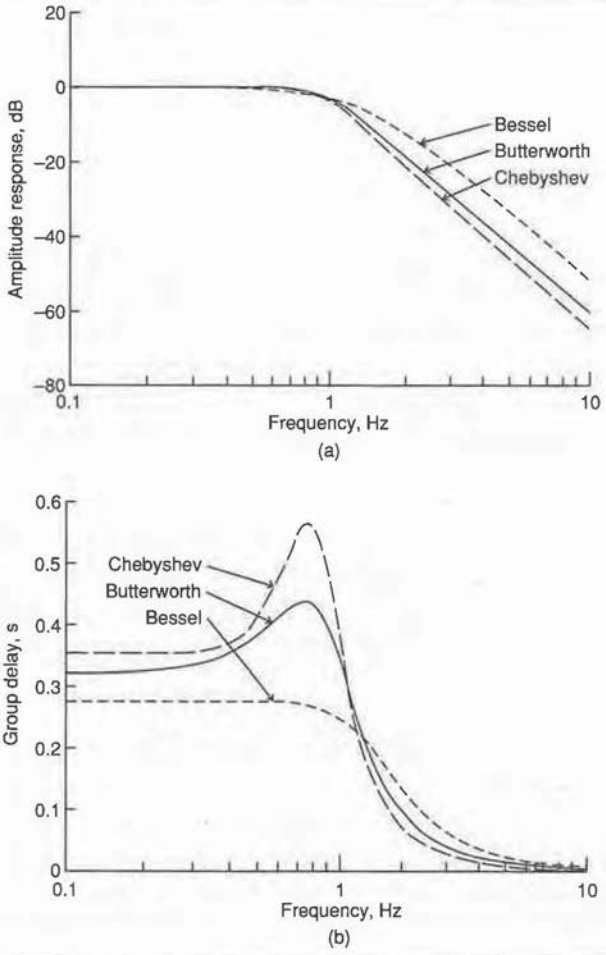
Relation of Pulse Resolution and Risetime to Bandwidth

In our consideration of signal distortion, we assumed bandlimited signal spectra. We found that the input signal to a filter is merely delayed and attenuated if the filter has constant amplitude response and linear phase response throughout the passband of the signal. But suppose the input signal is not bandlimited. What rule of thumb can we use to estimate the required bandwidth? This is a particularly important problem in pulse transmission, where the detection and resolution of pulses at a filter output are of interest.

A satisfactory definition for pulse duration and bandwidth, and the relationship between them, is obtained by consulting Figure 2.25. In Figure 2.25(a), a pulse with a single maximum, taken at $t = 0$ for convenience, is shown with a rectangular approximation of height $x(0)$ and duration T. It is required that the approximating pulse and $|x(t)|$ have equal areas. Thus

$$Tx(0) = \int_{-\infty}^{\infty} |x(t)|\, dt \geq \int_{-\infty}^{\infty} x(t)\, dt = X(0) \tag{2.223a}$$

FIGURE 2.24 Comparison of third-order Butterworth, Chebyshev (0.1-dB ripple), and Bessel filters. (a) Amplitude response. (b) Group delay. All filters are designed to have a 1-Hz, 3-dB bandwidth.



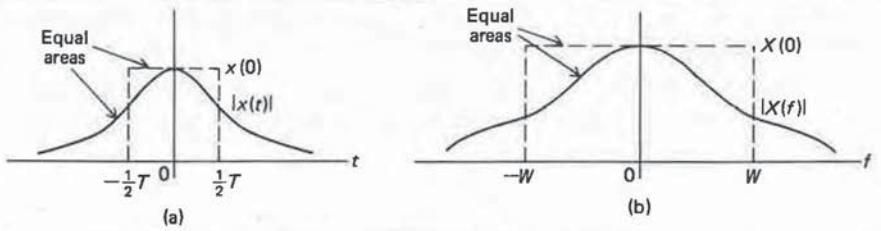
where we have used the relationship

$$X(0) = \mathcal{F}[x(t)]\Big|_{f=0} = \int_{-\infty}^{\infty} x(t)e^{-j2\pi t\cdot 0}\,dt \qquad (2.223b)$$

Turning to Figure 2.25(b), we obtain a similar inequality for the rectangular approximation to the pulse spectrum. Specifically, we may write

$$2WX(0) = \int_{-\infty}^{\infty} |X(f)|\,df \geq \int_{-\infty}^{\infty} X(f)\,df = x(0) \qquad (2.224)$$

FIGURE 2.25 Arbitrary pulse signal and spectrum. (a) Pulse and rectangular approximation. (b) Amplitude spectrum and rectangular approximation.



where we have used the relationship

$$x(0) = \mathcal{F}^{-1}[X(f)]\Big|_{t=0} = \int_{-\infty}^{\infty} X(f)e^{j2\pi f \cdot 0}\, df \tag{2.225}$$

Thus we have the pair of inequalities

$$\frac{x(0)}{X(0)} \geq \frac{1}{T} \quad \text{and} \quad 2W \geq \frac{x(0)}{X(0)} \tag{2.226}$$

which, when combined, result in the relationship of pulse duration and bandwidth

$$2W \geq \frac{1}{T} \tag{2.227a}$$

or

$$W \geq \frac{1}{2T}\ \text{Hz} \tag{2.227b}$$

Other definitions of pulse duration and bandwidth could have been used, but a relationship similar to (2.227) would have resulted.

This inverse relationship between pulse duration and bandwidth has been illustrated by all the examples involving pulse spectra that we have considered so far (for example, Examples 2.8, 2.11, 2.13). If pulses with bandpass spectra are considered, the relationship is

$$W \geq \frac{1}{T} \tag{2.228}$$

This is illustrated by Example 2.16.

A result similar to (2.227) also holds between the *risetime* T_R and bandwidth of a pulse. A suitable definition of *risetime* is the time required for a

pulse's leading edge to go from 10% to 90% of its final value. For the bandpass case, (2.228) holds with T replaced by T_R, where T_R is the risetime of the *envelope* of the pulse.

Risetime can be used as a measure of a system's distortion. To see how this is accomplished, we will express the step response of a filter in terms of its impulse response. From the superposition integral of (2.164b), with $x(t - \sigma) = u(t - \sigma)$, the step response of a filter with impulse response $h(t)$ is

$$y_s(t) = \int_{-\infty}^{\infty} h(\sigma)u(t - \sigma)\, d\sigma$$
$$= \int_{-\infty}^{t} h(\sigma)\, d\sigma \tag{2.229}$$

This follows because $u(t - \sigma) = 0$ for $\sigma > t$. Therefore, the step response of a linear system is the integral of its impulse response. This is not too surprising, since the unit step is the integral of a unit impulse.*

Examples 2.23 and 2.24 demonstrate how the risetime of a system's output due to a step input is a measure of the fidelity of the system.

▼ EXAMPLE 2.23 The impulse response of a lowpass RC filter is given by

$$h(t) = \frac{1}{RC} e^{-t/RC} u(t) \tag{2.230}$$

for which the step response is found to be

$$y_s(t) = (1 - e^{-2\pi f_3 t})u(t) \tag{2.231}$$

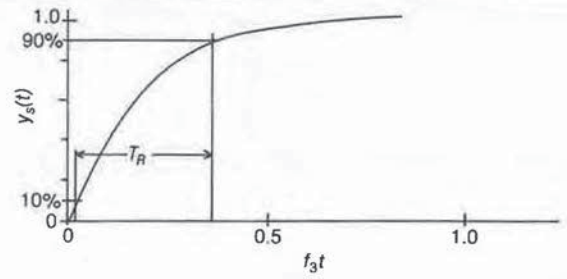
where the 3-dB bandwidth of the filter, defined following (2.175), has been used. The step response is plotted in Figure 2.26(a), where it is seen that the 10% to 90% risetime is approximately

$$T_R \cong \frac{0.35}{f_3} = 2.2RC \tag{2.232}$$

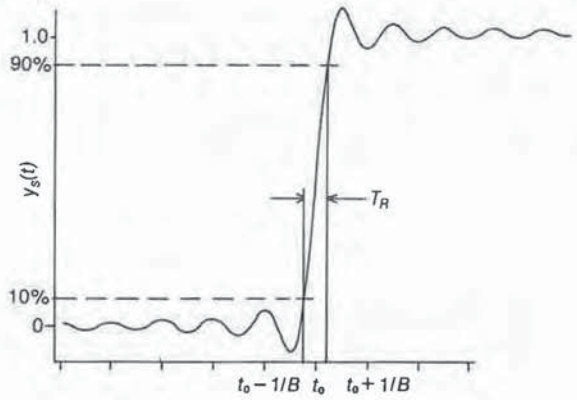
which demonstrates the inverse relationship between bandwidth and risetime. ▲

* This result is a special case of a more general result for an LTI system: if the response of a system to a given input is known and that input is modified through a linear operation, such as integration, then the output to the modified input is obtained by performing the same linear operation on the output to the original input.

FIGURE 2.26 Step response of (a) a lowpass RC filter and (b) an ideal lowpass filter, illustrating 10% to 90% risetime of each.



(a)



(b)

▼ **EXAMPLE 2.24** Using (2.210) with $H_0 = 1$, the step response of an ideal lowpass filter is

$$y_s(t) = \int_{-\infty}^{t} 2B \operatorname{sinc}[2B(\sigma - t_0)]\, d\sigma$$

$$= \int_{-\infty}^{t} 2B \frac{\sin[2\pi B(\sigma - t_0)]}{2\pi B(\sigma - t_0)}\, d\sigma \qquad (2.233)$$

By changing variables in the integrand to $u = 2\pi B(\sigma - t_0)$, the step response becomes

$$y_s(t) = \frac{1}{\pi} \int_{-\infty}^{2\pi B(t - t_0)} \frac{\sin u}{u}\, du = \frac{1}{2} + \frac{1}{\pi} \operatorname{Si}[2\pi B(t - t_0)] \qquad (2.234)$$

where $\text{Si}(x) = \int_0^x (\sin u/u)\, du = -\text{Si}(-x)$ is the sine-integral function.* A plot of $y_s(t)$ for an ideal lowpass filter, such as is shown in Figure 2.26(b), reveals that the 10% to 90% risetime is approximately

$$T_R \cong \frac{0.44}{B} \tag{2.235}$$

Again, the inverse relationship between bandwidth and risetime is demonstrated. ▲

2.8 SAMPLING THEORY

In many applications it is useful to represent a signal in terms of sample values taken at appropriately spaced intervals. Such sample-data systems find application in feedback control, digital computers, and pulse-modulation communication systems.

In this section we consider the representation of a signal $x(t)$ by a so-called *ideal instantaneous sampled waveform* of the form

$$x_\delta(t) = \sum_{n=-\infty}^{\infty} x(nT_s)\, \delta(t - nT_s) \tag{2.236}$$

where T_s is the sampling interval. Two questions to be answered in connection with such sampling are, "What are the restrictions on $x(t)$ and T_s to allow perfect recovery of $x(t)$ from $x_\delta(t)$?" and "How is $x(t)$ recovered from $x_\delta(t)$?" Both questions are answered by the *uniform sampling theorem for lowpass signals,* which may be stated as follows:

▷THEOREM

If a signal $x(t)$ contains no frequency components for frequencies above $f = W$ hertz, then it is completely described by instantaneous sample values uniformly spaced in time with period $T_s < 1/2W$. The signal can be exactly reconstructed from the sampled waveform given by (2.236) by passing it through an ideal lowpass filter with bandwidth B, where $W < B < f_s - W$ with $f_s = T_s^{-1}$. The frequency $2W$ is referred to as the *Nyquist frequency.*

To prove the sampling theorem, we find the spectrum of (2.236). Since $\delta(x)$ is zero everywhere except at $x = 0$, (2.236) can be written as

$$x_\delta(t) = x(t) \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \tag{2.237}$$

* See M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions,* Dover Publications, New York, 1972, pp. 238ff (Copy of the 10th National Bureau of Standards Printing).

Applying the multiplication theorem of Fourier transforms, (2.112), the Fourier transform of (2.237) is

$$X_\delta(f) = X(f) * \left[f_s \sum_{n=-\infty}^{\infty} \delta(f - nf_s) \right] \tag{2.238}$$

where the transform pair (2.127) has been used. Interchanging the orders of summation and convolution, and noting that

$$X(f) * \delta(f - nf_s) = \int_{-\infty}^{\infty} X(u)\, \delta(f - u - nf_s)\, du = X(f - nf_s) \tag{2.239}$$

by the sifting property of the delta function, we obtain

$$X_\delta(f) = f_s \sum_{n=-\infty}^{\infty} X(f - nf_s) \tag{2.240}$$

Thus, assuming that the spectrum of $x(t)$ is bandlimited to W Hz and that $f_s > 2W$ as stated in the sampling theorem, we may readily sketch $X_\delta(f)$. Figure 2.27 shows a typical choice for $X(f)$ and the corresponding $X_\delta(f)$. We note that sampling simply results in a periodic repetition of $X(f)$ in the frequency domain with a spacing f_s. If $f_s < 2W$, the separate terms in (2.240) overlap, and there is no apparent way to recover $x(t)$ from $x_\delta(t)$ without distortion. On the other hand, if $f_s > 2W$, the term in (2.240) for $n = 0$ is easily separated from the rest by ideal lowpass filtering. Assuming an ideal lowpass filter with the transfer function

$$H(f) = H_0 \Pi\left(\frac{f}{2B}\right) e^{-j2\pi f t_0}, \qquad W \le B \le f_s - W \tag{2.241}$$

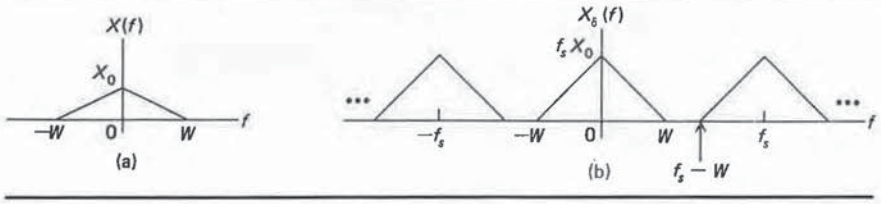
the output spectrum, with $x_\delta(t)$ at the input, is

$$Y(f) = f_s H_0 X(f) e^{-j2\pi f t_0} \tag{2.242}$$

and by the time-shift theorem, the output waveform is

$$y(t) = f_s H_0 x(t - t_0) \tag{2.243}$$

FIGURE 2.27 Signal spectra for lowpass sampling. (a) Assumed spectrum for $x(t)$. (b) Spectrum of sampled signal.



Thus, if the conditions of the sampling theorem are satisfied, we see that distortionless recovery of $x(t)$ from $x_\delta(t)$ is possible. Conversely, if the conditions of the sampling theorem are not satisfied, either because $x(t)$ is not bandlimited or because $f_s < 2W$, we see that distortion at the output of the reconstruction filter is inevitable. Such distortion, referred to as *aliasing*, is illustrated in Figure 2.28(a). It can be combated by filtering the signal before sampling or by increasing the sampling rate. A second type of error, illustrated in Figure 2.28(b), occurs in the reconstruction process and is due to the nonideal frequency response characteristics of practical filters. This type of error can be minimized by choosing reconstruction filters with sharper rolloff characteristics or by increasing the sampling rate. Note that the error due to aliasing and the error due to imperfect reconstruction filters are both *proportional to signal level*. Thus increasing the signal level does not improve the signal-to-error ratio.

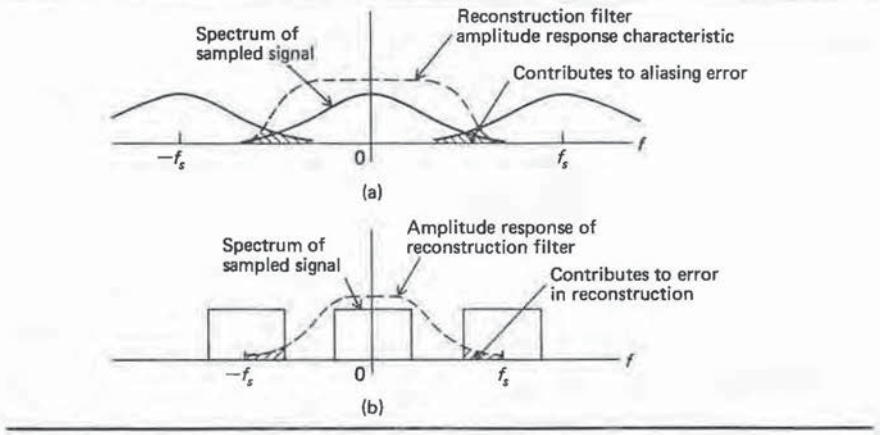
An alternative expression for the reconstructed output from the ideal lowpass filter can be obtained by noting that when (2.236) is passed through a filter with impulse response $h(t)$, the output is

$$y(t) = \sum_{n=-\infty}^{\infty} x(nT_s)h(t - nT_s) \tag{2.244}$$

But $h(t)$ corresponding to (2.241) is given by (2.210). Thus

$$y(t) = 2BH_0 \sum_{n=-\infty}^{\infty} x(nT_s)\ \text{sinc}\ 2B(t - t_0 - nT_s) \tag{2.245}$$

FIGURE 2.28 Spectra illustrating two types of errors encountered in reconstruction of sampled signals. (a) Illustration of aliasing error in the reconstruction of sampled signals. (b) Illustration of error due to nonideal reconstruction filter.



and we see that just as a periodic signal can be completely represented by its Fourier coefficients, a *bandlimited signal can be completely represented by its sample values.*

By setting $B = \frac{1}{2}f_s$, $H_0 = T_s$, and $t_0 = 0$ for simplicity, (2.245) becomes

$$y(t) = \sum_n x(nT_s)\ \text{sinc}\ (f_s t - n) \tag{2.246}$$

This expansion is equivalent to a generalized Fourier series of the form given by (2.36), for we may show that

$$\int_{-\infty}^{\infty} \text{sinc}\ (f_s t - n)\ \text{sinc}\ (f_s t - m)\ dt = K_n \delta_{nm} \tag{2.247}$$

Turning next to bandpass spectra, for which the upper limit on frequency f_u is much larger than the single-sided bandwidth W, one may naturally inquire as to the feasibility of sampling at rates less than $f_s > 2f_u$. The *uniform sampling theorem for bandpass spectra* gives the conditions for which this is possible, as follows:

▷ THEOREM

If a signal has a spectrum of bandwidth W Hz and upper frequency limit f_u, then a rate f_s at which the signal can be sampled is $2f_u/m$, where m is the largest integer not exceeding f_u/W. All higher sampling rates are not necessarily usable unless they exceed $2f_u$.

▼ EXAMPLE 2.25 Consider the bandpass signal $x(t)$ with the spectrum shown in Figure 2.29. According to the bandpass sampling theorem, it is possible to reconstruct $x(t)$ from sample values taken at a rate of

$$f_s = \frac{2f_u}{m} = \frac{2(3)}{2} = 3 \text{ samples per second} \tag{2.248}$$

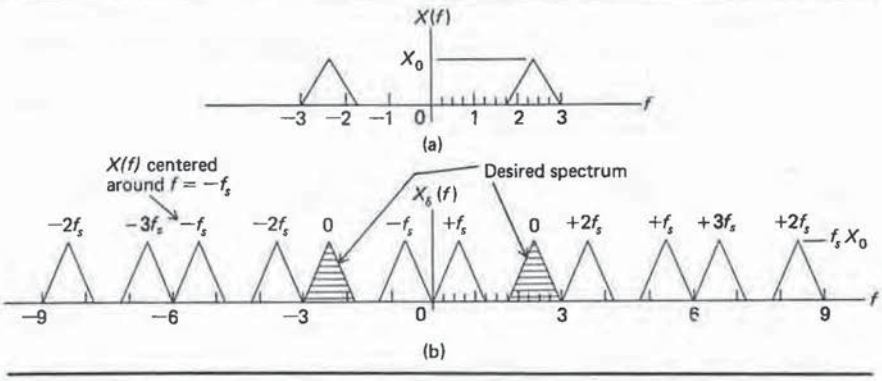
To show that this is possible, we sketch the spectrum of the sampled signal. According to (2.240), which holds in general,

$$X_\delta(f) = 3 \sum_{-\infty}^{\infty} X(f - 3n) \tag{2.249}$$

The resulting spectrum is shown in Figure 2.29(b), and we see that it is theoretically possible to recover $x(t)$ from $x_\delta(t)$ by bandpass filtering. ▲

Another way of sampling a bandpass signal of bandwidth W is to resolve it into two lowpass quadrature signals of bandwidth $\frac{1}{2}W$. Both of these may

FIGURE 2.29 Signal spectra for bandpass sampling. (a) Assumed bandpass signal spectrum. (b) Spectrum of sampled signal.



then be sampled at a minimum rate of $2(\frac{1}{2}W) = W$ samples per second, thus resulting in an overall minimum sampling rate of $2W$ samples per second.

2.9 THE HILBERT TRANSFORM*

Definition

Consider a filter that simply phase-shifts all frequency components of its input by $-\frac{1}{2}\pi$ radians; that is, its transfer function is

$$H(f) = -j \operatorname{sgn} f \tag{2.250}$$

where the sgn function (read "signum f") is defined as

$$\operatorname{sgn} f = \begin{cases} 1, & f > 0 \\ 0, & f = 0 \\ -1, & f < 0 \end{cases} \tag{2.251}$$

We note that $|H(f)| = 1$ and $\underline{/H(f)}$ is odd, as it must be. If $X(f)$ is the input spectrum to the filter, the output spectrum is $-j \operatorname{sgn}(f)X(f)$, and the corresponding time function is

$$\begin{aligned} \hat{x}(t) &= \mathcal{F}^{-1}[-j \operatorname{sgn}(f)X(f)] \\ &= h(t) * x(t) \end{aligned} \tag{2.252}$$

* It may be advantageous to postpone this section until consideration of single-sideband systems in Chapter 3.

where $h(t) = -j\mathcal{F}^{-1}[\operatorname{sgn} f]$ is the impulse response of the filter. To obtain $\mathcal{F}^{-1}[\operatorname{sgn} f]$ without resorting to contour integration, we consider the inverse transform of the function

$$G(f;\alpha) = \begin{cases} e^{-\alpha f}, & f > 0 \\ -e^{\alpha f}, & f < 0 \end{cases} \tag{2.253}$$

We note that $\lim_{\alpha \to 0} G(f;\alpha) = \operatorname{sgn} f$. Thus our procedure will be to inverse Fourier-transform $G(f;\alpha)$ and take the limit of the result as α approaches zero. Performing the inverse transformation, we obtain

$$\begin{aligned} g(t;\alpha) &= \mathcal{F}^{-1}[G(f;\alpha)] \\ &= \int_0^\infty e^{-\alpha f} e^{j2\pi ft}\, df - \int_{-\infty}^0 e^{\alpha f} e^{j2\pi ft}\, df = \frac{j4\pi t}{\alpha^2 + (2\pi t)^2} \end{aligned} \tag{2.254}$$

Taking the limit as α approaches zero, we get the transform pair

$$\frac{j}{\pi t} \leftrightarrow \operatorname{sgn} f \tag{2.255}$$

Using this result in (2.252), we obtain the output of the filter:

$$\hat{x}(t) = \int_{-\infty}^{\infty} \frac{x(t')}{\pi(t - t')}\, dt' = \int_{-\infty}^{\infty} \frac{x(t - t')}{\pi t'}\, dt' \tag{2.256}$$

The function $\hat{x}(t)$ is defined as the *Hilbert transform* of $x(t)$.

Since the Hilbert transform corresponds to a phase shift of $-\frac{1}{2}\pi$, we note that the Hilbert transform of $\hat{x}(t)$ corresponds to the transfer function $(-j \operatorname{sgn} f)^2 = -1$, or a phase shift of π radians. Thus

$$\hat{\hat{x}}(t) = -x(t) \tag{2.257}$$

▼ EXAMPLE 2.26 For the following input to a Hilbert transform filter:

$$x(t) = \cos 2\pi f_0 t \tag{2.258}$$

we obtain the output spectrum

$$\hat{X}(f) = \tfrac{1}{2}\delta(f - f_0)e^{-j\pi/2} + \tfrac{1}{2}\delta(f + f_0)e^{j\pi/2} \tag{2.259}$$

Taking the inverse Fourier transform, we find the output signal to be

$$\hat{x}(t) = \cos(2\pi f_0 t - \tfrac{1}{2}\pi) \quad \text{or} \quad \widehat{\cos 2\pi f_0 t} = \sin 2\pi f_0 t \tag{2.260}$$

Of course, the Hilbert transform could have been found directly in this case by adding $-\frac{1}{2}\pi$ to the argument of the cosine. Doing this for the signal

sin $\omega_0 t$, we find that

$$\widehat{\sin 2\pi f_0 t} = -\cos 2\pi f_0 t \tag{2.261}$$

We may use the two results obtained in Example 2.26 to show that

$$\widehat{e^{j2\pi f_0 t}} = -j \operatorname{sgn}(2\pi f_0) e^{j2\pi f_0 t} \tag{2.262}$$

This is done by considering the two cases $f_0 > 0$ and $f_0 < 0$ and using Euler's theorem in conjunction with the results of (2.260) and (2.261).

Properties

The Hilbert transform has several useful properties that will be used later. Three of these properties will be proved here:

1. The energy (or power) in a signal $x(t)$ and its Hilbert transform $\hat{x}(t)$ are equal.

   To show this, we consider the energy spectral densities at the input and output of a Hilbert transform filter. Since $H(f) = -j \operatorname{sgn} f$, these densities are related by

   $$|\hat{X}(f)|^2 \triangleq |\mathcal{F}[\hat{x}(t)]|^2 = |-j \operatorname{sgn} f|^2 \, |X(f)|^2 = |X(f)|^2 \tag{2.263}$$

   where $\hat{X}(f) = \mathcal{F}[\hat{x}(t)]$. Thus, since the energy spectral densities at input and output are equal, so are the total energies. A similar proof holds for power signals.
2. A signal and its Hilbert transform are orthogonal; that is,

   $$\int_{-\infty}^{\infty} x(t)\hat{x}(t)\, dt = 0 \qquad \text{(energy signals)} \tag{2.264a}$$

   or

   $$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)\hat{x}(t)\, dt = 0 \qquad \text{(power signals)} \tag{2.264b}$$

   Considering (2.264a), we note that the left-hand side can be written as

   $$\int_{-\infty}^{\infty} x(t)\hat{x}(t)\, dt = \int_{-\infty}^{\infty} X(f)\hat{X}^*(f)\, df \tag{2.265}$$

   by Parseval's theorem generalized, where $\hat{X}(f) = \mathcal{F}[\hat{x}(t)]$. It therefore follows that

   $$\int_{-\infty}^{\infty} x(t)\hat{x}(t)\, dt = \int_{-\infty}^{\infty} (+j \operatorname{sgn} f)|X(f)|^2\, df \tag{2.266}$$

But the integrand of the right-hand side of (2.266) is odd, being the product of the even function $|X(f)|^2$ and the odd function $+j$ sgn f. Therefore, the integral is zero, and (2.264a) is proved. A similar proof holds for (2.264b).

3. If $c(t)$ and $m(t)$ are signals with nonoverlapping spectra, where $m(t)$ is lowpass and $c(t)$ is highpass, then

$$\widehat{m(t)c(t)} = m(t)\hat{c}(t) \tag{2.267}$$

To prove this relationship, we use the Fourier integral to represent $m(t)$ and $c(t)$ in terms of their spectra, $M(f)$ and $C(f)$, respectively. Thus

$$m(t)c(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} M(f)C(f')\exp\left[j2\pi(f+f')t\right] df\, df' \tag{2.268}$$

where we assume $M(f) = 0$ for $|f| > W$ and $C(f') = 0$ for $|f'| < W$. The Hilbert transform of (2.268) is

$$\begin{aligned}\widehat{m(t)c(t)} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} M(f)C(f')\,\widehat{\exp\left[j2\pi(f+f')t\right] df\, df'} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} M(f)C(f')[-j\ \text{sgn}\ (f+f')]\exp\left[j2\pi(f+f')t\right] df\, df'\end{aligned} \tag{2.269}$$

where (2.262) has been used. However, the product $M(f)C(f')$ is nonvanishing only for $|f| < W$ and $|f'| > W$, and we may replace sgn $(f + f')$ by sgn f' in this case. Thus

$$\begin{aligned}&\widehat{m(t)c(t)} \\ &= \int_{-\infty}^{\infty} M(f)\exp\left(j2\pi ft\right) df \int_{-\infty}^{\infty} C(f')[-j\ \text{sgn}\, f'\exp\left(j2\pi f't\right)]\, df'\end{aligned} \tag{2.270}$$

However, the first integral on the right-hand side is just $m(t)$, and the second integral is $\hat{c}(t)$, since

$$c(t) = \int_{-\infty}^{\infty} C(f')\exp\left(j2\pi f't\right) df'$$

and

$$\begin{aligned}\hat{c}(t) &= \int_{-\infty}^{\infty} C(f')\,\widehat{\exp\left(j2\pi f't\right)}\, df' \\ &= \int_{-\infty}^{\infty} C(f')[-j\ \text{sgn}\, f'\exp\left(j2\pi f't\right)]\, df'\end{aligned} \tag{2.271}$$

Hence (2.270) is equivalent to (2.267), which was the relationship to be proved.

▼ EXAMPLE 2.27 Given that $m(t)$ is a lowpass signal with $M(f) = 0$ for $|f| > W$. We may directly apply (2.267) in conjunction with (2.266) and (2.261) to show that

$$\widehat{m(t) \cos \omega_0 t} = m(t) \sin \omega_0 t \tag{2.272}$$

and

$$\widehat{m(t) \sin \omega_0 t} = -m(t) \cos \omega_0 t \tag{2.273}$$

▲ if $f_0 = \omega_0/2\pi > W$.

Analytic Signals

The Hilbert transform is used to define the analytic signal $x_p(t)$ in terms of the real signal $x(t)$:

$$x_p(t) = x(t) + j\hat{x}(t) \tag{2.274}$$

We now consider several properties of the analytic signal.

We used the term *envelope* in connection with the ideal bandpass filter. The *envelope* of a signal is defined mathematically as the magnitude of the analytic signal $x_p(t)$. The concept of an envelope will acquire more importance when we discuss modulation in Chapter 3.

▼ EXAMPLE 2.28 In a previous section we showed that the impulse response of an ideal bandpass filter with bandwidth B, delay t_0, and center frequency f_0 is given by

$$h_{BP}(t) = 2H_0B \text{ sinc } B(t - t_0) \cos \omega_0(t - t_0) \tag{2.275}$$

Since $B < f_0$, we can use the result of Example 2.27 to determine the Hilbert transform of $h_{BP}(t)$:

$$\hat{h}_{BP}(t) = 2H_0B \text{ sinc } B(t - t_0) \sin \omega_0(t - t_0) \tag{2.276}$$

Thus the envelope is

$$|x_p(t)| = \sqrt{[2H_0B \text{ sinc } B(t - t_0)]^2[\cos^2 \omega_0(t - t_0) + \sin^2 \omega_0(t - t_0)]} \tag{2.277}$$

or

$$|x_p(t)| = 2H_0B\,|\text{sinc } B(t - t_0)| \tag{2.278}$$

as shown in Figure 2.22(b).

The envelope is obviously easy to identify if the signal is composed of a lowpass signal multiplied by a high-frequency sinusoid. Note, however, that the envelope is mathematically defined for any signal.

The spectrum of the analytic signal is also of interest. We will use it to advantage in Chapter 3 when we investigate single-sideband modulation. Since the analytic signal can be written as

$$x_p(t) = x(t) + j\hat{x}(t) \tag{2.279}$$

it follows that the Fourier transform of $x_p(t)$ is

$$X_p(f) = X(f) + j\{-j \operatorname{sgn}(f)X(f)\} \tag{2.280}$$

where the term in braces is the Fourier transform of $\hat{x}(t)$. Thus

$$X_p(f) = X(f)[1 + \operatorname{sgn} f] \tag{2.281}$$

or

$$X_p(f) = \begin{cases} 2X(f), & f > 0 \\ 0, & f < 0 \end{cases} \tag{2.282}$$

The subscript p is used to denote that the spectrum is nonzero only for positive frequencies.

Similarly, we can show that the signal

$$x_n(t) = x(t) - j\hat{x}(t) \tag{2.283}$$

is nonzero only for negative frequencies. Replacing $\hat{x}(t)$ by $-\hat{x}(t)$ in the preceding discussion results in

$$X_n(f) = X(f)[1 - \operatorname{sgn} f] \tag{2.284}$$

or

$$X_n(f) = \begin{cases} 0, & f > 0 \\ 2X(f), & f < 0 \end{cases} \tag{2.285}$$

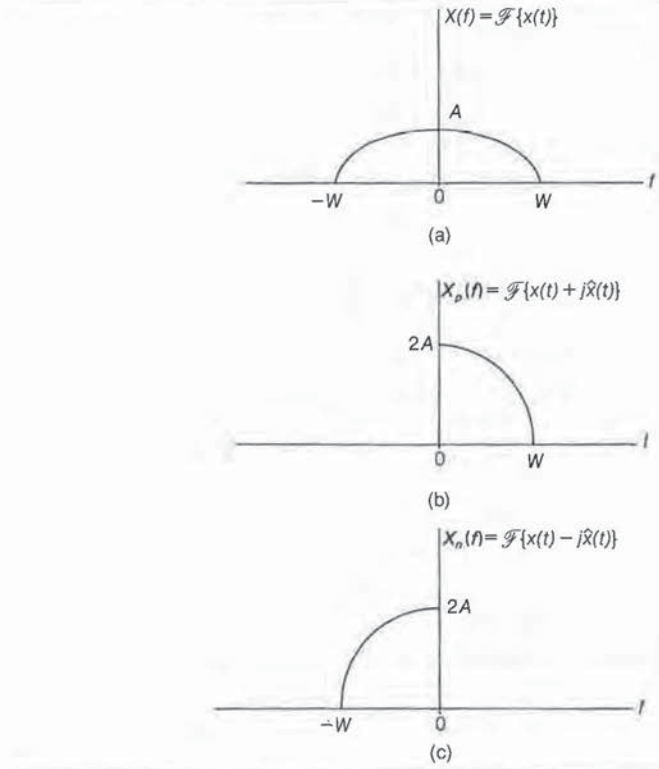
These spectra are shown in Figure 2.30.

Two observations must be made at this point. First, if $X(f)$ is nonzero at $f = 0$, then $X_p(f)$ and $X_n(f)$ will be discontinuous at $f = 0$. Also, we should not be confused that $|X_n(f)|$ and $|X_p(f)|$ are not even, since the corresponding time-domain signals are not real.

Complex Envelope Representation of Bandpass Signals

If $X(f)$ in (2.280) corresponds to a signal with a bandpass spectrum, as shown in Fig. 2.31(a), it then follows by (2.282) that $X_p(f)$ is just twice the positive-

FIGURE 2.30 Spectra of analytic signals. (a) Spectrum of $x(t)$. (b) Spectrum of $x(t) + j\hat{x}(t)$. (c) Spectrum of $x(t) - j\hat{x}(t)$.



frequency portion of $X(f) = \mathcal{F}\{x(t)\}$, as shown in Fig. 2.31(b). By the frequency-translation theorem, it follows that $x_p(t)$ can be written as

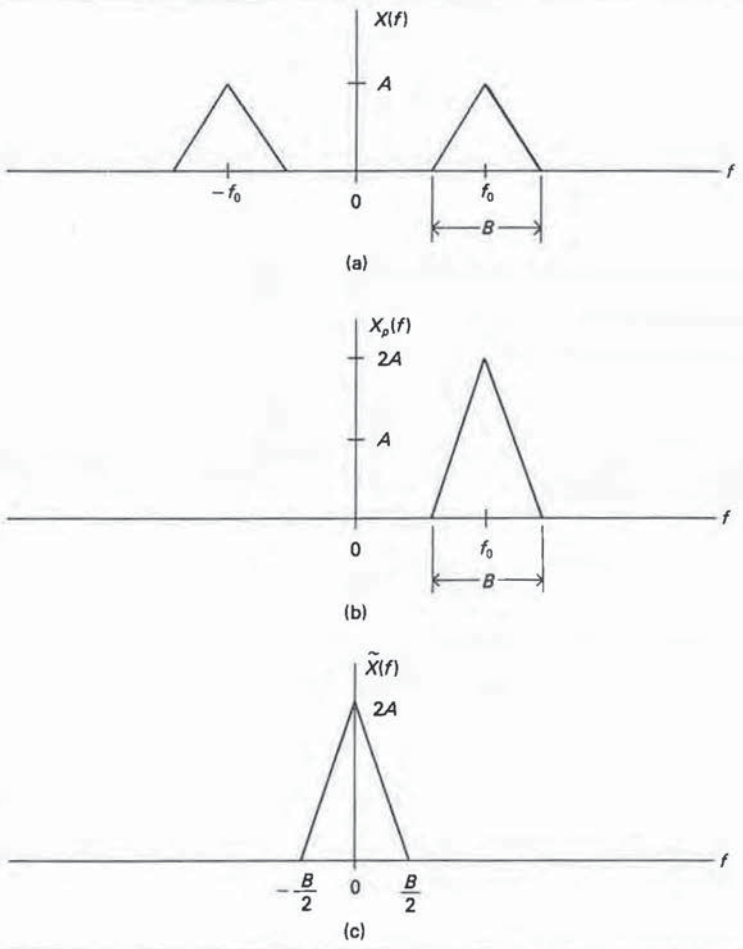
$$x_p(t) = \tilde{x}(t)e^{j2\pi f_0 t} \tag{2.286}$$

where $\tilde{x}(t)$ is a complex-valued lowpass signal (hereafter referred to as the *complex envelope*). The amplitude spectrum of $\tilde{x}(t)$ is shown in Fig. 2.31 (c).

To find $\tilde{x}(t)$, we may proceed along two paths. First, using (2.274), we find the analytic signal and then solve (2.286) for $\tilde{x}(t)$. Second, we can find $\tilde{x}(t)$ by using a frequency domain approach to obtain $X(f)$, which is then scaled by a factor of 2 to give $X_p(f)$ and translated to baseband. The inverse Fourier transform of this spectrum is then $\tilde{x}(t)$.

Since $x_p(t) = x(t) + j\hat{x}(t)$, where $x(t)$ and $\hat{x}(t)$ are the real and imaginary

FIGURE 2.31 Spectra pertaining to the formation of a complex envelope of a signal $x(t)$. (a) A bandpass signal spectrum. (b) Twice the positive-frequency portion of $X(f)$, corresponding to $\mathcal{F}[x(t) + j\hat{x}(t)]$. (c) Spectrum of $\hat{x}(t)$.



parts, respectively, of $x_p(t)$, it follows from (2.286) that

$$x_p(t) = \tilde{x}(t)e^{j2\pi f_0 t} \triangleq x(t) + j\hat{x}(t) \tag{2.287}$$

or

$$x(t) = \mathrm{Re}[\tilde{x}(t)e^{j2\pi f_0 t}] \tag{2.288a}$$

and

$$\hat{x}(t) = \mathrm{Im}[\tilde{x}(t)e^{j2\pi f_0 t}] \tag{2.288b}$$

Thus, from (2.288a), the real signal $x(t)$ can be expressed in terms of its complex envelope as

$$x(t) = x_R(t) \cos(2\pi f_0 t) - x_I(t) \sin(2\pi f_0 t) \tag{2.289}$$

where

$$\tilde{x}(t) \triangleq x_R(t) + jx_I(t) \tag{2.290}$$

The signals $x_R(t)$ and $x_I(t)$ are known as the *inphase* and *quadrature components* of $x(t)$.

Complex Envelope Representation of Bandpass Systems

Consider a bandpass system with impulse response $h(t)$ which is represented in terms of a complex envelope $\tilde{h}(t)$ as

$$h(t) = \text{Re}[\tilde{h}(t)e^{j2\pi f_0 t}] \tag{2.291}$$

where $\tilde{h}(t) = h_R(t) + jh_I(t)$. If the input is bandpass also with representation (2.288a), the output is

$$y(t) = x(t) * h(t) = \int_{-\infty}^{\infty} h(\lambda)x(t-\lambda)\, d\lambda \tag{2.292}$$

Since we can represent $h(t)$ and $x(t)$ as

$$h(t) = \tfrac{1}{2}\tilde{h}(t)e^{j2\pi f_0 t} + \text{c.c.} \tag{2.293a}$$

and

$$x(t) = \tfrac{1}{2}\tilde{x}(t)e^{j2\pi f_0 t} + \text{c.c.} \tag{2.293b}$$

where c.c. stands for the complex conjugate of the immediately preceding term, the output can be expressed as

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} [\tfrac{1}{2}\tilde{h}(t)e^{j2\pi f_0\lambda} + \text{c.c.}][\tfrac{1}{2}\tilde{x}(t-\lambda)e^{j2\pi f_0(t-\lambda)} + \text{c.c.}]\, d\lambda \\ &= \tfrac{1}{4}\int_{-\infty}^{\infty} \tilde{h}(\lambda)\tilde{x}(t-\lambda)\, d\lambda\, e^{j2\pi f_0 t} + \text{c.c} \\ &\quad + \tfrac{1}{4}\int_{-\infty}^{\infty} \tilde{h}(\lambda)\tilde{x}^*(t-\lambda)e^{j4\pi f_0\lambda}\, d\lambda\, e^{-j2\pi f_0 t} + \text{c.c.} \end{aligned} \tag{2.294}$$

The second pair of terms is approximately zero by virtue of the factor $e^{j4\lambda f_0\lambda}$ in the integrand ($\tilde{h}$ and $\tilde{x}$ are slowly varying with respect to this complex exponential, and therefore, the integrand cancels to zero half-cycle by

half-cycle). Thus

$$y(t) = \tfrac{1}{2}\mathrm{Re}\{[\tilde{h}(t) * \tilde{x}(t)]e^{j2\pi f_0 t}\} \tag{2.295}$$

where

$$\tilde{y}(t) = \tilde{h}(t) * \tilde{x}(t) = \mathcal{F}^{-1}[\tilde{H}(f)\tilde{X}(f)] \tag{2.296}$$

in which $\tilde{H}(f)$ and $\tilde{X}(f)$ are the respective Fourier transforms of $\tilde{h}(t)$ and $\tilde{x}(t)$.

▼ **EXAMPLE 2.29** As an example, consider the input

$$x(t) = \Pi(t/\tau)\cos(2\pi f_0 t) \tag{2.297}$$

to a filter with impulse response

$$h(t) = \alpha e^{-\alpha t}u(t)\cos(2\pi f_0 t) \tag{2.298}$$

Using the complex envelope analysis just developed with $\tilde{x}(t) = \Pi(t/\tau)$ and $\tilde{h}(t) = \alpha e^{-\alpha t}u(t)$, we have as the complex envelope of the filter output

$$\begin{aligned}\tilde{y}(t) &= \Pi(t/\tau) * \alpha e^{-\alpha t}u(t) \\ &= [1 - e^{-\alpha(t+\tau/2)}]u(t + \tau/2) - [1 - e^{-(t-\tau/2)}]u(t - \tau/2)\end{aligned} \tag{2.299}$$

Multiplying this by $\frac{1}{2}e^{j2\pi f_0 t}$ and taking the real part results in the output of the filter in accordance with (2.295).

To check this result, we convolve (2.297) and (2.298) directly. The superposition integral becomes

$$\begin{aligned}y(t) &= x(t) * h(t) \\ &= \int_{-\infty}^{\infty} \Pi(\lambda/\tau)\cos(2\pi f_0\lambda)\alpha e^{-\alpha(t-\lambda)}u(t - \lambda)\cos[2\pi f_0(t - \lambda)]\, d\lambda\end{aligned} \tag{2.300}$$

But

$$\begin{aligned}&\cos(2\pi f_0\lambda)\cos[2\pi f_0(t - \lambda)] \\ &\qquad = \tfrac{1}{2}\cos(2\pi f_0 t) + \tfrac{1}{2}\cos[2\pi f_0(t - 2\lambda)]\end{aligned} \tag{2.301}$$

so that the superposition integral becomes

$$\begin{aligned}y(t) = {} & \tfrac{1}{2}\int_{-\infty}^{\infty} \Pi(\lambda/\tau)\alpha e^{-\alpha(t-\lambda)}u(t - \lambda)\, d\lambda \cos(2\pi f_0 t) \\ & + \tfrac{1}{2}\int_{-\infty}^{\infty} \Pi(\lambda/\tau)\alpha e^{-\alpha(t-\lambda)}u(t - \lambda)\cos[2\pi f_0(t - 2\lambda)]\, d\lambda\end{aligned} \tag{2.302}$$

If $f_0^{-1} \ll \tau$ and $f_0^{-1} \ll \alpha^{-1}$, the second integral is approximately zero, so that we have only the first integral, which is $\Pi(t/\tau)$ and $\alpha e^{-\alpha t}u(t)$ convolved and the result multiplied by $\frac{1}{2}\cos(2\pi f_0 t)$, which is the same as before.

2.10 THE DISCRETE FOURIER TRANSFORM AND FAST FOURIER TRANSFORM

In order to compute the Fourier spectrum of a signal by means of a digital computer, the time-domain signal must be represented by sample values, and the spectrum must be computed at a discrete number of frequencies. It can be shown that the following sum gives an approximation to the Fourier spectrum of a signal at frequencies $k/(NT_s)$, $k = 0, 1, \ldots, N-1$:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-j2\pi nk/N}, \qquad k = 0, 1, \ldots, N-1 \tag{2.303}$$

where $x_0, x_1, x_2, \ldots, x_{N-1}$ are N sample values of the signal taken at T_s-second intervals for which the Fourier spectrum is desired. The sum (2.303) is called the *discrete Fourier transform (DFT)* of the sequence $\{x_n\}$. According to the sampling theorem, if the samples are spaced by T_s seconds, the spectrum repeats every $f_s = T_s^{-1}$ Hz. Since there are N frequency samples in this interval, it follows that the frequency resolution of (2.303) is $f_s/N = 1/(NT_s) \triangleq 1/T$. To obtain the sample sequence $\{x_n\}$ from the DFT sequence $\{X_k\}$, the sum

$$x_n = \frac{1}{N}\sum_{k=0}^{N-1} X_k e^{j2\pi nk/N}, \qquad n = 0, 1, 2, \ldots, N-1 \tag{2.304}$$

is used. That (2.303) and (2.304) form a transform pair can be shown by substituting (2.303) into (2.304) and using the sum formula for a geometric series:

$$S_N \equiv \sum_{k=0}^{N-1} x^k = \frac{1-x^N}{1-x}, \qquad x \neq 1 \tag{2.305a}$$

and

$$S_N = N, \qquad x = 1 \tag{2.305b}$$

As indicated above, the DFT and inverse DFT are approximations to the true Fourier spectrum of a signal $x(t)$, at the discrete set of frequencies $\{0, 1/T, 2/T, \ldots, (N-1)/T\}$. The error can be small if the DFT and its inverse are applied properly to a signal. To indicate the approximations

involved, we must visualize the spectrum of a sampled signal that is truncated to a finite number of sample values and whose spectrum is then sampled at a discrete number N of points. To see the approximations involved, we use the following Fourier transform theorems:

1. The Fourier transform of an ideal sampling waveform (Example 2.14):

$$y_s(t) = \sum_{m=-\infty}^{\infty} \delta(t - mT_s) \leftrightarrow f_s^{-1} \sum_{n=-\infty}^{\infty} \delta(f - nf_s), \qquad f_s = T_s^{-1}$$

2. The Fourier transform of a rectangular window function:

$$\Pi(t/T) \leftrightarrow T \operatorname{sinc}(fT)$$

3. The convolution theorem of Fourier transforms:

$$x_1(t) * x_2(t) \leftrightarrow X_1(f)X_2(f)$$

4. The multiplication theorem of Fourier transforms:

$$x_1(t)x_2(t) \leftrightarrow X_1(f) * X_2(f)$$

The approximations involved are illustrated by the following example.

▼ **EXAMPLE 2.30** An exponential signal is to be sampled, the samples truncated to a finite number, and the result represented by a finite number of samples of the Fourier spectrum of the sampled truncated signal. The continuous-time signal and its Fourier transform are

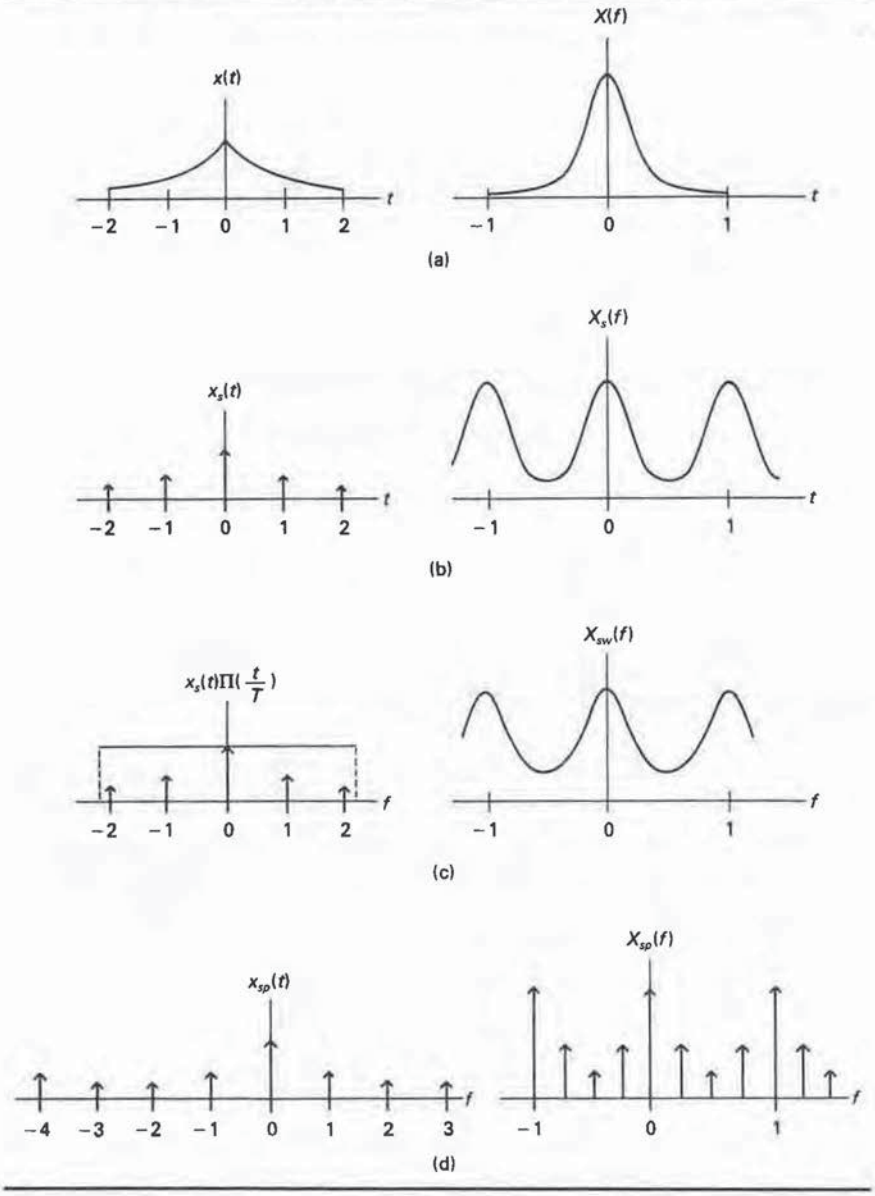
$$x(t) = e^{-|t|/\tau} \leftrightarrow X(f) = \frac{2\tau}{1 + 2(\pi f \tau)^2} \tag{2.306}$$

These signals are shown in Fig. 2.32(a). However, we are representing the signal by sample values spaced by T_s seconds, which entails multiplying the original signal by the ideal sampling waveform $y_s(t)$. The resulting spectrum of this signal is the convolution of $X(f)$ with $y_s(t) = 2\tau f_s \sum_{n=-\infty}^{\infty} \delta(f - nf_s)$. The result of this convolution, shown in Fig. 2.32(b), is

$$X_s(f) = f_s \sum_{n=-\infty}^{\infty} \frac{2\tau}{1 + [2\pi\tau(f - f_s)]^2} \tag{2.307}$$

In calculating the DFT, only a T-second chunk of $x(t)$ can be used (N samples spaced by $T_s = T/N$). This means that the sampled time domain signal is effectively multiplied by a window function $\Pi(t/T)$. In the frequency domain, this corresponds to convolution with the Fourier transform of the

FIGURE 2.32 Signals and spectra illustrating the computation of the DFT. (a) Signal to be sampled and its spectrum ($\tau = 1$ s). (b) Sampled signal and its spectrum ($f_s = 1$ Hz). (c) Windowed, sampled signal and its spectrum ($T = 4^+$ s). (d) Sampled signal spectrum and corresponding periodic repetition of the sampled, windowed signal.



rectangular window function, which is T sinc (fT). The resulting windowed, sampled signal and its spectrum are sketched in Fig. 2.32(c). Finally, the spectrum is available only at N discrete frequencies separated by the reciprocal of the window duration $1/T$. This corresponds to convolution in the time domain with a sequence of delta functions, and this result is shown in Fig. 2.32(d). It can be seen that unless one is careful, there is indeed a considerable likelihood that the DFT spectrum will look nothing like the spectrum of the original continuous-time signal. Means for minimizing these errors are discussed in several references on the subject.*

A little thought will indicate that to compute the complete DFT spectrum of a signal, approximately N^2 complex multiplications are required in addition to a number of complex additions. It is possible to find algorithms that allow the computation of the DFT spectrum of a signal using only approximately $N \log_2 N$ complex multiplications. Such algorithms are referred to as *fast Fourier transform (FFT) algorithms*.

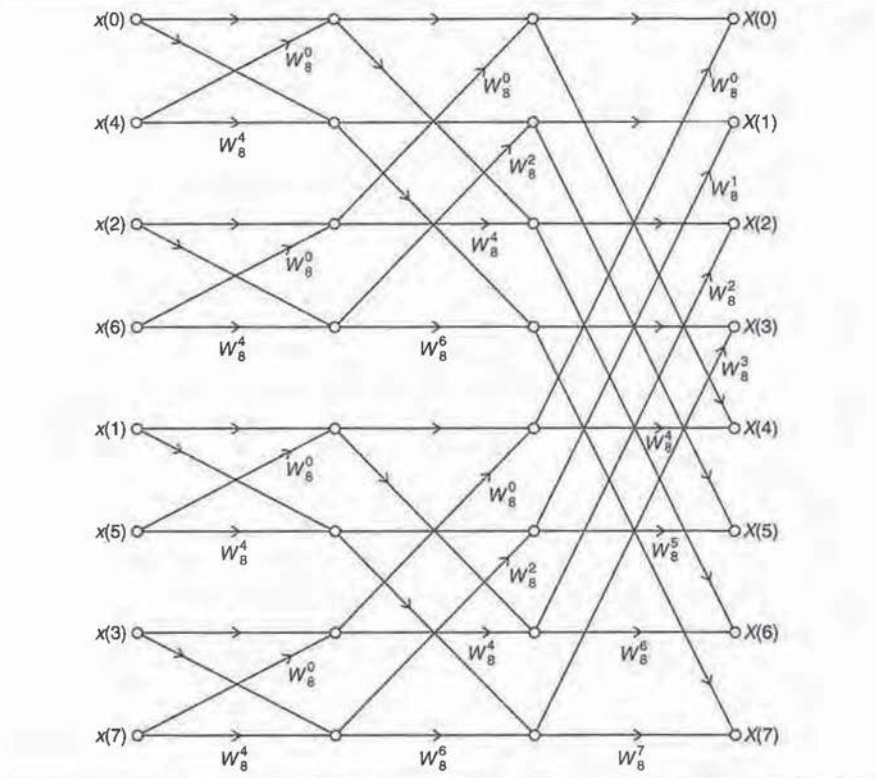
One such algorithm is based on a procedure called *decimation in time* (DIT). A flow graph for an eight-point DIT algorithm is shown in Figure 2.33, where the lower case x's denote input samples and the uppercase X's denote output samples. The arrows denote the flow of computation, with $W_N^p = e^{-j2\pi p/N}$ alongside the arrowhead indicating multiplication by this factor. Fortunately, FFT algorithms are included in most mathematics packages (see the footnote after the Computer Exercises for reference to some of these), so it is not necessary to design one's own FFT subroutine in doing computations requiring the FFT.

SUMMARY

1. Two general classes of signals are deterministic signals and random signals. A deterministic signal can be written as a completely known function of time whereas the amplitudes of random signals must be described probabilistically.
2. A periodic signal of period T_0 is one for which $x(t) = x(t + T_0)$.
3. A single-sided spectrum for a rotating phasor $\tilde{x}(t) = Ae^{j(\omega_0 t + \theta)}$ shows A (amplitude) and θ (phase) versus f (frequency). The real, sinusoidal signal corresponding to this phasor is obtained by taking the real part of $\tilde{x}(t)$.

* Ziemer, Tranter, and Fannin (1993), Chapter 10.

FIGURE 2.33 A flow graph for computation of an eight-point FFT using the DIT technique



A double-sided spectrum results if we think of forming $x(t) = \frac{1}{2}\tilde{x}(t) + \frac{1}{2}\tilde{x}^*(t)$. Graphs of amplitude and phase (two plots) of this rotating phasor sum versus f are known as the two-sided amplitude and phase spectra, respectively. Such spectral plots are referred to as frequency-domain representations of the signal $A \cos(\omega_0 t + \theta)$.

4. The unit impulse function, $\delta(t)$, can be thought of as a zero-width, infinite-height pulse with unity area. Many functions with these properties, in the limit as some parameter approaches zero, are suitable representations for a unit impulse. The sifting property, $\int_{-\infty}^{\infty} x(\lambda)\,\delta(\lambda - t_0)\,d\lambda = x(t_0)$, is a generalization of the defining relation for a unit impulse. The unit step function, $u(t)$, is the integral of a unit impulse.

5. A signal $x(t)$ for which $E = \int_{-\infty}^{\infty} |x(t)|^2 \, dt$ is finite is called an *energy signal*. If $x(t)$ is such that

$$P = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} |x(t)|^2 \, dt$$

is finite, the signal is known as a *power signal*.

6. A set of orthogonal functions, $\phi_1(t), \phi_2(t), \ldots, \phi_N(t)$, can be used as a series approximation of the form

$$x_a(t) = \sum_{n=1}^{N} X_n \phi_n(t)$$

to a signal $x(t)$, which is integral-square in the interval $(t_0, t_0 + T)$. The integral-squared error between $x_a(t)$ and $x(t)$ is minimized if the coefficients are chosen as

$$X_n = \frac{1}{c_n} \int_{t_0}^{t_0+T} x(t) \phi_n^*(t) \, dt$$

where

$$\int_{t_0}^{t_0+T} \phi_n(t) \phi_m^*(t) \, dt = c_n \delta_{nm}, \quad c_n = \text{real constant}$$

For a complete set of $\phi_n(t)$'s, the integral-squared error approaches zero as N approaches infinity, and Parseval's theorem then holds:

$$\int_T |x(t)|^2 \, dt = \sum_{n=1}^{\infty} c_n |X_n|^2$$

7. If $\phi_n(t) = e^{jn\omega_0 t}$, where $\omega_0 = 2\pi/T_0$ and T_0 is the expansion interval, is used in an orthogonal function series, the result is the complex exponential Fourier series. If $x(t)$ is periodic with period T_0, the exponential Fourier series represents $x(t)$ exactly for all t, except at points of discontinuity.

8. For exponential Fourier series of real signals, the Fourier coefficients obey $X_n = X_{-n}^*$, which implies that $|X_n| = |X_{-n}|$ and $\angle X_n = -\angle X_{-n}$. Plots of $|X_n|$ and $\angle X_n$ versus nf_0 are referred to as the discrete, double-sided amplitude and phase spectra, respectively, of $x(t)$. If $x(t)$ is real, the amplitude spectrum is even and the phase spectrum is odd as functions of nf_0.

9. Parseval's theorem for periodic signals is

$$\frac{1}{T_0}\int_{T_0} |x(t)|^2\, dt = \sum_{n=-\infty}^{\infty} |X_n|^2$$

10. The Fourier transform of a signal $x(t)$ is

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}\, dt$$

and the inverse Fourier transform is

$$x(t) = \int_{-\infty}^{\infty} X(f)e^{j2\pi ft}\, df$$

For real signals, $|X(f)| = |X(-f)|$ and $\underline{/X(f)} = -\underline{/X(-f)}$.

11. Plots of $|X(f)|$ and $\underline{/X(f)}$ versus f are referred to as the *double-sided amplitude* and *phase spectra,* respectively, of $x(t)$. Thus, as functions of frequency, the amplitude spectrum of a real signal is even, and the phase spectrum is odd.

12. The energy of a signal is

$$\int_{-\infty}^{\infty} |x(t)|^2\, dt = \int_{-\infty}^{\infty} |X(f)|^2\, df$$

This is known as *Rayleigh's energy theorem.* The energy spectral density of a signal is $G(f) = |X(f)|^2$. It is the density of energy with frequency of the signal.

13. The convolution of two signals, $x_1(t)$ and $x_2(t)$, is

$$x(t) = x_1 * x_2 = \int_{-\infty}^{\infty} x_1(\lambda)x_2(t-\lambda)\, d\lambda = \int_{-\infty}^{\infty} x_1(t-\lambda)x_2(\lambda)\, d\lambda$$

The convolution theorem of Fourier transforms states that $X(f) = X_1(f)X_2(f)$, where $X(f)$, $X_1(f)$, and $X_2(f)$ are the Fourier transforms of $x(t)$, $x_1(t)$, and $x_2(t)$, respectively.

14. The Fourier transform of a periodic signal can be obtained formally by Fourier-transforming its exponential Fourier series term by term, using $Ae^{j2\pi f_0 t} \leftrightarrow A\delta(f - f_0)$, even though, mathematically speaking, Fourier transforms of power signals do not exist. (The latter may be found as the limit of the Fourier transform of a properly Fourier-transformable function as a parameter approaches zero or infinity.)

15. The power spectrum $S(f)$ of a power signal $x(t)$ is a real, even, nonnegative function that integrates to give total average power: $\langle x^2(t) \rangle = \int_{-\infty}^{\infty} S(f)\, df$. The time-average autocorrelation function of a power signal is defined as $R(\tau) = \langle x(t)x(t + \tau) \rangle$. The Wiener-Khinchine theorem states that $S(f)$ and $R(\tau)$ are Fourier transform pairs.

16. A linear system, denoted operationally as $\mathcal{H}(\ )$, is one for which superposition holds; that is, if $y_1 = \mathcal{H}(x_1)$ and $y_2 = \mathcal{H}(x_2)$, then $\mathcal{H}(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 y_1 + \alpha_2 y_2$, where x_1 and x_2 are inputs and y_1 and y_2 are outputs. A system is fixed, or time-invariant, if, given $y(t) = \mathcal{H}[x(t)]$, the input $x(t - t_0)$ results in the output $y(t - t_0)$.

17. The impulse response $h(t)$ of a linear, fixed system is its response to an impulse applied at $t = 0$: $h(t) = \mathcal{H}[\delta(t)]$. The output of a fixed, linear system to an input $x(t)$ is given by $y(t) = h(t) * x(t)$.

18. A *causal system* is one which does not anticipate its input. For such a system, $h(t) = 0$ for $t < 0$. A *stable system* is one for which every bounded input results in a bounded output. A system is stable if and only if $\int_{-\infty}^{\infty} |h(t)|\, dt < \infty$.

19. The transfer function $H(f)$ of a linear fixed system is the Fourier transform of $h(t)$. The Fourier transform of the system output $y(t)$ due to an input $x(t)$ is $Y(f) = H(f)X(f)$, where $X(f)$ is the Fourier transform of the input. $|H(f)| = |H(-f)|$ is called the *amplitude response* of the system, and $\underline{/H(f)} = -\underline{/H(-f)}$ is called the *phase response*.

20. For a fixed linear system with a periodic input, the Fourier coefficients of the output are given by $Y_n = H(nf_0)X_n$, where X_n represents the Fourier coefficients of the input.

21. Input and output spectral densities for a fixed linear system are related by

$$G_y(f) = |H(f)|^2 G_x(f) \qquad \text{(energy signals)}$$
$$S_y(f) = |H(f)|^2 S_x(f) \qquad \text{(power signals)}$$

22. A system is distortionless if its output looks like its input except for a time delay and amplitude scaling: $y(t) = H_0 x(t - t_0)$. The transfer function of a distortionless system is $H(f) = H_0 e^{-j2\pi f t_0}$. Such a system's amplitude response is $|H(f)| = H_0$ and its phase response is $\underline{/H(f)} = -2\pi t_0 f$ over the band of frequencies occupied by the input. Three types of distortion that a system may introduce are amplitude, phase (or delay), and nonlinear, depending on whether $|H(f)| \neq$ constant,

$\underline{/H(f)} \neq -$ constant $\times f$, or the system is nonlinear, respectively. Two other important properties of a linear system are the group and phase delays. These are defined by

$$T_g(f) = -\frac{1}{2\pi}\frac{d}{df}\theta(f) \qquad \text{and} \qquad T_p(f) = -\frac{\theta(f)}{2\pi f}$$

respectively, in which $\theta(f)$ is the phase response of the network. Distortionless systems have equal group and phase delays.

23. Ideal filters are convenient in communication system analysis, even though they are unrealizable. Three types of ideal filters are lowpass, bandpass, and highpass. Throughout their passband, ideal filters have constant amplitude response and linear phase response. Outside their passband, ideal filters perfectly reject all spectral components of the input.

24. Approximations to ideal filters are Butterworth, Chebyshev, and Bessel filters. The first two are attempts at approximating the amplitude response of an ideal filter, and the latter is an attempt to approximate the linear phase response of an ideal filter.

25. An inequality relating the duration T of a pulse and its single-sided bandwidth W is $W \geq 1/2T$. Pulse risetime T_R and signal bandwidth are related approximately by $W = 1/2T_R$. These relationships hold for the lowpass case. For bandpass filters and signals, the required bandwidth is doubled, and the risetime is that of the envelope of the signal.

26. The sampling theorem for lowpass signals of bandwidth W states that a signal can be perfectly recovered by lowpass filtering from sample values taken at a rate of $f_s > 2W$ samples per second. The spectrum of an impulse-sampled signal is

$$X_\delta(f) = f_s \sum_{n=-\infty}^{\infty} X(f - nf_s)$$

where $X(f)$ is the spectrum of the original signal. For bandpass signals, lower sampling rates than specified by the lowpass sampling theorem may be possible.

27. The Hilbert transform $\hat{x}(t)$ of a signal $x(t)$ corresponds to a $-90°$ phase shift of all the signal's positive-frequency components. Mathematically,

$$\hat{x}(t) = \int_{-\infty}^{\infty} \frac{x(t')}{\pi(t - t')}\, dt'$$

In the frequency domain, $\hat{X}(f) = -j \operatorname{sgn}(f)X(f)$, where sgn (f) is the signum function, $X(f) = \mathcal{F}[x(t)]$, and $\hat{X}(f) = \mathcal{F}[\hat{x}(t)]$. The Hilbert transform of $\cos \omega_0 t$ is $\sin \omega_0 t$, and the Hilbert transform of $\sin \omega_0 t$ is $-\cos \omega_0 t$. The power (or energy) in a signal and its Hilbert transform are equal. A signal and its Hilbert transform are orthogonal in the range $(-\infty, \infty)$. If $m(t)$ is a lowpass signal and $c(t)$ is a highpass signal with nonoverlapping spectra,

$$\widehat{m(t)c(t)} = m(t)\hat{c}(t)$$

The Hilbert transform can be used to define the analytic signal

$$z(t) = x(t) \pm j\hat{x}(t)$$

The magnitude of the analytic signal, $|z(t)|$, is the envelope of the real signal $x(t)$. The Fourier transform of an analytic signal, $Z(f)$, is identically zero for $f < 0$ or $f > 0$, respectively, depending on if the $+$ sign or $-$ sign is chosen for the imaginary part of $z(t)$.

28. The complex envelope $\tilde{x}(t)$ of a bandpass signal is defined by

$$x(t) + j\hat{x}(t) = \tilde{x}(t)e^{j2\pi f_0 t}$$

where f_0 is the center frequency of the signal. Similarly, the complex envelope $\tilde{h}(t)$ of the impulse response of a bandpass system is defined by

$$h(t) + j\hat{h}(t) = \tilde{h}(t)e^{j2\pi f_0 t}$$

Its output is conveniently obtained in terms of the complex envelope of the output which can be found from either of the operations

$$\tilde{y}(t) = \tilde{h}(t) * \tilde{x}(t)$$

or

$$\tilde{y}(t) = \mathcal{F}^{-1}[\tilde{H}(f)\tilde{X}(f)]$$

where $\tilde{H}(f)$ and $\tilde{X}(f)$ are the Fourier transforms of $\tilde{h}(t)$ and $\tilde{x}(t)$, respectively. The actual (real) output is then given by

$$y(t) = \operatorname{Re}[\tilde{y}(t)e^{j2\pi f_0 t}]$$

29. The discrete Fourier transform (DFT) is defined as

$$X_k = \sum_{n=0}^{N-1} x_n e^{j2\pi nk/N} = \operatorname{DFT}[\{x_n\}], \qquad k = 0, 1, \ldots, N-1$$

and the inverse DFT can be found from

$$x_n = \frac{1}{N}\,\mathrm{DFT}[\{X_k^*\}]^*, \qquad n = 0, 1, \ldots, N-1$$

The DFT can be used to digitally compute spectra of sampled signals and to approximate operations carried out by the normal Fourier transform, for example, filtering.

30. FFT algorithms can be obtained in various ways. The flow graph for a *decimation in time* (DIT) algorithm is shown in Figure 2.33.

FURTHER READING

Fourier techniques from the standpoint of the electrical engineer are dealt with in many texts. Lathi (1991) and Stremler (1982) treat them in the context of communication systems. Bracewell (1986) and Papoulis (1977) are concerned exclusively with Fourier theory and applications.

Introductory reading on Fourier series, transfer functions, and impulse response is found in most introductory circuit-analysis texts. See, for example, Hayt and Kemmerly (1993), Cunningham and Stuller (1991), or Smith and Dorf (1992). Most of the books cited above provide good treatments of systems analysis. Ziemer, Tranter, and Fannin (1993) and Frederick and Carlson (1988) are devoted to continuous and discrete signal and system theory. Schwartz and Friedland (1965) provide an excellent exposition of system characterization at the senior-graduate level, especially of the relationship of transfer function and impulse response to state-variable formulations. Sampling theory and the Hilbert transform are treated by Carlson (1985) and Haykin (1983), among others, at about the same level as presented here. Treatments of the discrete Fourier and fast Fourier transforms are available in Ziemer, Tranter, and Fannin (1993) and Brigham (1988).

PROBLEMS

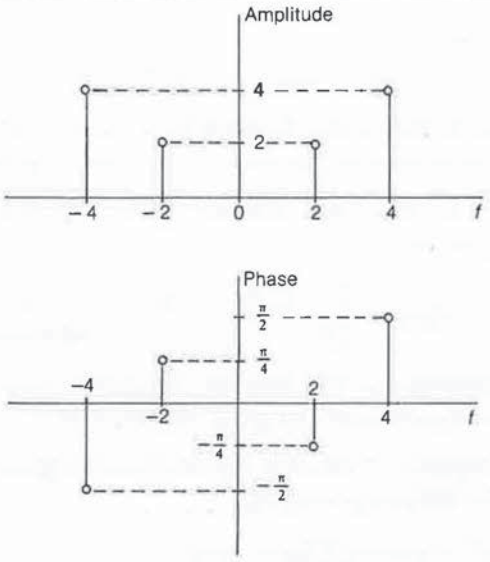
Section 2.1

2.1 Sketch the single-sided and double-sided amplitude and phase spectra of

$$x(t) = 10\cos(4\pi t + \pi/6) + 3\sin(10\pi t + 2\pi/3)$$

2.2 A signal has the double-sided amplitude and phase spectra shown in Figure 2.34. Write a time-domain expression for the signal.

FIGURE 2.34



2.3 The sum of two or more sinusoids may or may not be periodic depending on the relationship of their separate frequencies. For the sum of two sinusoids, let the frequencies of the individual terms be f_1 and f_2, respectively. For the sum to be periodic, f_1 and f_2 must be commensurable; i.e., there must be a number f_0 contained in each an integral number of times. Thus, if f_0 is the largest such number,

$$f_1 = n_1 f_0 \qquad \text{and} \qquad f_2 = n_2 f_0$$

where n_1 and n_2 are integers; f_0 is the fundamental frequency. Which of the signals given below are periodic? Find the periods of those that are periodic.

(a) $x_1(t) = 2\cos(2t) + 3\sin(5\pi t)$
(b) $x_2(t) = \cos(5t) + 5\cos(15t)$
(c) $x_3(t) = 3\sin(10\pi t) + \sin(12\pi t)$
(d) $x_4(t) = 4\cos(2\pi t) + 3\cos(4\pi t) + 5\sin(26\pi t)$

2.4 Sketch the single-sided and double-sided spectra of

(a) $x_a(t) = 6\cos(6\pi t - \pi/3)$
(b) $x_b(t) = 4\sin(10\pi t) + 2\cos(12\pi t)$

2.5 (a) Show that the function $\delta_\epsilon(t)$ sketched in Figure 2.4(b) has unity area.

(b) Show that

$$\delta_\epsilon(t) = \epsilon^{-1} e^{-t/\epsilon} u(t)$$

has unity area. Sketch this function for $\epsilon = 1, \frac{1}{2}$, and $\frac{1}{4}$. Comment on its suitability as an approximation for the unit impulse function.

2.6 Show that a suitable approximation for the unit impulse function as $\epsilon \to 0$ is given by

$$\delta_\epsilon(t) = \begin{cases} \epsilon^{-1}(1 - |t|/\epsilon), & |t| \leq \epsilon \\ 0, & \text{otherwise} \end{cases}$$

2.7 Use the properties of the Riemann integral to prove the properties of the unit impulse function given after (2.13b).

2.8 Use the properties of the unit impulse function given after (2.13b) to evaluate the following integrals.

(a) $\int_{-\infty}^{\infty} [t^2 + \cos(\pi t)]\, \delta(t - 8)\, dt$
(b) $\int_{-5^-}^{15^+} (t^2 + 1) \left[\sum_{n=-\infty}^{\infty} \delta(t - 5n)\right] dt$
(c) $\int_{-1}^{10} (t^2 + t)\, \delta(t + 3)\, dt$
(d) $\int_{-\infty}^{\infty} [e^{-\pi t} + \sin(10\pi t)]\, \delta(2t + 5)\, dt$
(e) $\int_{-\infty}^{\infty} [\cos(2\pi t) + e^{-10t}]\, \delta^{(2)}(t)\, dt$

2.9 Which of the following signals are periodic and which are aperiodic? Find the periods of those which are periodic. Sketch all signals.

(a) $x_1(t) = \cos(4\pi t) + \sin(7\pi t)$
(b) $x_2(t) = e^{-10t} u(t)$
(c) $x_3(t) = \sum_{n=-\infty}^{\infty} \delta(t - 6n)$
(d) $x_4(t) = \sin(3t) + \cos(2\pi t)$
(e) $x_5(t) = \sum_{n=-\infty}^{\infty} \Pi(t - 2n)$
(f) $x_6(t) = 1 + \cos(6\pi t)$

2.10 Write the signal $x(t) = 2\cos(5\pi t) + 4\sin(11\pi t)$ as

(a) The real part of a sum of rotating phasors.
(b) A sum of rotating phasors plus their complex conjugates.
(c) From your results in parts (a) and (b), sketch the single-sided and double-sided amplitude and phase spectra of $x(t)$.

Section 2.2

2.11 Find the normalized power for each signal below that is a power signal and the normalized energy for each signal that is an energy signal.

If a signal is neither a power signal nor an energy signal, so designate it. Sketch each signal (α is a positive constant).

(a) $x_1(t) = 4 \sin (6\pi t + \pi/6)$ (b) $x_2(t) = e^{-\alpha t}u(t)$
(c) $x_3(t) = e^{\alpha t}u(-t)$ (d) $x_4(t) = (\alpha^2 + t^2)^{-1/4}$
(e) $x_5(t) = e^{-\alpha|t|}$ (f) $x_6(t) = \cos (5\pi t)u(t)$

2.12 Classify each of the following signals as an energy signal or a power signal by calculating E, the energy, or P, the power (A, θ, ω, and τ are positive constants).

(a) $A|\sin (\omega t + \theta)|$ (b) $A\tau/\sqrt{\tau + jt}, \quad j = \sqrt{-1}$
(c) $At^2e^{-t/\tau}u(t)$ (d) $\Pi(t/\tau) + \Pi(t/2\tau)$

2.13 Sketch each of the following periodic waveforms and compute their average powers.

(a) $x_1(t) = \sum_{n=-\infty}^{\infty} \Pi[(t - 8n)/3]$
(b) $x_2(t) = \sum_{n=-\infty}^{\infty} \Lambda[(t - 4n)/2]$
(c) $x_3(t) = \sum_{n=-\infty}^{\infty} \Lambda[(t - 2n)/2]u(t - 2n)$
(d) $x_4(t) = |\sin (5\pi t)|$

2.14 For each of the following signals, determine both the normalized energy and power. (*Note:* 0 and ∞ are possible answers.)

(a) $x_1(t) = 8e^{(2+j4\pi)t}$ (b) $x_2(t) = \Pi[(t - 3)/2] + \Pi(t - 3)$
(c) $x_3(t) = 8e^{j4\pi t}u(t)$ (d) $x_4(t) = 2 + 3 \cos (8\pi t)$

2.15 Show that the signal

$$x(t) = t^{-1/4}u(t - 2)$$

is neither an energy nor a power signal.

2.16 (a) Prove Equation (2.23).
(b) Use Equation (2.23) to find the power in each of the signals given in Problem 2.13.

2.17 Show that the following are energy signals. Sketch each signal.

(a) $x_1(t) = \Pi(t/10) \cos (3\pi t)$
(b) $x_2(t) = e^{-|t|/2} \cos (10\pi t)$
(c) $x_3(t) = 2u(t) - 2u(t - 5)$
(d) $x_4(t) = \int_{-\infty}^{t} u(\lambda)\, d\lambda - 2\int_{-\infty}^{t-10} u(\lambda)\, d\lambda + \int_{-\infty}^{t-20} u(\lambda)\, d\lambda$

Section 2.3

2.18 (a) Fill in the steps for obtaining (2.30) from (2.29).
(b) Obtain (2.31) from (2.30).

2.19 (a) Given the set of orthogonal functions

$$\phi_n(t) = \Pi\left\{\frac{4[t - (2n - 1)T/8]}{T}\right\}, \quad n = 1, 2, 3, 4$$

sketch and dimension accurately these functions.

(b) Approximate the ramp signal

$$x(t) = (t/T)\Pi\left[\frac{(t - T/2)}{T}\right]$$

by a generalized Fourier series using this set.

(c) Do the same for the set

$$\phi_n(t) = \Pi\left\{\frac{2[t - (2n - 1)T/4]}{T}\right\}, \quad n = 1, 2$$

(d) Compute the integral-squared error for both part (b) and part (c). What do you conclude about the dependence of ϵ_N on N?

Section 2.4

2.20 Using the uniqueness property of the Fourier series, find exponential Fourier series for the following signals:

(a) $x_1(t) = \cos^2(\omega_0 t)$ (b) $x_2(t) = \cos(\omega_0 t) + \sin(\omega_0 t)$
(c) $x_3(t) = \sin^2(\omega_0 t)\cos(2\omega_0 t)$ (d) $x_4(t) = \cos^3(\omega_0 t)$

(*Hint:* Use appropriate trigonometric identities and Euler's theorem.)

2.21 Expand the signal $x(t) = 4|t|$ in a complex exponential Fourier series over the interval $|t| \leq 4$. Sketch the signal to which the Fourier series converges for all t.

2.22 Fill in all the steps to show that

(a) $|X_n| = |X_{-n}|$ and $\underline{/X_n} = -\underline{/X_{-n}}$ for $x(t)$ real.
(b) X_n is a real, even function of n for $x(t)$ real and even.
(c) X_n is imaginary and an odd function of n for $x(t)$ real and odd.
(d) Halfwave symmetry implies that $X_n = 0$, n even.

2.23 Obtain the complex exponential Fourier series coefficients for the (a) pulse train, (b) half-rectified sinewave, (c) full-rectified sinewave, and (d) triangular waveform given in Table 2.1.

2.24 Find the ratio of the power contained in a pulse train for $|nf_0| \leq \tau^{-1}$ to the total power for each of the following cases:

(a) $\tau/T_0 = \frac{1}{2}$ (b) $\tau/T_0 = \frac{1}{5}$

2.25 (a) If $x(t)$ has the Fourier series

$$x(t) = \sum_{n=-\infty}^{\infty} X_n e^{j2\pi n f_0 t}$$

and $y(t) = x(t - t_0)$, show that

$$Y_n = X_n e^{-j2\pi n f_0 t_0}$$

where the Y_n's are the Fourier coefficients for $y(t)$.

(b) Verify the theorem proved in part (a) by examining the Fourier coefficients for $x(t) = \cos(\omega_0 t)$ and $y(t) = \sin(\omega_0 t)$. *Hint:* Use the uniqueness property of the Fourier series.

2.26 Use the Fourier series expansion of a periodic triangular signal to find the sum of the following series:

(a) $1 + \frac{1}{9} + \frac{1}{25} + \frac{1}{49} + \cdots$

(b) $1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots$

2.27 Using the results given in Table 2.1 for the Fourier coefficients of a pulse train, plot the double-sided amplitude and phase spectra for the waveforms shown in Figure 2.35. *Hint:* Note that $x_b(t) = -x_a(t) + A$. How is a sign change and dc level shift manifested in the spectrum of the waveform?

2.28 (a) Plot the single-sided and double-sided amplitude and phase spectra of the square wave shown in Figure 2.36(a). How do they differ from those plotted in Problem 2.27 for Figure 2.35(b)?

(b) Obtain an expression relating the complex exponential Fourier series coefficients of the triangular waveform shown in Figure 2.36(b) and those of $x_a(t)$ shown in Figure 2.36(a). *Hint:* Note that $x_a(t) = K[dx_b(t)/dt]$, where K is an appropriate scale change.

(c) Plot the double-sided amplitude and phase spectra for $x_b(t)$.

FIGURE 2.35

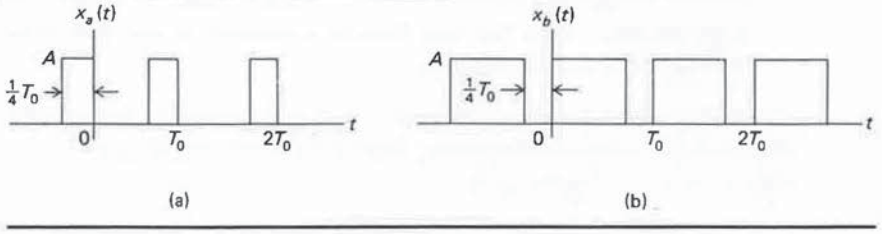
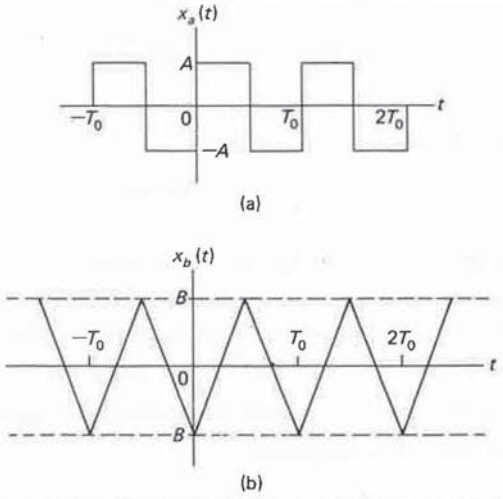


FIGURE 2.36



Section 2.5

2.29 Sketch each signal given below, and find its Fourier transform. Plot the amplitude and phase spectra of each signal (A and α are positive constants).

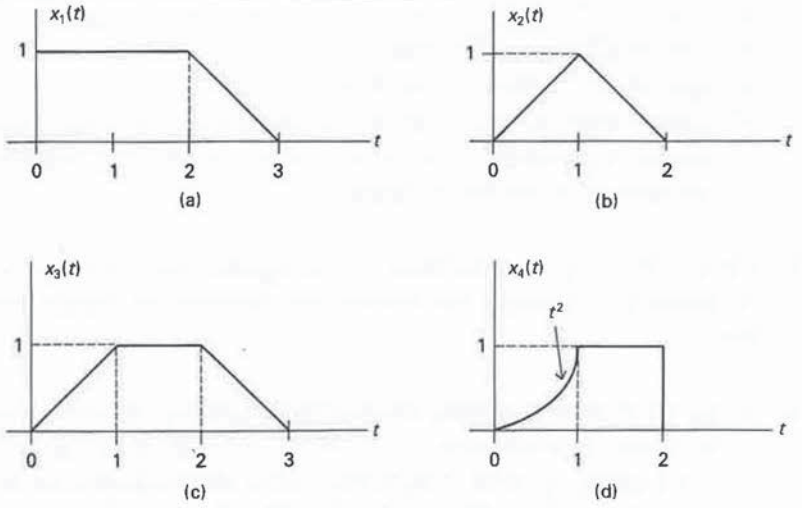
(a) $x_1(t) = Ae^{-\alpha t}u(t)$
(b) $x_2(t) = Ae^{\alpha t}u(-t)$
(c) $x_3(t) = x_1(t) - x_2(t)$
(d) Use the result of part (c) to find the Fourier transform of the signum function defined as

$$\text{sgn}(t) = \begin{cases} 1, & t > 0 \\ -1, & t < 0 \end{cases}$$

2.30 (a) Use the result of Problem 2.29(d) and the relation $u(t) = \frac{1}{2}[\text{sgn}(t) + 1]$ to find the Fourier transform of the unit step.
(b) Use the integration theorem and the Fourier transform of the unit impulse function to find the Fourier transform of the unit step. Compare the result with part (a).

2.31 Using only the Fourier transform of the unit impulse function and appropriate transform theorems, find the Fourier transforms of the signals shown in Figure 2.37.

FIGURE 2.37



2.32 Prove the multiplication theorem of Fourier transforms, showing all the steps.

2.33 Find the Fourier transforms of the signals given below. Discuss how any symmetry properties a given signal may have affect its Fourier transform in terms of being real or purely imaginary.

(a) $x_1(t) = \delta(t + 10) + 2\delta(t) + \delta(t - 10)$
(b) $x_2(t) = \delta(t + 5) - \delta(t - 5)$
(c) $x_3(t) = \Sigma_{n=0}^{4} (n + 1)\delta(t - 2n)$

2.34 Find and plot the energy spectral densities of the following signals. Dimension your plots fully.

(a) $x_1(t) = 2e^{-3|t|}$ (b) $x_2(t) = 20 \text{ sinc } (30t)$
(c) $x_3(t) = 4\Pi(5t)$ (d) $x_4(t) = 4\Pi(5t) \cos (40\pi t)$

2.35 Evaluate the following integrals using Rayleigh's energy theorem (Parseval's theorem for Fourier transforms).

(a) $I_1 = \int_{-\infty}^{\infty} \frac{df}{\alpha^2 + (2\pi f)^2}$ (b) $I_2 = \int_{-\infty}^{\infty} \text{sinc}^2 (\tau f)\, df$

(c) $I_3 = \int_{-\infty}^{\infty} \frac{df}{[\alpha^2 + (2\pi f)^2]^2}$ (d) $I_4 = \int_{-\infty}^{\infty} \text{sinc}^4 (\tau f)\, df$

2.36 Obtain and sketch the convolutions of the following signals.

(a) $y_1(t) = e^{-\alpha t}u(t)*\Pi(t - \tau)$, α and τ positive constants
(b) $y_2(t) = [\Pi(t/2) + \Pi(t)]*\Pi(t)$
(c) $y_3(t) = e^{-\alpha|t|}*\Pi(t)$, $\alpha > 0$
(d) $y_4(t) = x(t)*u(t)$, where $x(t)$ is any energy signal [you will have to assume a particular form for $x(t)$ to sketch this one, but obtain the general result before doing so].

2.37 Obtain the Fourier transforms of the signals $y_1(t)$, $y_2(t)$, and $y_3(t)$ in Problem 2.36 using the convolution theorem of Fourier transforms.

2.38 Given the following signals, suppose that all energy spectral components *outside* the bandwidth $|f| \leq W$ are removed by an ideal filter, while all energy spectral components within this bandwidth are kept. Find the ratio of energy to total energy kept in each case.

(a) $x_1(t) = e^{-\alpha t}u(t)$
(b) $x_2(t) = \Pi(t/\tau)$ (requires numerical integration)

2.39 (a) Find the Fourier transform of the cosine pulse

$$x(t) = A\Pi\left(\frac{2t}{T_0}\right)\cos \omega_0 t, \qquad \text{where } \omega_0 = \frac{2\pi}{T_0}$$

Express your answer in terms of the sinc function and sketch $|X(f)|$.

(b) Obtain the Fourier transform of the raised cosine pulse

$$y(t) = \tfrac{1}{2}A\Pi\left(\frac{2t}{T_0}\right)(1 + \cos 2\omega_0 t)$$

Sketch the amplitude spectrum, and compare with that of $x(t)$.

(c) Use Equation (2.139) with the result of part (a) to find the Fourier transform of the half-rectified cosine wave.

2.40 Show that the Fourier transform of $x(t) = e^{-\pi t^2}$ is $X(f) = e^{-\pi f^2}$. This is an example of a signal whose Fourier transform is of the same functional form as the signal itself.
[*Hint:* Complete the square in the exponent of the Fourier transform integral and use $\int_{-\infty}^{\infty} e^{-\alpha u^2}\, du = \sqrt{\pi/\alpha}$.]

Section 2.6

2.41 Show that the time-average correlation operation can be written in terms of convolution as

$$R(\tau) = \lim_{T\to\infty} \frac{1}{2T}[x(t) * x(-t)]\Big|_{t=\tau}$$

2.42 Obtain the time-average correlation function for a squarewave. Does it make any difference whether the squarewave has even or odd symmetry?

Section 2.7

2.43 A system is governed by the differential equation

$$\frac{dy}{dt} + ay = b\frac{dx}{dt} + cx$$

(a) Find $H(f)$.

(b) Find and plot $|H(f)|$ and $\underline{/H(f)}$ for $c = 0$.

(c) Find and plot $|H(f)|$ and $\underline{/H(f)}$ for $b = 0$.

2.44 For each of the following transfer functions, determine the unit-impulse response of the system.

(a) $H_1(f) = \dfrac{j2\pi f}{10 + j2\pi f}$ (b) $H_2(f) = \dfrac{3e^{-j6\pi f}}{4 + j20\pi f}$

2.45 A filter has transfer function $H(f) = \Pi(f/2B)$ and input $x(t) = 2W\,\text{sinc}\,(2Wt)$.

(a) Find the output $y(t)$ for $W < B$.

(b) Find the output $y(t)$ for $W > B$.

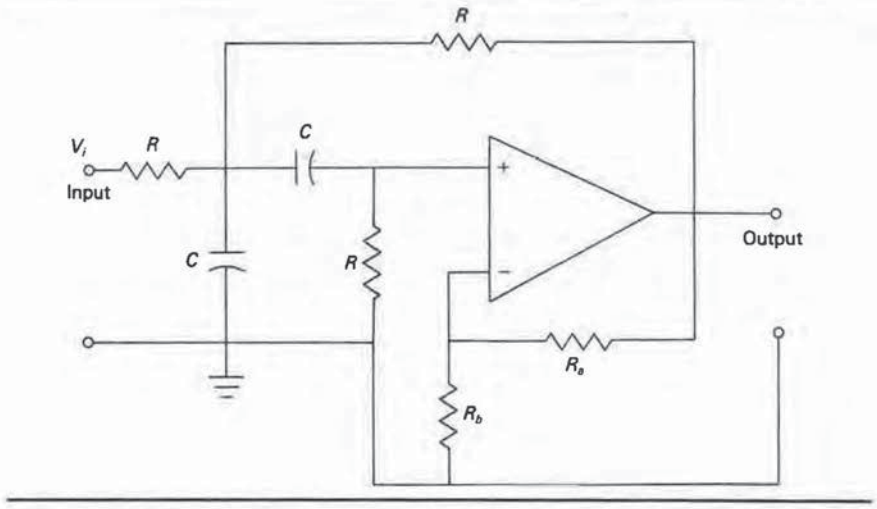
(c) In which case does the output suffer distortion? What influenced your answer?

2.46 A second-order active bandpass filter (BPF), known as a *Sallen-Key circuit*, is shown in Figure 2.38.

(a) Show that the transfer function of this filter is given by

$$H(j\omega) = \frac{(K\omega_0/\sqrt{2})(j\omega)}{-\omega^2 + (\omega_0/Q)(j\omega) + \omega_0^2}, \qquad \omega = 2\pi f$$

FIGURE 2.38



where

$$\omega_0 = \sqrt{2}(RC)^{-1}$$

$$Q = \frac{\sqrt{2}}{4 - K}$$

$$K = 1 + \frac{R_a}{R_b}$$

(b) Plot $|H(f)|$.

(c) Show that the 3-dB bandwidth of the filter can be expressed as $B = f_0/Q$, where $f_0 = \omega_0/2\pi$.

(d) Design a BPF using this circuit with center frequency $f_0 = 1000$ Hz and 3-dB bandwidth of 300 Hz. Find values of R_a, R_b, R, and C that will give these desired specifications.

2.47 For the two circuits shown in Figure 2.39, determine $H(f)$ and $h(t)$. Sketch accurately the amplitude and phase responses. Plot the amplitude response in decibels. Use a logarithmic frequency axis.

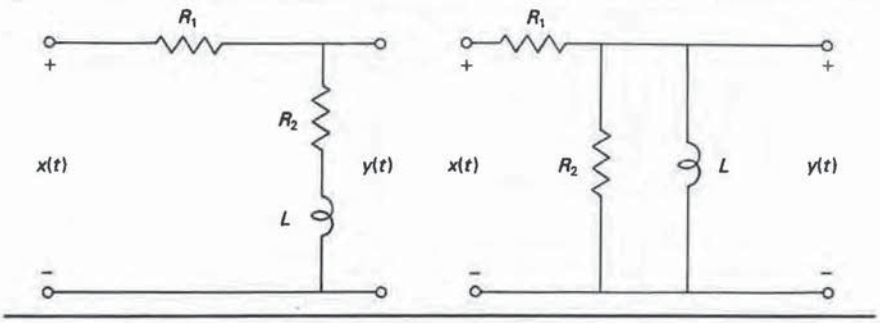
2.48 Using the Paley-Wiener criterion, show that

$$|H(f)| = e^{-\beta f^2}$$

is not a suitable amplitude response for a causal LTI.

2.49 Determine whether or not the filters shown in Figure 2.40 are BIBO stable.

FIGURE 2.39



2.50 Given a filter with transfer function

$$H(f) = \frac{20}{9 + (2\pi f)^2}$$

and input $x(t) = e^{-2t}u(t)$, obtain and plot accurately the energy spectral densities of the input and output.

2.51 A filter with transfer function

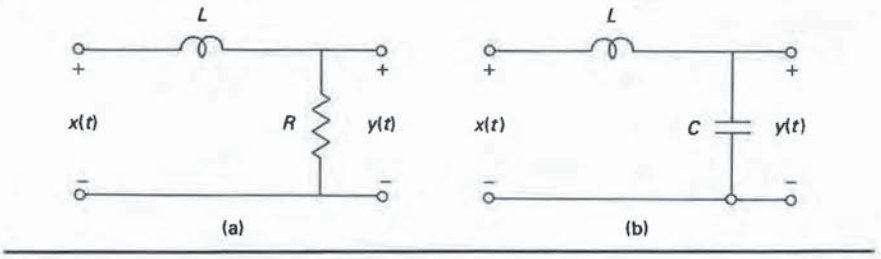
$$H(f) = 3\Pi(f/26)$$

has as an input a half-rectified cosine waveform of fundamental frequency 10 Hz. Determine the output of the filter.

2.52 Another definition of bandwidth for a signal is the 90% energy containment bandwidth. For a signal with energy spectral density $G(f) = |X(f)|^2$, it is given by B_{90} in the relation

$$0.9E_{\text{Total}} = \int_{-B_{90}}^{B_{90}} G(f)\,df = 2\int_{0}^{B_{90}} G(f)\,df$$

FIGURE 2.40



Obtain B_{90} for the following signals if it is defined. If it is not defined for a particular signal, state why it is not.

(a) $x_1(t) = \sqrt{2\alpha}\, e^{-\alpha t}u(t)$, where α is a positive constant
(b) $x_2(t) = 2W \operatorname{sinc}(2Wt)$
(c) $x_3(t) = \Pi(t/\tau)$ (requires numerical integration)

2.53 An ideal quadrature phase shifter has

$$H(f) = \begin{cases} e^{-j\pi/2}, & f > 0 \\ e^{+j\pi/2}, & f < 0 \end{cases}$$

If the input is a squarewave, sketch an approximation for the output waveform using the first four terms of its trigonometric Fourier series as obtained in Example 2.7.

2.54 A filter has amplitude response and phase shift shown in Figure 2.41. Find the output for each of the inputs given below. For which cases is the transmission distortionless? Tell what type of distortion is imposed for the others.

(a) $\cos 50\pi t + 5 \cos 120\pi t$ (b) $\cos 120\,\pi t + 0.5 \cos 160\pi t$
(c) $\cos 120\pi t + 3 \cos 140\pi t$ (d) $2 \cos 20\pi t + 4 \cos 40\pi t$

2.55 A system has the unit impulse response

$$h(t) = 4e^{-5t}u(t)$$

Determine and accurately plot, on the same set of axes, the group delay and the phase delay.

2.56 A system has the transfer function

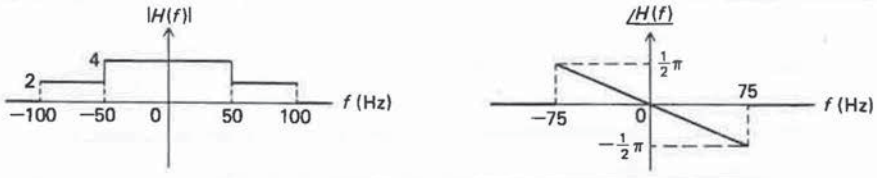
$$H(f) = \frac{3 + j2\pi f}{10 + j2\pi f}$$

Determine and accurately plot the group delay and the phase delay.

2.57 The nonlinear system defined by

$$y(t) = x(t) + 0.3x^2(t)$$

FIGURE 2.41



has an input signal with the bandpass spectrum

$$X(f) = 4\Pi\left(\frac{f-20}{10}\right) + 4\Pi\left(\frac{f+20}{10}\right)$$

Sketch the spectrum of the output, labeling all important frequencies and amplitudes.

2.58 (a) Consider a nonlinear device with the transfer characteristic $y(t) = x(t) + 0.1x^3(t)$. The frequency of the input signal $x(t) = \cos 2000\pi t$ is to be tripled by passing the signal through the nonlinearity and then through a second-order bandpass filter with a transfer function approximated by

$$H(f) = \frac{1}{1 + j2Q(f-3000)} + \frac{1}{1 + j2Q(f+3000)}$$

Neglecting negative frequency contributions, compute, in terms of the parameter Q, the *total harmonic distortion* (THD) at the tripler output, defined as

$$\text{THD} = \frac{\text{total power in all output distortion terms}}{\text{power in desired output component}} \times 100\%$$

Note that the desired output component in this case is the third harmonic of the input frequency.

(b) Find the minimum value of Q that will result in THD $\leq 0.01\%$.

2.59 A nonlinear device has $y(t) = a_0 + a_1x(t) + a_2x^2(t) + a_3x^3(t)$. If $x(t) = \cos \omega_1 t + \cos \omega_2 t$, list all the frequency components present in $y(t)$. Discuss the use of this device as a frequency multiplier.

2.60 (a) For a Butterworth filter, show that

$$\lim_{n\to\infty} |H_{\text{BU}}(f)| = \begin{cases} 1, & |f| < f_3 \\ 0, & |f| > f_3 \end{cases}$$

(b) For a Chebyshev filter, show that

$$\epsilon = \sqrt{10^{R_{dB}/10} - 1}$$

where R_{dB} is the ripple in dB.

2.61 Find the impulse response of an ideal highpass filter with the transfer function

$$H_{\text{HP}}(f) = H_0\left[1 - \Pi\left(\frac{f}{2W}\right)\right] e^{-j2\pi f t_0}$$

2.62 Sketch the amplitude response and the group delay for a second-order Chebyshev filter for $\epsilon = 0.5$, 0.707, and 1.0.

2.63 Verify the pulsewidth-bandwidth relationship of Equation (2.227) for the following signals. Sketch each signal and its spectrum.

(a) $x(t) = Ae^{-t^2/2\tau^2}$ (Gaussian pulse) (see Problem 2.40)
(b) $x(t) = Ae^{-\alpha|t|}$, $\alpha > 0$ (double-sided exponential)

Section 2.8

2.64 A sinusoidal signal of frequency 1 Hz is to be sampled periodically.

(a) Find the maximum allowable time interval between samples.
(b) Samples are taken at $\frac{1}{3}$-s intervals. Show graphically, to your satisfaction, that no other sinewave or signal with bandwidth less than 1.5 Hz can be represented by these samples.
(c) The samples are spaced $\frac{2}{3}$ s apart. Show graphically that these may represent another sinewave of frequency less than 1.5 Hz.

2.65 A flat-top sampler can be represented as the block diagram of Figure 2.42.

(a) Assuming $\tau \ll T_s$, sketch the output for a typical $x(t)$.
(b) Find the spectrum of the output, $Y(f)$, in terms of the spectrum of the input, $X(f)$. What must be the relationship between τ and T_s to minimize distortion in the recovered waveform?

2.66 Figure 2.43 illustrates so-called *zero-order-hold reconstruction*.

(a) Sketch $y(t)$ for a typical $x(t)$. Under what conditions is $y(t)$ a good approximation to $x(t)$?
(b) Find the spectrum of $y(t)$ in terms of the spectrum of $x(t)$. Discuss the approximation of $y(t)$ to $x(t)$ in terms of frequency-domain arguments.

2.67 The signal $x(t) = \cos 20\pi t$ is sampled at 15 samples per second. Show that the signal $\cos 10\pi t$ impersonates it as far as sample values are concerned.

FIGURE 2.42

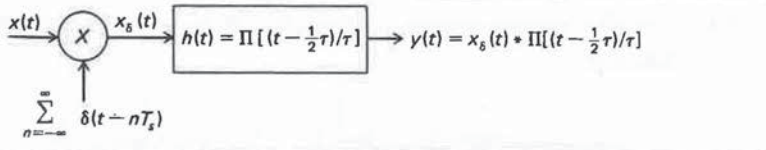
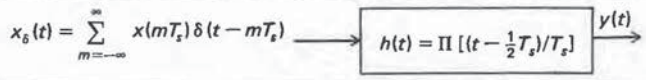


FIGURE 2.43



2.68 Determine the range of permissible cutoff frequencies for the ideal lowpass filter used to reconstruct the signal

$$x(t) = 10 \cos (600\pi t) \cos^2 (1600\pi t)$$

which is sampled at 4000 samples per second. Sketch $X(f)$ and $X_\delta(f)$. Find the minimum allowable sampling frequency.

2.69 Given the bandpass signal spectrum shown in Figure 2.44, sketch spectra for the following sampling rates f_s and indicate which ones are suitable.

(a) $2B$ (b) $2.5B$ (c) $3B$
(d) $4B$ (e) $5B$ (f) $6B$

Section 2.9

2.70 Using the integral expression for the Hilbert transform, Equation (2.256), show that

$$\widehat{\cos \omega_0 t} = \sin \omega_0 t$$

2.71 Using appropriate Fourier transform theorems and pairs, express the spectrum $Y(f)$ of

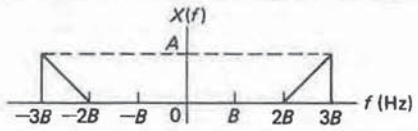
$$y(t) = x(t) \cos \omega_0 t + \hat{x}(t) \sin \omega_0 t$$

in terms of the spectrum $X(f)$ of $x(t)$, where $X(f)$ is lowpass with bandwidth

$$B < f_0 = \frac{\omega_0}{2\pi}$$

Sketch $Y(f)$ for a typical $X(f)$.

FIGURE 2.44



2.72 Show that $x(t)$ and $\hat{x}(t)$ are orthogonal for the following signals:

(a) $x_a(t) = \cos \omega_0 t$
(b) $x_b(t) = 2 \cos \omega_0 t + \sin \omega_0 t \cos^2 2\omega_0 t$
(c) $x_c(t) = A\Pi(t/\tau)$

2.73 (a) Using the results of Problem 2.71, sketch $Y(f)$ for $x(t) = \cos 10\pi t$ and $f_0 = \omega_0/2\pi = 20$ Hz.
(b) Obtain the spectrum for the same signal by using trigonometric identities and Equations (2.260) and (2.261).

2.74 Assume that the Fourier transform of $x(t)$ is real and has the shape shown in Figure 2.45. Determine and plot the spectrum of each of the following signals:

(a) $x_1(t) = \frac{3}{4}x(t) + \frac{1}{4}j\hat{x}(t)$
(b) $x_2(t) = [\frac{3}{4}x(t) + \frac{3}{4}j\hat{x}(t)]e^{j2\pi f_0 t}, \quad f_0 \gg W$
(c) $x_3(t) = [\frac{3}{4}x(t) + j\frac{1}{4}\hat{x}(t)]e^{j2\pi Wt}$
(d) $x_4(t) = [\frac{3}{4}x(t) - j\frac{1}{4}\hat{x}(t)]e^{j\pi Wt}$

2.75 Consider the signal

$$x(t) = 2W \operatorname{sinc}(2Wt) \cos(2\pi f_0 t)$$

(a) Obtain and sketch the spectrum of $x_p(t) = x(t) + j\hat{x}(t)$.
(b) Obtain and sketch the spectrum of the complex envelope, $\tilde{x}(t)$, where the complex envelope is defined by (2.286).
(c) Find the complex envelope $\tilde{x}(t)$.

2.76 Consider the input

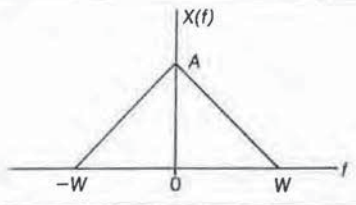
$$x(t) = \Pi(t/\tau) \cos[2\pi(f_0 + \Delta f)t], \qquad \Delta f \ll f_0$$

to a filter with impulse response

$$h(t) = \alpha e^{-\alpha t} \cos(2\pi f_0 t)u(t)$$

Find the output using complex envelope techniques.

FIGURE 2.45



Section 2.10

2.77 Prove that (2.303) and (2.304) are indeed inverses of each other by substituting (2.303) into (2.304), reversing the orders of summation, and using the sum formula for a geometric series to show that the final result is $x_n = x_n$.



COMPUTER EXERCISES*

1. (a) Write a computer program to obtain the generalized Fourier series for an energy signal using the orthonormal basis set

$$\Phi_n(t) = \Pi(t - 0.5 - n),\ n = 0, 1, 2, \cdots, T - 1,\ T \text{ integer}$$

where the signal extent is $(0, T)$ with T assumed to be integer-valued. Your program should compute the generalized Fourier coefficients, the integral-squared error, and make a plot of the signal being approximated and the approximating waveform. Test your program with the signal $e^{-2t}u(t)$, $0 \leq t \leq 5$.

(b) Repeat part (a) with the orthonormal basis set

$$\Phi_n(t) = \sqrt{2}\Pi\left(\frac{t - 0.5 - n}{0.5}\right),\ n = 0, 1, 2, \cdots, 2T - 1,\ T \text{ integer}$$

What is the integral-squared error now?

(c) Can you deduce whether or not the basis set resulting from repeatedly halving the pulse width and doubling the amplitude is complete?

2. Write a computer program to evaluate the coefficients of the complex exponential Fourier series of a signal. Include a plot of the amplitude and phase spectrum of the signal for which the Fourier series coefficients are evaluated. Check by evaluating the Fourier series coefficients of a squarewave. Plot the squarewave approximation by summing the series through the seventh harmonic.

3. (a) Write a computer program to evaluate the coefficients of the complex exponential Fourier series of a signal by using the fast Fourier

* When doing these computer exercises, we suggest that the student make use of a mathematics package such as Mathcad, MATHLAB, or Mathematica. Although the resulting programs may run slower that those written in C, Fortran, or some other high-level language, considerable time will be saved in being able to use the plotting capability of these mathematics packages.

transform (FFT). Note that you do not have to write a program to evaluate the FFT if the suggestion in the footnote is followed. Check it by evaluating the Fourier series coefficients of a squarewave and comparing your results with Computer Exercise 2.

(b) How would you use the same approach to evaluate the Fourier transform of a pulse-type signal. How do the two outputs differ? Compute an approximation to the Fourier transform of a square pulse signal 1 unit wide and compare with the theoretical result.

4. (a) Write a computer program to find the bandwidth of a lowpass energy signal that contains a certain specified percentage of its total energy, for example, 95%. In other words, write a program to find W in the equation

$$E_W = \frac{\int_0^W G_x(f)df}{\int_0^\infty G_x(f)df} \times 100\%$$

with E_W set equal to a specified value, where $G_x(f)$ is the energy spectral density of the signal.

(b) Write a computer program to find the time duration of a lowpass energy signal that contains a certain specified percentage of its total energy, for example, 95%. In other words, write a program to find T in the equation

$$E_T = \frac{\int_0^T |x(t)|^2\, dt}{\int_0^\infty |x(t)|^2\, dt} \times 100\%$$

with E_T set equal to a specified value, where it is assumed that the signal is zero for $t < 0$.

(c) Use the program you developed in parts (a) and (b) to check the pulse-duration bandwidth relationship (2.227).

5. Write a computer program to evaluate and plot the amplitude and phase response functions for a given transfer function. Use it to plot the amplitude and phase response functions for the Sallen-Key circuit of Problem 2.46 with the parameters chosen to realize a second-order Butterworth filter.

6. Using the FFT, write a program to find the output of a bandpass system for an arbitrary bandpass input using the complex envelope approach summarized by (2.295) and (2.296). Your program should include plots of the envelopes of the input and output as well as their spectra.

3 BASIC MODULATION TECHNIQUES

Before an information-bearing signal is transmitted through a communication channel, some type of modulation process is typically utilized to produce a signal that can easily be accommodated by the channel. In this chapter we will discuss various types of modulation techniques. The modulation process commonly translates an information-bearing signal, usually referred to as the *message signal,* to a new spectral location. For example, if the signal is to be transmitted through the atmosphere or free space, frequency translation is necessary to raise the signal spectrum to a frequency that can be radiated efficiently with antennas of reasonable size. If more than one signal utilizes a channel, modulation allows translation of the different signals to different spectral locations, thus allowing the receiver to select the desired signal. Multiplexing allows two or more message signals to be transmitted by a single transmitter and received by a single receiver simultaneously.

The logical choice of a modulation technique for a specific application is influenced by the characteristics of the message signal, the characteristics of the channel, the performance desired from the overall communication system, the use to be made of the transmitted data, and the economic factors that are always important in practical applications. This text will place emphasis on the first three of these factors. Through experience, the communications engineer develops sufficient insight to place the last two factors in proper perspective for a particular application.

The two basic types of analog modulation are continuous-wave modulation and pulse modulation. In *continuous-wave modulation,* a parameter of a high-frequency carrier is varied proportionally to the message signal such that a one-to-one correspondence exists between the parameter and the message signal. The carrier is usually assumed to be sinusoidal, but as will be illustrated, this is not a necessary restriction. For a sinusoidal carrier, a general modulated carrier can be represented mathematically as

$$x_c(t) = A(t) \cos [\omega_c t + \phi(t)] \tag{3.1}$$

where ω_c is known as the *carrier frequency*. Since a sinusoid is completely specified by its amplitude and argument, it follows that once the carrier frequency is specified, only two parameters are candidates to be varied in the modulation process: the instantaneous amplitude $A(t)$ and the instantaneous phase deviation $\phi(t)$. When the instantaneous amplitude $A(t)$ is linearly related to the modulating signal, the result is *linear modulation*. Letting $\phi(t)$ or the time derivative of $\phi(t)$ be linearly related to the modulating signal yields phase or frequency modulation, respectively. Collectively, phase and frequency modulation are referred to as *angle modulation*, since the phase angle of the modulated carrier conveys the information.

In *analog pulse modulation*, the message waveform is sampled at discrete time intervals and the amplitude, width, or position of a pulse is varied in one-to-one correspondence with the values of the samples. Since the samples are taken at discrete times, the periods between the samples are available for other uses, such as insertion of samples from other message signals. This is referred to as *time-division multiplexing*. If the value of each sample is quantized and encoded, *pulse-code modulation results*. We also briefly consider *delta modulation*. Pulse-code modulation and delta modulation are digital rather than analog modulation techniques, but are considered in this chapter for completeness.

3.1 LINEAR MODULATION

A general linearly modulated carrier is represented by setting the instantaneous phase deviation $\phi(t)$ in (3.1) equal to zero. Thus a linearly modulated carrier is represented by

$$x_c(t) = A(t) \cos \omega_c t \tag{3.2}$$

in which the carrier amplitude $A(t)$ varies in one-to-one correspondence with the message signal. We next discuss several different types of linear modulation as well as techniques that can be used for demodulation.

Double-Sideband Modulation

Double-sideband (DSB) modulation results when $A(t)$ is proportional to the message signal $m(t)$. Thus the output of a DSB modulator can be represented as

$$x_c(t) = A_c m(t) \cos \omega_c t \tag{3.3}$$

which illustrates that DSB modulation is simply the multiplication of a carrier, $A_c \cos \omega_c t$, by the message signal. It follows from the modulation the-

orem that the spectrum of a DSB signal is given by

$$X_c(f) = \tfrac{1}{2}A_cM(f + f_c) + \tfrac{1}{2}A_cM(f - f_c), \qquad f_c = \frac{\omega_c}{2\pi} \tag{3.4}$$

The process of DSB modulation is described in Figure 3.1. Figure 3.1(a) illustrates a DSB system and shows that a DSB signal is demodulated by multiplying the received signal, denoted by $x_r(t)$, by the demodulation carrier $2\cos\omega_c t$ and lowpass filtering. For the idealized system that we are considering here, the received signal, $x_r(t)$, is identical to the transmitted signal, $x_c(t)$. The output of the multiplier is

$$d(t) = 2A_c[m(t)\cos\omega_c t]\cos\omega_c t \tag{3.5}$$

or

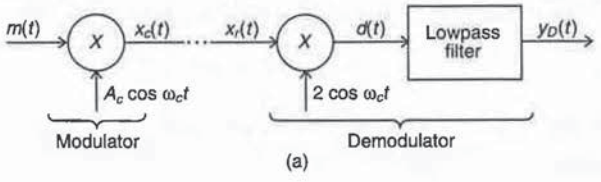
$$d(t) = A_cm(t) + A_cm(t)\cos 2\omega_c t \tag{3.6}$$

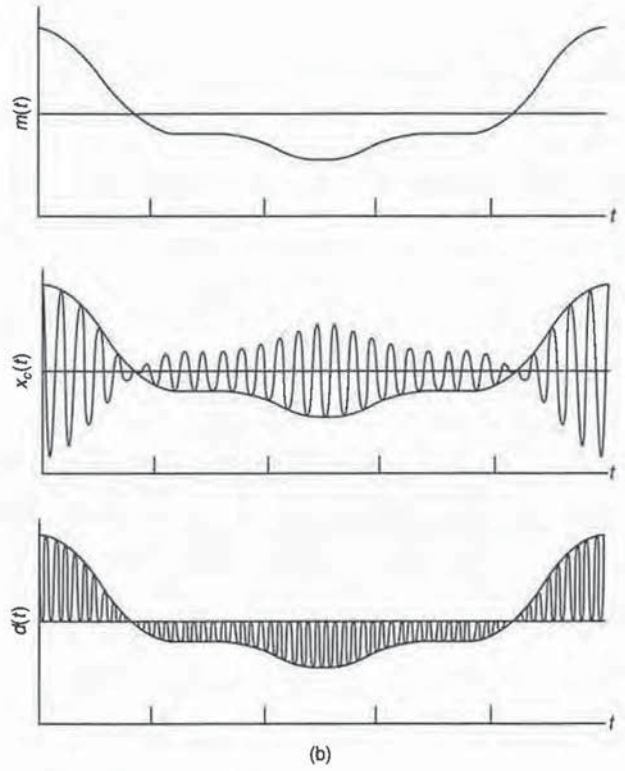
where we have used the trigonometric identity $2\cos^2 x = 1 + \cos 2x$.

The time-domain signals are shown in Figure 3.1(b) for an assumed $m(t)$. The message signal $m(t)$ forms the envelope of $x_c(t)$. The waveform for $d(t)$ can be best understood by realizing that since $\cos^2\omega_c t$ is nonnegative for all t, $d(t)$ is positive if $m(t)$ is positive and $d(t)$ is negative if $m(t)$ is negative. Also note that $m(t)$ (appropriately scaled) forms the envelope of $d(t)$ and that the frequency of the sinusoid under the envelope is $2f_c$ rather than f_c.

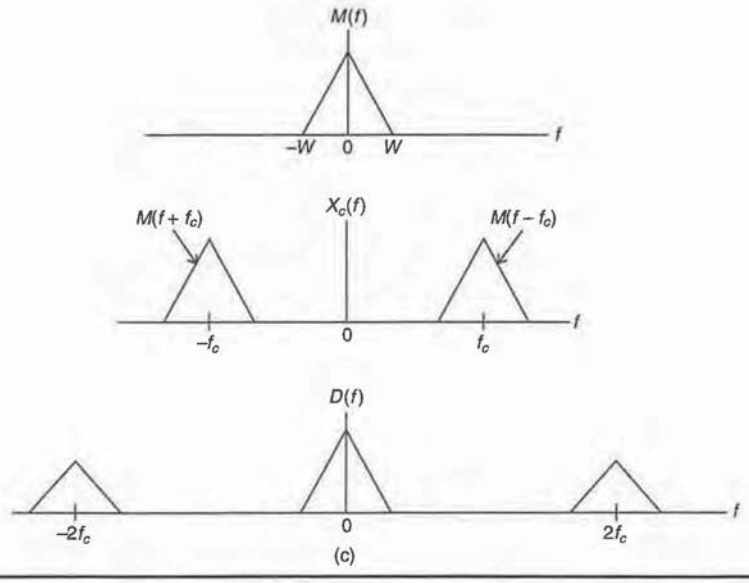
The spectra of the signals $m(t)$, $x_c(t)$, and $d(t)$ are shown in Figure 3.1(c) for an assumed $M(f)$ having a bandwidth W. The spectra $M(f + f_c)$ and $M(f - f_c)$ are simply the message spectrum translated to $f = \pm f_c$. The portion of $M(f - f_c)$ above the carrier frequency is called the *upper sideband* (USB), and the portion below the carrier frequency is called the *lower sideband* (LSB). Since the carrier frequency f_c is typically much greater than the bandwidth of the message signal W, the spectra of the two terms in $d(t)$ do not overlap. Thus $d(t)$ can be lowpass-filtered and amplitude-scaled by A_c to yield the demodulated output $y_D(t)$. In practice, any amplitude scaling factor can be used since, as we saw in Chapter 2, multiplication by a constant does not

FIGURE 3.1 Double-sideband (DSB) modulation. (a) System. (b) Waveforms. (c) Spectra.





(b)



(c)

induce amplitude distortion. Thus, for convenience, A_c is usually set equal to unity at the demodulator output. For this case, the demodulated output $y_D(t)$ will equal the message signal $m(t)$. The lowpass filter that removes the term at $2f_c$ must have a bandwidth greater than or equal to the bandwidth of the message signal W. We will see in Chapter 6 that when noise is present, this lowpass filter, known as the *postdetection* filter, should have the smallest possible bandwidth.

The basic difficulty with the use of DSB modulation is the need for a demodulation carrier at the receiver that has the same frequency as the original modulation carrier and is phase-coherent with it. Demodulation that requires a coherent reference is known as *synchronous* or *coherent demodulation*. The derivation of a coherent demodulation carrier requires careful attention, for if the demodulation carrier is out of synchronism by even a small amount, serious distortion of the demodulated message waveform can result. This effect will be thoroughly analyzed in Chapter 6, but a simplified analysis can be carried out by assuming a demodulation carrier in Figure 3.1(a) of the form $2 \cos [\omega_c t + \theta(t)]$, where $\theta(t)$ is a time-varying phase error. By making use of the trigonometric identity $\cos x \cos y = \frac{1}{2} \cos (x + y) + \frac{1}{2} \cos (x - y)$, it can be shown that this choice of demodulation carrier yields

$$d(t) = A_c m(t) \cos \theta(t) + A_c m(t) \cos [2\omega_c t + \theta(t)] \tag{3.7}$$

which, after lowpass filtering and amplitude scaling, becomes

$$y_D(t) = m(t) \cos \theta(t) \tag{3.8}$$

assuming, once again, that the spectra of the two terms of $d(t)$ do not overlap. If the phase error $\theta(t)$ is a constant, the effect of the phase error is an attenuation of the demodulated message signal. This does not represent distortion, since the effect of the phase error can be removed by amplitude scaling unless $\theta(t)$ is exactly 90°. However, if $\theta(t)$ is time-varying in an unknown and unpredictable manner, the effect of the phase error can be serious distortion of the demodulated output.

There are several techniques that can be utilized to generate a coherent demodulation carrier. For one technique, the first step is to square the received DSB signal, which yields

$$\begin{aligned} x_r^2(t) &= A_c^2 m^2(t) \cos^2 \omega_c t \\ &= \tfrac{1}{2} A_c^2 m^2(t) + \tfrac{1}{2} A_c^2 m^2(t) \cos 2\omega_c t \end{aligned} \tag{3.9}$$

If $m(t)$ is a power signal, $m^2(t)$ has a nonzero dc value. Thus, by the modulation theorem, $x_r^2(t)$ has a discrete frequency component at $2\omega_c$, which can be extracted from the spectrum of $x_r^2(t)$ using a narrowband bandpass filter.

The frequency of this component can be halved to yield the desired demodulation carrier. Later we will discuss a convenient technique for implementing the required frequency divider.

A very convenient technique for demodulating DSB signals is to use a Costas phase-lock loop. This system will be presented in Section 3.4.

The analysis of DSB illustrates that the spectrum of a DSB signal does not contain a discrete spectral component at the carrier frequency unless $m(t)$ has a dc component. For this reason, DSB systems with no carrier frequency component present are often referred to as *suppressed carrier systems*. However, if a carrier component is transmitted along with the DSB signal, demodulation can be simplified. The received carrier component can be extracted using a narrowband bandpass filter and can be used as the demodulation carrier. Alternatively, if the carrier amplitude is sufficiently large, the need for generating a demodulation carrier can be completely avoided. This naturally leads to the subject of amplitude modulation.

Amplitude Modulation

Amplitude modulation (AM) results when a dc bias A is added to $m(t)$ prior to the modulation process. The result of the dc bias is that a carrier component is present in the transmitted signal. For AM, the transmitted signal is defined as

$$x_c(t) = [A + m(t)]A_c' \cos \omega_c t \tag{3.10}$$

or

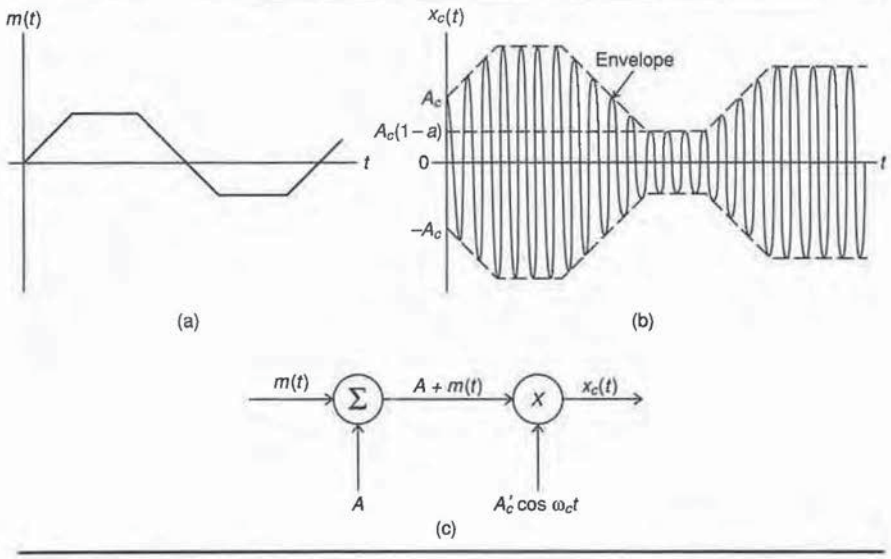
$$x_c(t) = A_c[1 + am_n(t)] \cos \omega_c t \tag{3.11}$$

In (3.11), $A_c = AA_c'$, $m_n(t)$ is the message signal normalized such that the minimum value of $m_n(t)$ is -1, and the parameter a is known as the modulation index.* We shall assume that $m(t)$ has zero dc value so that the carrier component in the transmitted signal arises entirely from the bias, A. The time-domain representation of AM is illustrated in Figures 3.2(a) and 3.2(b), and the block diagram of the modulator for producing AM is shown in Figure 3.2(c).

The fact that coherent demodulation can be used for AM is easily shown and is left to the problems (see Problem 3.2). The advantage of AM over DSB, however, is that a very simple technique, known as *envelope detection,* can be used if sufficient carrier power is transmitted. An envelope detector

* The parameter a as used here is sometimes called the *negative modulation factor*. Also, the quantity $a \times (100\%)$ is often referred to as the *percent modulation*.

FIGURE 3.2 AM modulation. (a) Message signal. (b) Modulator output for $a < 1$. (c) Modulator.



is implemented as shown in Figure 3.3(a). It can be seen from Figure 3.3(b) that as the carrier frequency is increased, the envelope, defined as $A_c[1 + am_n(t)]$, becomes easier to observe. It also follows from observation of Figure 3.3(b) that if $A_c[1 + am_n(t)]$ goes negative, distortion will result. The modulation index is defined such that if $a = 1$, the minimum value of $A_c[1 + am_n(t)]$ is zero. We therefore define

$$m_n(t) = \frac{m(t)}{|\min m(t)|} \tag{3.12}$$

and

$$a = \frac{|\min m(t)|}{A} \tag{3.13}$$

Clearly, any value of index less than 1 will result in $A_c[1 + am_n(t)] > 0$ for all t. It should be pointed out that most practical information-bearing message signals, such as speech or music signals, have nearly symmetrical maximum and minimum values.

In order for the envelope-detection process to operate properly, the RC time constant of the detector, shown in Figure 3.3(a), must be chosen carefully. The appropriate value for the time constant is related to the carrier frequency and to the bandwidth of $m(t)$. In practice, satisfactory operation

requires a carrier frequency of at least 10 times the bandwidth of $m(t)$, W. Also, the cutoff frequency of the RC circuit must lie between f_c and W and must be well separated from both. This is illustrated in Figure 3.3(c).

All information in the modulator output is contained in the sidebands. Thus the carrier component of (3.11), $A_c \cos \omega_c t$, is wasted power as far as information transfer is concerned. This fact can be of considerable importance in an environment where power is limited and can completely preclude the use of AM as a modulation technique.

The total (normalized) power contained in the AM modulator output is

$$\langle x_c^2(t) \rangle = \langle [A + m(t)]^2 (A_c')^2 \cos^2 \omega_c t \rangle \tag{3.14}$$

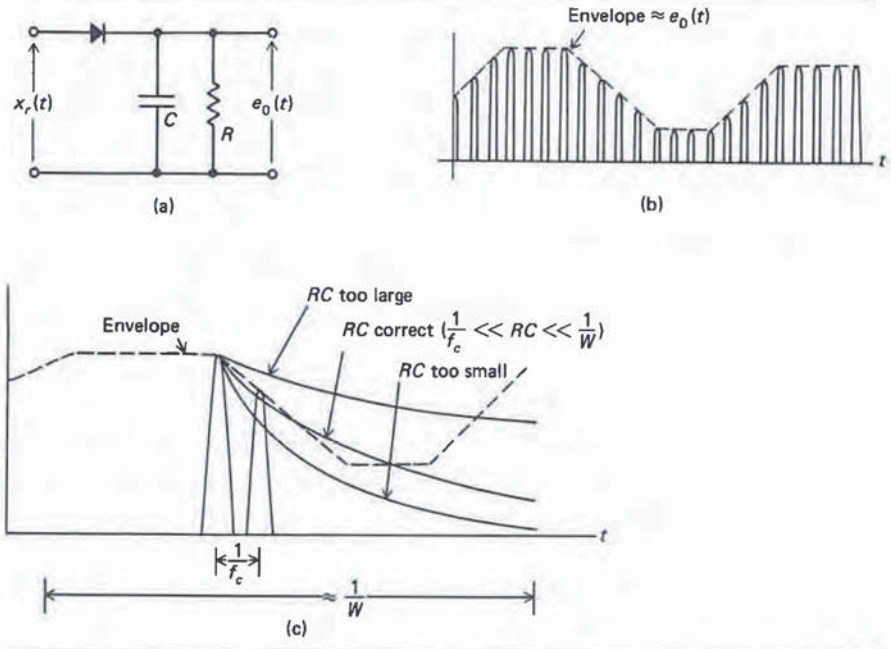
where $\langle\ \rangle$ denotes the time average value. If $m(t)$ is slowly varying with respect to $\cos \omega_c t$, then

$$\langle x_c^2(t) \rangle = [A^2 + 2A\langle m(t) \rangle + \langle m^2(t) \rangle] \tfrac{1}{2} (A_c')^2 \tag{3.15a}$$

which reduces to

$$\langle x_c^2(t) \rangle = [A^2 + \langle m^2(t) \rangle] \tfrac{1}{2} (A_c')^2 \tag{3.15b}$$

FIGURE 3.3 Envelope detection. (a) Circuit. (b) Waveforms. (c) Effect of RC time constant.



if the time average value of $m(t)$ is assumed zero. The *efficiency*, E, is defined as the percentage of total power that conveys information. Thus

$$E = \frac{\langle m^2(t) \rangle}{A^2 + \langle m^2(t) \rangle} (100\%) \tag{3.16}$$

Since $m(t) = aAm_n(t)$, efficiency can be written in terms of the modulation index as

$$E = \frac{a^2 \langle m_n^2(t) \rangle}{1 + a^2 \langle m_n^2(t) \rangle} (100\%) \tag{3.17}$$

If the message signal has symmetrical maximum and minimum values, such that $|\min m(t)|$ and $|\max m(t)|$ are equal, then $\langle m_n^2(t) \rangle \leq 1$. It follows that for $a \leq 1$, the maximum efficiency is 50% and is achieved for squarewave-type message signals. If $m(t)$ is a sinewave, $\langle m_n^2(t) \rangle = \frac{1}{2}$ and the efficiency is 33.3% for $a = 1$. Efficiency obviously declines rapidly as the index is reduced below unity. If the message signal does not have symmetrical maximum and minimum values, then higher values of efficiency can be achieved (see Problem 3.6).

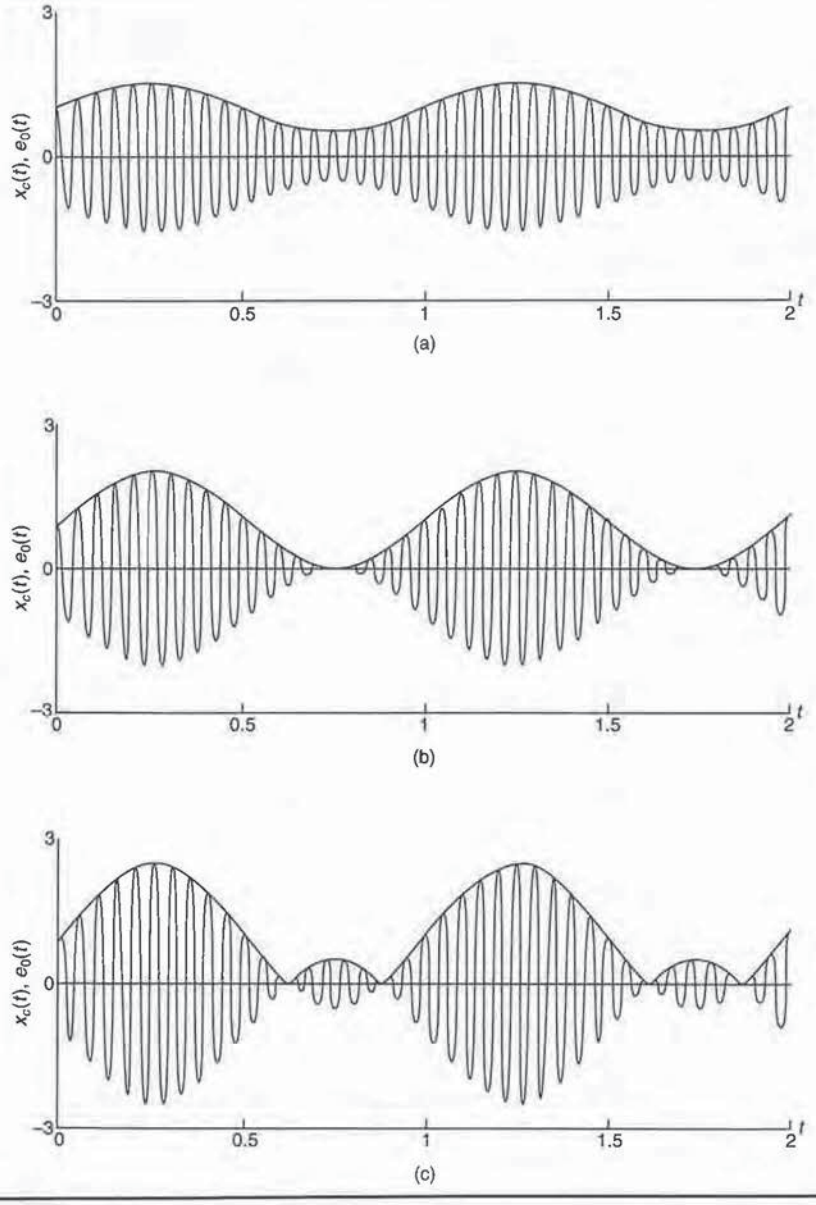
The main advantage of AM is that since a coherent reference is not needed for demodulation, the demodulator becomes simple and inexpensive. In many applications, such as commercial radio, this fact alone is sufficient to justify its use.

The AM modulator output $x_c(t)$ is shown in Figure 3.4 for three values of the modulation index: $a = 0.5$, $a = 1.0$, and $a = 1.5$. The message signal $m(t)$ is assumed to be a unity amplitude sinusoid with a frequency of 1 Hz. A unity amplitude carrier is also assumed. The envelope detector output, $e_o(t)$, as identified in Figure 3.3, is also shown for each value of the modulation index. Note that for $a = 0.5$ the envelope is always positive. For $a = 1.0$ the envelope reaches zero. Thus, envelope detection can be used for both of these cases. For $a = 1.5$ the envelope goes negative and $e_o(t)$, which is the absolute value of the envelope, is a badly distorted replica of the message signal.

▼ **EXAMPLE 3.1** Let us determine the efficiency and the output spectrum for an AM modulator operating with a modulation index of 0.5. The carrier power is 50 W, and the message signal is

$$m(t) = 4 \cos \left[2\pi f_m t - \frac{\pi}{9} \right] + 2 \sin [4\pi f_m t] \tag{3.18}$$

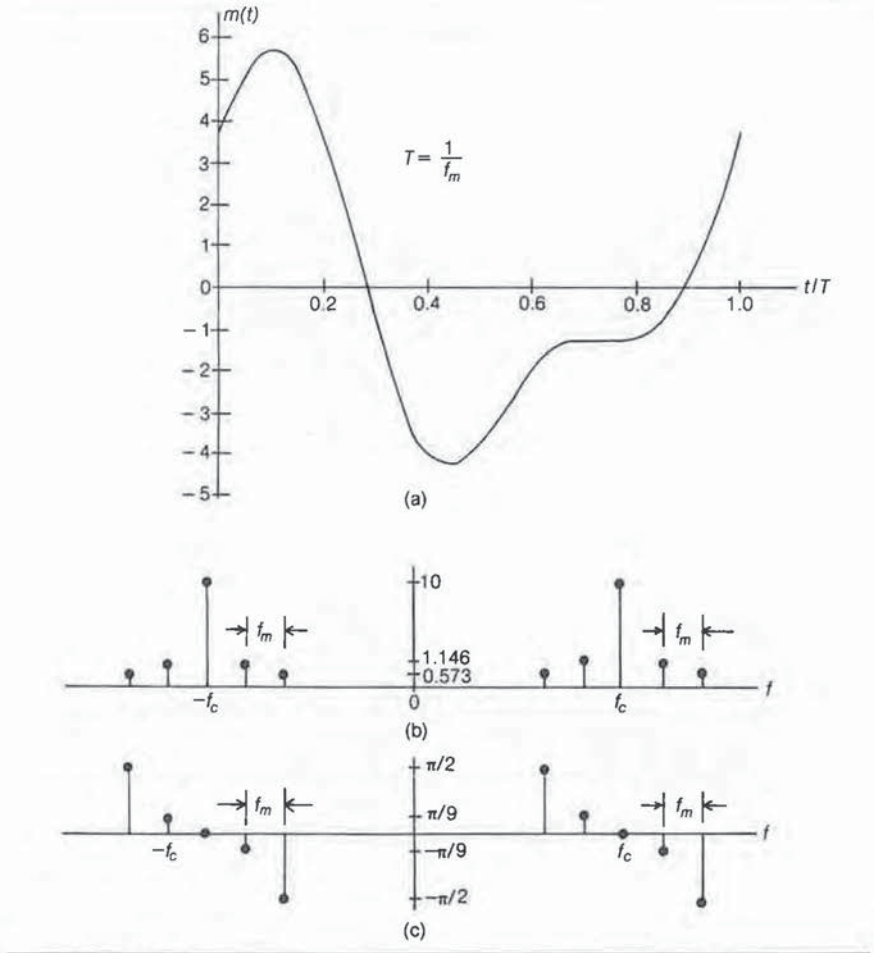
FIGURE 3.4 Modulated carrier and envelope detector outputs for various values of the modulation index. (a) $a = 0.5$. (b) $a = 1.0$. (c) $a = 1.5$.



SOLUTION The first step is to determine the minimum value of $m(t)$. The message signal is shown in Figure 3.5(a). The minimum value of $m(t)$ is -4.364 and falls at $f_m t = 0.435$, as shown. The normalized message signal is given by

$$m_n(t) = \frac{1}{4.364}\left\{4 \cos\left[2\pi f_m t - \frac{\pi}{9}\right] + 2 \sin\left[4\pi f_m t\right]\right\} \qquad (3.19)$$

FIGURE 3.5 Waveform and spectra for Example 3.1 (a) Message signal. (b) Amplitude spectrum of modulator output. (c) Phase spectrum of modulator output.



or

$$m_n(t) = 0.9166 \cos\left[2\pi f_m t - \frac{\pi}{9}\right] + 0.4583 \sin\left[4\pi f_m t\right] \tag{3.20}$$

The mean-square value of $m_n(t)$ is

$$\langle m_n^2(t)\rangle = \tfrac{1}{2}(0.9166)^2 + \tfrac{1}{2}(0.4583)^2 = 0.5251 \tag{3.21}$$

Thus the efficiency is

$$E = \frac{(0.25)(0.5251)}{1 + (0.25)(0.5251)} \tag{3.22}$$

which yields

$$E = 11.60\% \tag{3.23}$$

Since the carrier power is 50 W, we have

$$\tfrac{1}{2}(A_c)^2 = 50 \tag{3.24}$$

from which

$$A_c = 10 \tag{3.25}$$

Also, since $\sin x = \cos(x - \pi/2)$, we can write $x_c(t)$ as

$$x_c(t) = 10\left\{1 + 0.5\left[0.9166 \cos\left(2\pi f_m t - \frac{\pi}{9}\right) + 0.4583 \cos\left(4\pi f_m t - \frac{\pi}{2}\right)\right]\right\} \cos(2\pi f_c t) \tag{3.26}$$

In order to plot the spectrum of $x_c(t)$, we write the preceding equation as

$$\begin{aligned} x_c(t) = {} & 10 \cos(2\pi f_c t) \\ & + 2.292\left\{\cos\left[2\pi(f_c + f_m)t - \frac{\pi}{9}\right] + \cos\left[2\pi(f_c - f_m)t + \frac{\pi}{9}\right]\right\} \\ & + 1.146\left\{\cos\left[2\pi(f_c + 2f_m)t - \frac{\pi}{2}\right] + \cos\left[2\pi(f_c - 2f_m)t + \frac{\pi}{2}\right]\right\} \end{aligned} \tag{3.27}$$

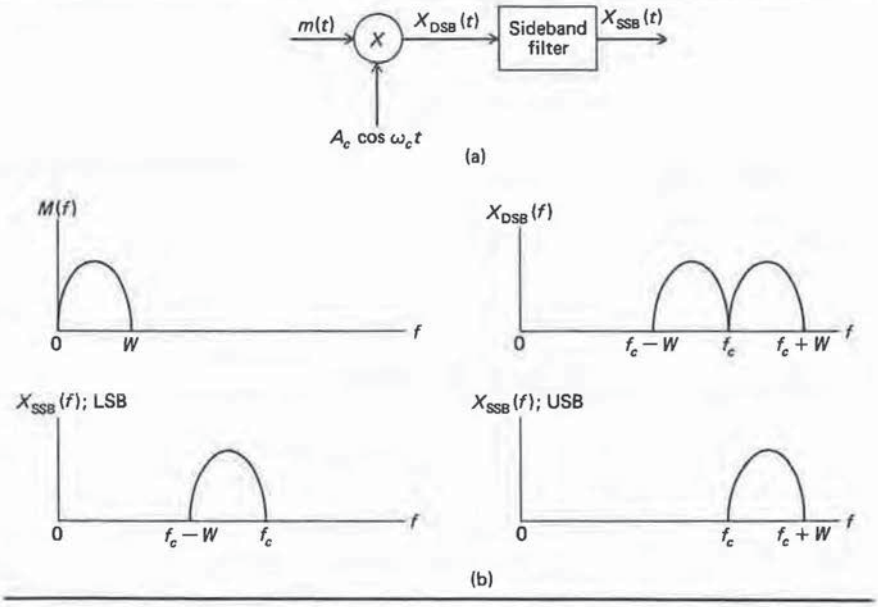
Parts (b) and (c) of Figure 3.5 show the amplitude and phase spectra of $x_c(t)$. Note that the amplitude spectrum has even symmetry about the carrier frequency and that the phase spectrum has odd symmetry about the carrier frequency for $f > 0$ or $f < 0$. Of course, the overall amplitude spectrum is even about $f = 0$, and the overall phase spectrum is odd about $f = 0$.

Single-Sideband Modulation

In our development of DSB, we saw that the upper sideband (USB) and lower sideband (LSB) have even amplitude and odd phase symmetry about the carrier frequency. Thus transmission of both sidebands is not necessary, since either sideband contains sufficient information to reconstruct the message signal $m(t)$. Elimination of one of the sidebands prior to transmission results in single sideband (SSB), which reduces the bandwidth of the modulator output from $2W$ to W, where W is the bandwidth of $m(t)$. However, this bandwidth savings is accompanied by a considerable increase in complexity.

On the following pages, two different methods are used to derive the time-domain expression for the signal at the output of an SSB modulator. Although the two methods are equivalent, they do present different view-

FIGURE 3.6 Generation of SSB by sideband filtering. (a) SSB modulator. (b) Spectra (single-sided).



points. In the first method, the transfer function of the filter used to generate an SSB signal from a DSB signal is derived using the Hilbert transform. The second method derives the SSB signal directly from $m(t)$ using the results illustrated in Figure 2.30 and the frequency-translation theorem.

The generation of an SSB signal by sideband filtering is illustrated in Figure 3.6. First, a DSB signal, $x_{\text{DSB}}(t)$, is formed. Sideband filtering of the DSB signal then yields an upper-sideband or a lower-sideband SSB signal, depending on the filter passband selected.

The filtering process that yields lower-sideband SSB is illustrated in detail in Figure 3.7. A lower-sideband SSB signal can be generated by passing a DSB signal through an ideal filter that passes the lower sideband and rejects the upper sideband. It follows from Figure 3.7(b) that the transfer function of this filter is

$$H_L(f) = \tfrac{1}{2}[\text{sgn}\,(f + f_c) - \text{sgn}\,(f - f_c)] \tag{3.28}$$

Since the Fourier transform of a DSB signal is

$$X_{\text{DSB}}(f) = \tfrac{1}{2}A_cM(f + f_c) + \tfrac{1}{2}A_cM(f - f_c) \tag{3.29}$$

the transform of the lower-sideband SSB signal is

$$\begin{aligned} X_c(f) = {} & \tfrac{1}{4}A_c[M(f + f_c)\,\text{sgn}\,(f + f_c) + M(f - f_c)\,\text{sgn}\,(f + f_c)] \\ & - \tfrac{1}{4}A_c[M(f + f_c)\,\text{sgn}\,(f - f_c) + M(f - f_c)\,\text{sgn}\,(f - f_c)] \end{aligned} \tag{3.30}$$

which is

$$\begin{aligned} X_c(f) = {} & \tfrac{1}{4}A_c[M(f + f_c) + M(f - f_c)] \\ & + \tfrac{1}{4}A_c[M(f + f_c)\,\text{sgn}\,(f + f_c) - M(f - f_c)\,\text{sgn}\,(f - f_c)] \end{aligned} \tag{3.31}$$

From our study of DSB, we know that

$$\tfrac{1}{2}A_cm(t)\cos\omega_ct \leftrightarrow \tfrac{1}{4}A_c[M(f + f_c) + M(f - f_c)] \tag{3.32}$$

and from our study of Hilbert transforms in Chapter 2, we recall that

$$\hat{m}(t) \leftrightarrow -j\,\text{sgn}\,(f)M(f)$$

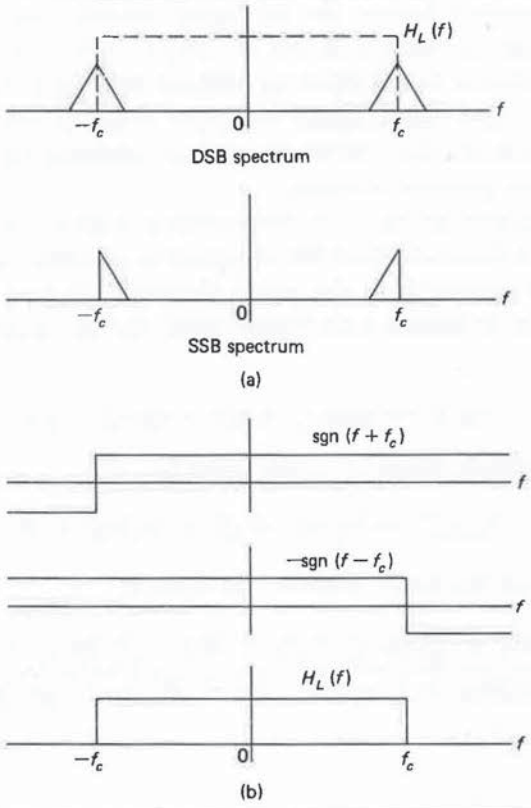
By the frequency-translation theorem, we have

$$m(t)e^{\pm j2\pi f_ct} \leftrightarrow M(f \mp f_c) \tag{3.33}$$

Replacing $m(t)$ by $\hat{m}(t)$ in the foregoing yields

$$\hat{m}(t)e^{\pm j2\pi f_ct} \leftrightarrow -jM(f \mp f_c)\,\text{sgn}\,(f \mp f_c) \tag{3.34}$$

FIGURE 3.7 Generation of lower-sideband SSB. (a) Sideband filtering process. (b) Generation of lower-sideband filter.



Thus

$$\mathcal{F}^{-1}\{\tfrac{1}{4}A_c[M(f+f_c)\ \text{sgn}\ (f+f_c) - M(f-f_c)\ \text{sgn}\ (f-f_c)]\}$$

$$= -A_c\frac{1}{4j}\hat{m}(t)e^{-j2\pi f_c t} + A_c\frac{1}{4j}\hat{m}(t)e^{+j2\pi f_c t}$$

$$= \tfrac{1}{2}A_c\hat{m}(t)\sin\omega_c t \tag{3.35}$$

Combining (3.32) and (3.35), we get the general form of a lower-sideband SSB signal:

$$x_c(t) = \tfrac{1}{2}A_c m(t)\cos\omega_c t + \tfrac{1}{2}A_c\hat{m}(t)\sin\omega_c t \tag{3.36}$$

A similar development can be carried out for upper-sideband SSB. The

result is

$$x_c(t) = \tfrac{1}{2}A_c m(t) \cos \omega_c t - \tfrac{1}{2}A_c \hat{m}(t) \sin \omega_c t \tag{3.37}$$

which shows that LSB and USB modulators have the same defining equations except for the sign of the term representing the Hilbert transform of the modulation. Observation of the spectrum of an SSB signal illustrates that SSB systems do not have dc response.

The generation of SSB by the method of sideband filtering the output of DSB modulators requires the use of filters that are very nearly ideal if low-frequency information is contained in $m(t)$. These filters are typically implemented using crystal filter technology.

Another method for generating an SSB signal, known as *phase-shift modulation,* is illustrated in Figure 3.8. This system is a term-by-term realization of (3.36) or (3.37). Like the ideal filters required for sideband filtering, the ideal wideband phase shifter, which performs the Hilbert transforming operation, is impossible to implement exactly. However, since the frequency at which the discontinuity occurs is $f = 0$ instead of $f = f_c$, ideal phase shifters can be closely approximated.

An alternative derivation of $x_c(t)$ for an SSB signal is based on the discussion surrounding Figure 2.30. As shown in Figure 3.9, the positive-frequency portion of $M(f)$ is given by

$$M_p(f) = \tfrac{1}{2}\mathcal{F}\{m(t) + j\hat{m}(t)\} \tag{3.38}$$

and the negative-frequency portion of $M(f)$ is given by

$$M_n(f) = \tfrac{1}{2}\mathcal{F}\{m(t) - j\hat{m}(t)\} \tag{3.39}$$

FIGURE 3.8 Phase-shift modulator

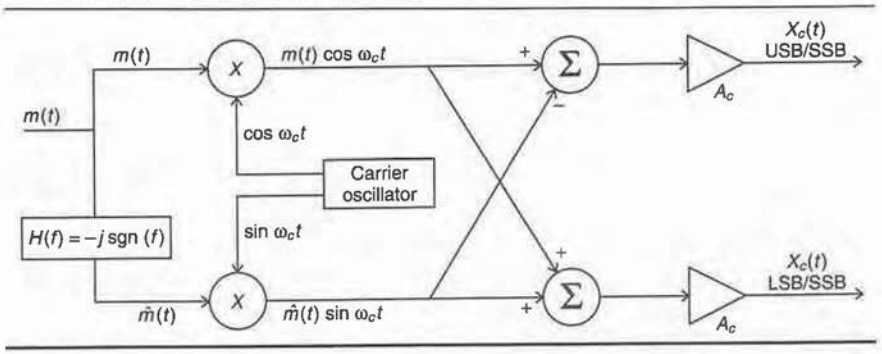
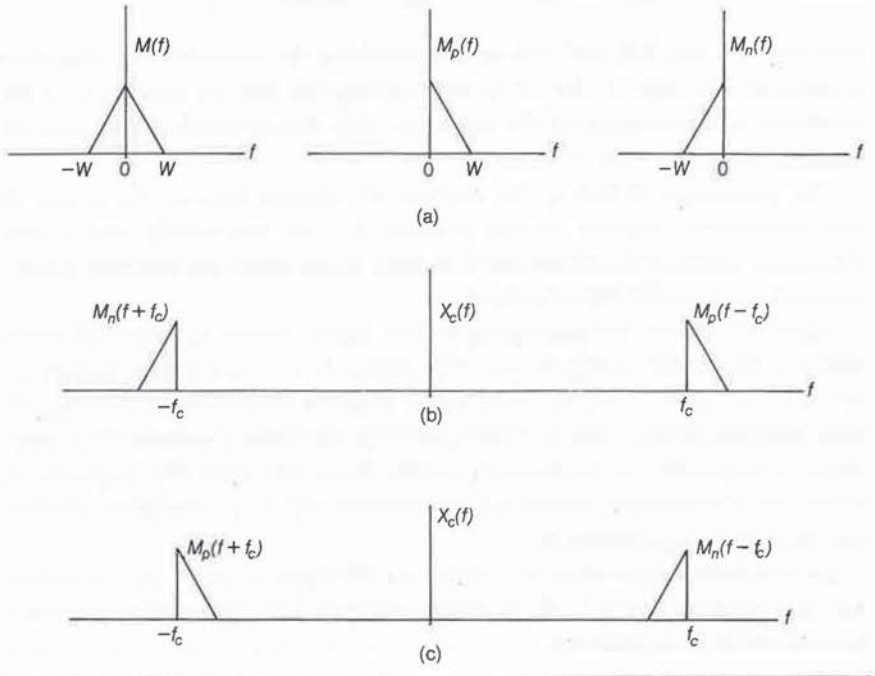


FIGURE 3.9 Alternative derivation of SSB signals. (a) $M(f)$, $M_p(f)$, and $M_n(f)$. (b) Upper-sideband SSB signal. (c) Lower-sideband SSB signal.



By definition, an upper-sideband SSB signal is given in the frequency domain by

$$X_c(f) = \tfrac{1}{2}A_c M_p(f - f_c) + \tfrac{1}{2}A_c M_n(f + f_c) \tag{3.40}$$

Inverse Fourier-transforming yields

$$x_c(t) = \tfrac{1}{4}A_c[m(t) + j\hat{m}(t)]e^{j2\pi f_c t} + \tfrac{1}{4}A_c[m(t) - j\hat{m}(t)]e^{-j2\pi f_c t} \tag{3.41}$$

which is

$$x_c(t) = \tfrac{1}{4}A_c m(t)[e^{j2\pi f_c t} + e^{-j2\pi f_c t}] + j\tfrac{1}{4}A_c \hat{m}(t)[e^{j2\pi f_c t} - e^{-j2\pi f_c t}] \tag{3.42}$$

The preceding expression is clearly equivalent to (3.37).

The lower-sideband SSB signal is derived in a similar manner. By definition, for a lower-sideband SSB signal,

$$X_c(f) = \tfrac{1}{2}A_c M_p(f + f_c) + \tfrac{1}{2}A_c M_n(f - f_c) \tag{3.43}$$

This becomes, after inverse Fourier-transforming,

$$x_c(t) = \tfrac{1}{4}A_c[m(t) + j\hat{m}(t)]e^{-j2\pi f_c t} + \tfrac{1}{4}A_c[m(t) - j\hat{m}(t)]e^{j2\pi f_c t} \quad (3.44)$$

which can be written as

$$x_c(t) = \tfrac{1}{4}A_c m(t)[e^{j2\pi f_c t} + e^{-j2\pi f_c t}] - j\tfrac{1}{4}A_c\hat{m}(t)[e^{j2\pi f_c t} - e^{-j2\pi f_c t}] \quad (3.45)$$

This expression is clearly equivalent to (3.36). Figures 3.9(b) and 3.9(c) show the four signals used in this development: $M_p(f + f_c)$, $M_p(f - f_c)$, $M_n(f + f_c)$, and $M_n(f - f_c)$.

There are several methods that can be employed to demodulate SSB. The simplest technique is to multiply $x_c(t)$ by a demodulation carrier and lowpass filter the result, as illustrated in Figure 3.1(a). We assume a demodulation carrier having a phase error $\theta(t)$ that yields

$$d(t) = [\tfrac{1}{2}A_c m(t)\cos\omega_c t \pm \tfrac{1}{2}A_c\hat{m}(t)\sin\omega_c t]\{4\cos[\omega_c t + \theta(t)]\} \quad (3.46)$$

where the factor of 4 is chosen for mathematical convenience. The preceding expression can be written as

$$\begin{aligned} d(t) = {} & A_c m(t)\cos\theta(t) + A_c m(t)\cos[2\omega_c t + \theta(t)] \\ & \mp A_c\hat{m}(t)\sin\theta(t) \pm A_c\hat{m}(t)\sin[2\omega_c t + \theta(t)] \end{aligned} \quad (3.47)$$

Lowpass filtering and amplitude scaling yield

$$y_D(t) = m(t)\cos\theta(t) \mp \hat{m}(t)\sin\theta(t) \quad (3.48)$$

for the demodulated output. Observation of (3.48) illustrates that for $\theta(t)$ equal to zero, the demodulated output is the desired message signal. However, if $\theta(t)$ is nonzero, the output consists of the sum of two terms. The first term is a time-varying attenuation of the message signal and is the output present in a DSB system operating in a similar manner. The second term is a crosstalk term and can represent serious distortion if $\theta(t)$ is not small.

Another useful technique for demodulating an SSB signal is carrier reinsertion, which is illustrated in Figure 3.10. The output of a local oscillator is added to the received signal $x_r(t)$. This yields

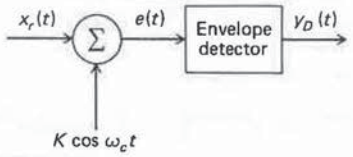
$$e(t) = [\tfrac{1}{2}A_c m(t) + K]\cos\omega_c t \pm \tfrac{1}{2}A_c\hat{m}(t)\sin\omega_c t \quad (3.49)$$

which is the input to the envelope detector.

The output of the envelope detector must next be computed. This is slightly more difficult for signals of the form of (3.49) than for signals of the form of (3.10) because both cosine and sine terms are present. In order to derive the desired result, consider the signal

$$x(t) = a(t)\cos\omega_c t - b(t)\sin\omega_c t \quad (3.50)$$

FIGURE 3.10 Demodulation using carrier reinsertion



which can be represented as illustrated in Figure 3.11. Figure 3.11 shows the amplitude of the direct component $a(t)$, the amplitude of the quadrature component $b(t)$, and the resultant $R(t)$. It follows from Figure 3.11 that

$$a(t) = R(t) \cos \theta(t) \qquad \text{and} \qquad b(t) = R(t) \sin \theta(t)$$

This yields

$$x(t) = R(t)[\cos \theta(t) \cos \omega_c t - \sin \theta(t) \sin \omega_c t] \tag{3.51}$$

which is

$$x(t) = R(t) \cos [\omega_c t + \theta(t)] \tag{3.52}$$

where

$$\theta(t) = \tan^{-1} \frac{b(t)}{a(t)} \tag{3.53}$$

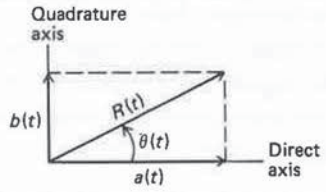
The instantaneous amplitude $R(t)$, which is the envelope of the signal, is given by

$$R(t) = \sqrt{a^2(t) + b^2(t)} \tag{3.54}$$

and will be the output of an envelope detector with $x(t)$ on the input if $a(t)$ and $b(t)$ are slowly varying with respect to $\cos \omega_c t$.

A comparison of (3.49) and (3.54) illustrates that the envelope of an SSB

FIGURE 3.11 Signal representation



signal, after carrier reinsertion, is given by

$$y_D(t) = \sqrt{[\tfrac{1}{2}A_c m(t) + K]^2 + [\tfrac{1}{2}A_c \hat{m}(t)]^2} \tag{3.55}$$

which is the demodulated output $y_D(t)$ in Figure 3.10. If K is chosen large enough such that

$$[\tfrac{1}{2}A_c m(t) + K]^2 \gg [\tfrac{1}{2}A_c \hat{m}(t)]^2$$

the output of the envelope detector becomes

$$y_D(t) \cong \tfrac{1}{2}A_c m(t) + K \tag{3.56}$$

from which the message signal can easily be extracted. The development shows that carrier reinsertion requires that the locally generated carrier must be phase-coherent with the original modulation carrier. This is easily accomplished in speech-transmission systems. The frequency and phase of the demodulation carrier can be manually adjusted until intelligibility is obtained.

▼ EXAMPLE 3.2 As we saw in the preceding analysis, the concept of single sideband is probably best understood by using frequency-domain analysis. However, the SSB time-domain waveforms are also interesting and are the subject of this example. Assume that the message signal is given by

$$m(t) = \cos(\omega_1 t) - 0.4\cos(2\omega_1 t) + 0.9\cos(3\omega_1 t) \tag{3.57}$$

The Hilbert transform of $m(t)$ is

$$\hat{m}(t) = \sin(\omega_1 t) - 0.4\sin(2\omega_1 t) + 0.9\sin(3\omega_1 t) \tag{3.58}$$

These two waveforms are shown in Figures 3.12(a) and (b). As we have seen, the SSB signal is given by

$$x_c(t) = \frac{A_c}{2}[m(t)\cos\omega_c t \pm \hat{m}(t)\sin\omega_c t] \tag{3.59}$$

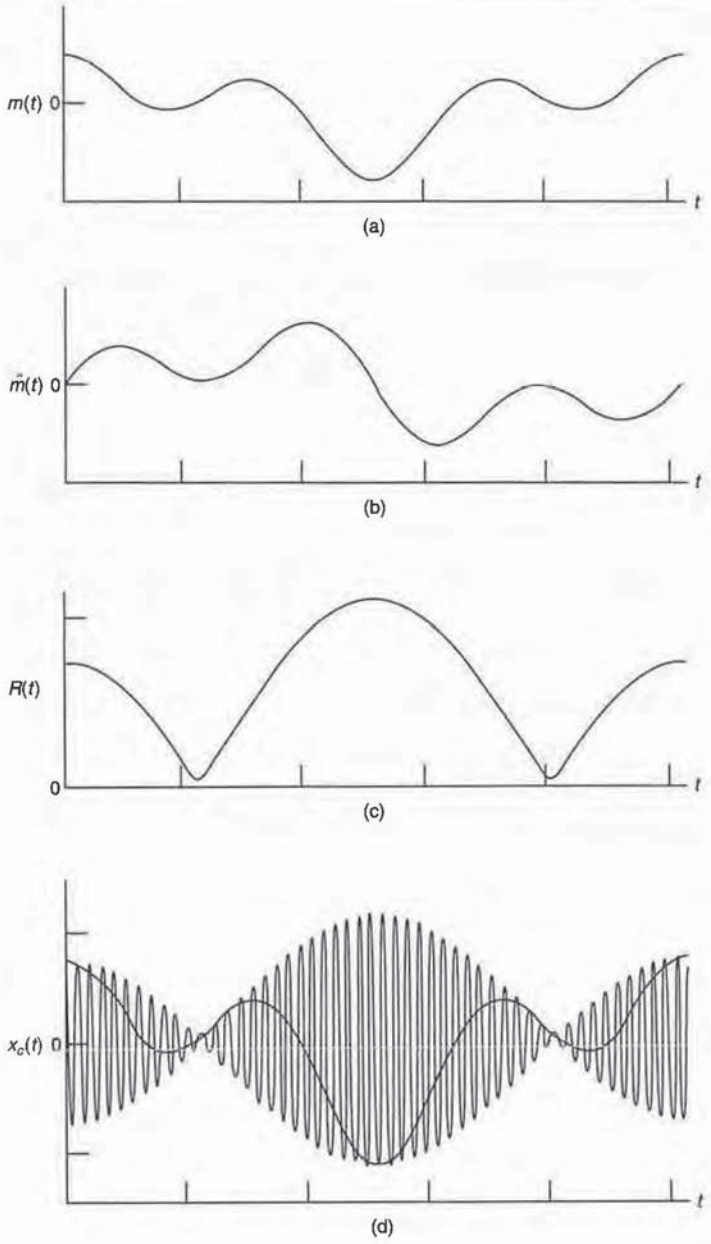
with the choice of sign depending upon the sideband to be used for transmission. Using (3.39), (3.40), and (3.41), we can place $x_c(t)$ in the standard form of (3.1). This gives

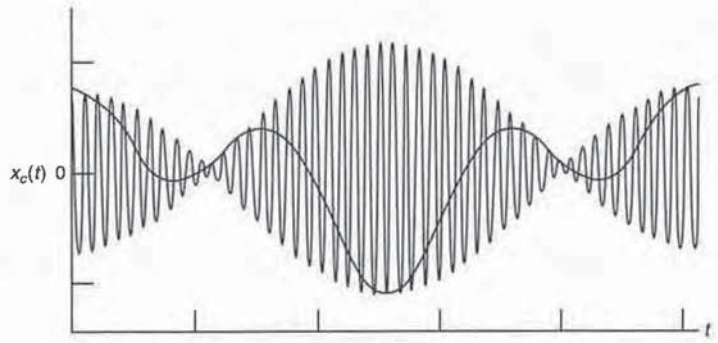
$$x_c(t) = R(t)\cos[\omega_c t + \theta(t)] \tag{3.60}$$

where the envelope $R(t)$ is

$$R(t) = \frac{A_c}{2}\sqrt{m^2(t) + \hat{m}^2(t)} \tag{3.61}$$

FIGURE 3.12 Time-domain signals for SSB system. (a) Message signal. (b) Hilbert transform of message signal. (c) Envelope of SSB signal. (d) Upper-sideband SSB signal with message signal. (e) Lower-sideband SSB signal with message signal.





(e)

and $\theta(t)$, which is the phase deviation of $x_c(t)$, is given by

$$\theta(t) = \pm \tan^{-1}\left[\frac{\hat{m}(t)}{m(t)}\right] \tag{3.62}$$

The instantaneous frequency of $x_c(t)$ is therefore

$$\frac{d}{dt}\{\omega_c t + \theta(t)\} = \omega_c \pm \frac{d}{dt}\left\{\tan^{-1}\left[\frac{\hat{m}(t)}{m(t)}\right]\right\} \tag{3.63}$$

From (3.61) we see that the envelope of the SSB signal is independent of the choice of the sideband. The instantaneous frequency, however, is a rather complicated function of the message signal and also depends upon the choice of sideband. We therefore see that the message signal $m(t)$ affects both the envelope and phase of the modulated carrier $x_c(t)$. In DSB and AM the message signal affected only the envelope of $x_c(t)$.

The envelope of the SSB signal is shown in Figure 3.12(c). The upper-sideband SSB signal is illustrated in Figure 3.12(d) and the lower-sideband SSB signal is shown in Figure 3.12(c). It is easily seen that both the upper-sideband and lower-sideband SSB signals have the envelope shown in Figure 3.12(c). The message signal $m(t)$ is also shown in parts (d) and (e) in Figure 3.12. ▲

Vestigial-Sideband Modulation

Vestigial-sideband (VSB) modulation overcomes two of the difficulties present in SSB modulation. By allowing a vestige of the unwanted sideband to appear at the output of an SSB modulator, we may simplify the design of the sideband filter, since the need for sharp cutoff at the carrier frequency is elim-

inated. In addition, a VSB system has improved low-frequency response and can even have dc response.

An example will illustrate the technique. For simplicity, let the message signal be the sum of two sinusoids:

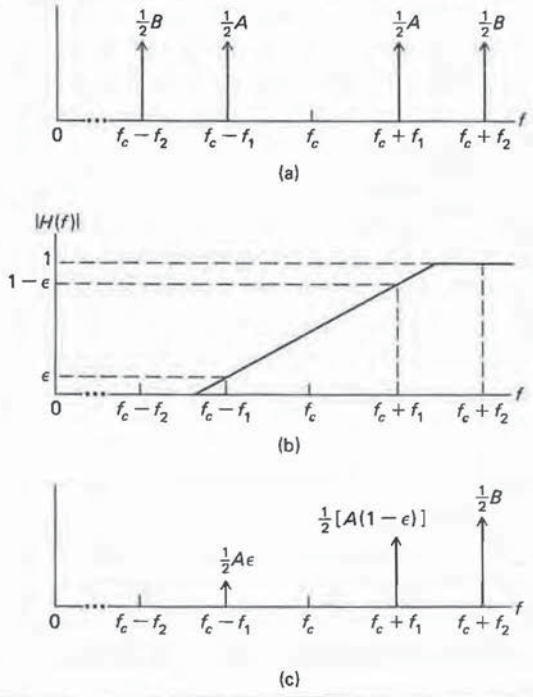
$$m(t) = A \cos \omega_1 t + B \cos \omega_2 t \tag{3.64}$$

This message signal is then multiplied by a carrier, $\cos \omega_c t$, to form the DSB signal

$$\begin{aligned} e_{\text{DSB}}(t) = {} & \tfrac{1}{2}A \cos (\omega_c + \omega_1)t + \tfrac{1}{2}A \cos (\omega_c - \omega_1)t \\ & + \tfrac{1}{2}B \cos (\omega_c + \omega_2)t + \tfrac{1}{2}B \cos (\omega_c - \omega_2)t \end{aligned} \tag{3.65}$$

Figure 3.13(a) shows the single-sided spectrum of this signal. A vestigial sideband filter is then used to generate the VSB signal. Figure 3.13(b) shows the assumed form of the VSB filter. The skirt of the VSB filter must be sym-

FIGURE 3.13 Generation of vestigial sideband. (a) DSB spectrum (single-sided). (b) VSB filter characteristic near f_c. (c) VSB spectrum.



metrical about the carrier as shown. Figure 3.13(c) shows the single-sided spectrum of the filter output. This is the spectrum of the VSB signal.

The spectrum shown in Figure 3.13(c) corresponds to the VSB signal

$$x_c(t) = \tfrac{1}{2}A\epsilon \cos(\omega_c - \omega_1)t + \tfrac{1}{2}A(1-\epsilon)\cos(\omega_c + \omega_1)t + \tfrac{1}{2}B\cos(\omega_c + \omega_2)t \quad (3.66)$$

This signal can be demodulated by multiplying by $4\cos \omega_c t$ and lowpass filtering. The result is

$$e(t) = A\epsilon \cos \omega_1 t + A(1-\epsilon)\cos \omega_1 t + B\cos \omega_2 t \quad (3.67)$$

or

$$e(t) = A\cos \omega_1 t + B\cos \omega_2 t \quad (3.68)$$

which is the assumed message signal.

The slight increase in bandwidth required for VSB over that required for SSB is often more than offset by the resulting electronic simplifications. As a matter of fact, if a carrier component is added to a VSB signal, envelope detection can be used. The development of this technique is similar to the development of envelope detection of SSB with carrier reinsertion and is relegated to the problems. The process, however, is demonstrated in the following example.

▼ **EXAMPLE 3.3** In this example we consider the time domain waveforms corresponding to VSB modulation and consider demodulation using envelope detection or carrier reinsertation.

We assume the same message signal as was assumed in the previous example. In other words

$$m(t) = \cos \omega_1 t - 0.4\cos 2\omega_1 t + 0.9\cos 3\omega_1 t \quad (3.69)$$

The message signal $m(t)$ is shown in Figure 3.14(a).

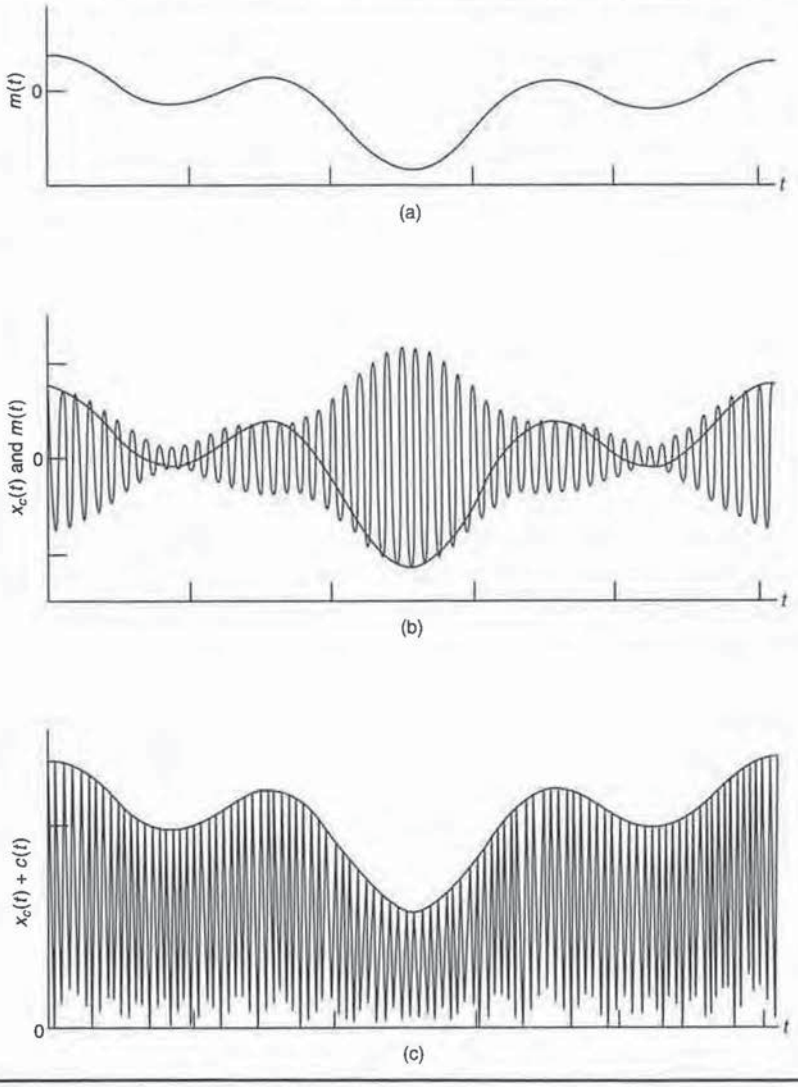
The VSB signal can be expressed as

$$\begin{aligned} x_c(t) = A_c[&\epsilon_1 \cos(\omega_c + \omega_1)t + (1-\epsilon_1)\cos(\omega_c - \omega_1)t \\ &-0.4\epsilon_2 \cos(\omega_c + 2\omega_1)t - 0.4(1-\epsilon_2)\cos(\omega_c - 2\omega_1)t \\ &+0.9\epsilon_3 \cos(\omega_c + 3\omega_1)t + 0.9(1-\epsilon_3)\cos(\omega_c - 3\omega_1)] \end{aligned} \quad (3.70)$$

The modulated carrier, along with the message signal, is shown in Figure 3.14(b) for $\epsilon_1 = 0.64$, $\epsilon_2 = 0.78$, and $\epsilon_3 = 0.92$.

The result of carrier reinsertion and envelope detection is shown in Figure 3.14(c). The message signal, biased by the amplitude of the carrier component, is clearly shown and will be the output of an envelope detector. ▲

FIGURE 3.14 Time-domain signals for VSB system. (a) Message signal. (b) VSB signal and message signal. (c) Sum of VSB signal and carrier signal.



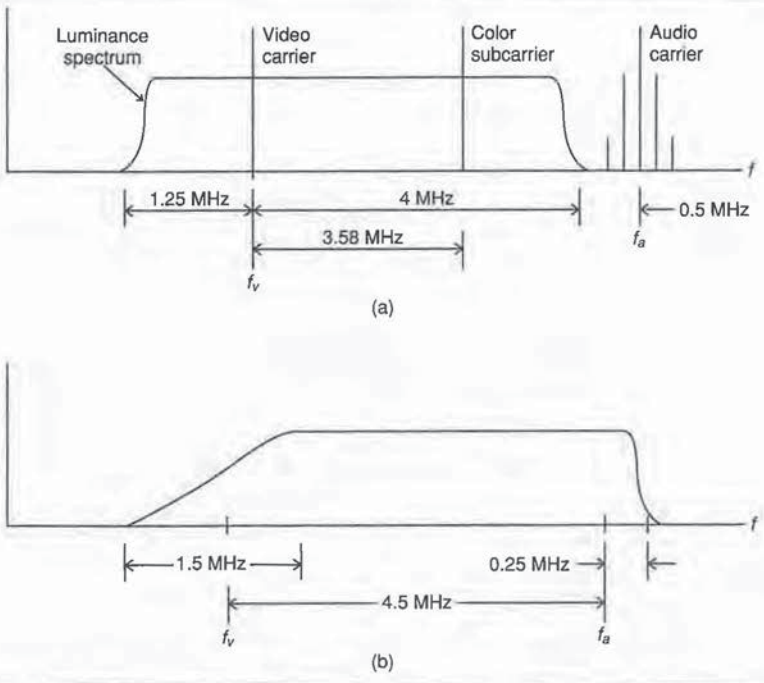
VSB is currently used in the United States for transmission of the video signal in commercial television broadcasting. However, exact shaping of the vestigial sideband is not carried out at the transmitter, but at the receiver, where signal levels are low. The filter in the transmitter simply bandlimits the video signal, as shown in Figure 3.15(a). The video carrier frequency is

denoted f_v, and, as can be seen, the bandwidth of the video signal is approximately 5.25 MHz. The spectrum of the audio signal is centered about the audio carrier, which is 4.5 MHz above the video carrier. The modulation method used for audio transmission is FM. When we study FM in the following sections, you will understand the shape of the audio spectrum. Since the spectrum centered on the audio carrier is a line spectrum in Figure 3.15(a), a periodic audio signal is implied. This was done for clarity, and in practice the audio signal will have a continuous spectrum. Figure 3.15(b) shows the amplitude response of the receiver VSB filter.

Also shown in Figure 3.15(a), at a frequency 3.58 MHz above the video carrier, is the color carrier. Quadrature multiplexing, which we shall study in Section 3.6, is used with the color subcarrier so that two different signals are transmitted with the color carrier. These two signals, commonly referred to as the I-channel and Q-channel signals, carry luminance and chrominance (color) information necessary to reconstruct the image at the receiver.

One problem faced by the designers of a system for the transmission of a color TV signal was that the transmitted signal was required to be compatible

FIGURE 3.15 Transmitted spectrum and VSB filtering for television. (a) Spectrum of transmitted signal. (b) VSB filter in receiver.



with existing black-and-white television receivers. Such a consideration is a significant constraint in the design process. A similar problem was faced by the designers of stereophonic radio receivers. We shall study this system in Section 3.6 and, since the audio system is simpler than a color TV system, we can see how the compatibility problem was solved.

Switching Modulators

In all our previous developments, the carrier has been assumed to be sinusoidal. However, if the modulator is followed by a bandpass filter, any periodic waveform can serve as a carrier. For example, if the carrier is periodic with fundamental frequency, f_c, it can be represented by the Fourier series

$$c(t) = \sum_{n=-\infty}^{\infty} C_n e^{jn2\pi f_c t} \tag{3.71}$$

Multiplying this carrier by the message signal $m(t)$ yields

$$m(t)c(t) = \sum_{n=-\infty}^{\infty} C_n m(t) e^{jn2\pi f_c t} \tag{3.72}$$

From the frequency-translation theorem, the Fourier transform of $m(t)c(t)$ is

$$\mathcal{F}[m(t)c(t)] = \sum_{n=-\infty}^{\infty} C_n M(f - nf_c) \tag{3.73}$$

which shows that the effect of multiplying $m(t)$ by a periodic waveform is to translate the spectrum of $m(t)$ to a carrier frequency and all harmonics of the carrier frequency. A DSB signal can be obtained from (3.73) using a bandpass filter with center frequency f_c and bandwidth $2W$, where W is the highest frequency present in the message signal.

Considerable simplification of the product device results if a squarewave carrier, which varies between zero and unity, is used. When the carrier is unity, the multiplier output is $m(t)$, and when the carrier is zero, the multiplier output is zero. Thus the multiplier can be replaced by a switch that opens and closes at the carrier frequency. Switching modulators are much easier to implement than are analog product devices, especially at higher frequencies. A switching modulator is illustrated in Figure 3.16(c), together with appropriate spectra.

Switching modulators are often implemented using diode bridges. Two examples are the shunt and series modulators illustrated in Figure 3.17. The carrier waveform, having frequency f_c, alternately short-circuits and open-

FIGURE 3.16 Modulation using periodic carriers. (a) Modulation with a periodic carrier. (b) Spectrum of multiplier output. (c) Switching modulator.

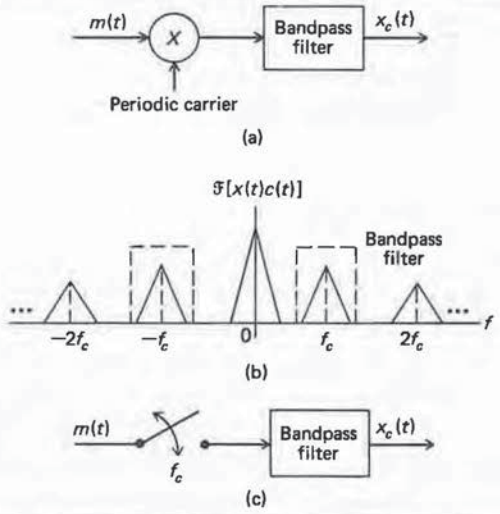
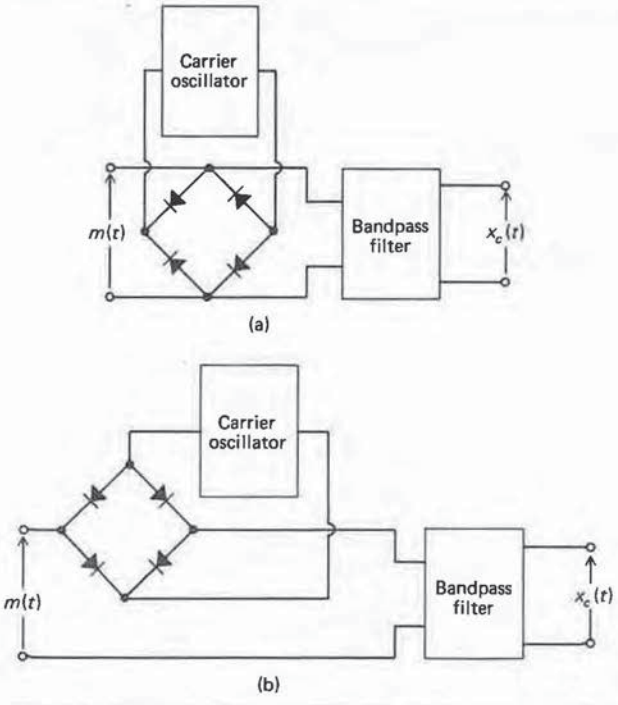


FIGURE 3.17 (a) Shunt modulator. (b) Series modulator.



circuits the diodes so that they operate as a switch. If DSB is to be generated, care must be exercised to ensure that the diodes are matched, thereby balancing the carrier component. Also, the carrier amplitude must be large, so that diode switching occurs properly.

The switching modulator is an example of the use of nonlinear elements to implement linear modulation. Additional examples are presented in the problems.

Frequency Translation and Mixing

We have seen that multiplying a lowpass signal by a high-frequency periodic signal translates the spectrum of the lowpass signal to all frequencies present in the periodic signal. Quite often it is desirable to translate a bandpass signal to a new center frequency. This process also can be accomplished by multiplication of the bandpass signal by a periodic signal and is called *mixing* or *converting*. A block diagram of a mixer is given in Figure 3.18. As an example, the bandpass signal $m(t) \cos \omega_1 t$ can be translated to a new carrier frequency ω_2 by multiplying it by a local oscillator signal of the form $2 \cos (\omega_1 \pm \omega_2)t$. By using appropriate trigonometric identities, we can easily show that the result of the multiplication is

$$e(t) = m(t) \cos \omega_2 t + m(t) \cos (2\omega_1 \pm \omega_2)t \tag{3.74}$$

The undesired term is removed by filtering. The filter should have a bandwidth at least $2W$ for the assumed DSB modulation, where W is the bandwidth of $m(t)$.

A common problem with mixers is that inputs of the form $k(t) \cos (\omega_1 \pm 2\omega_2)t$ are also translated to ω_2, since

$$\begin{aligned} &2k(t) \cos [(\omega_1 \pm 2\omega_2)t] \cos [(\omega_1 \pm \omega_2)t] \\ &\quad = k(t) \cos \omega_2 t + k(t) \cos (2\omega_1 \pm 3\omega_2)t \end{aligned} \tag{3.75}$$

FIGURE 3.18 Mixer

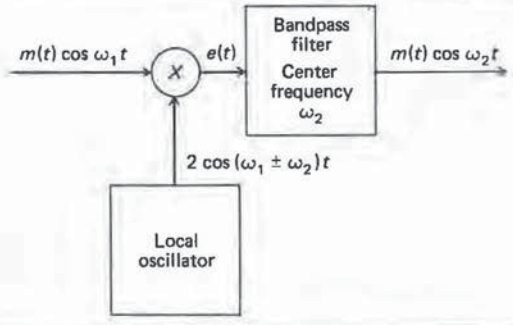
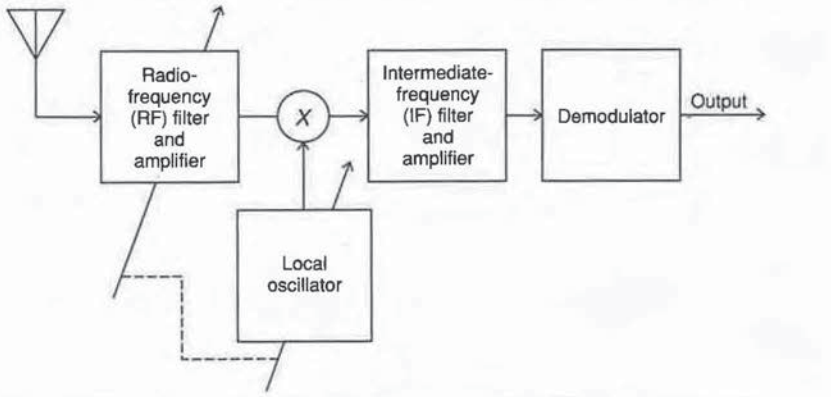


FIGURE 3.19 Superheterodyne receiver

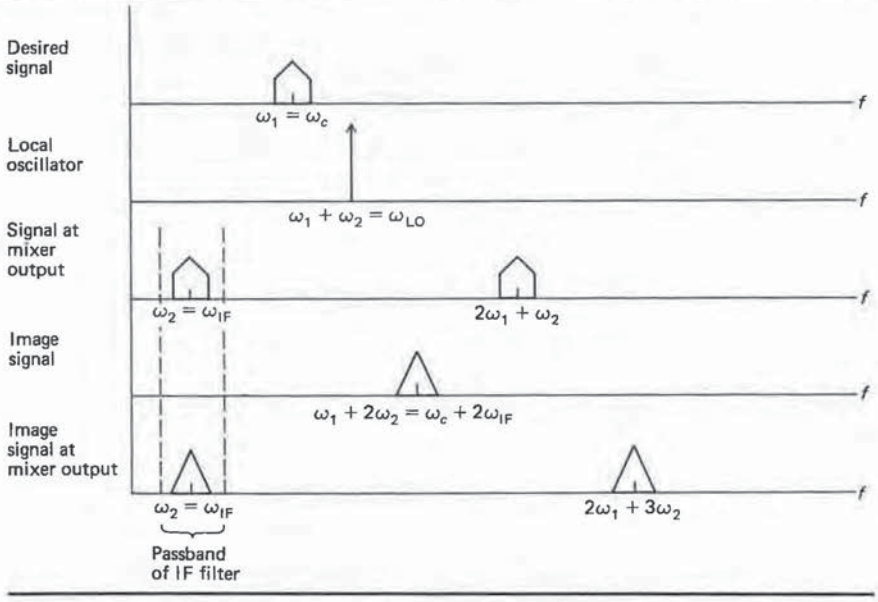


In (3.75), all three signs must be plus or all three must be minus. The input frequency $\omega_1 \pm 2\omega_2$, which results in an output at ω_2, is referred to as an *image frequency* of the desired frequency ω_1.

To realize that image frequencies are a problem, consider the superheterodyne receiver shown in Figure 3.19. The carrier frequency of the signal to be demodulated is ω_c, and the intermediate-frequency (IF) filter is a bandpass filter with center frequency ω_{IF}. The superheterodyne receiver has good sensitivity (the ability to detect weak signals) and selectivity (the ability to separate closely spaced signals). This results because the IF filter, which provides most of the predetection filtering, need not be tunable. Thus it can be a rather complex filter. Tuning of the receiver is accomplished by varying the frequency of the local oscillator. The superheterodyne receiver of Figure 3.19 is the mixer of Figure 3.18 with $\omega_c = \omega_1$ and $\omega_{IF} = \omega_2$. The mixer translates the input frequency ω_c to the IF frequency ω_{IF}. As shown previously, the image frequency $\omega_c \pm 2\omega_{IF}$, where the sign depends on the choice of local oscillator frequency, also will appear at the IF output. This means that if we are attempting to receive a signal having carrier frequency ω_c, we can also receive a signal at $\omega_c + 2\omega_{IF}$ if the local oscillator frequency is $\omega_c + \omega_{IF}$ or a signal at $\omega_c - 2\omega_{IF}$ if the local oscillator frequency is $\omega_c - \omega_{IF}$. There is only one image frequency, and it is always separated from the desired frequency by $2\omega_{IF}$. Figure 3.20 shows the desired signal and image signal for a local oscillator having the frequency

$$\omega_{LO} = \omega_c + \omega_{IF} \tag{3.76}$$

FIGURE 3.20 Illustration of image frequency (high-side tuning)



The image frequency can be eliminated by the radio-frequency (RF) filter. A standard IF frequency for AM radio is 455 kHz. Thus the image frequency is separated from the desired signal by almost 1 MHz. This shows that the RF filter need not be narrowband. Furthermore, since the AM broadcast band occupies the frequency range 540 kHz to 1.6 MHz, it is apparent that a tunable RF filter is not required, provided that stations at the high end of the band are not located geographically near stations at the low end of the band. Some inexpensive receivers take advantage of this fact. Additionally, if the RF filter is made tunable, it need be tunable only over a narrow range of frequencies.

One decision to be made when designing a superheterodyne receiver is whether the frequency of the local oscillator is to be below the frequency of the input carrier *(low-side tuning)* or above the frequency of the input carrier *(high-side tuning)*. A simple example based on the standard AM broadcast band illustrates one major consideration. The standard AM broadcast band extends from 540 kHz to 1600 kHz. For this example, let us choose a common intermediate frequency, 455 kHz. As shown in Table 3.1, for low-side tuning, the frequency of the local oscillator must be variable from 85 to 1600 kHz, which represents a frequency range in excess of 13 to 1. If high-side tuning is used, the frequency of the local oscillator must be variable from 995 to 2055 kHz, which represents a frequency range slightly in excess of 2 to 1. Oscillators whose frequency must vary over a large ratio

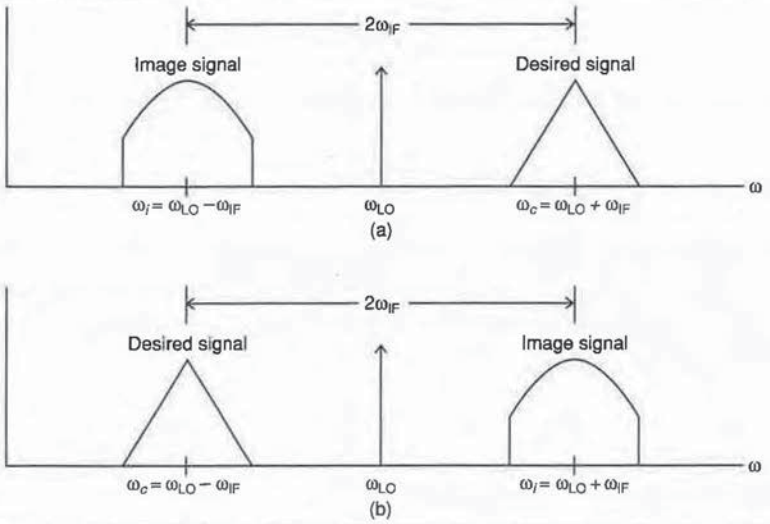
TABLE 3.1 Comparison of Low-Side and High-Side Tuning for AM Broadcast Band with f_{IF} = 455 kHz

	Lower Frequency	Upper Frequency	Tuning Range of Local Oscillator
Standard AM broadcast band	540 kHz	1600 kHz	—
Frequencies of local oscillator for low-side tuning	540 kHz − 455 kHz = 85 kHz	1600 kHz − 455 kHz = 1145 kHz	13.47 to 1
Frequencies of local oscillator for high-side tuning	540 kHz + 455 kHz = 995 kHz	1600 kHz + 455 kHz = 2055 kHz	2.07 to 1

are much more difficult to implement than are those whose frequency varies over a small ratio.

The relationship between the desired signal to be demodulated and the image signal is summarized in Figure 3.21 for low-side and high-side tuning. The desired signal to be demodulated has a carrier frequency of ω_c and the image signal has a carrier frequency of ω_i.

FIGURE 3.21 Relationship between ω_c and ω_i for (a) low-side tuning and (b) high-side tuning



3.2 ANGLE MODULATION

To generate angle modulation, the amplitude of the modulated carrier is held constant and either the phase or the time derivative of the phase of the carrier is varied linearly with the message signal $m(t)$. Thus the general angle-modulated signal is given by

$$x_c(t) = A_c \cos [\omega_c t + \phi(t)] \tag{3.77}$$

The instantaneous phase of $x_c(t)$ is defined as

$$\theta_i(t) = \omega_c t + \phi(t) \tag{3.78}$$

and the instantaneous frequency is defined as

$$\omega_i(t) = \frac{d\theta_i}{dt} = \omega_c + \frac{d\phi}{dt} \tag{3.79}$$

The functions $\phi(t)$ and $d\phi/dt$ are known as the *phase deviation* and *frequency deviation*, respectively.

The two basic types of angle modulation are phase modulation (PM) and frequency modulation (FM). *Phase modulation* implies that the phase deviation of the carrier is proportional to the message signal. Thus, for phase modulation,

$$\phi(t) = k_p m(t) \tag{3.80}$$

where k_p is the *deviation constant* in radians per unit of $m(t)$. Similarly, *frequency modulation* implies that the frequency deviation of the carrier is proportional to the modulating signal. This yields

$$\frac{d\phi}{dt} = k_f m(t) \tag{3.81}$$

The phase deviation of a frequency-modulated carrier is given by

$$\phi(t) = k_f \int_{t_0}^{t} m(\alpha)\, d\alpha + \phi_0 \tag{3.82}$$

in which ϕ_0 is the phase deviation at $t = t_0$. It follows from (3.81) that k_f is the frequency-deviation constant, expressed in radians per second per unit of $m(t)$. Since it is often more convenient to measure frequency deviation in hertz, we define

$$k_f = 2\pi f_d \tag{3.83}$$

where f_d is known as the *frequency-deviation constant* of the modulator and is expressed in hertz per unit of $m(t)$.

With these definitions, the phase modulator output is

$$x_c(t) = A_c \cos [\omega_c t + k_p m(t)] \tag{3.84}$$

and the frequency modulator output is

$$x_c(t) = A_c \cos \left[\omega_c t + 2\pi f_d \int^t m(\alpha)\, d\alpha \right] \tag{3.85}$$

The lower limit of the integral is typically not specified, since to do so would require the inclusion of an initial condition as shown in (3.82).

Figures 3.22 and 3.23 illustrate the outputs of PM and FM modulators for a step function and a sinusoidal message signal, respectively. With a unit-step message signal, the instantaneous frequency of the PM modulator output is f_c for both $t < t_0$ and $t > t_0$. The phase of the unmodulated carrier is advanced by $k_p = \pi/2$ radians for $t > t_0$, giving rise to a signal that is discontinuous at $t = t_0$. The frequency of the output of the FM modulator is f_c for $t < t_0$ and $f_c + f_d$ for $t > t_0$. The modulator output phase is, however, continuous at $t = t_0$.

With a sinusoidal message signal, the phase deviation of the PM modulator output is proportional to $m(t)$. The frequency deviation is proportional to the derivative of the phase deviation. Thus the instantaneous frequency of the output of the PM modulator is maximum when the *slope* of $m(t)$ is maximum and minimum when the *slope* of $m(t)$ is minimum. The frequency deviation of the FM modulator output is proportional to $m(t)$. Thus the instantaneous frequency of the FM modulator ouput is maximum when $m(t)$ is maximum and minimum when $m(t)$ is minimum. It should be noted that if $m(t)$ were not shown along with the modulator outputs, it would not be possible to distinguish the PM and FM modulator outputs. In the following sections we will devote considerable attention to the case in which $m(t)$ is sinusoidal.

Narrowband Angle Modulation

An angle-modulated carrier can be represented in exponential form by writing (3.77) as

$$x_c(t) = \text{Re}\,(A_c e^{j\omega_c t} e^{j\phi(t)}) \tag{3.86}$$

where Re $(\cdot)$ implies that the real part of the argument is to be taken. Expanding $e^{j\phi(t)}$ in a power series yields

$$x_c(t) = \text{Re}\left\{ A_c e^{j\omega_c t} \left[1 + j\phi(t) - \frac{\phi^2(t)}{2!} - \cdots \right] \right\} \tag{3.87}$$

If the maximum value of $|\phi(t)|$ is much less than unity, the modulated carrier can be approximated as

$$x_c(t) \cong \text{Re}\,[A_c e^{j\omega_c t} + A_c \phi(t) j e^{j\omega_c t}]$$

Taking the real part yields

$$x_c(t) \cong A_c \cos \omega_c t - A_c \phi(t) \sin \omega_c t \tag{3.88}$$

The form of (3.88) is reminiscent of AM. The modulator output contains a carrier component and a term in which a function of $m(t)$ multiplies a 90°

FIGURE 3.22 Comparison of PM and FM modulator outputs for a unit-step input. (a) Message signal. (b) Unmodulated carrier. (c) Phase modulator ouput ($k_p = \frac{1}{2\pi}$). (d) Frequency modulator output.

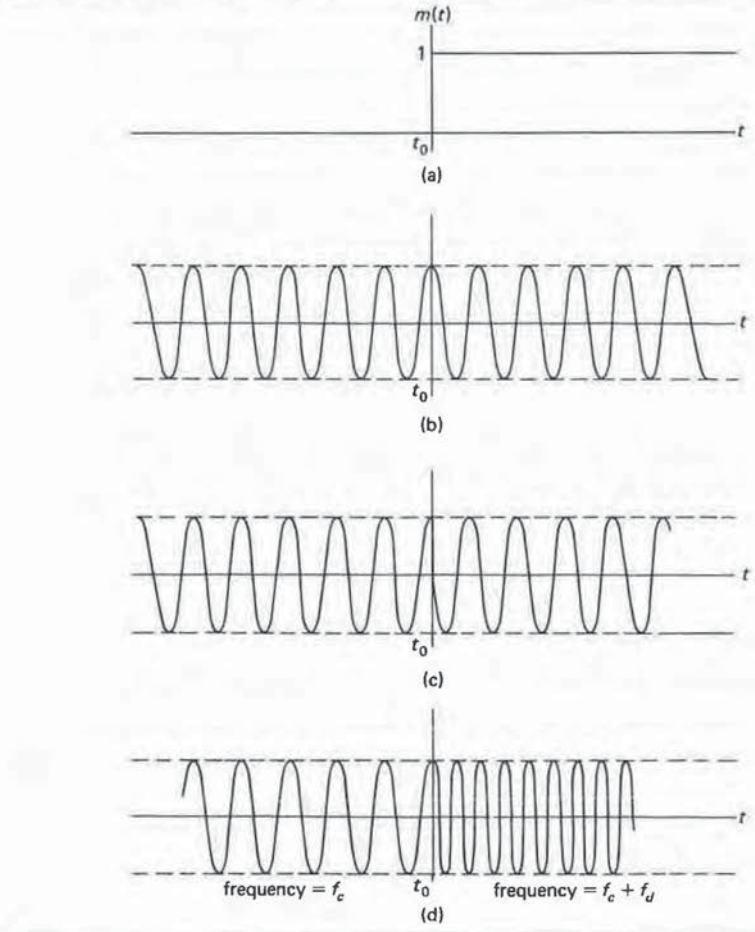
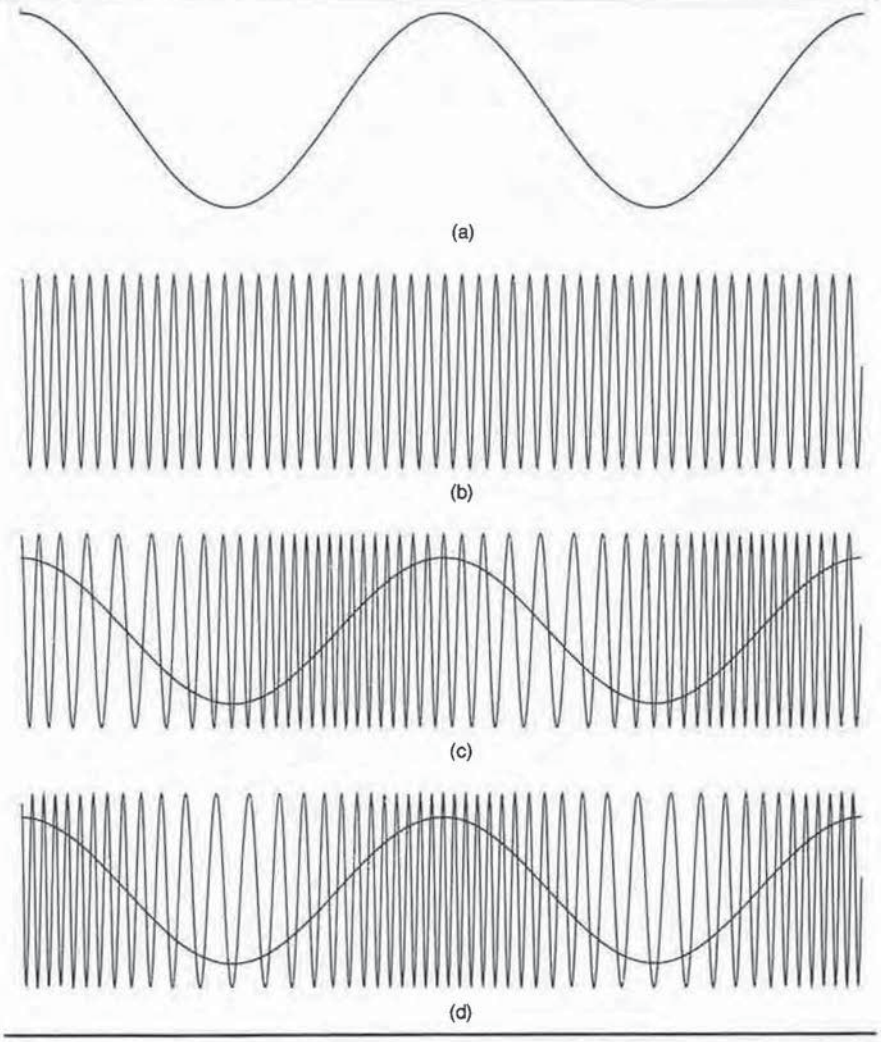
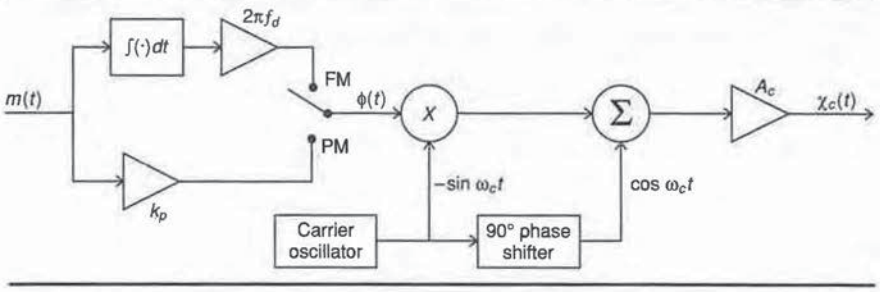


FIGURE 3.23 Angle modulation with sinusoidal message signal. (a) Message signal $m(t)$. (b) Unmodulated carrier $A_c \cos 2\pi f_c t$. (c) Output of phase modulator with $m(t)$. (d) Output of frequency modulator with $m(t)$.



phase-shifted carrier. This multiplication generates a pair of sidebands. Thus, if $\phi(t)$ has a bandwidth W, the bandwidth of a narrowband angle modulator output is $2W$. It is important to note, however, that the carrier and the *resultant* of the sidebands for narrowband angle modulation with sinusoidal modulation are in phase quadrature, whereas for AM they are not. This will be illustrated in Example 3.4.

FIGURE 3.24 Generation of narrowband angle modulation



The generation of narrowband angle modulation is easily accomplished using the method shown in Figure 3.24. The switch allows for the generation of either narrowband FM or narrowband PM. We will show later that narrowband angle modulation is useful for the generation of angle-modulated signals that are not necessarily narrowband through a process called narrowband to wideband conversion.

▼ **EXAMPLE 3.4** Consider an FM system operating with

$$m(t) = A \cos \omega_m t \tag{3.89}$$

From (3.82), with t_0 equal to zero,

$$\phi(t) = k_f \int_0^t A \cos \omega_m \alpha \, d\alpha = \frac{Ak_f}{\omega_m} \sin \omega_m t = \frac{Af_d}{f_m} \sin \omega_m t \tag{3.90}$$

so that

$$x_c(t) = A_c \cos \left(\omega_c t + \frac{Af_d}{f_m} \sin \omega_m t \right) \tag{3.91}$$

If $f_d/f_m \ll 1$, the modulator output can be approximated as

$$x_c(t) = A_c \left(\cos \omega_c t - \frac{Af_d}{f_m} \sin \omega_c t \sin \omega_m t \right) \tag{3.92}$$

which is

$$x_c(t) = A_c \cos \omega_c t + \frac{A_c}{2} \frac{Af_d}{f_m} [\cos (\omega_c + \omega_m)t - \cos (\omega_c - \omega_m)t] \tag{3.93}$$

Thus $x_c(t)$ can be written as

$$x_c(t) = A_c \operatorname{Re}\left\{e^{j2\pi f_c t}\left[1 + \frac{Af_d}{2f_m}\left(e^{j2\pi f_m t} - e^{-j2\pi f_m t}\right)\right]\right\} \tag{3.94}$$

It is interesting to compare this result with the equivalent result for an AM signal. Since sinusoidal modulation is assumed, the AM signal can be written as

$$x_c(t) = A_c(1 + a\cos 2\pi f_m t)\cos 2\pi f_c t \tag{3.95}$$

where a is the modulation index. Combining the two cosine terms yields

$$x_c(t) = A_c \cos 2\pi f_c t + \frac{A_c}{2} a[\cos 2\pi(f_c + f_m)t + \cos 2\pi(f_c - f_m)t] \tag{3.96}$$

This can be written in exponential form as

$$x_c(t) = A_c \operatorname{Re}\left\{e^{j2\pi f_c t}\left[1 + \frac{a}{2}a(e^{j2\pi f_m t} + e^{-j2\pi f_m t})\right]\right\} \tag{3.97}$$

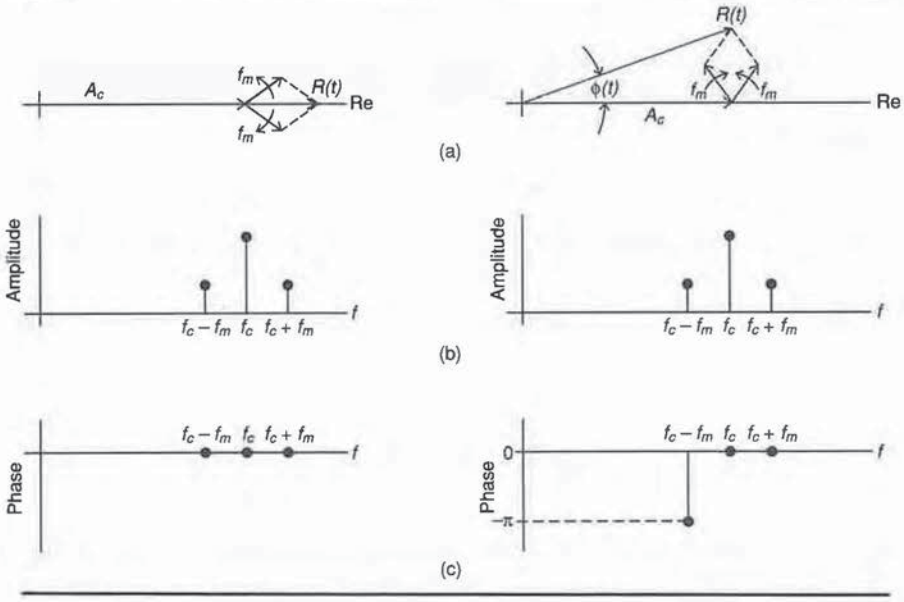
Comparing (3.94) and (3.97) illustrates the similarity between the two signals. The first, and most important, difference is the sign of the term at frequency $f_c - f_m$, which represents the lower sideband. The other difference is that the index a in the AM signal is replaced by f_d/f_m in the narrowband FM signal. We will see in the following section that f_d/f_m determines the index for an FM signal. Thus these two parameters are in a sense equivalent.

Additional insight is gained by sketching the phasor diagrams and the amplitude and phase spectra for both signals. These are given in Figure 3.25. The phasor diagrams are drawn using the carrier phase as a reference. The difference between AM and narrowband angle modulation with a sinusoidal message signal lies in the fact that the phasor resulting from the lower-sideband (LSB) and upper-sideband (USB) phasors adds to the carrier for AM but is in phase quadrature with the carrier for angle modulation. This difference results from the minus sign in the LSB component and is also clearly seen in the phase spectra of the two signals. The amplitude spectra are equivalent. ▲

Spectrum of an Angle-Modulated Signal

The derivation of the spectrum of an angle-modulated signal is typically a very difficult task. However, if the message signal is sinusoidal, the instantaneous phase deviation of the modulated carrier is sinusoidal for both FM and PM, and the spectrum can be obtained with ease. Thus this is the case

FIGURE 3.25 Comparison of AM and narrowband angle modulation. (a) Phasor diagrams. (b) Single-sided amplitude spectra. (c) Single-sided phase spectra.



we will consider. Even though we are restricting our attention to a very special case, the results provide much insight into the frequency-domain behavior of angle modulation.

In order to compute the spectrum of an angle-modulated signal with a sinusoidal message signal, we assume that

$$\phi(t) = \beta \sin \omega_m t \tag{3.98}$$

The parameter β is known as the *modulation index* and is the maximum value of phase deviation for both FM and PM. The signal

$$x_c(t) = A_c \cos (\omega_c t + \beta \sin \omega_m t) \tag{3.99}$$

can be expressed as

$$x_c(t) = A_c \text{ Re } (e^{j\omega_c t} e^{j\beta \sin \omega_m t}) \tag{3.100}$$

The function $e^{j\beta \sin \omega_m t}$ is periodic with frequency ω_m and can therefore be expanded in a Fourier series. The Fourier coefficients are given by

$$\frac{\omega_m}{2\pi} \int_{-\pi/\omega_m}^{\pi/\omega_m} e^{j\beta \sin \omega_m t} e^{-jn\omega_m t}\, dt = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-j(nx - \beta \sin x)}\, dx \tag{3.101}$$

This integral cannot be evaluated in closed form. However, it has been well tabulated. The integral is a function of n and β and is known as the *Bessel function* of the first kind of order n and argument β. It is denoted $J_n(\beta)$ and is tabulated for several values of n and β in Table 3.2. The significance of the underlining of various values in the table will be explained later.

Thus, with the aid of Bessel functions, the Fourier series for $e^{j\beta\sin\omega_m t}$ can be written as

$$e^{j\beta\sin\omega_m t} = \sum_{n=-\infty}^{\infty} J_n(\beta)e^{jn\omega_m t} \tag{3.102}$$

which allows the modulated carrier to be written as

$$x_c(t) = A_c \,\mathrm{Re}\left[e^{j\omega_c t} \sum_{n=-\infty}^{\infty} J_n(\beta)e^{jn\omega_m t} \right] \tag{3.103}$$

Taking the real part yields

$$x_c(t) = A_c \sum_{n=-\infty}^{\infty} J_n(\beta) \cos(\omega_c + n\omega_m)t \tag{3.104}$$

from which the spectrum of $x_c(t)$ can be determined by inspection. The spectrum has components at the carrier frequency and has an infinite number of sidebands separated from the carrier frequency by integer multiples of the modulation frequency ω_m. The amplitude of each spectral component can be determined from a table of values of the Bessel function. Such tables typically give $J_n(\beta)$ only for positive values of n. However, from the definition of $J_n(\beta)$, it can be determined that

$$J_{-n}(\beta) = J_n(\beta) \tag{3.105}$$

for n even and

$$J_{-n}(\beta) = -J_n(\beta) \tag{3.106}$$

for n odd. The single-sided spectra of (3.104) are shown in Figure 3.26.

A useful relationship between values of $J_n(\beta)$ for various values of n is the recursion formula

$$J_{n-1}(\beta) + J_{n+1}(\beta) = \frac{2n}{\beta} J_n(\beta) \tag{3.107}$$

Thus we can compute $J_{n+1}(\beta)$ from knowledge of $J_n(\beta)$ and $J_{n-1}(\beta)$. This enables us to compute a table of values of the Bessel function, similar to Table 3.2, up to any value of n from $J_0(\beta)$ and $J_1(\beta)$.

TABLE 3.2 Table of Bessel Functions

n	$\beta = 0.05$	$\beta = 0.1$	$\beta = 0.2$	$\beta = 0.3$	$\beta = 0.5$	$\beta = 0.7$	$\beta = 1.0$	$\beta = 2.0$	$\beta = 3.0$	$\beta = 5.0$	$\beta = 7.0$	$\beta = 8.0$	$\beta = 10.0$
0	0.999	0.998	0.990	0.978	0.938	0.881	0.765	0.224	−0.260	−0.178	0.300	0.172	−0.246
1	0.025	0.050	0.100	0.148	0.242	0.329	0.440	0.577	0.339	−0.328	−0.005	0.235	0.043
2		0.001	0.005	0.011	0.031	0.059	0.115	0.353	0.486	0.047	−0.301	−0.113	0.255
3				0.001	0.003	0.007	0.020	0.129	0.309	0.365	−0.168	−0.291	0.058
4						0.001	0.002	0.034	0.132	0.391	0.158	−0.105	−0.220
5								0.007	0.043	0.261	0.348	0.186	−0.234
6								0.001	0.011	0.131	0.339	0.338	−0.014
7									0.003	0.053	0.234	0.321	0.217
8										0.018	0.128	0.223	0.318
9										0.006	0.059	0.126	0.292
10										0.001	0.024	0.061	0.207
11											0.008	0.026	0.123
12											0.003	0.010	0.063
13											0.001	0.003	0.029
14												0.001	0.012
15													0.005
16													0.002
17													0.001

FIGURE 3.26 Spectra of an angle-modulated signal. (a) Single-sided amplitude spectrum. (b) Single-sided phase spectrum.

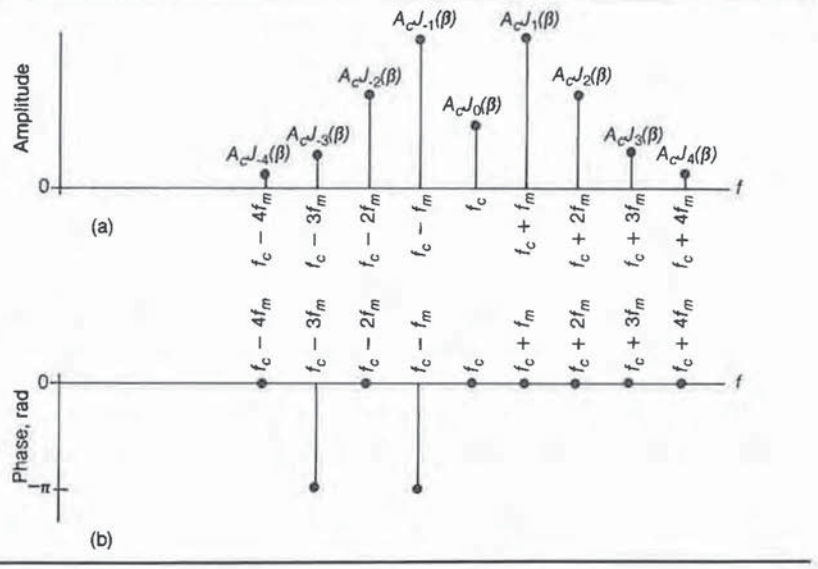
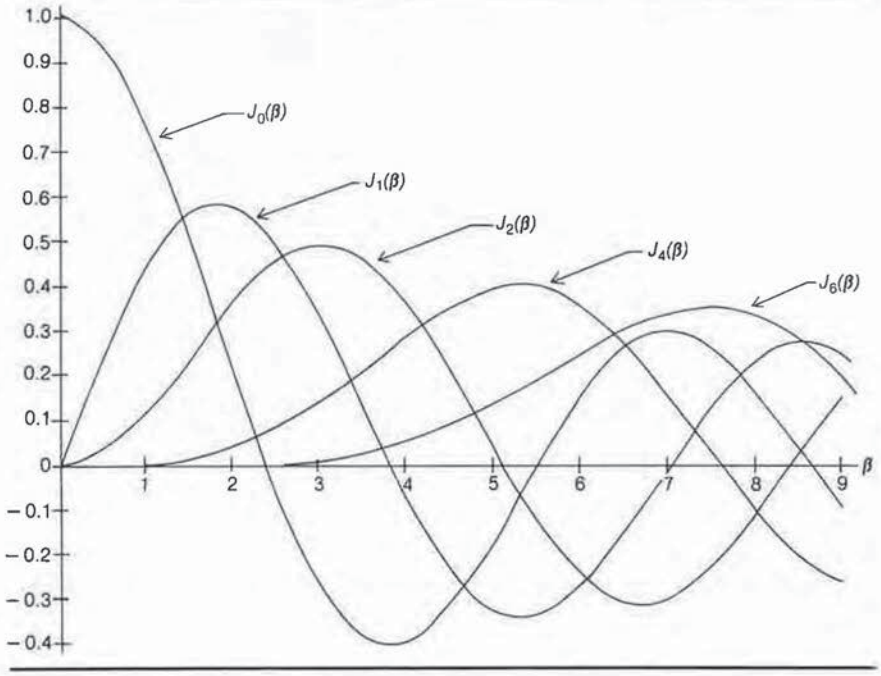


Figure 3.27 illustrates the behavior of the Fourier Bessel coefficients $J_n(\beta)$, for $n = 0, 1, 2, 4$, and 6 with $0 \leq \beta \leq 9$. Several interesting observations can be made. First, for $\beta \ll 1$, it is clear that $J_0(\beta)$ predominates, giving rise to narrowband angle modulation. It also can be seen that $J_n(\beta)$ oscillates for increasing β, but that the amplitude of oscillation decreases with increasing β. Also of interest is the fact that the maximum value of $J_n(\beta)$ decreases with increasing n.

As Figure 3.27 shows, $J_n(\beta)$ is equal to zero at several values of β. Denoting these values of β by β_{nk}, where $k = 0, 1, 2$, we have the results in Table 3.3. As an example, $J_0(\beta)$ is zero for β equal to 2.4048, 5.5201, and 8.6537. Of course, there are an infinite number of points at which $J_n(\beta)$ is zero for any n, but consistent with Figure 3.27, only the values in the range $0 \leq \beta \leq 9$ are shown in Table 3.3. It follows that since $J_0(\beta)$ is zero at β equal to 2.4048, 5.5201, and 8.6537, the spectrum of the modulator output will not contain a component at the carrier frequency for these values of the modulation index. These points are referred to as *carrier nulls*. In a similar manner, the components at $f = f_c \pm f_m$ are zero if $J_1(\beta)$ is zero. The values of the modulation index giving rise to this condition are 0, 3.8317, and 7.0156. It should be obvious why only $J_0(\beta)$ is nonzero at $\beta = 0$. If the modulation index is zero, then either $m(t)$ is zero or the deviation constant f_d is zero. In

FIGURE 3.27 $J_n(\beta)$ as a function of β



either case, the modulator output is the unmodulated carrier, which has frequency components only at the carrier frequency.

In computing the spectrum of the modulator output, our starting point was the assumption that

$$\phi(t) = \beta \sin \omega_m t \tag{3.108}$$

We did not specify the modulator type. The assumed $\phi(t)$ could represent the phase deviation of a PM modulator with $m(t) = A \sin \omega_m t$ and an index $\beta = k_p A$.

TABLE 3.3 Values of β for which $J_n(\beta) = 0$ for $0 \leq \beta \leq 9$

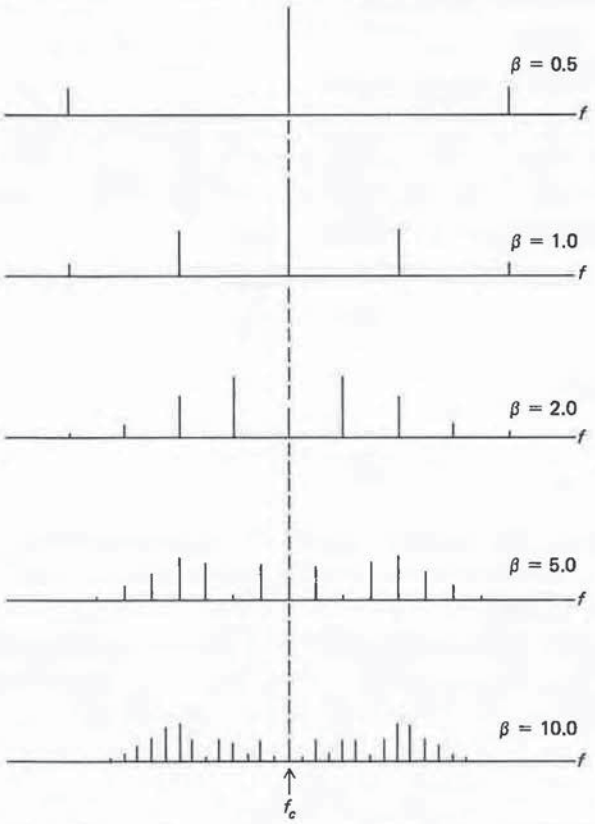
n		β_{n0}	β_{n1}	β_{n2}
0	$J_0(\beta) = 0$	2.4048	5.5201	8.6537
1	$J_1(\beta) = 0$	0.0000	3.8317	7.0156
2	$J_2(\beta) = 0$	0.0000	5.1356	8.4172
4	$J_4(\beta) = 0$	0.0000	7.5883	—
6	$J_6(\beta) = 0$	0.0000	—	—

Alternatively, an FM modulator with $m(t) = A \cos \omega_m t$ yields the assumed $\phi(t)$ with

$$\beta = \frac{2\pi f_d A}{\omega_m} = \frac{f_d A}{f_m} \tag{3.109}$$

Thus the modulation index for FM is a function of the modulation frequency. This is not the case for PM. The behavior of the spectrum of an FM signal is illustrated in Figure 3.28 as f_m is decreased while holding Af_d constant. For large values of f_m, the signal is narrowband FM, since only two sidebands are significant. For small values of f_m, many sidebands have significant value.

FIGURE 3.28 Amplitude spectrum of an FM signal as β is increased by decreasing f_m



Power in an Angle-Modulated Signal

The power in an angle-modulated signal is easily computed from (3.77). Squaring (3.77) and taking the time-average value yields

$$\langle x_c^2(t)\rangle = A_c^2\langle \cos^2[\omega_c t + \phi(t)]\rangle \tag{3.110}$$

which can be written as

$$\langle x_c^2(t)\rangle = \tfrac{1}{2}A_c^2 + \tfrac{1}{2}A_c^2\langle \cos 2[\omega_c t + \phi(t)]\rangle \tag{3.111}$$

If the carrier frequency is large so that $x_c(t)$ has negligible frequency content in the region of dc, the second term in (3.111) is negligible and

$$\langle x_c^2(t)\rangle = \tfrac{1}{2}A_c^2 \tag{3.112}$$

Thus the power contained in the output of an angle modulator is independent of the message signal. This is one important difference between angle modulation and linear modulation.

Bandwidth of Angle-Modulated Signals

Strictly speaking, the bandwidth of an angle-modulated signal is infinite, since angle modulation of a carrier results in the generation of an infinite number of sidebands. However, it can be seen from the series expansion of $J_n(\beta)$ (Appendix C, Table C.4) that for large n

$$J_n(\beta) \approx \frac{\beta^n}{2^n n!} \tag{3.113}$$

Thus, for fixed β,

$$\lim_{n\to\infty} J_n(\beta) = 0 \tag{3.114}$$

This behavior can also be seen from the values of $J_n(\beta)$ given in Table 3.2.

Since the values of $J_n(\beta)$ become negligible for sufficiently large n, the bandwidth of an angle-modulated signal can be defined by considering only those terms that contain significant power. The power ratio, P_r, is defined as the ratio of the power contained in the carrier ($n = 0$) component and the k components each side of the carrier to the total power in $x_c(t)$. Thus

$$P_r = \frac{\frac{1}{2}A_c^2 \sum_{n=-k}^{k} J_n^2(\beta)}{\frac{1}{2}A_c^2} \tag{3.115a}$$

or simply

$$P_r = J_0^2(\beta) + 2\sum_{n=1}^{k} J_n^2(\beta) \tag{3.115b}$$

Bandwidth for a particular application can be determined by defining an acceptable power ratio, solving for the required value of k using a table of Bessel functions, and then recognizing that the resulting bandwidth is

$$B = 2kf_m \tag{3.116}$$

The acceptable value of the power ratio is dictated by the particular application of the system. Two power ratios are depicted in Table 3.2: $P_r \geq 0.7$ and $P_r \geq 0.98$. The value of n corresponding to k for $P_r \geq 0.7$ is indicated by a single underscore, and the value of n corresponding to k for $P_r \geq 0.98$ is indicated by a double underscore. For $P_r \geq 0.98$, it is noted that n is equal to the integer part of $1 + \beta$, so that

$$B \cong 2(\beta + 1)f_m \tag{3.117}$$

The preceding expression assumes sinusoidal modulation, since the modulation index β is defined only for sinusoidal modulation. For arbitrary $m(t)$, a generally accepted expression for bandwidth results if the deviation ratio D is defined as

$$D = \frac{\text{peak frequency deviation}}{\text{bandwidth of } m(t)} \tag{3.118}$$

which is

$$D = \frac{f_d}{W}[\max\ |m(t)|] \tag{3.119}$$

The deviation ratio plays the same role for nonsinusoidal modulation as the modulation index plays for sinusoidal systems. Replacing β by D and replacing f_m by W in (3.117), we obtain

$$B = 2(D + 1)W \tag{3.120}$$

This expression for bandwidth is generally referred to as *Carson's rule*. If $D \ll 1$, the bandwidth is approximately $2W$, and the signal is known as a *narrowband angle-modulated signal*. Conversely, if $D \gg 1$, the bandwidth is approximately $2DW = 2f_d[\max\ |m(t)|]$, which is twice the peak frequency deviation. Such a signal is known as a *wideband angle-modulated signal*.

EXAMPLE 3.5 In this example we consider an FM modulator with output

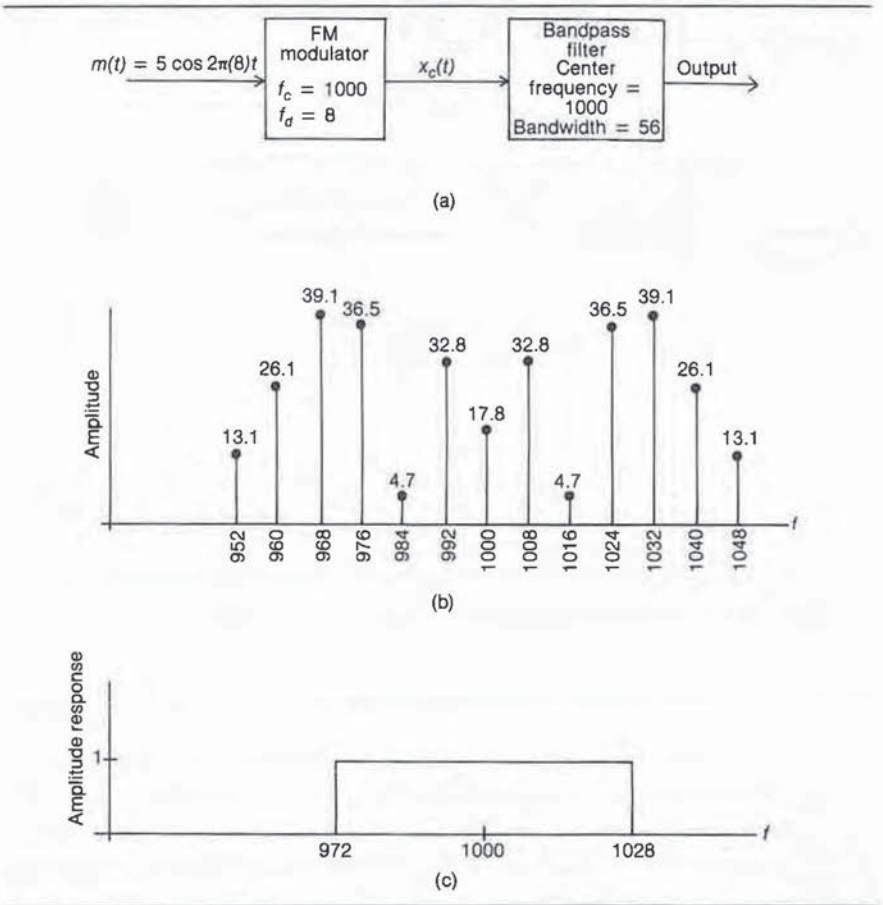
$$x_c(t) = 100 \cos [2\pi(1000)t + \phi(t)] \tag{3.121}$$

The modulator operates with $f_d = 8$ and has the input message signal

$$m(t) = 5 \cos 2\pi(8)t \tag{3.122}$$

The modulator is followed by a bandpass filter with a center frequency of 1000 Hz and a bandwidth of 56 Hz, as shown in Figure 3.29(a). Our problem is to determine the power at the filter output.

FIGURE 3.29 System and spectra for Example 3.5. (a) FM system. (b) Single-sided spectrum of modulator output. (c) Amplitude response of bandpass filter.



SOLUTION The peak deviation is $5f_d$ or 40 Hz, and $f_m = 8$ Hz. Thus the modulation index is

$$\beta = \frac{40}{8} = 5 \tag{3.123}$$

This yields the single-sided amplitude spectrum shown in Figure 3.29(b).

Figure 3.31(c) shows the passband of the bandpass filter. The filter passes the component at the carrier frequency and three components on each side of the carrier. Thus the power ratio is

$$P_r = J_0^2(5) + 2[J_1^2(5) + J_2^2(5) + J_3^2(5)] \tag{3.124}$$

which is

$$P_r = (0.178)^2 + 2[(0.328)^2 + (0.047)^2 + (0.365)^2] \tag{3.125}$$

This yields

$$P_r = 0.518 \tag{3.126}$$

The power at the output of the modulator is

$$\overline{x_c^2} = \tfrac{1}{2}A_c^2 = \tfrac{1}{2}(100)^2 = 5000 \text{ W} \tag{3.127}$$

The power at the filter output is the power of the modulator output multiplied by the power ratio. Thus the power at the filter output is

$$P_r\overline{x_c^2} = 2589 \text{ W} \tag{3.128}$$

▲

▼ **EXAMPLE 3.6** In the development of the spectrum of an angle-modulated signal, it was assumed that the message signal was a single sinusoid. We now consider a somewhat more general problem in which the message signal is the sum of two sinusoids. Let the message signal be

$$m(t) = A \cos \omega_1 t + B \cos \omega_2 t \tag{3.129}$$

For FM modulation the phase deviation is therefore given by

$$\phi(t) = \beta_1 \sin \omega_1 t + \beta_2 \sin \omega_2 t \tag{3.130}$$

where $\beta_1 = Af_d/f_1$ and $\beta_2 = Bf_d/f_2$. The modulator output for this case becomes

$$x_c(t) = A_c \cos (\omega_c t + \beta_1 \sin \omega_1 t + \beta_2 \sin \omega_2 t) \tag{3.131}$$

which can be expressed

$$x_c(t) = A_c \text{Re}\left\{e^{j\omega_c t}\, e^{j\beta_1 \sin\omega_1 t}\, e^{j\beta_2 \sin\omega_2 t}\right\} \tag{3.132}$$

Using (3.102) we can write

$$e^{j\beta_1 \sin\omega_1 t} = \sum_{n=-\infty}^{\infty} J_n(\beta_1) e^{jn\omega_1 t} \tag{3.133a}$$

and

$$e^{j\beta_2 \sin\omega_2 t} = \sum_{m=-\infty}^{\infty} J_m(\beta_2) e^{jm\omega_2 t} \tag{3.133b}$$

The modulator output can therefore be written

$$x_c(t) = A_c \mathrm{Re}\left\{ e^{j\omega_c t} \sum_{n=-\infty}^{\infty} J_n(\beta_1) e^{jn\omega_1 t} \sum_{m=-\infty}^{\infty} J_m(\beta_2) e^{jm\omega_2 t} \right\} \tag{3.134}$$

which, upon taking the real part, can be expressed

$$x_c(t) = A_c \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} J_n(\beta_1) J_m(\beta_2) \cos(\omega_c + n\omega_1 + m\omega_2)t \tag{3.135}$$

Examination of the signal $x_c(t)$ shows that it not only contains frequency components at $\omega_c + n\omega_1$ and $\omega_c + m\omega_2$, but contains frequency components at $\omega_c + n\omega_1 + m\omega_2$ for all combinations of n and m. Therefore, the spectrum of the modulator output due to a message signal consisting of the sum of two sinusoids contains additional components over the spectrum formed by the superposition of the two spectra resulting from the individual message components. This example therefore illustrates the nonlinear nature of angle modulation. The spectrum resulting from a message signal consisting of the sum of two sinusoids is shown in Figure 3.30 for the case in which $\beta_1 = \beta_2$ and $f_2 = 12f_1$. ▲

FIGURE 3.30 Aplitude spectrum for (3.135) with $\beta_1 = \beta_2 = 2$ and $f_2 = 12f_1$.

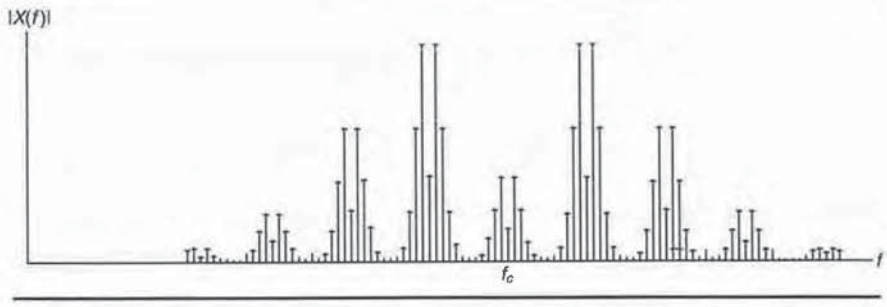
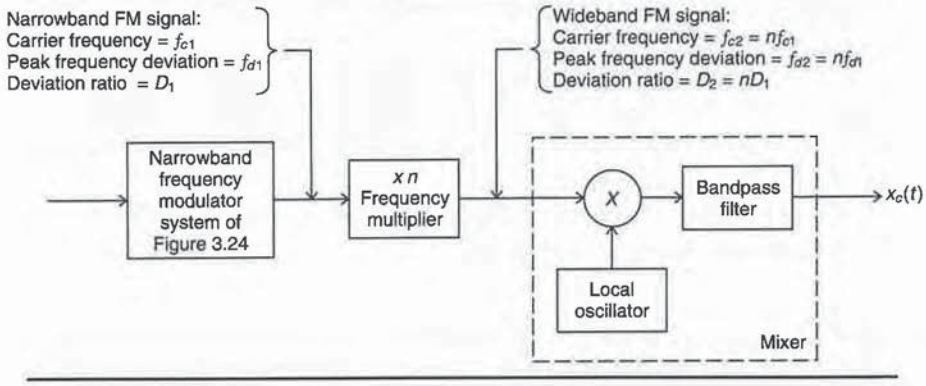


FIGURE 3.31 Frequency modulation utilizing narrowband-to-wideband conversion



Narrowband-to-Wideband Conversion

One technique for generating wideband FM is illustrated in Figure 3.31. The carrier frequency of the narrowband frequency modulator is f_{c1}, and the peak frequency deviation is f_{d1}. The frequency multiplier multiplies the argument of the input sinusoid by n. In other words, if the input of a frequency multiplier is

$$x(t) = A_c \cos [\omega_0 t + \phi(t)] \tag{3.136}$$

the output of the frequency multiplier is

$$y(t) = A_c \cos [n\omega_0 t + n\phi(t)] \tag{3.137}$$

Assuming that the output of the local oscillator is

$$e_{LO}(t) = 2 \cos \omega_{LO} t \tag{3.138}$$

results in

$$e(t) = A_c \cos [(n\omega_0 + \omega_{LO})t + n\phi(t)] + A_c \cos [(n\omega_0 - \omega_{LO})t + n\phi(t)] \tag{3.139}$$

for the multiplier output. This signal is then filtered, using a bandpass filter having center frequency ω_c, given by

$$\omega_c = n\omega_0 + \omega_{LO} \quad \text{or} \quad \omega_c = n\omega_0 - \omega_{LO}$$

This yields the output

$$x_c(t) = A_c \cos [\omega_c t + n\phi(t)] \tag{3.140}$$

The bandwidth of the bandpass filter is chosen in order to pass the desired term in (3.139). One can use Carson's rule to determine the bandwidth of the bandpass filter if the transmitted signal is to contain 98% of the power in $x_c(t)$.

The central idea in narrowband-to-wideband conversion is that the frequency multiplier changes both the carrier frequency and the deviation ratio by a factor of n, whereas the mixer changes the effective carrier frequency but does not affect the deviation ratio.

This technique of implementing wideband frequency modulation is known as *indirect frequency modulation*.

▼ **EXAMPLE 3.7** A narrowband-to-wideband converter is implemented as shown in Figure 3.31. The output of the narrowband frequency modulator is given by (3.136) with ω_0 given by 2π (100,000). The peak frequency deviation of $\phi(t)$ is 50 Hz and the bandwidth of $\phi(t)$ is 500 Hz. The wideband output $x_c(t)$ is to have a carrier frequency of 85 MHz and a deviation ratio of 5. Determine the frequency multiplier factor, n. Also determine two possible local oscillator frequencies. Finally, determine the center frequency and the bandwidth of the bandpass filter.

SOLUTION The deviation ratio at the output of the narrowband modulator is

$$D_1 = \frac{f_{d1}}{W} = \frac{50}{500} = 0.1 \tag{3.141}$$

The frequency multiplier factor is

$$n = \frac{D_2}{D_1} = \frac{5}{0.1} = 50 \tag{3.142}$$

Thus, the carrier frequency at the output of the narrowband modulator is

$$nf_{c1} = 50(100{,}000) = 5 \text{ MHz} \tag{3.143}$$

The two permissible frequencies for the local oscillator are

$$85 + 5 = 90 \text{ MHz} \tag{3.144a}$$

and

$$85 - 5 = 80 \text{ MHz} \tag{3.144b}$$

The center frequency of the bandpass filter must be equal to the desired carrier frequency of the wideband output. Thus the center frequency of

bandpass filter is 85 MHz. The bandwidth of the bandpass filter is established using Carson's rule. From (3.120) we have

$$B = 2(D + 1)\,W = 2(5 + 1)(500) \tag{3.145a}$$

Thus

$$B = 6{,}000 \text{ Hz} \tag{3.145b}$$

Demodulation of Angle-Modulated Signals

The demodulation of an FM signal requires a circuit that yields an output proportional to the frequency deviation of the input. Such circuits are known as *discriminators*. If the input to an ideal discriminator is the angle-modulated signal

$$x_r(t) = A_c \cos [\omega_c t + \phi(t)] \tag{3.146}$$

the output of the ideal discriminator is

$$y_D(t) = \frac{1}{2\pi} K_D \frac{d\phi}{dt} \tag{3.147}$$

For FM, $\phi(t)$ is given by

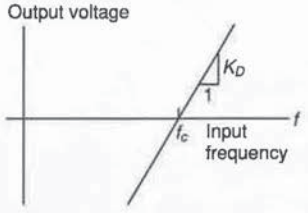
$$\phi(t) = 2\pi f_d \int^t m(\alpha)\, d\alpha \tag{3.148}$$

so that (3.147) becomes

$$y_D(t) = K_D f_d m(t) \tag{3.149}$$

The constant K_D is known as the *discriminator constant* and has units of volts per hertz. Since an ideal discriminator yields an output signal proportional to the frequency deviation from a carrier, it has a linear frequency-to-voltage transfer function, which passes through zero at $f = f_c$. This is illustrated in Figure 3.32.

FIGURE 3.32 Ideal discriminator characteristic



The system characterized by Figure 3.32 also can be used to demodulate PM signals. Since $\phi(t)$ is proportional to $m(t)$ for PM, $y_D(t)$ given by (3.147) is proportional to the time derivative of $m(t)$ for PM inputs. Integration of the discriminator output yields a signal proportional to $m(t)$. Thus a demodulator for PM can be implemented as an FM discriminator followed by an integrator. We define the output of a PM discriminator as

$$y_D(t) = K_D k_p m(t) \tag{3.150}$$

It will be clear from the text whether $y_D(t)$ and K_D refer to an FM or a PM system.

An approximation to the characteristic illustrated in Figure 3.32 can be obtained by the use of a differentiator followed by an envelope detector, as shown in Figure 3.33. If the input to the differentiator is

$$x_r(t) = A_c \cos [\omega_c t + \phi(t)] \tag{3.151}$$

the output of the differentiator is

$$e(t) = -A_c \left(\omega_c + \frac{d\phi}{dt} \right) \sin [\omega_c t + \phi(t)] \tag{3.152}$$

This is exactly the same form as an AM signal, except for the phase deviation $\phi(t)$. Thus, after differentiation, envelope detection can be used to recover the message signal. The envelope of $e(t)$ is

$$y(t) = A_c \left(\omega_c + \frac{d\phi}{dt} \right) \tag{3.153}$$

and is always positive if

$$\omega_c > -\frac{d\phi}{dt}, \text{ all } t$$

which is certainly the typical case. With this assumption, the output of the envelope detector is

$$y_D(t) = A_c \frac{d\phi}{dt} = 2\pi A_c f_d \, m(t) \tag{3.154}$$

FIGURE 3.33 FM discriminator

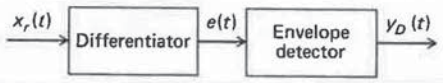
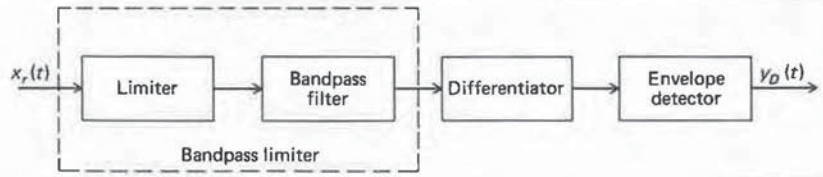


FIGURE 3.34 FM discriminator with bandpass limiter



assuming that the dc term, $A_c\omega_c$, is removed. Comparing (3.154) and (3.149) shows that the discriminator constant for this discriminator is

$$K_D = 2\pi A_c \tag{3.155}$$

We will see later that interference and channel noise perturb the amplitude A_c of $x_r(t)$. In order to ensure that the amplitude at the input to the differentiator is constant, a *limiter* is placed before the differentiator. The output of the limiter is a signal of squarewave type, which is K sgn $[x_r(t)]$. A bandpass filter having center frequency ω_c is then placed after the limiter to convert the signal back to the sinusoidal form required by the differentiator to yield the response defined by (3.152). The cascade combination of a limiter and a bandpass filter is known as a *bandpass limiter*. The complete discriminator is illustrated in Figure 3.34.

The process of differentiation can often be realized using a time-delay implementation, as shown in Figure 3.35. The signal $e(t)$, which is the input to the envelope detector, is given by

$$e(t) = x_r(t) - x_r(t - \tau) \tag{3.156}$$

which can be written

$$\frac{e(t)}{\tau} = \frac{x_r(t) - x_r(t - \tau)}{\tau} \tag{3.157}$$

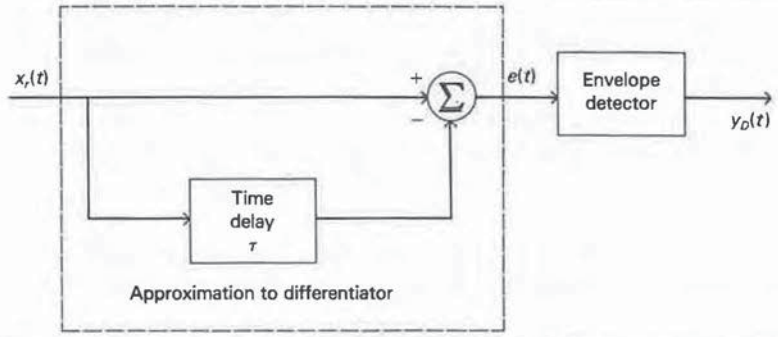
Since, by definition,

$$\lim_{\tau \to 0} \frac{e(t)}{\tau} = \lim_{\tau \to 0} \frac{x_r(t) - x_r(t - \tau)}{\tau} = \frac{dx_r(t)}{dt} \tag{3.158}$$

it follows that for small τ,

$$e(t) \cong \tau \frac{dx_r(t)}{dt} \tag{3.159}$$

FIGURE 3.35 Discriminator implementation using delay and envelope detection



This is, except for the constant factor τ, identical to the envelope detector input shown in Figure 3.33 and defined by (3.152). The resulting discriminator constant K_D is $2\pi A_c \tau$.

There are many other techniques that can be used to implement a discriminator. In Section 3.4 we will examine the phase-lock loop, which is an especially attractive implementation.

▼ **EXAMPLE 3.8** Consider the simple RC network shown in Figure 3.36(a). The transfer function is

$$H(f) = \frac{R}{R + \dfrac{1}{j2\pi fC}} = \frac{j2\pi fRC}{1 + j2\pi fRC} \tag{3.160}$$

The amplitude response is shown in Figure 3.36(b). If all frequencies present in the input are low, so that

$$f \ll \frac{1}{2\pi RC}$$

the transfer function can be approximated by

$$H(f) = j2\pi fRC \tag{3.161}$$

Thus, for small f, the RC network has the linear amplitude-frequency characteristic required of an ideal discriminator.

Equation (3.161) illustrates that for small f, the RC filter acts as a differ-

entiator with gain RC. Thus, the RC network can be used in place of the differentiator in Figure 3.34 to yield a discriminator with

$$K_D = 2\pi A_c RC \tag{3.162}$$

Example 3.8 illustrates the essential components of a frequency discriminator, a circuit that has an amplitude response linear with frequency and an envelope detector. However, a highpass filter does not in general yield a practical implementation. This can be seen from the expression for K_D. Clearly the 3-dB frequency of the filter, $1/2\pi RC$, must exceed the carrier frequency f_c. In commercial FM broadcasting, the carrier frequency at the discriminator input, i.e., the IF frequency, is on the order of 10 MHz. As a result, the discriminator constant K_D is very small indeed.

A solution to the problem of a very small K_D is to use a bandpass filter, as illustrated in Figure 3.37. But, as shown in Figure 3.37(a), the region of linear operation is often unacceptably small. In addition, use of a bandpass filter results in a dc bias on the discriminator output. This dc bias could of

FIGURE 3.36 Implementation of a simple discriminator. (a) RC network. (b) Transfer function. (c) Simple discriminator.

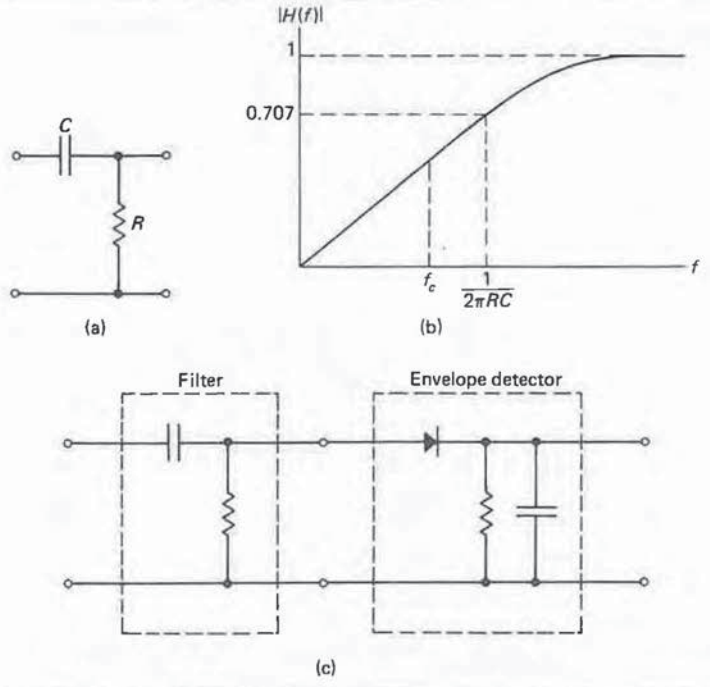
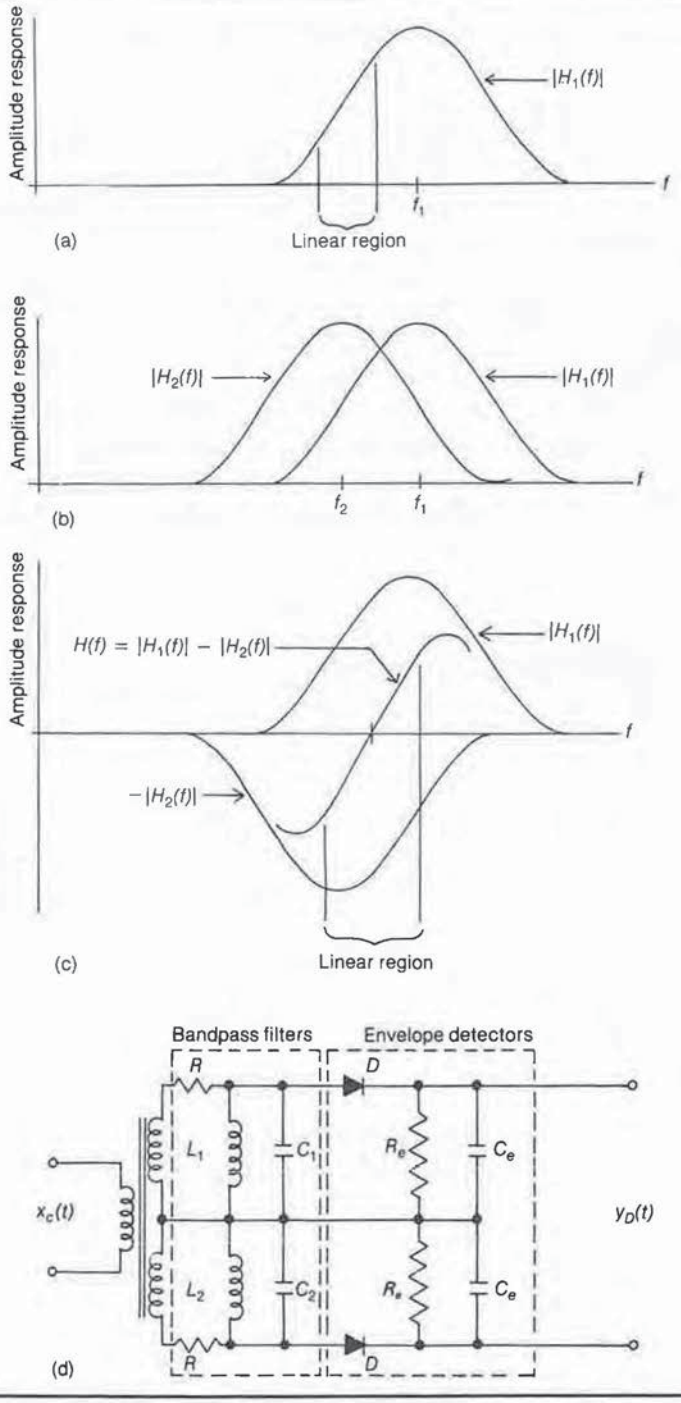


FIGURE 3.37 Derivation of balanced discriminator. (a) Bandpass filter. (b) Stagger-tuned bandpass filters. (c) Amplitude response $H(f)$ of balanced discriminator. (d) Balanced discriminator.



course be removed by a blocking capacitor, but the blocking capacitor would negate an inherent advantage of FM—namely, that FM has dc response. One can solve these problems by using two filters with staggered center frequencies f_1 and f_2, as shown in Figure 3.37(b). The magnitudes of the envelope detector outputs following the two filters are proportional to $|H_1(f)|$ and $|H_2(f)|$. Subtracting these two outputs yields the overall characteristic

$$H(f) = |H_1(f)| - |H_2(f)| \tag{3.163}$$

as shown in Figure 3.37(c). The combination is linear over a wider frequency range than would be the case for either filter used alone, and it is clearly possible to make $H(f_c) = 0$.

There are several techniques that can be used to combine the outputs of two envelope detectors. A differential amplifier can be used, for example. Another alternative, using a strictly passive circuit, is shown in Figure 3.37(d). A center-tapped transformer supplies the input signal $x_c(t)$ to the inputs of the two bandpass filters. The center frequencies of the two bandpass filters are given by

$$f_i = \frac{1}{2\pi\sqrt{L_iC_i}} \tag{3.164}$$

for $i = 1, 2$. The envelope detectors are formed by the diodes and the resistor-capacitor combinations R_eC_e. The output of the upper envelope detector is proportional to $|H_1(f)|$, and the output of the lower envelope detector is proportional to $|H_2(f)|$. The output of the upper envelope detector is the positive portion of its input envelope, and the output of the lower envelope detector is the negative portion of its input envelope. Thus $y_D(t)$ is proportional to $|H_1(f)| - |H_2(f)|$. This system is known as a *balanced discriminator* because the response to the undeviated carrier is balanced so that the net response is zero.

3.3 INTERFERENCE

We now consider the effect of interference in communication systems. In real-world systems interference occurs from various sources, such as radio-frequency emissions from transmitters having carrier frequencies close to that of the carrier being demodulated. We also study interference because the analysis of systems in the presence of interference provides us with important insights into the behavior of systems operating in the presence of noise, which is the topic of Chapter 6. In this section we consider both linear

modulation and angle modulation. It is important to understand the very different manner in which these two systems behave in the presence of interference.

Interference in Linear Modulation

As a simple case of linear modulation in the presence of interference, we consider the received signal having the spectrum (single-sided) shown in Figure 3.38. The received signal consists of three components: a carrier component, a pair of sidebands representing a sinusoidal message signal and an undesired interfering tone of frequency $f_c + f_i$. The input to the demodulator is therefore

$$x_c(t) = A_c \cos \omega_c t + A_i \cos (\omega_c + \omega_i)t + A_m \cos \omega_m t \cos \omega_c t \quad (3.165)$$

Multiplying $x_c(t)$ by 2 cos $\omega_c t$ and lowpass-filtering (coherent demodulation) yields

$$y_D(t) = A_m \cos \omega_m t + A_i \cos \omega_i t \quad (3.166)$$

where we have assumed that the interference component is passed by the filter and that the dc term resulting from the carrier is blocked. From this simple example we see that the signal and interference are additive at the receiver output if the interference is additive at the receiver input. This result was obtained because the coherent demodulator operates as a linear demodulator.

The effect of interference with envelope detection is quite different because of the nonlinear nature of the envelope detector. The analysis with envelope detection is much more difficult than the coherent demodulation case. Some insight can be gained by writing $x_c(t)$ in a form that leads to the phasor diagram. In order to develop the phasor diagram we write (3.165) in the form

$$x_c(t) = \text{Re}[A_c e^{j\omega_c t} + A_i e^{j\omega_c t} e^{j\omega_i t} + \tfrac{1}{2}A_m e^{j\omega_c t} e^{j\omega_m t} + \tfrac{1}{2}A_m e^{j\omega_c t} e^{-j\omega_m t}] \quad (3.167)$$

FIGURE 3.38 Assumed received-signal spectrum

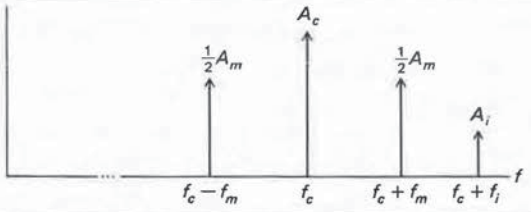
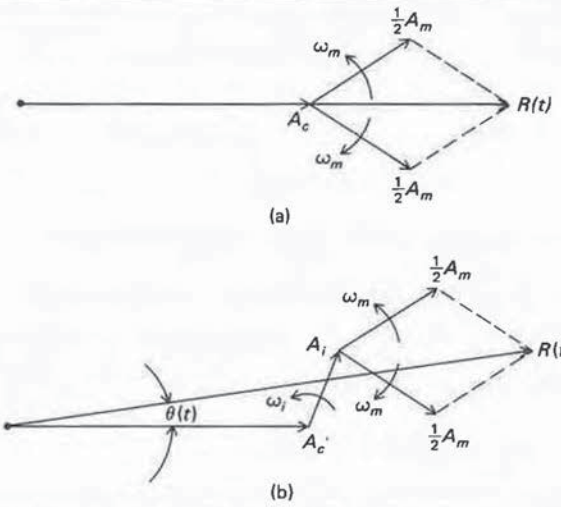


FIGURE 3.39 Phasor diagrams illustrating interference. (a) Phasor diagram without interference. (b) Phasor diagram with interference.



which also can be written as

$$x_r(t) = \text{Re}\left\{e^{j\omega_c t}[A_c + A_i e^{j\omega_i t} + \tfrac{1}{2}A_m e^{j\omega_m t} + \tfrac{1}{2}A_m e^{-j\omega_m t}]\right\} \tag{3.168}$$

The phasor diagram is constructed with respect to the carrier by taking the carrier frequency as equal to zero. The phasor diagrams are illustrated in Figure 3.39, both with and without interference. The output of an ideal envelope detector is $R(t)$ in both cases. The phasor diagrams illustrate that interference induces both an amplitude distortion and a phase deviation.

The effect of interference with envelope detection is determined by writing (3.165) as

$$\begin{aligned} x_r(t) &= A_c \cos \omega_c t + A_m \cos \omega_m t \cos \omega_c t \\ &\quad + A_i[\cos \omega_c t \cos \omega_i t - \sin \omega_c t \sin \omega_i t] \end{aligned} \tag{3.169}$$

which is

$$x_r(t) = [A_c + A_m \cos \omega_m t + A_i \cos \omega_i t] \cos \omega_c t - A_i \sin \omega_i t \sin \omega_c t \tag{3.170}$$

If $A_c \gg A_i$, which is the usual case of interest, the second term in (3.170) is negligible compared to the first term. For this case, the output of the envelope detector is

$$y_D(t) \cong A_m \cos \omega_m t + A_i \cos \omega_i t \tag{3.171}$$

assuming that the dc term is blocked. Thus, for the small interference case, envelope detection and coherent demodulation are essentially equivalent.

If $A_c \ll A_i$, the assumption cannot be made that the second term of (3.170) is negligible, and the output is significantly different. To show this, (3.165) is rewritten as

$$x_r(t) = A_c \cos(\omega_c + \omega_i - \omega_i)t + A_i \cos(\omega_c + \omega_i)t + A_m \cos \omega_m t \cos(\omega_c + \omega_i - \omega_i)t \tag{3.172}$$

which, when we use appropriate trigonometric identities, becomes

$$x_r(t) = A_c[\cos(\omega_c + \omega_i)t \cos \omega_i t + \sin(\omega_c + \omega_i)t \sin \omega_i t] + A_i \cos(\omega_c + \omega_i)t + A_m \cos \omega_m t[\cos(\omega_c + \omega_i)t \cos \omega_i t + \sin(\omega_c + \omega_i)t \sin \omega_i t] \tag{3.173}$$

Equation (3.173) can also be written as

$$x_r(t) = [A_i + A_c \cos \omega_i t + A_m \cos \omega_m t \cos \omega_i t] \cos(\omega_c + \omega_i)t + [A_c \sin \omega_i t + A_m \cos \omega_m t \sin \omega_i t] \sin(\omega_c + \omega_i)t \tag{3.174}$$

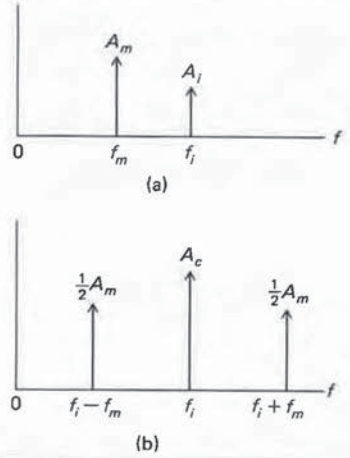
If $A_i \gg A_c$, the second term in (3.174) is negligible with respect to the first term. It follows that the envelope detector output is approximated by

$$y_D(t) \cong A_c \cos \omega_i t + A_m \cos \omega_m t \cos \omega_i t \tag{3.175}$$

At this point, several observations are in order. In envelope detectors, the largest high-frequency component is treated as the carrier. If $A_c \gg A_i$, the effective demodulation carrier has a frequency ω_c, whereas if $A_i \gg A_c$, the *effective* carrier frequency becomes the interference frequency $\omega_c + \omega_i$.

The spectra of the envelope detector output are illustrated in Figure 3.40 for $A_c \gg A_i$ and for $A_c \ll A_i$. For $A_c \gg A_i$ the interfering tone simply appears as a sinusoidal component, having frequency f_i, at the output of the envelope detector. This illustrates that for $A_c \gg A_i$, the envelope detector performs as a linear demodulator. The situation is much different for $A_c \ll A_i$, as can be seen from (3.175) and Figure 3.40(b). For this case we see that the sinusoidal message signal, having frequency f_m, modulates the interference tone. The output of the envelope detector has a spectrum that reminds us of the spectrum of an AM signal with carrier frequency f_i and sideband components at $f_i + f_m$ and $f_i - f_m$. The message signal is effectively lost. This degradation of the desired signal is called the *threshold effect* and is a consequence of the nonlinear nature of the envelope detector. We shall study the threshold effort in detail in Chapter 6 when we investigate the effect of noise in analog systems.

FIGURE 3.40 Envelope detector output spectra. (a) $A_c \gg A_i$, (b) $A_c \ll A_i$.



Interference in Angle Modulation

We now consider the effect of interference in angle modulation. We will see that the effect of interference in angle modulation is quite different from what was observed in linear modulation. Furthermore, we will see that the effect of interference in an FM system can be reduced by placing a lowpass filter at the discriminator output. We will consider this problem in considerable detail since the results will provide significant insight into the behavior of FM discriminators operating in the presence of noise.

Assume that the input to a PM or FM ideal discriminator is an unmodulated carrier plus an interfering tone at frequency $\omega_c + \omega_i$. Thus the input to the discriminator is assumed to have the form

$$x_t(t) = A_c \cos \omega_c t + A_i \cos (\omega_c + \omega_i)t \tag{3.176}$$

which can be written as

$$x_t(t) = A_c \cos \omega_c t + A_i \cos \omega_i t \cos \omega_c t - A_i \sin \omega_i \sin \omega_c t \tag{3.177}$$

Using Equations (3.50) through (3.54), the preceding expression can be written as

$$x_r(t) = R(t) \cos [\omega_c t + \psi(t)] \tag{3.178}$$

where

$$R(t) = \sqrt{(A_c + A_i \cos \omega_i t)^2 + (A_i \sin \omega_i t)^2} \tag{3.179}$$

and

$$\psi(t) = \tan^{-1} \frac{A_i \sin \omega_i t}{A_c + A_i \cos \omega_i t} \tag{3.180}$$

If $A_c \gg A_i$, Equations (3.179) and (3.180) can be approximated

$$R(t) = A_c + A_i \cos \omega_i t \tag{3.181}$$

and

$$\psi(t) = \frac{A_i}{A_c} \sin \omega_i t \tag{3.182}$$

Thus (3.178) is

$$x_r(t) = A_c\left(1 + \frac{A_i}{A_c} \cos \omega_i t\right) \cos\left(\omega_c t + \frac{A_i}{A_c} \sin \omega_i t\right) \tag{3.183}$$

Since the instantaneous phase deviation $\psi(t)$ is given by

$$\psi(t) = \frac{A_i}{A_c} \sin \omega_i t \tag{3.184}$$

the ideal discriminator output for PM is

$$y_D(t) = K_D \frac{A_i}{A_c} \sin \omega_i t \tag{3.185}$$

and for FM is

$$\begin{aligned} y_D(t) &= \frac{1}{2\pi} K_D \frac{d}{dt} \frac{A_i}{A_c} \sin \omega_i t \\ &= K_D \frac{A_i}{A_c} f_i \cos \omega_i t \end{aligned} \tag{3.186}$$

For both cases, the discriminator output is a sinusoid of frequency f_i. The amplitude of the discriminator output, however, is proportional to the frequency f_i for the FM case. It can be seen that for small f_i, the interfering tone has less effect on the FM system than on the PM system and that the opposite is true for large values of f_i. Values of $f_i > W$, the bandwidth of $m(t)$, are of little interest, since they can be removed by a lowpass filter following the discriminator.

For larger values of A_i the assumption that $A_i \ll A_c$ cannot be made and (3.186) no longer can describe the discriminator output. If the condition

$A_i \ll A_c$ does not hold, the discriminator is not operating above threshold and the analysis becomes much more difficult. Some insight into this case can be obtained from the phasor diagram, which is obtained by writing (3.176) in the form

$$x_r(t) = \mathrm{Re}\{[A_c + A_i\, e^{j\omega_i t}]\, e^{j\omega_c t}\} \tag{3.187}$$

The term in square brackets defines the phasor. The phasor diagram is shown in Figure 3.41(a). The carrier phase is taken as the reference and the interference phase is

$$\theta(t) = \omega_i t \tag{3.188}$$

Approximations to the phase of the resultant $\psi(t)$ can be determined using the phasor diagram.

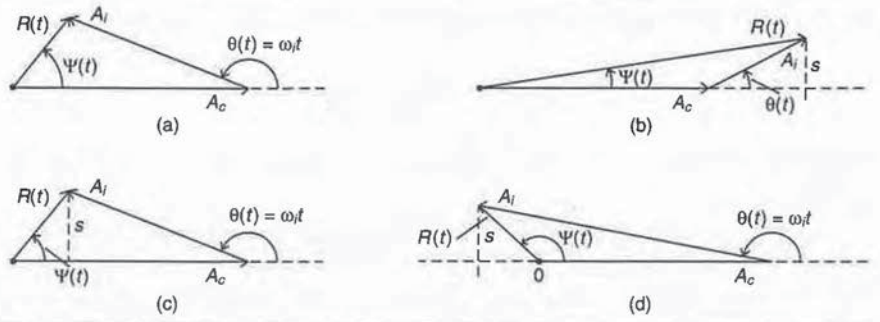
From Figure 3.41(b) we see that the magnitude of the discriminator output will be small when $\theta(t)$ is near zero. This results because for $\theta(t)$ near zero, a given change in $\theta(t)$ will result in a much smaller change in $\psi(t)$. Using the relationship between arc length, s; angle, θ, and radius, r, which is $s = \theta r$, we obtain

$$s = \theta(t) A_i \approx (A_c + A_i)\, \psi(t), \qquad \theta(t) \approx 0 \tag{3.189}$$

Solving for $\psi(t)$ yields

$$\psi(t) \approx \frac{A_i}{A_c + A_i}\, \omega_i t \tag{3.190}$$

FIGURE 3.41 Phasor diagram for carrier plus single-tone interference. (a) Phasor diagram for general $\theta(t)$. (b) Phasor diagram for $\theta(t) \approx 0$. (c) Phasor diagram for $\theta(t) \approx \pi$ and $A_i \lesssim A_c$. (d) Phasor diagram for $\theta(t) \approx \pi$ and $A_i \gtrsim A_c$.



Since the discriminator output is defined by

$$y_D(t) = \frac{K_D}{2\pi}\frac{d\psi}{dt} \tag{3.191}$$

we have

$$y_D(t) = K_D \frac{A_i}{A_c + A_i} f_i, \qquad \theta(t) \approx 0 \tag{3.192}$$

This is a positive quantity for $f_i > 0$ and a negative quantity for $f_i < 0$.

If A_i is slightly less than A_c, denoted $A_i \lesssim A_c$, and $\theta(t)$ is near π, a small positive change in $\theta(t)$ will result in a large negative change in $\psi(t)$. The result will be a negative spike appearing at the discriminator output. From Figure 3.41(c) we can write

$$s = A_i(\pi - \theta(t)) \approx (A_c - A_i)\psi(t), \qquad \theta(t) \approx \pi \tag{3.193}$$

which can be expressed

$$\psi(t) \approx \frac{A_i(\pi - \omega_i t)}{A_c - A_i} \tag{3.194}$$

Using (3.191) we see that the discriminator output is

$$y_D(t) \approx -K_D \frac{A_i}{A_c - A_i} f_i, \qquad \theta(t) \approx \pi \tag{3.195}$$

This is a negative quantity for $f_i > 0$ and a positive quantity for $f_i < 0$.

If A_i is slightly greater than A_c, denoted $A_i \gtrsim A_c$, and $\theta(t)$ is near π, a small positive change in $\theta(t)$ will result in a large positive change in $\psi(t)$. The result will be a positive spike appearing at the discriminator output. From Figure 3.41(d) we can write

$$s = A_i[\pi - \theta(t)] \approx (A_i - A_c)[\pi - \psi(t)], \quad \theta(t) \approx \pi \tag{3.196}$$

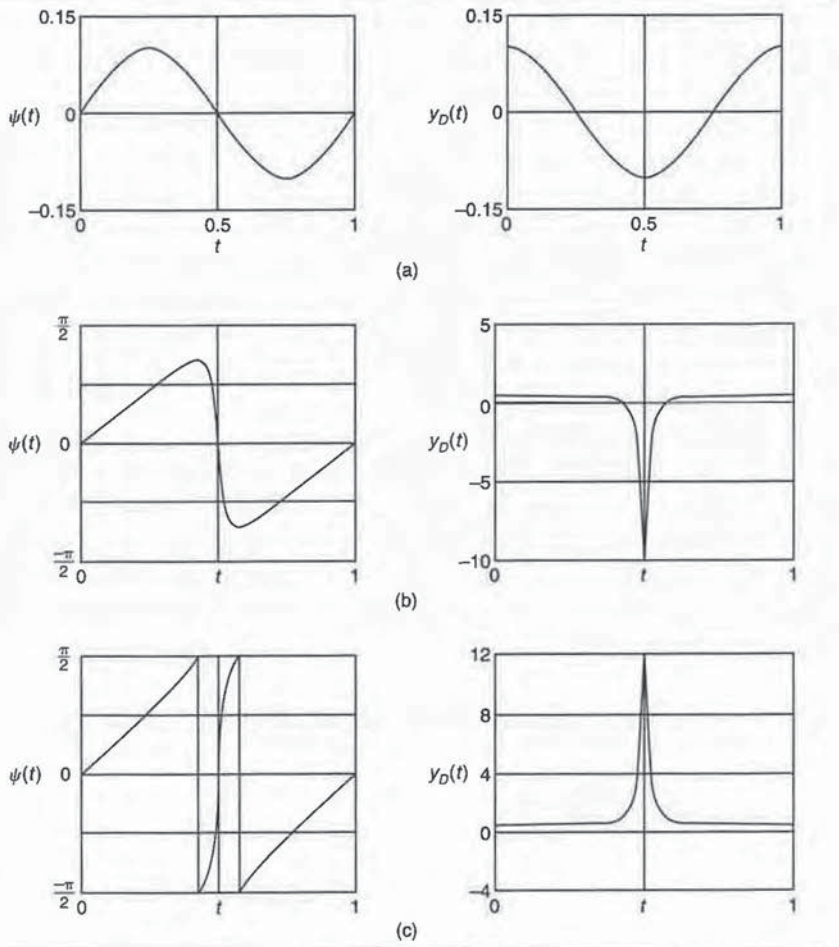
Solving for $\psi(t)$ and differentiating gives the discriminator output

$$y_D(t) \approx K_D \frac{A_i}{A_i - A_c} f_i \tag{3.197}$$

Note that this is a positive quantity for $f_i > 0$ and a negative quantity for $f_i < 0$.

The phase deviation and discriminator output waveforms are shown in Figure 3.42 for $A_i = 0.1A_c$, $A_i = 0.9A_c$, and $A_i = 1.1A_c$. Figure 3.42(a) illustrates that for small A_i the phase deviation and the discriminator output are nearly sinusoidal as predicted by the results of the small interference analysis given in (3.184) and (3.185). For $A_i = 0.9A_c$, we see that we have a negative

FIGURE 3.42 Phase deviation and discriminator outputs due to interference. (a) Phase deviation and discriminator output for $A_i = 0.2A_c$. (b) Phase deviation and discriminator output for $A = 0.9A_c$. (c) Phase deviation and discriminator output for $A = 1.1A_c$.



spike at the discriminator output as predicted by (3.195). For $A_c = 1.1A_c$, we have a positive spike at the discriminator output as predicted by (3.197).

Note that for $A_i > A_c$ the origin of the phasor diagram is encircled as $\theta(t)$ goes from 0 to 2π. In other words $\psi(t)$ goes from 0 to 2π as $\theta(t)$ goes from 0 to 2π. The origin is not encircled if $A_i < A_c$. Thus the integral

$$\int_T \left(\frac{d\psi}{dt}\right) dt = \begin{cases} 2\pi, & A_i > A_c \\ 0, & A_i < A_c \end{cases} \tag{3.198}$$

where T is the time required for $\theta(t)$ to go from $\theta(t) = 0$ to $\theta(t) = 2\pi$. In other words $T = 2\pi/\omega_i$. Thus the area under the discriminator output curve is 0 for parts (a) and (b) of Figure 3.42 and $2\pi K_D$ for the discriminator output curve in Figure 3.42(c).

The origin encirclement phenomenon will be revisited in Chapters 5 and 6. An understanding of the results presented here will prove beneficial when we examine noise effects in FM demodulation.

For operation above threshold $A_i \ll A_c$, the severe effect of interference on FM for large f_i can be reduced by placing a filter, called a *deemphasis filter,* at the FM discriminator output. This filter is typically a simple RC lowpass filter with a 3-dB frequency considerably less than the modulation bandwidth W. The deemphasis filter effectively reduces the interference for large f_i, as shown in Figure 3.43. For large frequencies, the magnitude of the transfer function of a first-order filter is approximately $1/f$. Since the amplitude of the interference increases linearly with f_i for FM, the output is constant for large f_i, as shown in Figure 3.43.

Since $f_3 < W$, the deemphasis filter distorts the message signal in addition to combating interference. The distortion can be avoided by passing the message through a *preemphasis filter* that has a transfer function equal to the reciprocal of the transfer function of the deemphasis filter. Since the transfer function of the cascade combination of the preemphasis and deemphasis filters is unity, there is no detrimental effect on the modulation. This yields the system shown in Figure 3.44.

FIGURE 3.43 Amplitude of discriminator output due to interference

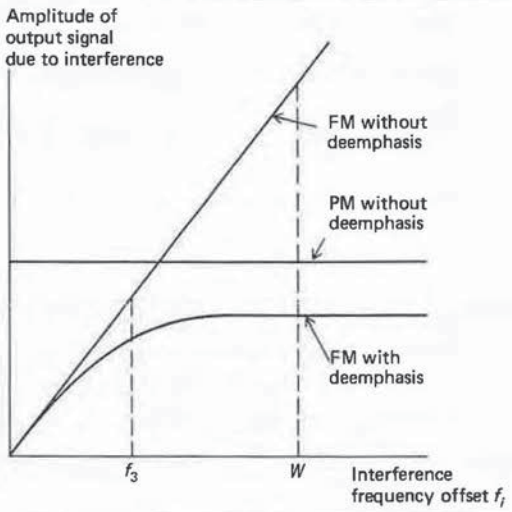
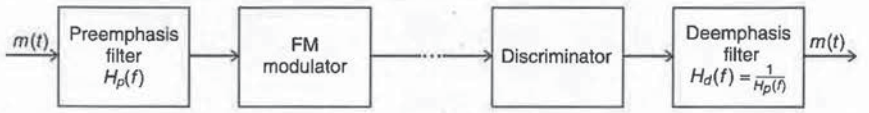


FIGURE 3.44 FM system with preemphasis and deemphasis



The improvement offered by the use of preemphasis and deemphasis is not gained without a price. The preemphasis filter amplifies the high-frequency components, and this results in increased deviation. The end result is an increase in bandwidth. However, many signals of practical interest have little power in the higher-frequency components, so the increased deviation resulting from the use of preemphasis is negligible. We shall see in Chapter 6, when noise is studied, that the use of preemphasis and deemphasis often provides sufficient improvement to be definitely worth the price.

3.4 FEEDBACK DEMODULATORS

We have already studied the technique of FM to AM conversion for demodulating an angle-modulated signal. We shall see in Chapter 6 that improved performance in the presence of noise can be gained by utilizing a feedback demodulator. In this section we examine the basic operation of several types of feedback demodulators. Such systems are widely used in today's communication systems not only because of their superior performance but also because of their ease of implementation using inexpensive integrated circuits. The phase-lock loop is the basic building block.

Phase-Lock Loops for FM Demodulation

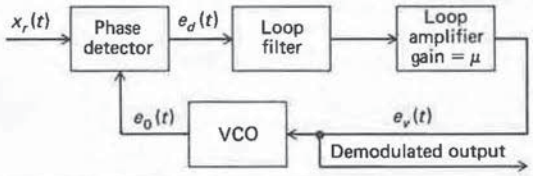
A block diagram of a phase-lock loop (PLL) is shown in Figure 3.45. Such a system generally contains four basic elements:

1. A phase detector
2. A loop filter
3. A loop amplifier
4. A voltage-controlled oscillator (VCO)

In order to understand the operation of a PLL, assume that the input signal is given by

$$x_r(t) = A_c \cos [\omega_c t + \phi(t)] \tag{3.199}$$

FIGURE 3.45 Phase-lock loop



and that the VCO output signal is given by

$$e_0(t) = A_v \sin [\omega_c t + \theta(t)] \tag{3.200}$$

There are many different types of phase detectors, all having different operating properties. For our application, we assume that the phase detector is a multiplier followed by a lowpass filter to remove the second harmonic of the carrier. We also assume that an inverter is present to remove the minus sign resulting from the multiplication. With these assumptions, the output of the phase detector becomes

$$e_d(t) = \tfrac{1}{2} A_c A_v K_d \sin [\phi(t) - \theta(t)] \tag{3.201}$$

where K_d is the constant associated with the multiplier in the phase detector.

The output of the phase detector is filtered, amplified, and applied to the VCO. A VCO is essentially a frequency modulator, the frequency deviation of the output, $d\theta/dt$, being proportional to the input signal. In other words,

$$\frac{d\theta}{dt} = K_v e_v(t) \quad \text{rad/s} \tag{3.202}$$

which yields

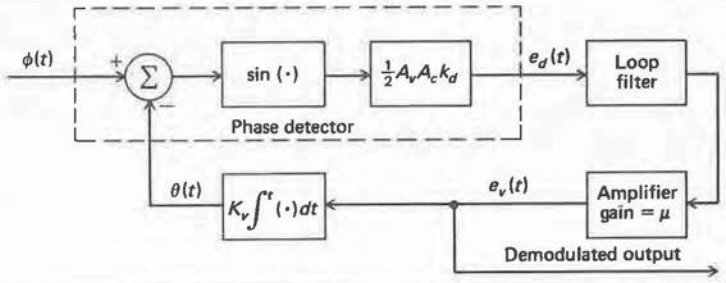
$$\theta(t) = K_v \int^t e_v(\alpha)\, d\alpha \tag{3.203}$$

The parameter K_v is known as the *VCO constant* and is measured in radians per second per unit of input.

Since the output of the phase detector is determined by the phase deviation of the PLL input and VCO output, we can model a PLL without regard to the carrier frequency ω_c. Such a model is shown in Figure 3.46. This model is known as the *nonlinear model* because of the sinusoidal nonlinearity in the phase detector. When the PLL is operating *in lock,* the VCO phase $\theta(t)$ is a good estimate of the input-phase deviation $\phi(t)$. For this mode of operation, the phase error $\phi(t) - \theta(t)$ is small, and

$$\sin [\phi(t) - \theta(t)] \cong \phi(t) - \theta(t) \tag{3.204}$$

FIGURE 3.46 Nonlinear PLL model



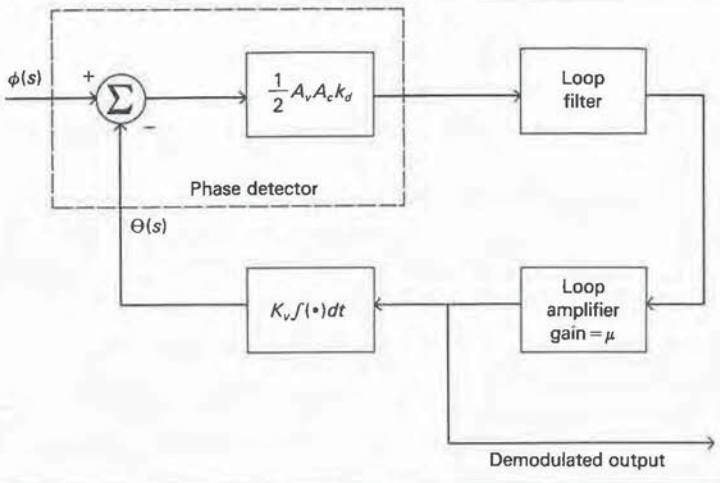
Thus, if the phase error is small, the sinusoidal nonlinearity can be neglected and the PLL becomes a linear feedback control system, which is easily analyzed. The linear model that results is illustrated in Figure 3.47. While both the nonlinear and linear models involve $\theta(t)$ and $\phi(t)$ rather than $x_r(t)$ and $e_0(t)$, knowledge of $\theta(t)$ and $\phi(t)$ fully determines $x_r(t)$ and $e_0(t)$, as can be seen from (3.199) and (3.200).

If $\theta(t) \cong \phi(t)$, it follows that

$$\frac{d\theta(t)}{dt} \cong \frac{d\phi(t)}{dt} \tag{3.205}$$

and the VCO frequency deviation is a good estimate of the input frequency deviation. For an FM system, the input frequency deviation is proportional

FIGURE 3.47 Linear PLL model



to the message signal $m(t)$. Since the VCO frequency deviation is proportional to the VCO input $e_v(t)$, it follows that $e_v(t)$ is proportional to $m(t)$ if (3.205) is satisfied. Thus $e_v(t)$ is the demodulated output for FM systems.

We will now show that the phase error signal tends to drive the PLL into lock. In order to simplify the analysis, we assume that the loop filter is replaced by a short circuit. For this case,

$$e_v(t) = \tfrac{1}{2}\mu A_c A_v K_d \sin [\phi(t) - \theta(t)] \tag{3.206}$$

which yields

$$\theta(t) = K_t \int^t \sin [\phi(\alpha) - \theta(\alpha)]\, d\alpha \tag{3.207}$$

where K_t is the total effective loop gain, which is

$$K_t = \tfrac{1}{2}\mu A_c A_v K_d K_v \tag{3.208}$$

The expression for $\theta(t)$ can be differentiated to yield

$$\frac{d\theta}{dt} = K_t \sin [\phi(t) - \theta(t)] \tag{3.209}$$

Assume that the input to the FM modulator is a unit step so that the frequency deviation $d\phi/dt$ is a unit step of magnitude $\Delta\omega$. Let the phase error $\phi(t) - \theta(t)$ be denoted $\psi(t)$. This yields

$$\frac{d\theta}{dt} = \frac{d\phi}{dt} - \frac{d\psi}{dt} = \Delta\omega - \frac{d\psi}{dt} = K_t \sin \psi(t), \qquad t \geq 0 \tag{3.210}$$

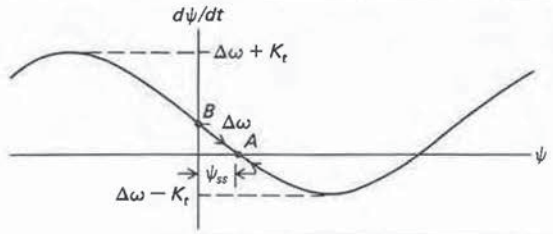
or

$$\frac{d\psi}{dt} + K_t \sin \psi(t) = \Delta\omega \tag{3.211}$$

This equation is sketched in Figure 3.48. It relates the frequency error and the phase error.

A plot of the derivative of a function versus the function is known as a *phase-plane plot* and tells us much about the operation of a nonlinear system. The PLL must operate with a phase error $\psi(t)$ and a frequency error $d\psi/dt$ that are consistent with (3.211). To demonstrate that the PLL achieves lock, assume that it is operating with zero phase and frequency errors prior to the application of the frequency step. When the step in frequency is applied, the frequency error becomes $\Delta\omega$. This establishes the initial operating point, point B in Figure 3.46, assuming $\Delta\omega > 0$. In order to determine the trajectory of the operating point, we need only recognize that since dt, a time

FIGURE 3.48 Phase-plane plot



increment, is always a positive quantity, $d\psi$ must be positive if $d\psi/dt$ is positive. Thus, in the upper half plane, ψ increases. In other words, the operating point moves from left to right in the upper half plane. In the same manner, the operating point moves from right to left in the lower half plane, the region for which $d\psi/dt$ is less than zero. Thus the operating point must move from point B to point A. When the operating point attempts to move from point A by a small amount, it is forced back to point A. Thus point A is a stable operating point and is the steady-state operating point of the system. The steady-state phase error is ψ_{ss}, and the steady-state frequency error is zero.

The preceding analysis illustrates that the loop locks only if there is an intersection of the operating curve with the $d\psi/dt = 0$ axis. Thus, if the loop is to lock, $\Delta\omega$ must be less than K_t. For this reason, K_t is shown as the *lock range* for the first-order PLL.

The phase-plane plot for a first-order PLL with a frequency-step input is illustrated in Figure 3.49. The loop gain is $2\pi(50)$, and four values for the frequency step are shown: $\Delta f = 12$, 24, 48, and 55 Hz. The steady-state phase errors are indicated by A, B, and C for frequency-step values of 12, 24, and 48 Hz, respectively. For $\Delta f = 55$, the loop does not lock, as shown.

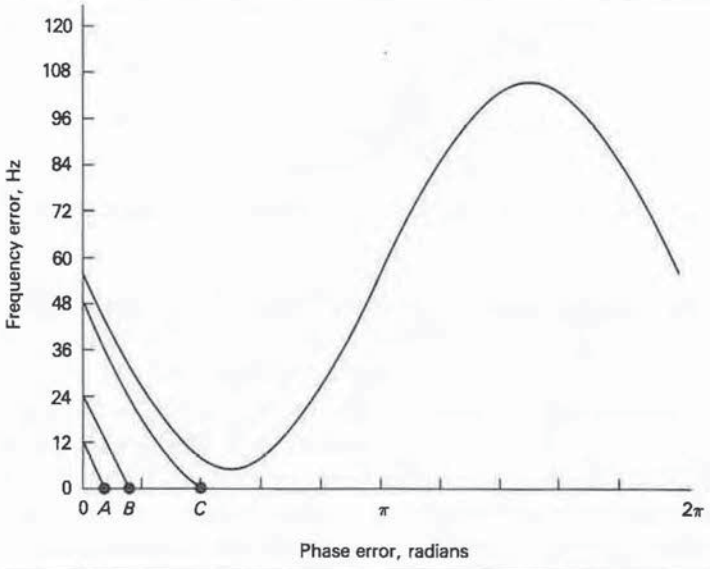
From the previous paragraphs we saw that the steady-state phase errors resulting from changes in the input frequency can be made as small as desired by increasing the total loop gain K_t. For $K_t \gg \Delta\omega$, the phase errors will be small and the linear model can be used. The relationship between $\theta(t)$ and $\phi(t)$ is, from (3.207) with the linear approximation made,

$$\theta(t) = K_t \int^t [\phi(\alpha) - \theta(\alpha)]\, d\alpha \tag{3.212}$$

which yields the differential equation

$$\frac{d\theta(t)}{dt} + K_t\theta(t) = K_t\phi(t) \tag{3.213}$$

FIGURE 3.49 Phase-plane plot of first-order PLL for several initial frequency errors



The loop transfer function, in terms of the Laplace transform variable s, is

$$H(s) = \frac{\Theta(s)}{\Phi(s)} = \frac{K_t}{s + K_t} \tag{3.214}$$

The unit impulse response is therefore

$$h(t) = K_t e^{-K_t t} u(t) \tag{3.215}$$

The limit of $h(t)$ as the loop gain K_t tends to infinity satisfies all properties of the delta function. Therefore,

$$\lim_{K_t \to \infty} K_t e^{-K_t t} u(t) = \delta(t) \tag{3.216}$$

which proves that for large loop gain $\theta(t) = \phi(t)$.

The PLL also can be used as a demodulator for phase-modulated signals by integrating the VCO input. Since the VCO input signal is proportional to the frequency deviation of the PLL input, the integral of this signal is proportional to the phase deviation of the PLL input.

We also have seen that operation of the PLL as a discriminator requires a large loop gain. This implies a large bandwidth. In general, very large values of loop gain cannot be used in practical applications without difficulty. How-

ever, the use of appropriate loop filters allows good performance to be achieved with reasonable values of loop gain and bandwidth. These filters make the analysis more complicated than our simple example, as we shall soon see.

There is one final observation we should make. The expression for loop gain shows that it is a function of the amplitude of the input signal. This means that a PLL must be designed for a given signal level. If that signal level changes, a new design may be necessary. In most practical applications, the dependence on loop gain is removed by placing a limiter on the loop input.

▼ **EXAMPLE 3.9** The input to an FM modulator is $m(t) = Au(t)$. The resulting modulated carrier

$$x_c(t) = A_c \cos\left[\omega_c t + k_f A \int^t u(\alpha)\, d\alpha\right] \tag{3.217}$$

is to be demodulated using a first-order PLL. Determine the demodulated output.

SOLUTION This problem will be solved using linear analysis and the Laplace transform. The Laplace transform of the loop transfer function, Equation (3.214), is

$$\frac{\Theta(s)}{\Phi(s)} = \frac{K_t}{K_t + s} \tag{3.218}$$

The phase deviation $\phi(t)$ is

$$\phi(t) = Ak_f \int^t u(\alpha)\, d\alpha \tag{3.219}$$

which yields

$$\Phi(s) = \frac{Ak_f}{s^2} \tag{3.220}$$

This gives

$$\Theta(s) = \frac{AK_t k_f}{s^2(s + K_t)} \tag{3.221}$$

The Laplace transform of the defining equation of the VCO, Equation (3.203), yields

$$E_v(s) = \frac{s}{K_v}\Theta(s) \tag{3.222}$$

so that

$$E_v(s) = \frac{Ak_f}{K_v}\frac{K_t}{s(s + K_t)} \tag{3.223}$$

Partial fraction expansion yields

$$E_v(s) = \frac{Ak_f}{K_v}\left(\frac{1}{s} - \frac{1}{s + K_t}\right) \tag{3.224}$$

Thus the demodulated output is given by

$$e_v(t) = \frac{Ak_f}{K_v}(1 - e^{-K_t t})u(t) \tag{3.225}$$

which is sketched in Figure 3.50 for k_f equal to K_v.

As the loop gain K_t is increased, the demodulated output clearly becomes a closer approximation to $m(t)$. It should be remembered that we have assumed that linear analysis is valid. This requires that the phase error be small. The validity of this assumption is easily checked by solving for the phase error, an exercise that is left to the problems. ▲

The first-order PLL has a number of drawbacks that limit its use for most applications. Among these drawbacks are the limited lock range and the nonzero steady-state phase error to a step-frequency input. Both these problems can be solved by using a second-order PLL, which is obtained by using a loop filter of the form

$$F(s) = \frac{s + a}{s} \tag{3.226}$$

FIGURE 3.50 Demodulated output for $m(t) = Au(t)$

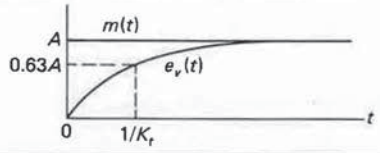
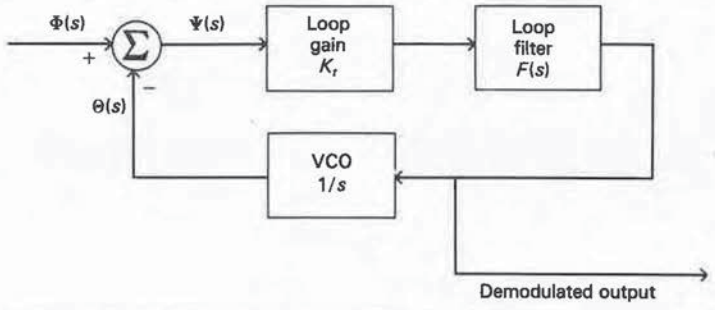


FIGURE 3.51 Linear PLL model



This choice of loop filter results in what is generally referred to as a *perfect second-order PLL.*

If the phase error is small, the linear model applies and the transfer function of the second-order PLL is easily derived. The linear model is redrawn in Figure 3.51 in terms of the Laplace transform variable s. From Figure 3.51, it follows that

$$\Theta(s) = \frac{K_t F(s)}{s}[\Phi(s) - \Theta(s)] \tag{3.227}$$

so that the PLL transfer function is

$$H(s) = \frac{\Theta(s)}{\Phi(s)} = \frac{K_t F(s)}{s + K_t F(s)} \tag{3.288}$$

With $F(s)$ given by (3.226), the transfer function becomes

$$\frac{\Theta(s)}{\Phi(s)} = \frac{K_t(s + a)}{s^2 + K_t s + K_t a} \tag{3.229}$$

We also can write the relationship between the phase error $\Psi(s)$ and the input phase $\Phi(s)$ from Figure 3.51. The result is

$$\frac{\Psi(s)}{\Phi(s)} = \frac{s}{s + K_t F(s)} \tag{3.230}$$

which for the loop filter defined by (3.226) yields

$$\frac{\Psi(s)}{\Phi(s)} = \frac{s^2}{s^2 + K_t a s + K_t a} \tag{3.231}$$

We now place this in the standard form for a second-order system. The result is

$$\frac{\Psi(s)}{\Phi(s)} = \frac{s^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{3.232}$$

in which ζ is the damping factor and ω_n is the natural frequency. It follows that the natural frequency is

$$\omega_n = \sqrt{K_t a} \tag{3.233}$$

and that the damping factor is

$$\zeta = \frac{1}{2}\sqrt{\frac{K_t}{a}} \tag{3.234}$$

A typical value of the damping factor is 0.707.

Using the linear model, we can easily show that the steady-state phase error due to a step in the input frequency is zero. Since phase is the integral of frequency and integration is equivalent to division by s, the input phase due to a step in frequency of magnitude $\Delta\omega$ is

$$\Phi(s) = \frac{\Delta\omega}{s^2} \tag{3.235}$$

From (3.232) we see that the Laplace transform of the phase error $\psi(t)$ is

$$\psi(s) = \frac{\Delta\omega}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{3.236}$$

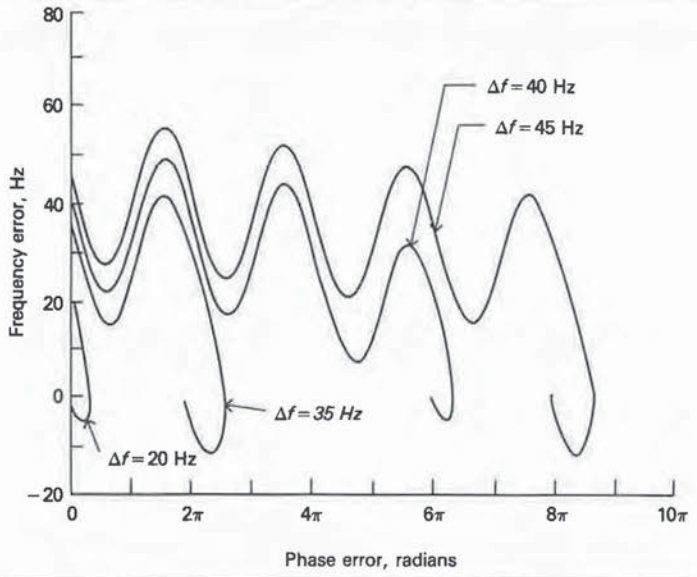
Inverse transforming yields, for $\zeta < 1$,

$$\psi(t) = \frac{\Delta\omega}{\omega_n\sqrt{1 - \zeta^2}}\, e^{-\zeta\omega_n t} \sin\left(\omega_n\sqrt{1 - \zeta^2}\, t\right) \tag{3.237}$$

and we see that $\psi(t) \to 0$ as $t \to \infty$. Thus the steady-state phase error is zero. We now turn our attention to the lock range by examining the phase-plane plot of the second-order PLL.

A mathematical development of the phase-plane plot of a second-order PLL is well beyond the level of our treatment here. However, the phase-plane plot was obtained, using computer simulation, for a second-order PLL having a damping factor ζ of 0.707 and a natural frequency f_n of 10 Hz. For these parameters, the loop gain K_t is 88.9 and the filter parameter a is 44.4. The input to the PLL is assumed to be a step change in frequency at

FIGURE 3.52 Phase-plane plot for second-order PLL



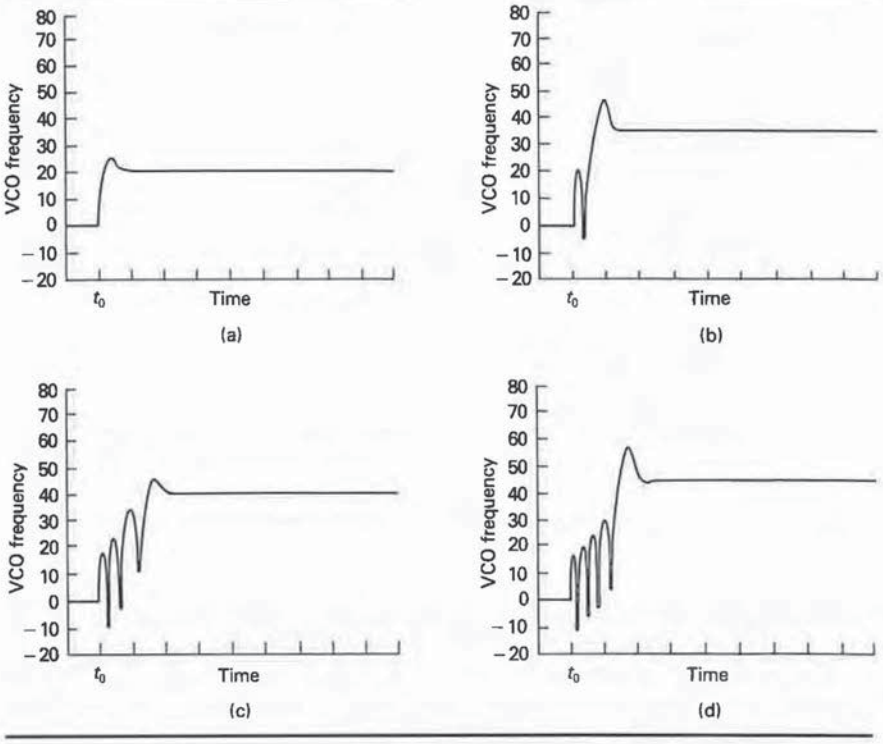
time $t = t_0$. Four values were used for the step change in frequency $\Delta\omega = 2\pi(\Delta f)$. These were $\Delta f = 20$, 35, 40, and 45 Hz.

The results are illustrated in Figure 3.52. Note that for $\Delta f = 20$ Hz, the operating point returns to a steady-state value for which the frequency and phase error are both zero. For $\Delta f = 35$ Hz, the phase-plane is somewhat more complicated. The steady-state frequency error is zero, but the steady-state phase error is 2π rad. We say that the PLL has slipped one cycle. The cycle-slipping phenomenon accounts for the nonzero steady-state phase error. The responses for $\Delta f = 40$ and 45 Hz illustrate that three and four cycles are slipped, respectively. The instantaneous VCO frequency is shown in Figure 3.53 for these four cases. The cycle-slipping behavior is clearly shown.

The second-order PLL does indeed have an infinite lock range, and cycle slipping occurs until the phase error is within π rad of the steady-state value.

Phase-lock loops also allow for simple implementation of frequency multipliers and dividers. There are two basic schemes. In the first scheme, harmonics of the input are generated and the VCO tracks one of these harmonics. This scheme is most useful for implementing frequency multipliers. The second scheme is to generate harmonics of the VCO output and to phase-lock one of these frequency components to the input. This scheme

FIGURE 3.53 VCO frequency for four values of input frequency step. (a) VCO frequency for $\Delta f = 20$ Hz. (b) VCO frequency for $\Delta f = 35$ Hz. (c) VCO frequency for $\Delta f = 40$ Hz. (d) VCO frequency for $\Delta f = 45$ Hz.



can be used to implement either frequency multipliers or frequency dividers.

Figure 3.54 illustrates the first technique. The limiter is a nonlinear device and therefore generates harmonics of the input frequency. If the input is sinusoidal, the output of the limiter is a squarewave; therefore, odd harmonics are present. In the example illustrated, the VCO quiescent frequency [VCO output frequency f_c with $e_v(t)$ equal to zero] is set equal to $5f_0$. The result is that the VCO phase-locks to the fifth harmonic of the input. Thus the system shown multiplies the input frequency by 5.

Figure 3.55 illustrates frequency division by a factor of 2. The VCO quiescent frequency is $\frac{1}{2}f_0$ Hz, but the VCO output waveform is a narrow pulse that has the spectrum shown. The component at frequency f_0 phase-locks to the input. A bandpass filter can be used to select the component desired from the VCO output spectrum. For the example shown, the center fre-

FIGURE 3.54 Phase-lock loop used as a frequency multiplier

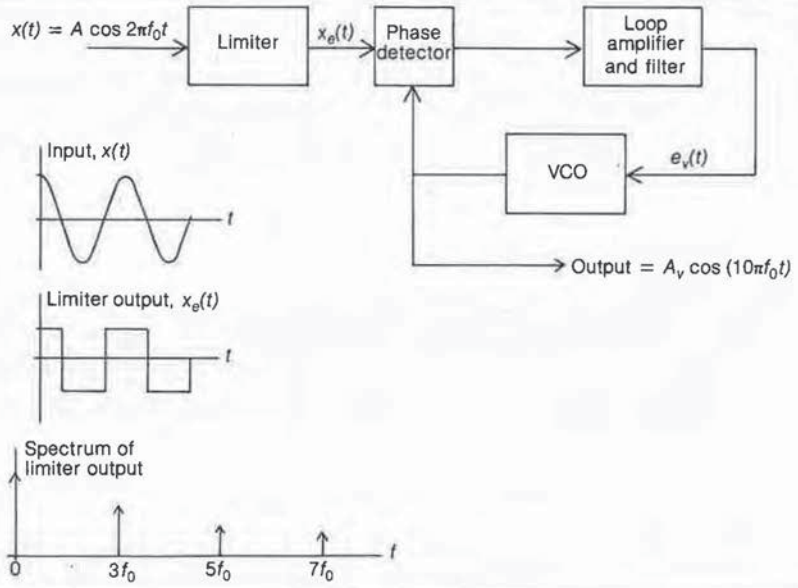
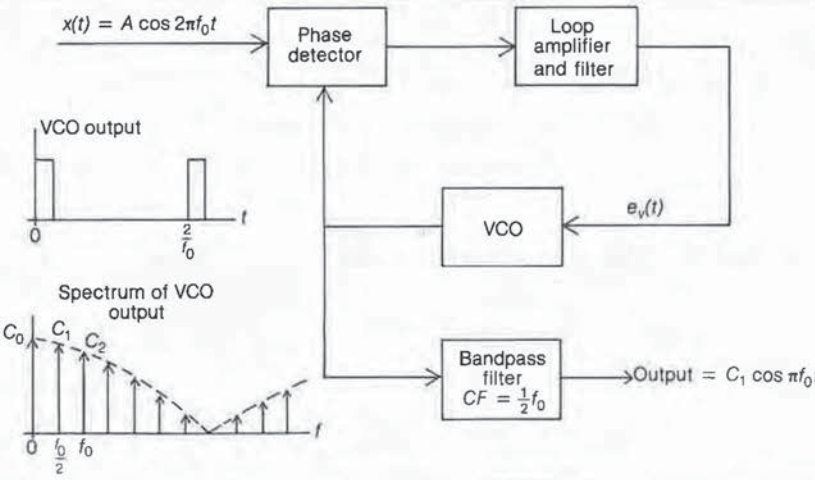


FIGURE 3.55 Phase-lock loop used as a frequency divider



quency of the bandpass filter should be $\frac{1}{2}f_0$. The bandwidth of the bandpass filter must be less than the spacing between the components in the VCO output spectrum; in this case, this spacing is $\frac{1}{2}f_0$. It is worth noting that the system shown in Figure 3.55 could be used to multiply the input frequency by 5 by setting the center frequency of the bandpass filter to $5f_0$. Thus this system could also serve as a $\times 5$ frequency multiplier, like the first example. Many variations are possible.

Frequency-Compressive Feedback

Another system that can be used to demodulate FM signals is the frequency-compressive feedback demodulator, which is illustrated in Figure 3.56. The system is similar to a PLL except for the bandpass filter and discriminator in the loop and the offset in the VCO center frequency. As in the case of the PLL, assume that the input is

$$x_r(t) = A_c \cos [\omega_c t + \phi(t)] \tag{3.238}$$

Also assume that the center frequency of the VCO is $\omega_c - \omega_0$. The frequency ω_0 is the VCO offset frequency and the center frequency of the discriminator. The VCO output is given by

$$e_0(t) = A_v \sin [(\omega_c - \omega_0)t + \theta(t)] \tag{3.239}$$

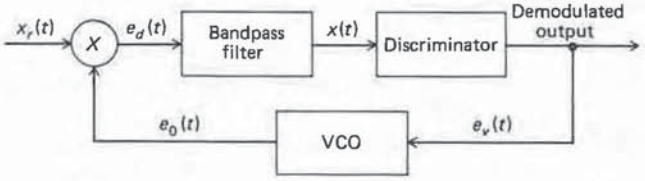
The quantity $\theta(t)$ is the phase deviation of the VCO and is related to the VCO input $e_v(t)$:

$$\theta(t) = K_v \int^t e_v(\alpha)\, d\alpha \tag{3.240}$$

where K_v is the VCO constant. It should be noted that except for the VCO offset, the expressions for $x_r(t)$, $e_0(t)$, and $\theta(t)$ are identical to the equivalent expressions for the PLL. The multiplier output $e_d(t)$ is given by

$$\begin{aligned} e_d(t) &= \tfrac{1}{2}A_c A_v \sin [(2\omega_c - \omega_0)t + \phi(t) + \theta(t)] \\ &\quad - \tfrac{1}{2}A_c A_v \sin [\omega_0 t + \phi(t) - \theta(t)] \end{aligned} \tag{3.241}$$

FIGURE 3.56 Frequency-compressive feedback demodulator



The phase deviation of the discriminator input is

$$\phi(t) - K_v \int^t e_v(\alpha)\, d\alpha \tag{3.242}$$

Thus the discriminator output can be written as

$$e_v(t) = \frac{1}{2\pi} K_D \frac{d}{dt}\left[\phi(t) - K_v \int^t e_v(\alpha)\, d\alpha\right] \tag{3.243}$$

which is

$$e_v(t) = \frac{(1/2\pi)K_D}{1 + (1/2\pi)K_vK_D}\frac{d\phi}{dt} = \frac{K_D f_d}{1 + (1/2\pi)K_vK_D} m(t) \tag{3.244}$$

Thus the system is an FM demodulator.

The advantage of this technique can be understood by using (3.244) to write an expression for $x(t)$. This yields

$$x(t) = -\tfrac{1}{2}A_cA_v \sin\left[\omega_0 t + \frac{1}{1 + (1/2\pi)K_DK_v}\phi(t)\right] \tag{3.245}$$

This shows that for large values of the product K_DK_v, the phase deviation can be made small, thereby reducing significantly the bandwidth of the signal at the discriminator input. It is even possible to compress the bandwidth of a wideband FM signal to that of a narrowband FM signal at the input to the discriminator. It is this bandwidth compression that gives the scheme the name *frequency-compressive feedback*. The advantages of this bandwidth compression will be examined in Chapter 6.

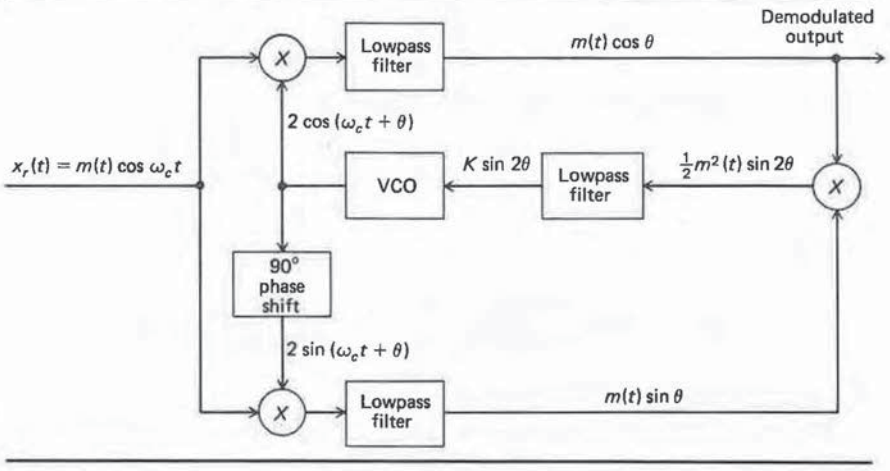
Costas Phase-Lock Loops

We have seen that systems utilizing feedback can be used to demodulate angle-modulated carriers. A feedback system also can be used to generate the coherent demodulation carrier necessary for the demodulation of DSB signals. One system that accomplishes this is the Costas phase-lock loop illustrated in Figure 3.57. The input to the loop is the assumed DSB signal

$$x_r(t) = m(t) \cos \omega_c t \tag{3.246}$$

The signals at the various points within the loop are easily derived from the assumed input and VCO output and are included in Figure 3.57. The lowpass filter preceding the VCO is assumed sufficiently narrow so that the output is $K \sin 2\theta$, essentially the dc value of the input. This signal drives the VCO such that θ is reduced. For sufficiently small θ, the output of the top lowpass filter is the demodulated output.

FIGURE 3.57 Costas phase-lock loop



We will see in Chapter 7 that the Costas phase-lock loop is useful in the implementation of digital data systems.

3.5 PULSE MODULATION

In Section 2.8 we saw that continuous bandlimited signals can be represented by a sequence of discrete samples and that the continuous signal can be reconstructed with negligible error if the sampling rate is sufficiently high. Consideration of sampled signals immediately leads us to the topic of pulse modulation.

Analog Pulse Modulation

Analog pulse modulation results when some characteristic of a pulse is made to vary in one-to-one correspondence with the message signal. We shall see that since a pulse is characterized by three quantities—amplitude, width, and position—there are three types of analog pulse modulation. These are pulse-amplitude modulation, pulse-width modulation, and pulse-position modulation.

Pulse-Amplitude Modulation (PAM) As illustrated in Figure 3.58, a PAM waveform consists of a sequence of flat-topped pulses. The amplitude of each pulse corresponds to the value of the message signal $m(t)$ at the leading edge of the pulse. Thus this type of modulation is essentially a sampling