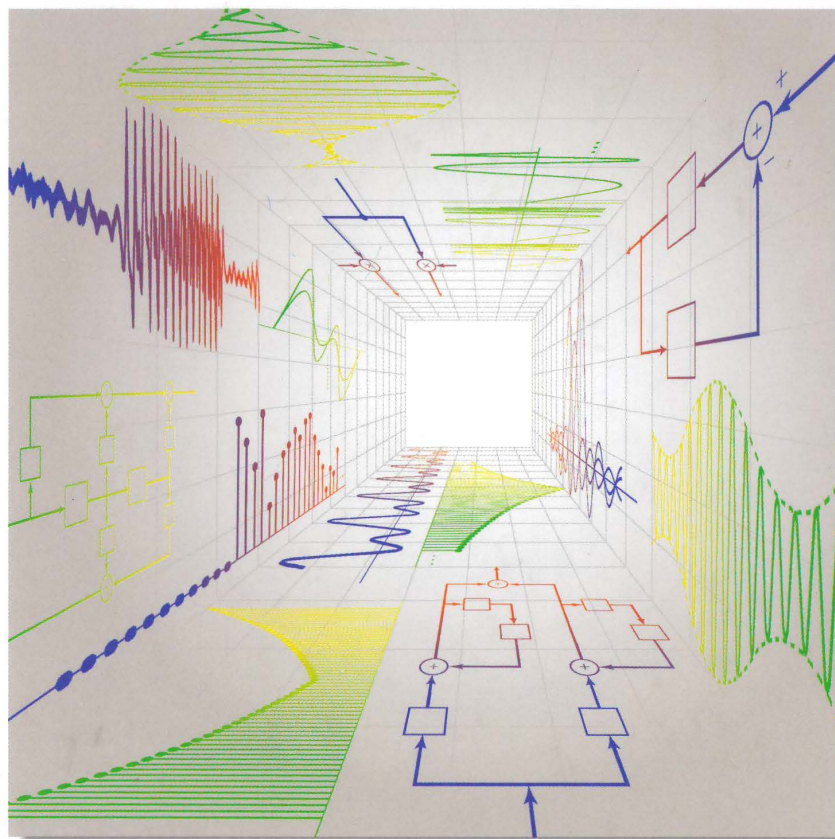


SIGNALS & SYSTEMS

SECOND EDITION



ALAN V. OPPENHEIM
ALAN S. WILLSKY
WITH S. HAMID NAWAB

PRENTICE HALL SIGNAL PROCESSING SERIES
ALAN V. OPPENHEIM, SERIES EDITOR



SIGNALS & SYSTEMS

SECOND EDITION

SIGNALS & SYSTEMS

ALAN V. OPPENHEIM

ALAN S. WILLSKY

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

WITH

S. HAMID NAWAB

BOSTON UNIVERSITY



PRENTICE HALL

UPPER SADDLE RIVER, NEW JERSEY 07458

Library of Congress Cataloging-in-Publication Data

Oppenheim, Alan V.

Signals and systems / Alan V. Oppenheim, Alan S. Willsky, with
S. Hamid Nawab. — 2nd ed..

p. cm. — Prentice-Hall signal processing series

Includes bibliographical references and index.

ISBN 0-13-814757-4

1. System analysis. 2. Signal theory (Telecommunication)

I. Willsky, Alan S. II. Nawab, Syed Hamid. III. Title.

IV. Series.

QA402.O63 1996

621.382'23—dc20

96-19945

CIP

Acquisitions editor: **Tom Robbins**

Production service: **TKM Productions**

Editorial/production supervision: **Sharyn Vitrano**

Copy editor: **Brian Baker**

Interior and cover design: **Patrice Van Acker**

Art director: **Amy Rosen**

Managing editor: **Bayani Mendoza DeLeon**

Editor-in-Chief: **Marcia Horton**

Director of production and manufacturing: **David W. Riccardi**

Manufacturing buyer: **Donna Sullivan**

Editorial assistant: **Phyllis Morgan**



© 1997 by Alan V. Oppenheim and Alan S. Willsky

© 1983 by Alan V. Oppenheim, Alan S. Willsky, and Ian T. Young

Published by Prentice-Hall, Inc.

Simon & Schuster / A Viacom Company

Upper Saddle River, New Jersey 07458

All rights reserved. No part of this book may be
reproduced, in any form or by any means,
without permission in writing from the publisher.

The author and publisher of this book have used their best efforts in preparing this book. These efforts include the development, research, and testing of the theories and programs to determine their effectiveness. The author and publisher make no warranty of any kind, expressed or implied, with regard to these programs or the documentation contained in this book. The author and publisher shall not be liable in any event for incidental or consequential damages in connection with, or arising out of, the furnishing, performance, or use of these programs.

Printed in the United States of America

10 9 8 7

ISBN 0-13-814757-4

Prentice-Hall International (UK) Limited, London

Prentice-Hall of Australia Pty. Limited, Sydney

Prentice-Hall Canada Inc., Toronto

Prentice-Hall Hispanoamericana, S.A., Mexico

Prentice-Hall of India Private Limited, New Delhi

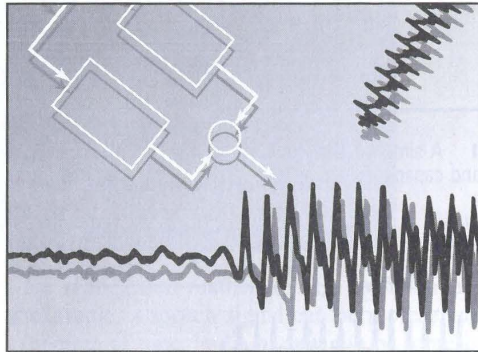
Prentice-Hall of Japan, Inc., Tokyo

Simon & Schuster Asia Pte. Ltd., Singapore

Editora Prentice-Hall do Brasil, Ltda., Rio de Janeiro

1

SIGNALS AND SYSTEMS



1.0 INTRODUCTION

As described in the Foreword, the intuitive notions of signals and systems arise in a rich variety of contexts. Moreover, as we will see in this book, there is an analytical framework—that is, a language for describing signals and systems and an extremely powerful set of tools for analyzing them—that applies equally well to problems in many fields. In this chapter, we begin our development of the analytical framework for signals and systems by introducing their mathematical description and representations. In the chapters that follow, we build on this foundation in order to develop and describe additional concepts and methods that add considerably both to our understanding of signals and systems and to our ability to analyze and solve problems involving signals and systems that arise in a broad array of applications.

1.1 CONTINUOUS-TIME AND DISCRETE-TIME SIGNALS

1.1.1 Examples and Mathematical Representation

Signals may describe a wide variety of physical phenomena. Although signals can be represented in many ways, in all cases the information in a signal is contained in a pattern of variations of some form. For example, consider the simple circuit in Figure 1.1. In this case, the patterns of variation over time in the source and capacitor voltages, v_s and v_c , are examples of signals. Similarly, as depicted in Figure 1.2, the variations over time of the applied force f and the resulting automobile velocity v are signals. As another example, consider the human vocal mechanism, which produces speech by creating fluctuations in acoustic pressure. Figure 1.3 is an illustration of a recording of such a speech signal, obtained by

1

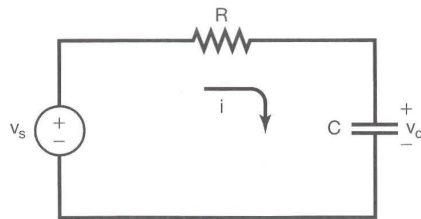


Figure 1.1 A simple RC circuit with source voltage v_s and capacitor voltage v_c .

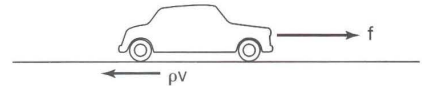


Figure 1.2 An automobile responding to an applied force f from the engine and to a retarding frictional force ρv proportional to the automobile's velocity v .

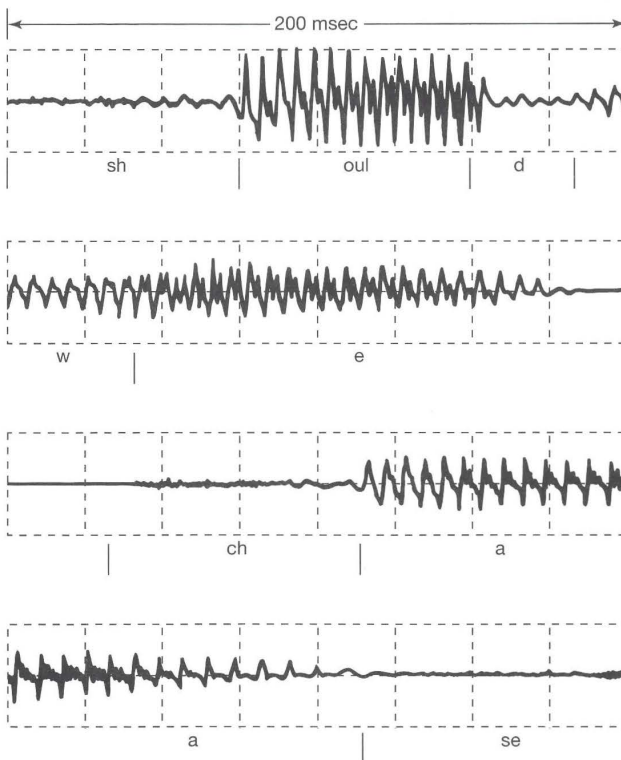


Figure 1.3 Example of a recording of speech. [Adapted from *Applications of Digital Signal Processing*, A.V. Oppenheim, ed. (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1978), p. 121.] The signal represents acoustic pressure variations as a function of time for the spoken words "should we chase." The top line of the figure corresponds to the word "should," the second line to the word "we," and the last two lines to the word "chase." (We have indicated the approximate beginnings and endings of each successive sound in each word.)

using a microphone to sense variations in acoustic pressure, which are then converted into an electrical signal. As can be seen in the figure, different sounds correspond to different patterns in the variations of acoustic pressure, and the human vocal system produces intelligible speech by generating particular sequences of these patterns. Alternatively, for the monochromatic picture, shown in Figure 1.4, it is the pattern of variations in brightness across the image that is important.

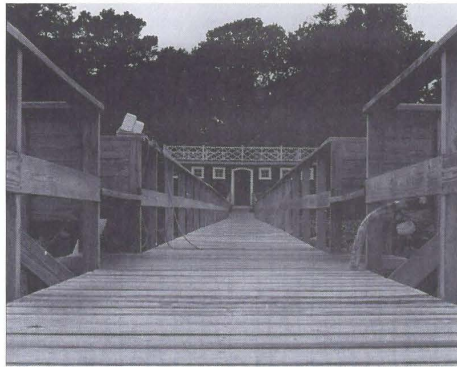


Figure 1.4 A monochromatic picture.

Signals are represented mathematically as functions of one or more independent variables. For example, a speech signal can be represented mathematically by acoustic pressure as a function of time, and a picture can be represented by brightness as a function of two spatial variables. In this book, we focus our attention on signals involving a single independent variable. For convenience, we will generally refer to the independent variable as time, although it may not in fact represent time in specific applications. For example, in geophysics, signals representing variations with depth of physical quantities such as density, porosity, and electrical resistivity are used to study the structure of the earth. Also, knowledge of the variations of air pressure, temperature, and wind speed with altitude are extremely important in meteorological investigations. Figure 1.5 depicts a typical example of annual average vertical wind profile as a function of height. The measured variations of wind speed with height are used in examining weather patterns, as well as wind conditions that may affect an aircraft during final approach and landing.

Throughout this book we will be considering two basic types of signals: continuous-time signals and discrete-time signals. In the case of continuous-time signals the independent variable is continuous, and thus these signals are defined for a continuum of values

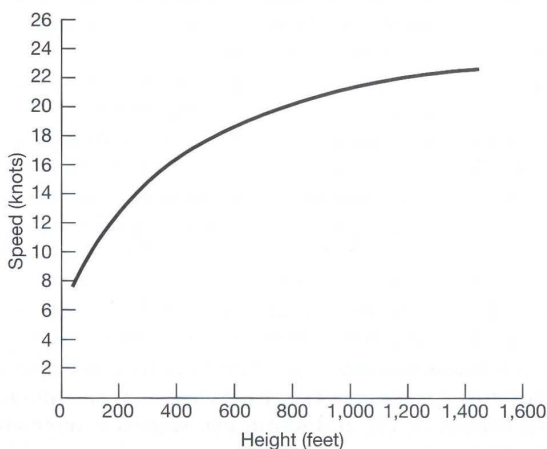


Figure 1.5 Typical annual vertical wind profile. (Adapted from Crawford and Hudson, National Severe Storms Laboratory Report, ESSA ERLTM-NSSL 48, August 1970.)

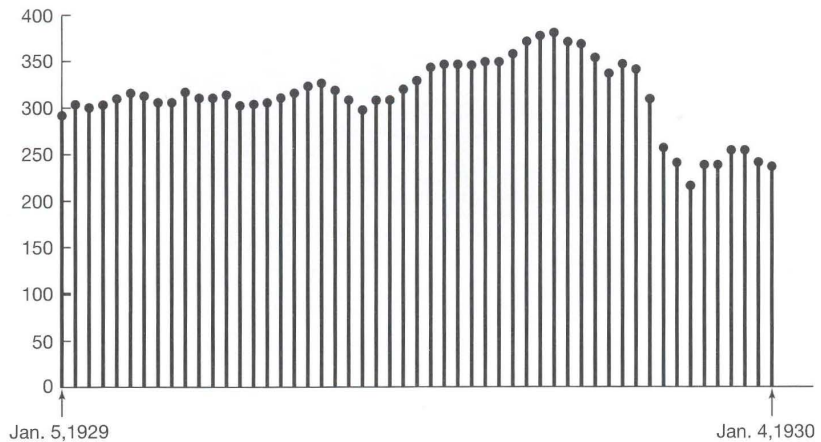


Figure 1.6 An example of a discrete-time signal: The weekly Dow-Jones stock market index from January 5, 1929, to January 4, 1930.

of the independent variable. On the other hand, discrete-time signals are defined only at discrete times, and consequently, for these signals, the independent variable takes on only a discrete set of values. A speech signal as a function of time and atmospheric pressure as a function of altitude are examples of continuous-time signals. The weekly Dow-Jones stock market index, as illustrated in Figure 1.6, is an example of a discrete-time signal. Other examples of discrete-time signals can be found in demographic studies in which various attributes, such as average budget, crime rate, or pounds of fish caught, are tabulated against such discrete variables as family size, total population, or type of fishing vessel, respectively.

To distinguish between continuous-time and discrete-time signals, we will use the symbol t to denote the continuous-time independent variable and n to denote the discrete-time independent variable. In addition, for continuous-time signals we will enclose the independent variable in parentheses ($\cdot$), whereas for discrete-time signals we will use brackets $[\cdot]$ to enclose the independent variable. We will also have frequent occasions when it will be useful to represent signals graphically. Illustrations of a continuous-time signal $x(t)$ and a discrete-time signal $x[n]$ are shown in Figure 1.7. It is important to note that the discrete-time signal $x[n]$ is defined *only* for integer values of the independent variable. Our choice of graphical representation for $x[n]$ emphasizes this fact, and for further emphasis we will on occasion refer to $x[n]$ as a discrete-time *sequence*.

A discrete-time signal $x[n]$ may represent a phenomenon for which the independent variable is inherently discrete. Signals such as demographic data are examples of this. On the other hand, a very important class of discrete-time signals arises from the *sampling* of continuous-time signals. In this case, the discrete-time signal $x[n]$ represents successive samples of an underlying phenomenon for which the independent variable is continuous. Because of their speed, computational power, and flexibility, modern digital processors are used to implement many practical systems, ranging from digital autopilots to digital audio systems. Such systems require the use of discrete-time sequences representing sampled versions of continuous-time signals—e.g., aircraft position, velocity, and heading for an

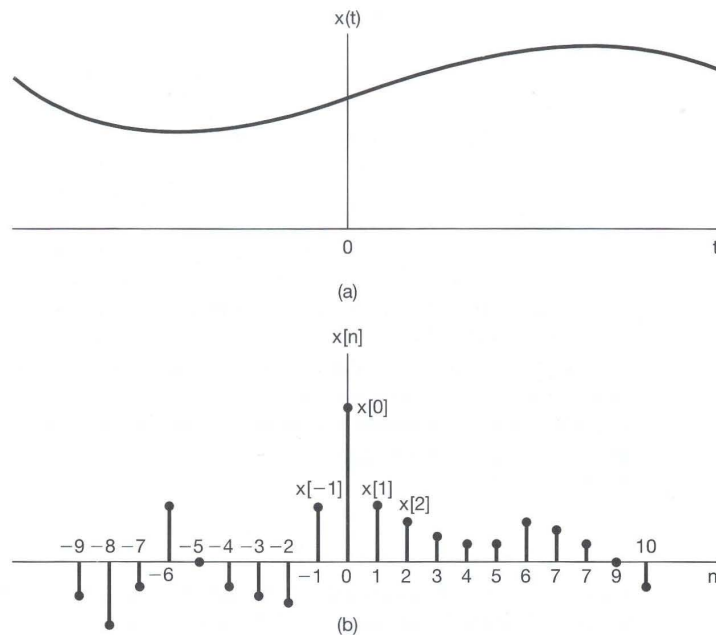


Figure 1.7 Graphical representations of (a) continuous-time and (b) discrete-time signals.

autopilot or speech and music for an audio system. Also, pictures in newspapers—or in this book, for that matter—actually consist of a very fine grid of points, and each of these points represents a sample of the brightness of the corresponding point in the original image. No matter what the source of the data, however, the signal $x[n]$ is defined only for integer values of n . It makes no more sense to refer to the $3\frac{1}{2}$ th sample of a digital speech signal than it does to refer to the average budget for a family with $2\frac{1}{2}$ family members.

Throughout most of this book we will treat discrete-time signals and continuous-time signals separately but in parallel, so that we can draw on insights developed in one setting to aid our understanding of another. In Chapter 7 we will return to the question of sampling, and in that context we will bring continuous-time and discrete-time concepts together in order to examine the relationship between a continuous-time signal and a discrete-time signal obtained from it by sampling.

1.1.2 Signal Energy and Power

From the range of examples provided so far, we see that signals may represent a broad variety of phenomena. In many, but not all, applications, the signals we consider are directly related to physical quantities capturing power and energy in a physical system. For example, if $v(t)$ and $i(t)$ are, respectively, the voltage and current across a resistor with resistance R , then the instantaneous power is

$$p(t) = v(t)i(t) = \frac{1}{R}v^2(t). \quad (1.1)$$

The total *energy* expended over the time interval $t_1 \leq t \leq t_2$ is

$$\int_{t_1}^{t_2} p(t) dt = \int_{t_1}^{t_2} \frac{1}{R} v^2(t) dt, \quad (1.2)$$

and the *average power* over this time interval is

$$\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} p(t) dt = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \frac{1}{R} v^2(t) dt. \quad (1.3)$$

Similarly, for the automobile depicted in Figure 1.2, the instantaneous power dissipated through friction is $p(t) = bv^2(t)$, and we can then define the total energy and average power over a time interval in the same way as in eqs. (1.2) and (1.3).

With simple physical examples such as these as motivation, it is a common and worthwhile convention to use similar terminology for power and energy for *any* continuous-time signal $x(t)$ or *any* discrete-time signal $x[n]$. Moreover, as we will see shortly, we will frequently find it convenient to consider signals that take on complex values. In this case, the total energy over the time interval $t_1 \leq t \leq t_2$ in a continuous-time signal $x(t)$ is defined as

$$\int_{t_1}^{t_2} |x(t)|^2 dt, \quad (1.4)$$

where $|x|$ denotes the magnitude of the (possibly complex) number x . The time-averaged power is obtained by dividing eq. (1.4) by the length, $t_2 - t_1$, of the time interval. Similarly, the total energy in a discrete-time signal $x[n]$ over the time interval $n_1 \leq n \leq n_2$ is defined as

$$\sum_{n=n_1}^{n_2} |x[n]|^2, \quad (1.5)$$

and dividing by the number of points in the interval, $n_2 - n_1 + 1$, yields the average power over the interval. It is important to remember that the terms “power” and “energy” are used here independently of whether the quantities in eqs. (1.4) and (1.5) actually are related to physical energy.¹ Nevertheless, we will find it convenient to use these terms in a general fashion.

Furthermore, in many systems we will be interested in examining power and energy in signals over an infinite time interval, i.e., for $-\infty < t < +\infty$ or for $-\infty < n < +\infty$. In these cases, we define the total energy as limits of eqs. (1.4) and (1.5) as the time interval increases without bound. That is, in continuous time,

$$E_\infty \triangleq \lim_{T \rightarrow \infty} \int_{-T}^T |x(t)|^2 dt = \int_{-\infty}^{+\infty} |x(t)|^2 dt, \quad (1.6)$$

and in discrete time,

$$E_\infty \triangleq \lim_{N \rightarrow \infty} \sum_{n=-N}^{+N} |x[n]|^2 = \sum_{n=-\infty}^{+\infty} |x[n]|^2. \quad (1.7)$$

¹Even if such a relationship does exist, eqs. (1.4) and (1.5) may have the wrong dimensions and scalings. For example, comparing eqs. (1.2) and (1.4), we see that if $x(t)$ represents the voltage across a resistor, then eq. (1.4) must be divided by the resistance (measured, for example, in ohms) to obtain units of physical energy.

Note that for some signals the integral in eq. (1.6) or sum in eq. (1.7) might not converge—e.g., if $x(t)$ or $x[n]$ equals a nonzero constant value for all time. Such signals have infinite energy, while signals with $E_\infty < \infty$ have finite energy.

In an analogous fashion, we can define the time-averaged power over an infinite interval as

$$P_\infty \triangleq \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T |x(t)|^2 dt \quad (1.8)$$

and

$$P_\infty \triangleq \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^{+N} |x[n]|^2 \quad (1.9)$$

in continuous time and discrete time, respectively. With these definitions, we can identify three important classes of signals. The first of these is the class of signals with finite total energy, i.e., those signals for which $E_\infty < \infty$. Such a signal must have zero average power, since in the continuous time case, for example, we see from eq. (1.8) that

$$P_\infty = \lim_{T \rightarrow \infty} \frac{E_\infty}{2T} = 0. \quad (1.10)$$

An example of a finite-energy signal is a signal that takes on the value 1 for $0 \leq t \leq 1$ and 0 otherwise. In this case, $E_\infty = 1$ and $P_\infty = 0$.

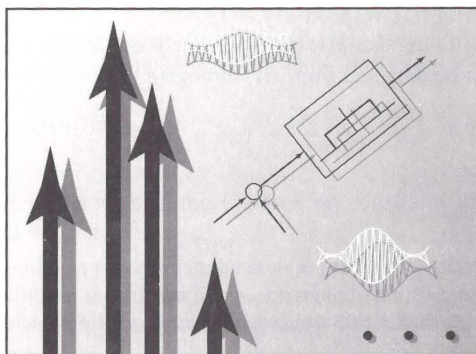
A second class of signals are those with finite average power P_∞ . From what we have just seen, if $P_\infty > 0$, then, of necessity, $E_\infty = \infty$. This, of course, makes sense, since if there is a nonzero average energy per unit time (i.e., nonzero power), then integrating or summing this over an infinite time interval yields an infinite amount of energy. For example, the constant signal $x[n] = 4$ has infinite energy, but average power $P_\infty = 16$. There are also signals for which neither P_∞ nor E_∞ are finite. A simple example is the signal $x(t) = t$. We will encounter other examples of signals in each of these classes in the remainder of this and the following chapters.

1.2 TRANSFORMATIONS OF THE INDEPENDENT VARIABLE

A central concept in signal and system analysis is that of the transformation of a signal. For example, in an aircraft control system, signals corresponding to the actions of the pilot are transformed by electrical and mechanical systems into changes in aircraft thrust or the positions of aircraft control surfaces such as the rudder or ailerons, which in turn are transformed through the dynamics and kinematics of the vehicle into changes in aircraft velocity and heading. Also, in a high-fidelity audio system, an input signal representing music as recorded on a cassette or compact disc is modified in order to enhance desirable characteristics, to remove recording noise, or to balance the several components of the signal (e.g., treble and bass). In this section, we focus on a very limited but important class of elementary signal transformations that involve simple modification of the independent variable, i.e., the time axis. As we will see in this and subsequent sections of this chapter, these elementary transformations allow us to introduce several basic properties of signals and systems. In later chapters, we will find that they also play an important role in defining and characterizing far richer and important classes of systems.

8

COMMUNICATION SYSTEMS



8.0 INTRODUCTION

Communication systems play a key role in our modern world in transmitting information between people, systems, and computers. In general terms, in all communication systems the information at the source is first processed by a transmitter or modulator to change it into a form suitable for transmission over the communication channel. At the receiver, the signal is then recovered through appropriate processing. This processing is required for a variety of reasons. In particular, quite typically, any specific communication channel has associated with it a frequency range over which it is best suited for transmitting a signal and outside of which communication is severely degraded or impossible. For example, the atmosphere will rapidly attenuate signals in the audible frequency range (10 Hz to 20 kHz), whereas it will propagate signals at a higher frequency range over longer distances. Thus, in transmitting audio signals such as speech or music over a communication channel that relies on propagation through the atmosphere, the transmitter first embeds the signal through an appropriate process into another, higher frequency signal.

Many of the concepts and techniques we have developed in the earlier chapters of this text play a central role in the analysis and design of communication systems. As with any concept that is closely tied to a wide variety of important applications, there are a large number of detailed issues to be considered, and, as indicated in the bibliography, there are many excellent texts on the subject. While a full and detailed analysis of communication systems is well beyond the scope of our discussions here, with the background of the previous chapters we are now in a position to introduce some of the basic principles and issues encountered in the design and analysis of these systems.

The general process of embedding an information-bearing signal into a second signal is typically referred to as *modulation*. Extracting the information-bearing signal

is known as *demodulation*. As we will see, modulation techniques not only allow us to embed information into signals that can be transmitted effectively, but also make possible the simultaneous transmission of more than one signal with overlapping spectra over the same channel, through a concept referred to as *multiplexing*.

There are a wide variety of modulation methods used in practice, and in this chapter we examine several of the most important of these. One large class of modulation methods relies on the concept of *amplitude modulation* or AM in which the signal we wish to transmit is used to modulate the amplitude of another signal. A very common form of amplitude modulation is *sinusoidal amplitude modulation*, which we explore in some detail in Sections 8.1–8.4 together with the related concepts of frequency-division multiplexing. Another important class of AM systems involves the modulation of the amplitude of a pulsed signal, and in Sections 8.5 and 8.6 we examine this form of modulation as well as the concept of time-division multiplexing. In Section 8.7 we then examine a different form of modulation, namely *sinusoidal frequency modulation* in which the information-bearing signal is used to vary the frequency of a sinusoidal signal.

All of the discussion up through Section 8.7 focuses attention on continuous-time signals, since most transmission media, such as the atmosphere, are best thought of as continuous-time phenomena. Nevertheless, not only is it possible to develop analogous techniques for discrete-time signals, but it is of considerable practical importance to consider modulation concepts involving such signals, and in Section 8.8 we examine some of the basic ideas behind the communication of discrete-time signals.

8.1 COMPLEX EXPONENTIAL AND SINUSOIDAL AMPLITUDE MODULATION

Many communication systems rely on the concept of sinusoidal amplitude modulation, in which a complex exponential or sinusoidal signal $c(t)$ has its amplitude multiplied (modulated) by the information-bearing signal $x(t)$. The signal $x(t)$ is typically referred to as the *modulating signal* and the signal $c(t)$ as the *carrier signal*. The modulated signal $y(t)$ is then the product of these two signals:

$$y(t) = x(t)c(t)$$

As we discussed in Section 8.0, an important objective in modulation is to produce a signal whose frequency range is suitable for transmission over the communication channel to be used. In telephone transmission systems, for example, long-distance transmission is often accomplished over microwave or satellite links. The individual voice signals are in the frequency range 200 Hz to 4 kHz, whereas a microwave link requires signals in the range 300 megahertz (MHz) to 300 gigahertz (GHz), and communication satellite links operate in the frequency range from a few hundred MHz to over 40 GHz. Thus, for transmission over these channels, the information in a voice signal must be shifted into these higher ranges of frequency. As we will see in this section, sinusoidal amplitude modulation achieves such a shift in frequency in a very simple manner.

8.1.1 Amplitude Modulation with a Complex Exponential Carrier

There are two common forms of sinusoidal amplitude modulation, one in which the carrier signal is a complex exponential of the form

$$c(t) = e^{j(\omega_c t + \theta_c)} \quad (8.1)$$

and the second in which the carrier signal is sinusoidal and of the form

$$c(t) = \cos(\omega_c t + \theta_c). \quad (8.2)$$

In both cases, the frequency ω_c is referred to as the *carrier frequency*. Let us consider first the case of a complex exponential carrier, and for convenience, let us choose $\theta_c = 0$, so that the modulated signal is

$$y(t) = x(t)e^{j\omega_c t}. \quad (8.3)$$

From the multiplication property (Section 4.5), and with $X(j\omega)$, $Y(j\omega)$, and $C(j\omega)$ denoting the Fourier transforms of $x(t)$, $y(t)$, and $c(t)$, respectively,

$$Y(j\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X(j\theta)C(j(\omega - \theta))d\theta. \quad (8.4)$$

For $c(t)$ a complex exponential as given in eq. (8.1),

$$C(j\omega) = 2\pi\delta(\omega - \omega_c), \quad (8.5)$$

and hence,

$$Y(j\omega) = X(j\omega - j\omega_c). \quad (8.6)$$

Thus, the spectrum of the modulated output $y(t)$ is simply that of the input, shifted in frequency by an amount equal to the carrier frequency ω_c . For example, with $X(j\omega)$ band limited with highest frequency ω_M (and bandwidth $2\omega_M$), as depicted in Figure 8.1(a), the output spectrum $Y(j\omega)$ is that shown in Figure 8.1(c).

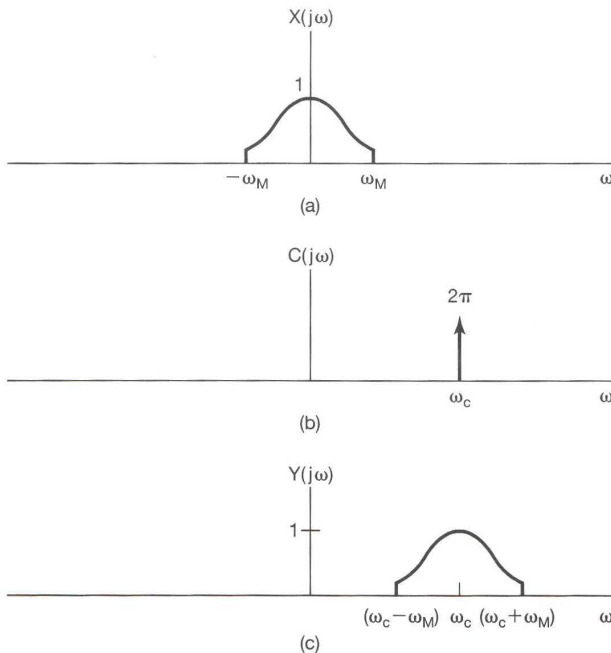


Figure 8.1 Effect in the frequency domain of amplitude modulation with a complex exponential carrier: (a) spectrum of modulating signal $x(t)$; (b) spectrum of carrier $c(t) = e^{j\omega_c t}$; (c) spectrum of amplitude-modulated signal $y(t) = x(t)e^{j\omega_c t}$.

From eq. (8.3), it is clear that $x(t)$ can be recovered from the modulated signal $y(t)$ by multiplying by the complex exponential $e^{-j\omega_c t}$; that is,

$$x(t) = y(t)e^{-j\omega_c t}. \quad (8.7)$$

In the frequency domain, this has the effect of shifting the spectrum of the modulated signal back to its original position on the frequency axis. The process of recovering the original signal from the modulated signal is referred to as *demodulation*, a topic we discuss at more length in Section 8.2.

Since $e^{j\omega_c t}$ is a complex signal, eq. (8.3) can be rewritten as

$$y(t) = x(t) \cos \omega_c t + jx(t) \sin \omega_c t. \quad (8.8)$$

Implementation of eq. (8.7) or (8.8) with $x(t)$ real utilizes two separate multipliers and two sinusoidal carrier signals that have a phase difference of $\pi/2$, as depicted in Figure 8.2 for $c(t)$ given by eq. (8.1). In Section 8.4 we give an example of one of the applications in which there are particular advantages to using a system, such as in Figure 8.2, employing two sinusoidal carriers with a phase difference of $\pi/2$.

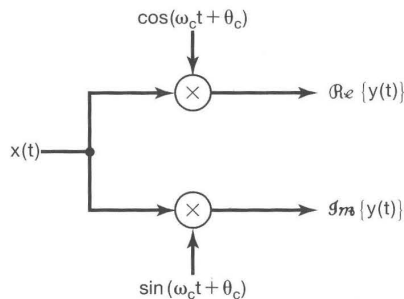


Figure 8.2 Implementation of amplitude modulation with a complex exponential carrier $c(t) = e^{j(\omega_c t + \theta_c)}$.

8.1.2 Amplitude Modulation with a Sinusoidal Carrier

In many situations, using a sinusoidal carrier of the form of eq. (8.2) is often simpler than and equally as effective as using a complex exponential carrier. In effect, using a sinusoidal carrier corresponds to retaining only the real or imaginary part of the output of Figure 8.2. A system that uses a sinusoidal carrier is depicted in Figure 8.3.

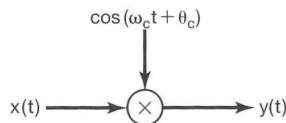


Figure 8.3 Amplitude modulation with a sinusoidal carrier.

The effect of amplitude modulation with a sinusoidal carrier in the form of eq. (8.2) can be analyzed in a manner identical to that in the preceding subsection. Again, for

convenience we choose $\theta_c = 0$. In this case, the spectrum of the carrier signal is

$$C(j\omega) = \pi[\delta(\omega - \omega_c) + \delta(\omega + \omega_c)], \quad (8.9)$$

and thus, from eq. (8.4),

$$Y(j\omega) = \frac{1}{2}[X(j\omega - j\omega_c) + X(j\omega + j\omega_c)]. \quad (8.10)$$

With $X(j\omega)$ as depicted in Figure 8.4(a), the spectrum of $y(t)$ is that shown in Figure 8.4(c). Note that there is now a replication of the spectrum of the original signal, centered around both $+\omega_c$ and $-\omega_c$. As a consequence, $x(t)$ is recoverable from $y(t)$ only if $\omega_c > \omega_M$, since otherwise the two replications will overlap in frequency. This is in contrast to the case of a complex exponential carrier, for which a replication of the spectrum of the original signal is centered only around ω_c . Specifically, as we saw in Section 8.1.1, in the case of amplitude modulation with a complex exponential carrier, $x(t)$ can always be recovered from $y(t)$ for any choice of ω_c by shifting the spectrum back to its original location by multiplying by $e^{-j\omega_c t}$, as in eq. (8.7). With a sinusoidal carrier, on the other hand, as we see from Figure 8.4, if $\omega_c < \omega_M$, then there will be an overlap between the two replications of $X(j\omega)$. For example, Figure 8.5 depicts $Y(j\omega)$ for $\omega_c = \omega_M/2$. Clearly, the spectrum of $x(t)$ is no longer replicated in $Y(j\omega)$, and thus, it may no longer be possible to recover $x(t)$ from $y(t)$.

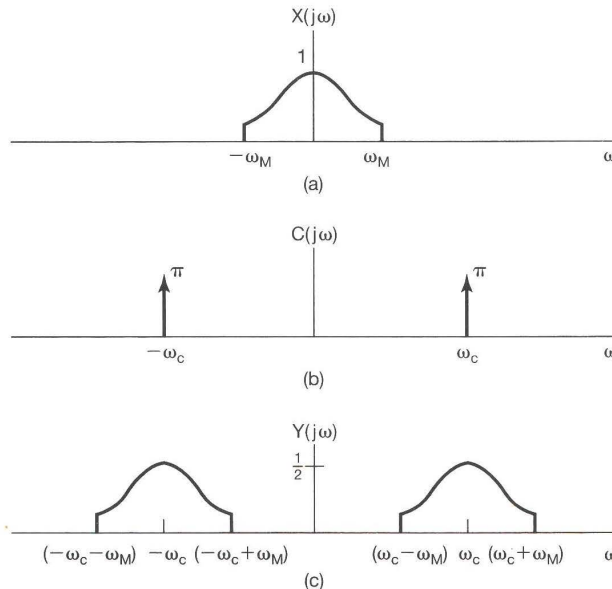


Figure 8.4 Effect in the frequency domain of amplitude modulation with a sinusoidal carrier: (a) spectrum of modulating signal $x(t)$; (b) spectrum of carrier $c(t) = \cos \omega_c t$; (c) spectrum of amplitude-modulated signal.

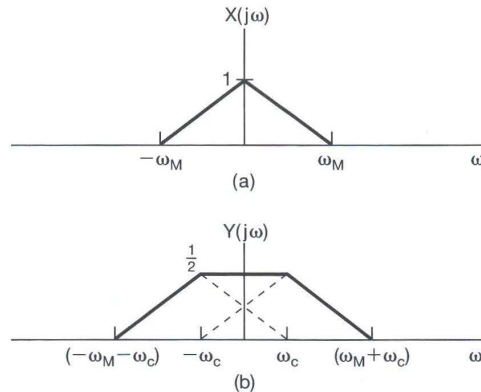


Figure 8.5 Sinusoidal amplitude modulation with carrier $\cos \omega_c t$ for which $\omega_c = \omega_M/2$: (a) spectrum of modulating signal; (b) spectrum of modulated signal.

8.2 DEMODULATION FOR SINUSOIDAL AM

At the receiver in a communication system, the information-bearing signal $x(t)$ is recovered through demodulation. In this section, we examine the process of demodulation for sinusoidal amplitude modulation, as introduced in the previous section. There are two commonly used methods for demodulation, each with its own advantages and disadvantages. In Section 8.2.1 we discuss the first of these, a process referred to as *synchronous demodulation*, in which the transmitter and receiver are synchronized in phase. In Section 8.2.2, we describe an alternative method referred to as *asynchronous demodulation*.

8.2.1 Synchronous Demodulation

Assuming that $\omega_c > \omega_M$, demodulation of a signal that was modulated with a sinusoidal carrier is relatively straightforward. Specifically, consider the signal

$$y(t) = x(t) \cos \omega_c t. \quad (8.11)$$

As was suggested in Example 4.21, the original signal can be recovered by modulating $y(t)$ with the same sinusoidal carrier and applying a lowpass filter to the result. To see this, consider

$$w(t) = y(t) \cos \omega_c t. \quad (8.12)$$

Figure 8.6 shows the spectra of $y(t)$ and $w(t)$, and we observe that $x(t)$ can be recovered from $w(t)$ by applying an ideal lowpass filter with a gain of 2 and a cutoff frequency that is greater than ω_M and less than $2\omega_c - \omega_M$. The frequency response of the lowpass filter is indicated by the dashed line in Figure 8.6(c).

The basis for using eq. (8.12) and a lowpass filter to demodulate $y(t)$ can also be seen algebraically. From eqs. (8.11) and (8.12), it follows that

$$w(t) = x(t) \cos^2 \omega_c t,$$

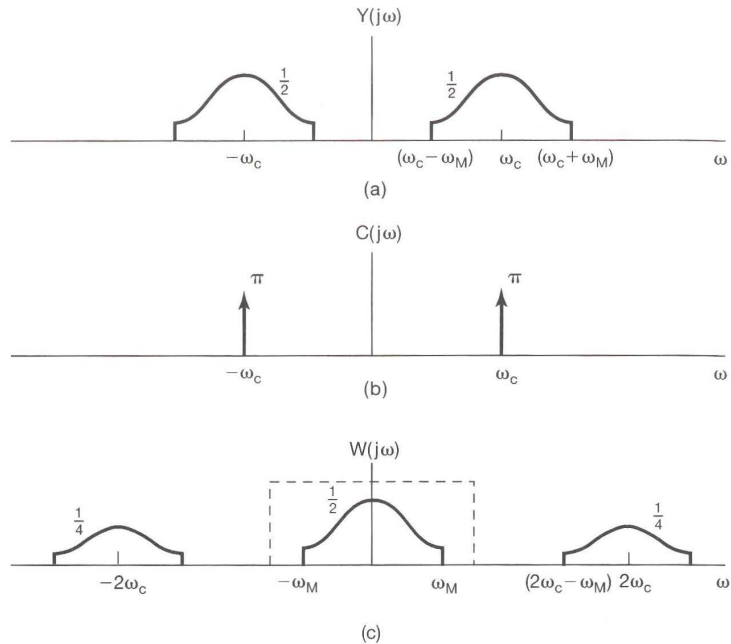


Figure 8.6 Demodulation of an amplitude-modulated signal with a sinusoidal carrier: (a) spectrum of modulated signal; (b) spectrum of carrier signal; (c) spectrum of modulated signal multiplied by the carrier. The dashed line indicates the frequency response of a lowpass filter used to extract the demodulated signal.

or, using the trigonometric identity

$$\cos^2 \omega_c t = \frac{1}{2} + \frac{1}{2} \cos 2\omega_c t,$$

we can rewrite $w(t)$ as

$$w(t) = \frac{1}{2}x(t) + \frac{1}{2}x(t) \cos 2\omega_c t. \quad (8.13)$$

Thus, $w(t)$ consists of the sum of two terms, namely one-half the original signal and one-half the original signal modulated with a sinusoidal carrier at twice the original carrier frequency ω_c . Both of these terms are apparent in the spectrum shown in Figure 8.6(c). Applying the lowpass filter to $w(t)$ corresponds to retaining the first term on the right-hand side of eq. (8.13) and eliminating the second term.

The overall system for amplitude modulation and demodulation using a complex exponential carrier is depicted in Figure 8.7, and the overall system for modulation and demodulation using a sinusoidal carrier is depicted in Figure 8.8. In these figures, we have indicated the more general case in which, for both the complex exponential and the sinusoidal carrier, a carrier phase θ_c is included. The modification of the preceding analysis so as to include θ_c is straightforward and is considered in Problem 8.21.

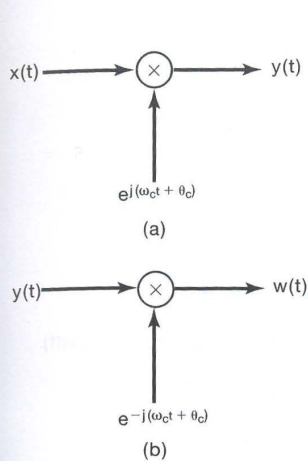


Figure 8.7 System for amplitude modulation and demodulation using a complex exponential carrier: (a) modulation; (b) demodulation.

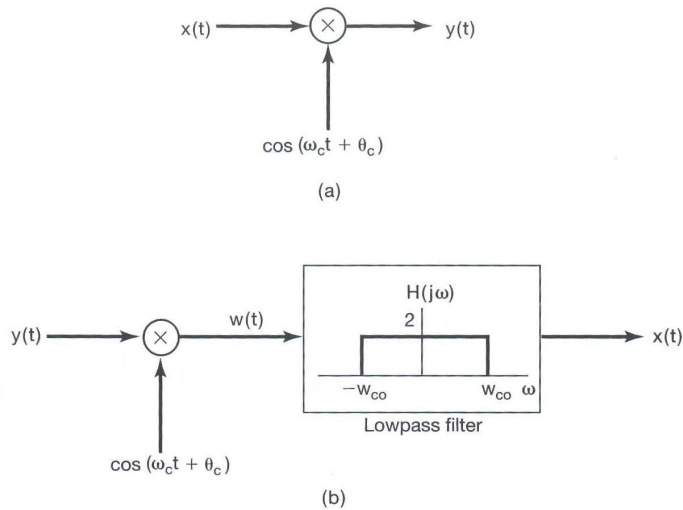


Figure 8.8 Amplitude modulation and demodulation with a sinusoidal carrier: (a) modulation system; (b) demodulation system. The lowpass filter cut-off frequency ω_{co} is greater than ω_M and less than $2\omega_c - \omega_M$.

In the systems of Figures 8.7 and 8.8, the demodulating signal is assumed to be synchronized in phase with the modulating signal, and consequently the process is referred to as *synchronous demodulation*. Suppose, however, that the modulator and demodulator are not synchronized in phase. For the case of the complex exponential carrier, with θ_c denoting the phase of the modulating carrier and ϕ_c the phase of the demodulating carrier,

$$y(t) = e^{j(\omega_c t + \theta_c)} x(t), \quad (8.14)$$

$$w(t) = e^{-j(\omega_c t + \phi_c)} y(t), \quad (8.15)$$

and consequently,

$$w(t) = e^{j(\theta_c - \phi_c)} x(t). \quad (8.16)$$

Thus, if $\theta_c \neq \phi_c$, $w(t)$ will have a complex amplitude factor. For the particular case in which $x(t)$ is positive, $x(t) = |w(t)|$, and thus $x(t)$ can be recovered by taking the magnitude of the demodulated signal.

For the sinusoidal carrier, again let θ_c and ϕ_c denote the phases of the modulating and demodulating carriers, respectively, as indicated in Figure 8.9. The input to the lowpass filter is now

$$w(t) = x(t) \cos(\omega_c t + \theta_c) \cos(\omega_c t + \phi_c), \quad (8.17)$$

or, using the trigonometric identity

$$\cos(\omega_c t + \theta_c) \cos(\omega_c t + \phi_c) = \frac{1}{2} \cos(\theta_c - \phi_c) + \frac{1}{2} \cos(2\omega_c t + \theta_c + \phi_c), \quad (8.18)$$

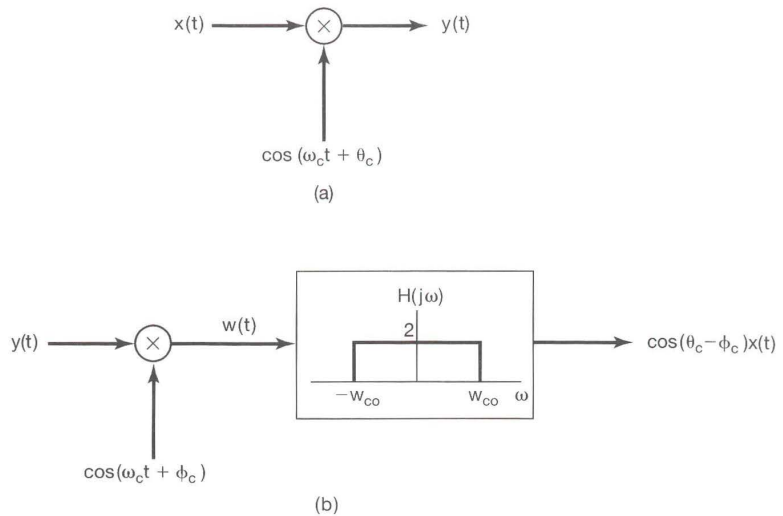


Figure 8.9 Sinusoidal amplitude modulation and demodulation system for which the carrier signals and the modulator and demodulator are not synchronized: (a) modulator; (b) demodulator.

we have

$$w(t) = \frac{1}{2} \cos(\theta_c - \phi_c) x(t) + \frac{1}{2} x(t) \cos(2\omega_c t + \theta_c + \phi_c), \quad (8.19)$$

and the output of the lowpass filter is then $x(t)$ multiplied by the amplitude factor $\cos(\theta_c - \phi_c)$. If the oscillators in the modulator and demodulator are in phase, $\theta_c = \phi_c$, and the output of the lowpass filter is $x(t)$. On the other hand, if these oscillators have a phase difference of $\pi/2$, the output will be zero. In general, for a maximum output signal, the oscillators should be in phase. Of even more importance, the phase relation between the two oscillators must be maintained over time, so that the amplitude factor $\cos(\theta_c - \phi_c)$ does not vary. This requires careful synchronization between the modulator and the demodulator, which is often difficult, particularly when they are geographically separated, as is typical in a communication system. The corresponding effects of, and the need for, synchronization not only between the phase of the modulator and demodulator, but between the frequencies of the carrier signals used in both, are explored in detail in Problem 8.23.

8.2.2 Asynchronous Demodulation

In many systems that employ sinusoidal amplitude modulation, an alternative demodulation procedure referred to as *asynchronous demodulation* is commonly used. Asynchronous demodulation avoids the need for synchronization between the modulator and demodulator. In particular, suppose that $x(t)$ is always positive and that the carrier frequency ω_c is much higher than ω_M , the highest frequency in the modulating signal. The modulated signal $y(t)$ will then have the general form illustrated in Figure 8.10.

In particular, the *envelope* of $y(t)$ —that is, a smooth curve connecting the peaks in $y(t)$ —would appear to be a reasonable approximation to $x(t)$. Thus, $x(t)$ could be approximately recovered through the use of a system that tracks these peaks to extract the envelope. Such a system is referred to as an *envelope detector*. One example of a simple circuit that acts as an envelope detector is shown in Figure 8.11(a). This circuit is generally followed by a lowpass filter to reduce the variations at the carrier frequency, which are evident in Figure 8.11(b) and which will generally be present in the output of an envelope detector of the type indicated in Figure 8.11(a).

The two basic assumptions required for asynchronous demodulation are that $x(t)$ be positive and that $x(t)$ vary slowly compared to ω_c , so that the envelope is easily tracked. The second condition is satisfied, for example, in audio transmission over a radio-frequency (RF) channel, where the highest frequency present in $x(t)$ is typically 15 to 20 kHz and $\omega_c/2\pi$ is in the range 500 kHz to 2 MHz. The first condition, that $x(t)$ be positive, can be satisfied by simply adding an appropriate constant value to $x(t)$ or, equivalently, by a simple change in the modulator, as shown in Figure 8.12. The output of the envelope detector then approximates $x(t) + A$, from which $x(t)$ is easily obtained.

To use the envelope detector for demodulation, we require that A be sufficiently large so that $x(t) + A$ is positive. Let K denote the maximum amplitude of $x(t)$; that is, $|x(t)| \leq K$. For $x(t) + A$ to be positive, we require that $A > K$. The ratio K/A is commonly referred to as the *modulation index* m . Expressed in percent, it is referred to as the *percent modulation*. An illustration of the output of the modulator of Figure 8.12 for $x(t)$ sinusoidal and for $m = 0.5$ (50% modulation) and $m = 1.0$ (100% modulation), is shown in Figure 8.13.

In Figure 8.14, we show a comparison of the spectra associated with the modulated signal when synchronous demodulation and when asynchronous demodulation are used. We note in particular that the output of the modulator for the asynchronous system in Figure 8.12 has an additional component $A \cos \omega_c t$ that is neither present nor necessary in the synchronous system. This is represented in the spectrum of Figure 8.14(c) by the presence of impulses at $+\omega_c$ and $-\omega_c$. For a fixed maximum amplitude K of the modulating signal, as A is decreased the relative amount of carrier present in the modulated output decreases. Since the carrier component in the output contains no information, its presence

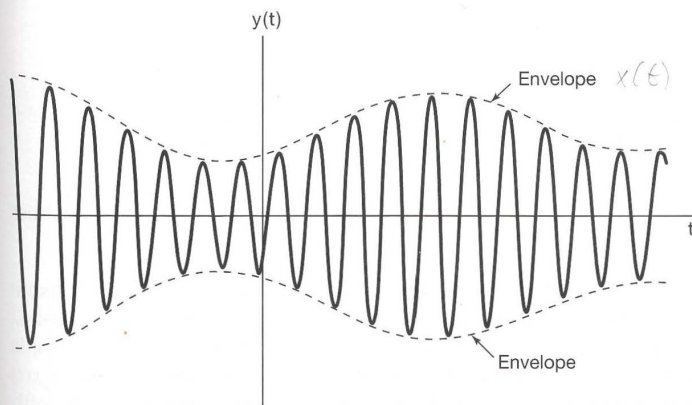


Figure 8.10 Amplitude-modulated signal for which the modulating signal is positive. The dashed curve represents the envelope of the modulated signal.

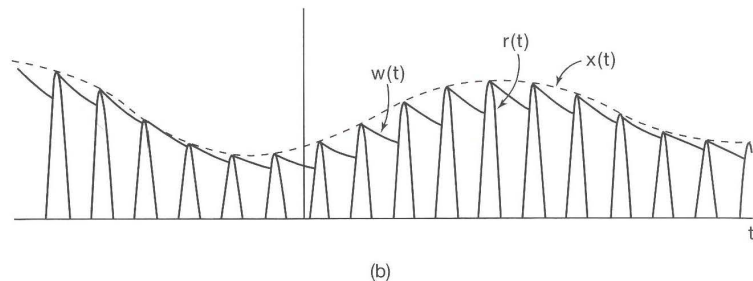
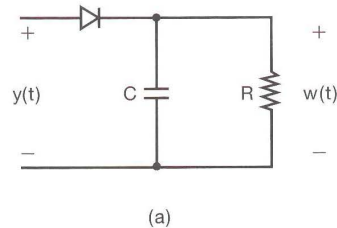


Figure 8.11 Demodulation by envelope detection: (a) circuit for envelope detection using half-wave rectification; (b) waveforms associated with the envelope detector in (a): $r(t)$ is the half-wave rectified signal, $x(t)$ is the true envelope, and $w(t)$ is the envelope obtained from the circuit in (a). The relationship between $x(t)$ and $w(t)$ has been exaggerated in (b) for purposes of illustration. In a practical asynchronous demodulation system, $w(t)$ would typically be a much closer approximation to $x(t)$ than depicted here.

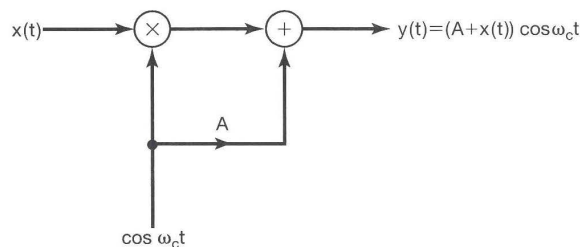


Figure 8.12 Modulator for an asynchronous modulation-demodulation system.

represents an inefficiency—for example, in the amount of power required to transmit the modulated signal—and thus, in one sense it is desirable to make the ratio K/A —i.e., the modulation index m —as large as possible. On the other hand, the ability of a simple envelope detector such as that in Figure 8.11 to follow the envelope and thus extract $x(t)$ improves as the modulation index decreases. Hence, there is a trade-off between the effi-

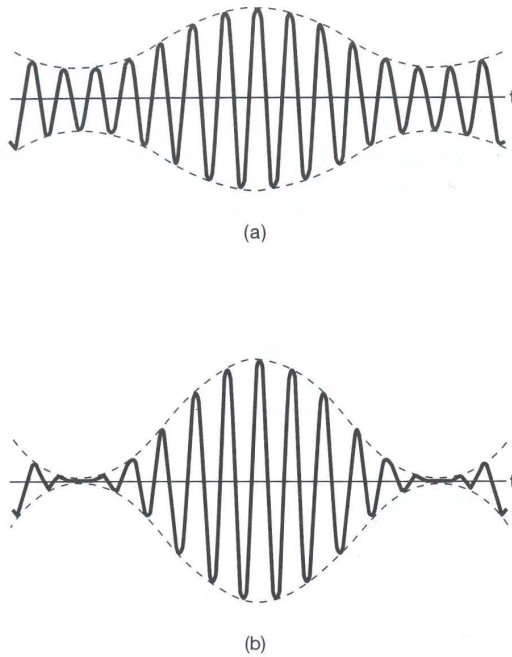


Figure 8.13 Output of the amplitude modulation system of Figure 8.12: (a) modulation index $m = 0.5$; (b) modulation index $m = 1.0$.

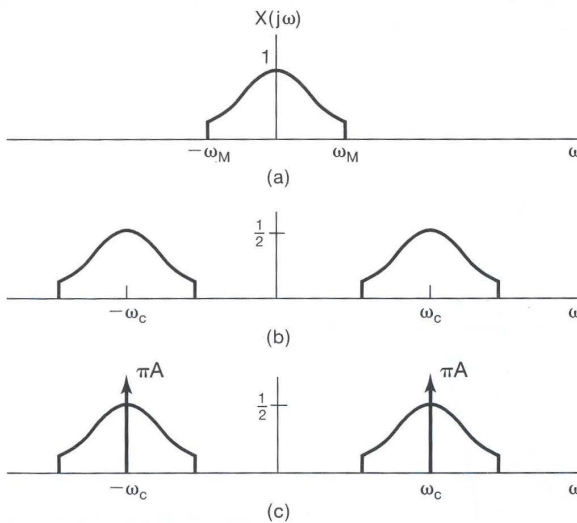


Figure 8.14 Comparison of spectra for synchronous and asynchronous sinusoidal amplitude modulation systems: (a) spectrum of modulating signal; (b) spectrum of $x(t) \cos \omega_c t$ representing modulated signal in a synchronous system; (c) spectrum of $[x(t) + A] \cos \omega_c t$ representing modulated signal in an asynchronous system.

ciency of the system in terms of the power in the output of the modulator and the quality of the demodulated signal.

There are a number of advantages and disadvantages to the asynchronous modulation-demodulation system of Figures 8.11 and 8.12, compared with the synchronous system of Figure 8.8. The synchronous system requires a more sophisticated demodulator because the oscillator in the demodulator must be synchronized with the oscillator in the modulator, both in phase and in frequency. On the other hand, the asynchronous modulator in general requires transmitting more power than the synchronous modulator, since, for the envelope detector to operate properly, the envelope must be positive, or equivalently, there must be a carrier component present in the transmitted signal. This is often preferable in cases such as that associated with public radio broadcasting, in which it is desirable to mass-produce large numbers of receivers (demodulators) at moderate cost. The additional cost in transmitted power is then offset by the savings in cost for the receiver. On the other hand, in situations in which transmitter power requirements are at a premium, as in satellite communication, the cost of implementing a more sophisticated synchronous receiver is warranted.

8.3 FREQUENCY-DIVISION MULTIPLEXING

Many systems used for transmitting signals provide more bandwidth than is required for any one signal. For example, a typical microwave link has a total bandwidth of several gigahertz, which is considerably greater than the bandwidth required for one voice channel. If the individual voice signals, which are overlapping in frequency, have their frequency content shifted by means of sinusoidal amplitude modulation so that the spectra of the modulated signals no longer overlap, they can be transmitted *simultaneously* over a single wideband channel. The resulting concept is referred to as *frequency-division multiplexing* (FDM). Frequency-division multiplexing using a sinusoidal carrier is illustrated in Figure 8.15. The individual signals to be transmitted are assumed to be band limited and

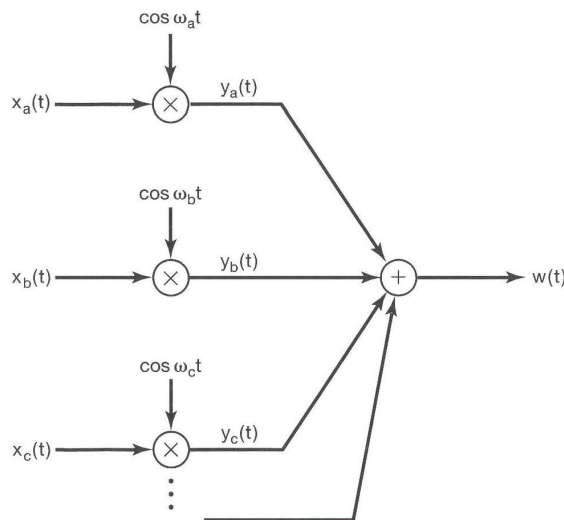


Figure 8.15 Frequency-division multiplexing using sinusoidal amplitude modulation.

are modulated with different carrier frequencies. The modulated signals are then summed and transmitted simultaneously over the same communication channel. The spectra of the individual subchannels and the composite multiplexed signal are illustrated in Figure 8.16. Through this multiplexing process, the individual input signals are allocated distinct segments of the frequency band. To recover the individual channels in the demultiplexing process requires two basic steps: bandpass filtering to extract the modulated signal corresponding to a specific channel, followed by demodulation to recover the original signal. This is illustrated in Figure 8.17 to recover channel a , where, for purposes of illustration, synchronous demodulation is assumed.

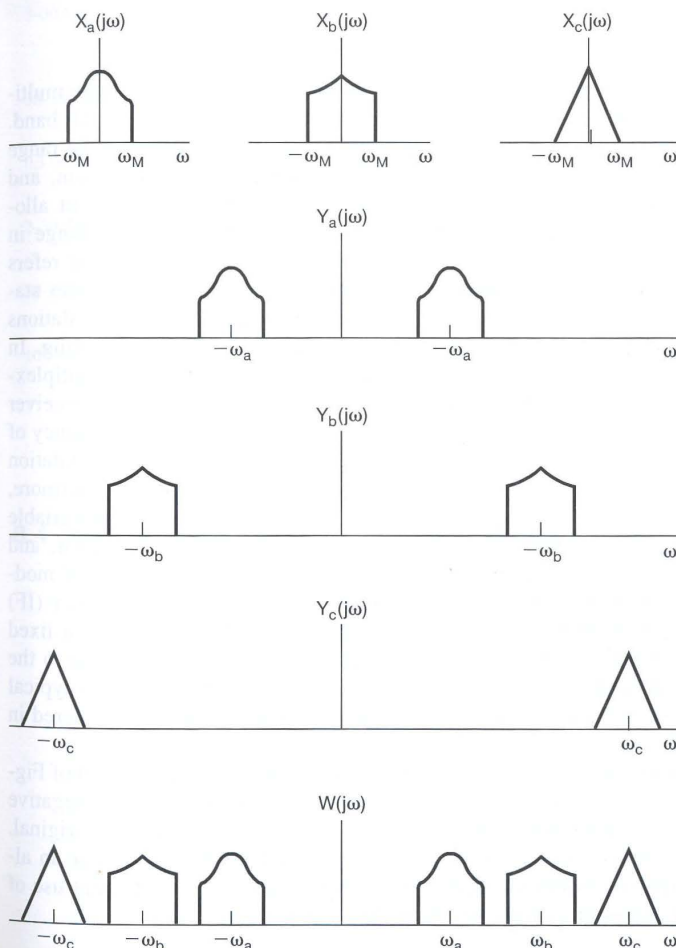


Figure 8.16 Spectra associated with the frequency-division multiplexing system of Figure 8.15.

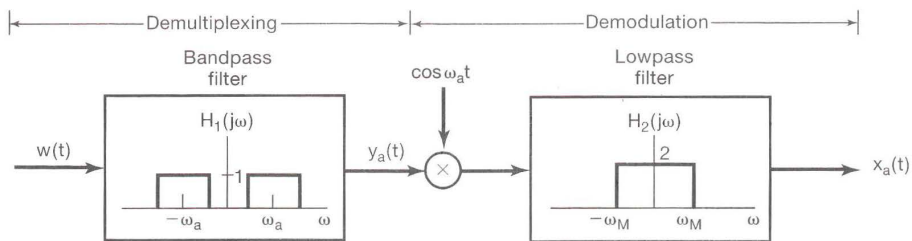


Figure 8.17 Demultiplexing and demodulation for a frequency-division multiplexed signal.

Telephone communication is one important application of frequency-division multiplexing. Another is the transmission of signals through the atmosphere in the RF band. In the United States, the use of radio frequencies for transmitting signals over the range 10 kHz to 275 GHz is controlled by the Federal Communications Commission, and different portions of the range are allocated for different purposes. The current allocation of frequencies is shown in Figure 8.18. As indicated, the frequency range in the neighborhood of 1 MHz is assigned to the AM broadcast band, where *AM* refers specifically to the use of sinusoidal amplitude modulation. Individual AM radio stations are assigned specific frequencies within the AM band, and thus, many stations can broadcast simultaneously through this use of frequency-division multiplexing. In principle, at the receiver, an individual radio station can be selected by demultiplexing and demodulating, as illustrated in Figure 8.17. The tuning dial on the receiver would then control both the center frequency of the bandpass filter and the frequency of the demodulating oscillator. In fact, for public broadcasting, asynchronous modulation and demodulation are used to simplify the receiver and reduce its cost. Furthermore, the demultiplexing in Figure 8.17 requires a sharp cutoff bandpass filter with variable center frequency. Variable frequency-selective filters are difficult to implement, and consequently, a fixed filter is implemented instead, and an intermediate stage of modulation and filtering [referred to in a radio receiver as the *intermediate-frequency* (IF) stage] is used. The use of modulation to slide the spectrum of the signal past a fixed bandpass filter replaces the use of a variable bandpass filter in a manner similar to the procedure discussed in Section 4.5.1. This basic procedure is incorporated into typical home AM radio receivers. Some of the more detailed issues involved are considered in Problem 8.36.

As illustrated in Figure 8.16, in the frequency-division multiplexing system of Figure 8.15 the spectrum of each individual signal is replicated at both positive and negative frequencies, and thus the modulated signal occupies twice the bandwidth of the original. This represents an inefficient use of bandwidth. In the next section we consider an alternative form of sinusoidal amplitude modulation, which leads to more efficient use of bandwidth at the cost of a more complicated modulation system.

Frequency range	Designation	Typical uses	Propagation method	Channel features
30–300 Hz	ELF (extremely low frequency)	Macrowave, submarine communication	Megametric waves	Penetration of conducting earth and seawater
0.3–3 kHz	VF (voice frequency)	Data terminals, telephony	Copper wire	
3–30 kHz	VLf (very low frequency)	Navigation, telephone, telegraph, frequency and timing standards	Surface ducting (ground wave)	Low attenuation, little fading, extremely stable phase and frequency, large antennas
30–300 kHz	LF (low frequency)	Industrial (power line) communication, aeronautical and maritime long-range navigation, radio beacons	Mostly surface ducting	Slight fading, high atmospheric pulse
0.3–3 MHz	MF (medium frequency)	Mobile, AM broadcasting, amateur, public safety	Ducting and ionospheric reflection (sky wave)	Increased fading, but reliable
3–30 MHz	HF (high frequency)	Military communication, aeronautical mobile, international fixed, amateur and citizen's band, industrial	Ionospheric reflecting sky wave, 50–400 km layer altitudes	Intermittent and frequency-selective fading, multipath
30–300 MHz	VHF (very high frequency)	FM and TV broadcast, land transportation (taxis, buses, railroad)	Sky wave (ionospheric and tropospheric scatter)	Fading, scattering, and multipath
0.3–3 GHz	UHF (ultra high frequency)	UHF TV, space telemetry, radar, military	Transhorizon tropospheric scatter and line-of-sight relaying	
3–30 GHz	SHF (super high frequency)	Satellite and space communication, common carrier (CC), microwave	Line-of-sight ionosphere penetration	Ionospheric penetration, extraterrestrial noise, high directly
30–300 GHz	EHF (extremely high frequency)	Experimental, government, radio astronomy	Line of sight	Water vapor and oxygen absorption
10^3 – 10^7 GHz	Infrared, visible light, ultraviolet	Optical communications	Line of sight	

Figure 8.18 Allocation of frequencies in the RF spectrum.

8.4 SINGLE-SIDEBAND SINUSOIDAL AMPLITUDE MODULATION

For the sinusoidal amplitude modulation systems discussed in Section 8.1, the total bandwidth of the original signal $x(t)$ is $2\omega_M$, including both positive and negative frequencies, where ω_M is the highest frequency present in $x(t)$. With the use of a complex exponential carrier, the spectrum is translated to ω_c , and the total width of the frequency band over which there is energy from the signal is still $2\omega_M$, although the modulated signal is now complex. With a sinusoidal carrier, on the other hand, the spectrum of the signal is shifted to $+\omega_c$ and $-\omega_c$, and thus, twice the bandwidth is required. This suggests that there is a basic redundancy in the modulated signal with a sinusoidal carrier. Using a technique referred to as *single-sideband modulation*, we can remove the redundancy.

The spectrum of $x(t)$ is illustrated in Figure 8.19(a), in which we have shaded the positive and negative frequency components differently to distinguish them. The spectrum

in Figure 8.19(b) results from modulation with a sinusoidal carrier, where we identify an upper and lower sideband for the portion of the spectrum centered at $+\omega_c$ and that centered at $-\omega_c$. Comparing Figures 8.19(a) and (b), we see that $X(j\omega)$ can be recovered if only the upper sidebands at positive and negative frequencies are retained, or alternatively, if only the lower sidebands at positive and negative frequencies are retained. The resulting spectrum if only the upper sidebands are retained is shown in Figure 8.19(c), and the resulting spectrum if only the lower sidebands are retained is shown in Figure 8.19(d). The conversion of $x(t)$ to the form corresponding to Figure 8.19(c) or (d) is referred to as *single-sideband modulation* (SSB), in contrast to the *double-sideband modulation* (DSB) of Figure 8.19(b), in which both sidebands are retained.

There are several methods by which the single-sideband signal can be obtained. One is to apply a sharp cutoff bandpass or highpass filter to the double-sideband signal of Figure 8.19(b), as illustrated in Figure 8.20, to remove the unwanted sideband. Another is to use a procedure that utilizes phase shifting. Figure 8.21 depicts a system designed

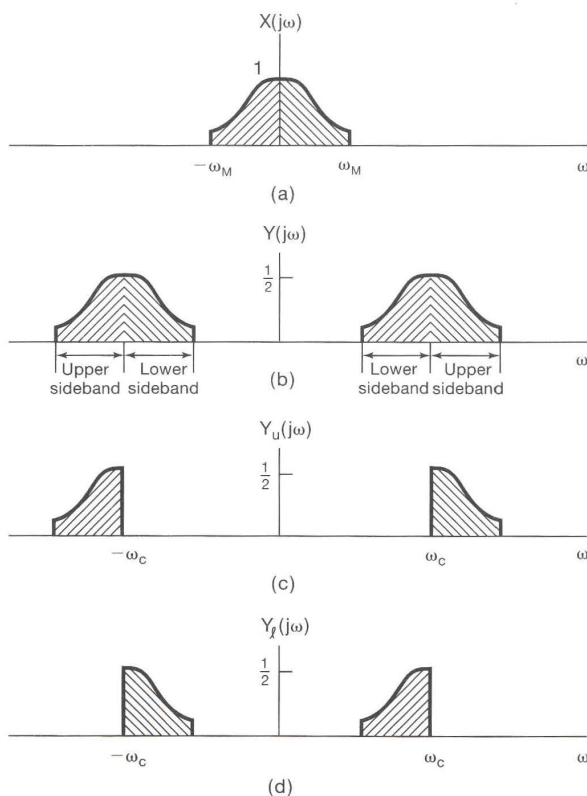


Figure 8.19 Double- and single-sideband modulation: (a) spectrum of modulating signal; (b) spectrum after modulation with a sinusoidal carrier; (c) spectrum with only the upper sidebands; (d) spectrum with only the lower sidebands.

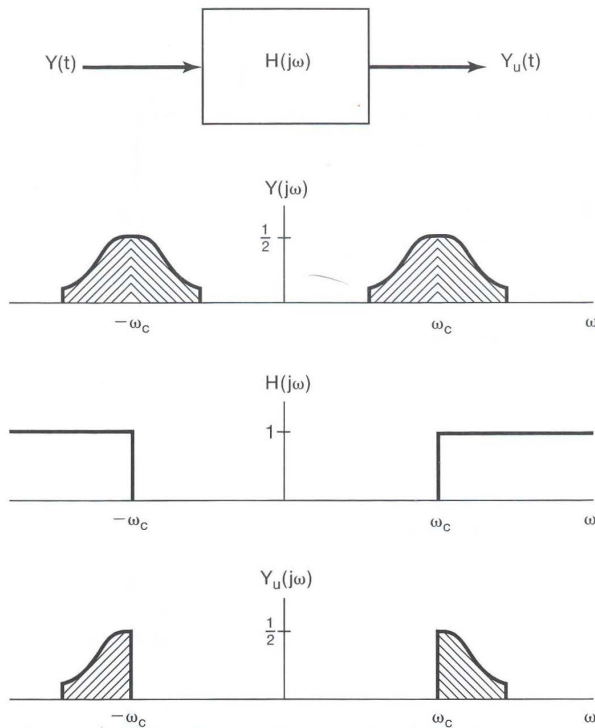


Figure 8.20 System for retaining the upper sidebands using ideal high-pass filtering.

to retain the lower sidebands. The system $H(j\omega)$ in the figure is referred to as a “90° phase-shift network,” for which the frequency response is of the form

$$H(j\omega) = \begin{cases} -j, & \omega > 0 \\ j, & \omega < 0 \end{cases} \quad (8.20)$$

The spectra of $x(t)$, $y_1(t) = x(t) \cos \omega_c t$, $y_2(t) = x_p(t) \sin \omega_c t$, and $y(t)$ are illustrated in Figure 8.22. As is examined in Problem 8.28, to retain the upper sidebands instead of the lower sidebands, the phase characteristic of $H(j\omega)$ is reversed so that

$$H(j\omega) = \begin{cases} j, & \omega > 0 \\ -j, & \omega < 0 \end{cases} \quad (8.21)$$

As is explored in Problem 8.29, synchronous demodulation of single-sideband systems can be accomplished in a manner identical to synchronous demodulation of double-sideband systems. The price paid for the increased efficiency of single-sideband systems is added complexity in the modulator.

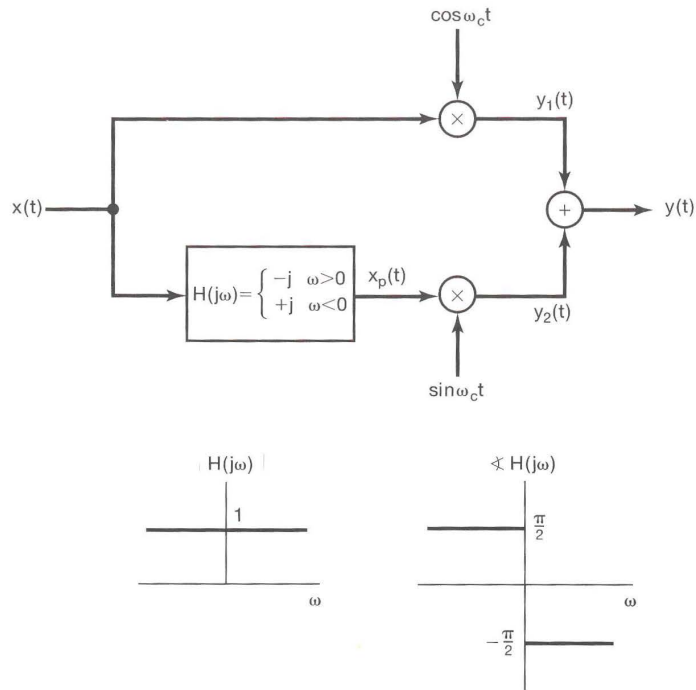


Figure 8.21 System for single-sideband amplitude modulation, using a 90° phase-shift network, in which only the lower sidebands are retained.

In summary, in Sections 8.1 through 8.4 we have seen a number of variations of complex exponential and sinusoidal amplitude modulation. With asynchronous demodulation, discussed in Section 8.2.2, a constant must be added to the modulating signal so that it is positive. This results in the presence of the carrier signal as a component in the modulated output, requiring more power for transmission, but resulting in a simpler demodulator than is required in a synchronous system. Alternatively, only the upper or lower sidebands in the modulated output may be retained, which makes more efficient use of bandwidth and transmitter power, but requires a more sophisticated modulator. Sinusoidal amplitude modulation with both sidebands and the presence of a carrier is typically abbreviated as AM-DSB/WC (amplitude modulation, double sideband/with carrier) and, when the carrier is suppressed or absent, as AM-DSB/SC (amplitude modulation, double-sideband/suppressed carrier). The corresponding single-sideband systems are abbreviated AM-SSB/WC and AM-SSB/SC.

Sections 8.1 through 8.4 are intended to provide an introduction to many of the basic concepts associated with sinusoidal amplitude modulation. There are many variations in details and implementation, and the reader is referred to the bibliography for an indication of the numerous excellent books that explore this topic further.

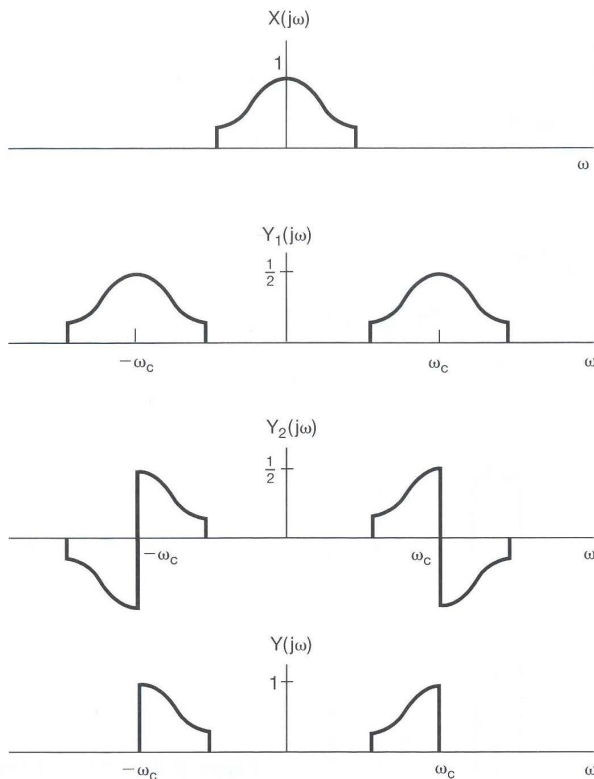


Figure 8.22 Spectra associated with the single-sideband system of Figure 8.21.

8.5 AMPLITUDE MODULATION WITH A PULSE-TRAIN CARRIER

8.5.1 Modulation of a Pulse-Train Carrier

In previous sections, we examined amplitude modulation with a sinusoidal carrier. Another important class of amplitude modulation techniques corresponds to the use of a carrier signal that is a pulse train, as illustrated in Figure 8.23; amplitude modulation of this type effectively corresponds to transmitting equally spaced time slices of $x(t)$. In general, we would not expect that an arbitrary signal could be recovered from such a set of time slices. However, our examination of the concept of sampling in Chapter 7 suggests that this should be possible if $x(t)$ is band limited and the pulse repetition frequency is high enough.

From Figure 8.23,

$$y(t) = x(t)c(t); \quad (8.22)$$

i.e., the modulated signal $y(t)$ is the product of $x(t)$ and the carrier $c(t)$. With $Y(j\omega)$, $X(j\omega)$, and $C(j\omega)$ representing the Fourier transforms of each of these signals, it follows from

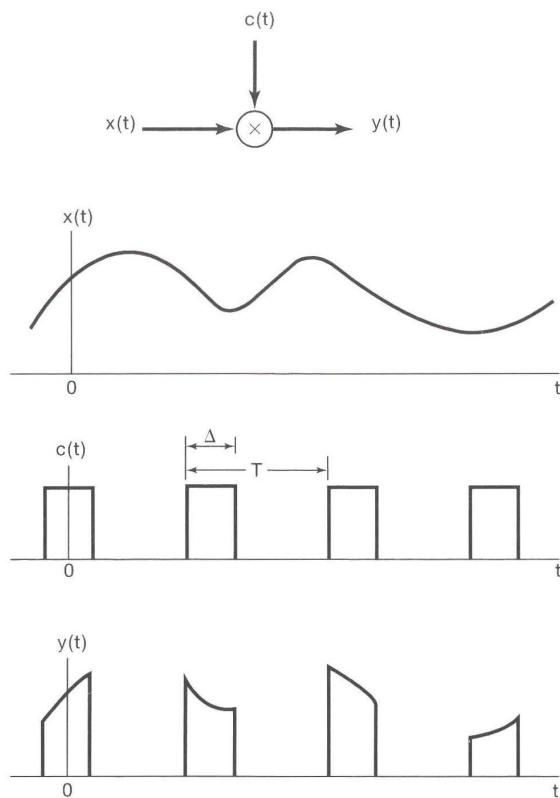


Figure 8.23 Amplitude modulation of a pulse train.

the multiplication property that

$$Y(j\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X(j\theta) C(j(\omega - \theta)) d\theta. \quad (8.23)$$

Since $c(t)$ is periodic with period T , $C(j\omega)$ consists of impulses in frequency spaced by $2\pi/T$; that is,

$$C(j\omega) = 2\pi \sum_{k=-\infty}^{+\infty} a_k \delta(\omega - k\omega_c), \quad (8.24)$$

where $\omega_c = 2\pi/T$ and the coefficients a_k are the Fourier series coefficients of $c(t)$, which, from Example 3.5, are

$$a_k = \frac{\sin(k\omega_c \Delta/2)}{\pi k}. \quad (8.25)$$

The spectrum of $c(t)$ is shown in Figure 8.24(b). With the spectrum of $x(t)$ as illustrated in Figure 8.24(a), the resulting spectrum of the modulated signal $y(t)$ is shown in Figure 8.24(c). From eqs. (8.23) and (8.24), $Y(j\omega)$ is a sum of scaled and shifted replicas of $X(j\omega)$:

$$Y(j\omega) = \sum_{k=-\infty}^{+\infty} a_k X(j(\omega - k\omega_c)). \quad (8.26)$$

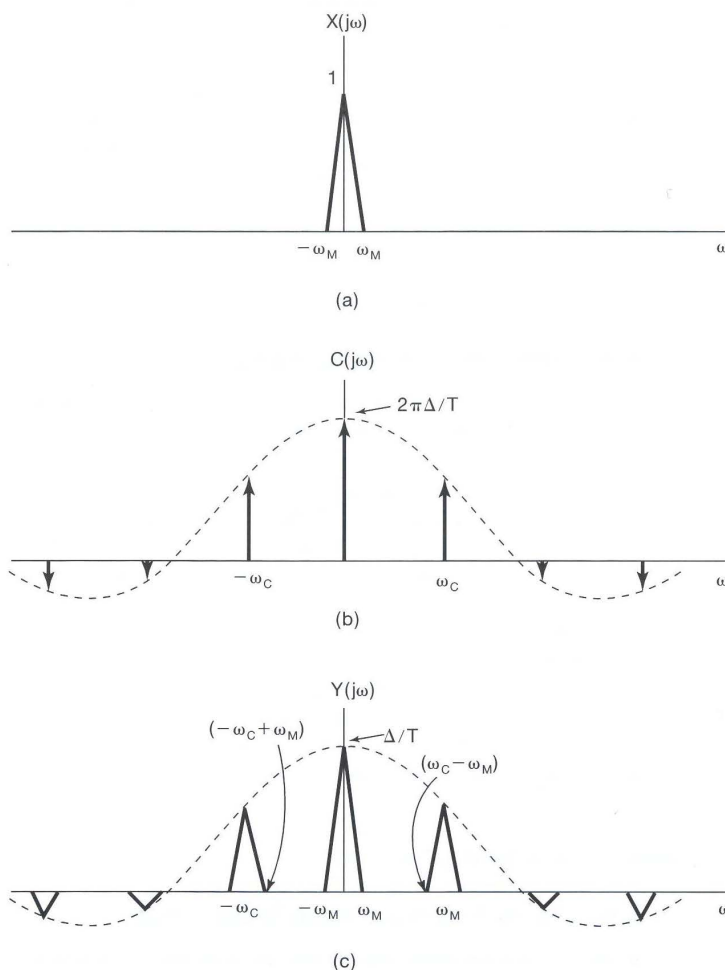


Figure 8.24 Spectra associated with amplitude modulation of a pulse train: (a) spectrum of a bandlimited-signal $x(t)$; (b) spectrum of the pulse carrier signal $c(t)$ in Figure 8.23; (c) spectrum of the modulated pulse train $y(t)$.

Comparing eq. (8.26) with eq. (7.6) and Figure 8.24 with Figure 7.3(c), we see that the spectrum of $y(t)$ is very similar in form to the spectrum resulting from sampling with a periodic impulse train, the only difference being the values of the Fourier coefficients of the pulse train. For the periodic impulse train used in Chapter 7, all of the Fourier coefficients are equal to $1/T$ in value, while for the pulse train $c(t)$ in Figure 8.23, the Fourier coefficients are given by eq. (8.25). Consequently, the replicas of $X(j\omega)$ do not overlap as long as $\omega_c > 2\omega_M$, which corresponds to the condition of the Nyquist sampling theorem. If this constraint is satisfied, then, as with impulse-train sampling, $x(t)$ can be recovered from $y(t)$ through the use of a lowpass filter with cutoff frequency greater than ω_M and less than $\omega_c - \omega_M$.

Note that the same conclusion holds for a wide variety of other pulselike carrier waveforms: If $c(t)$ is any periodic signal with Fourier transform as in eq. (8.24) for some set of Fourier coefficients a_k , then $Y(j\omega)$ is given by eq. (8.26). Then, as long as $\omega_c = 2\pi/T > 2\omega_M$, the replicas of $X(j\omega)$ do not overlap, allowing us to recover $x(t)$ by lowpass filtering, provided that the DC Fourier coefficient a_0 is nonzero. As shown in Problem 8.11, if a_0 is zero or unacceptably small, then, by using a bandpass filter to select one of the shifted replicas of $X(j\omega)$ with a larger value of a_k , we obtain a sinusoidal AM signal with a scaled version of $x(t)$ as the modulating signal. Using the demodulation methods described in Section 8.2, we can then recover $x(t)$.

8.5.2 Time-Division Multiplexing

Amplitude modulation with a pulse-train carrier is often used to transmit several signals over a single channel. As indicated in Figure 8.23, the modulated output signal $y(t)$ is nonzero only when the carrier signal $c(t)$ is on (i.e., is nonzero). During the intervals in which $c(t)$ is off, other similarly modulated signals can be transmitted. Two equivalent representations of this process are shown in Figure 8.25. In this technique for transmitting several signals over a single channel, each signal is in effect assigned a set of time slots of duration Δ that repeat every T seconds and that do not overlap with the slots assigned to other signals. The smaller the ratio Δ/T , the larger the number of signals that can be transmitted over the channel. This procedure is referred to as *time-division multiplexing* (TDM). Whereas frequency-division multiplexing, as discussed in Section 8.3, assigns different *frequency* intervals to individual signals, time-division multiplexing assigns different *time* intervals to individual signals. Demultiplexing the individual signals from the composite signal in Figure 8.25 is accomplished by time gating, to select the particular time slots associated with each individual signal.

8.6 PULSE-AMPLITUDE MODULATION

8.6.1 Pulse-Amplitude Modulated Signals

In Section 8.5 we described a modulation system in which a continuous-time signal $x(t)$ modulates a periodic pulse train, corresponding to transmitting time slices of $x(t)$ of duration Δ seconds every T seconds. As we saw both in that discussion and in our investigation of sampling in Chapter 7, our ability to recover $x(t)$ from these time slices depends not on their duration Δ , but rather on their frequency $2\pi/T$, which must exceed the Nyquist rate

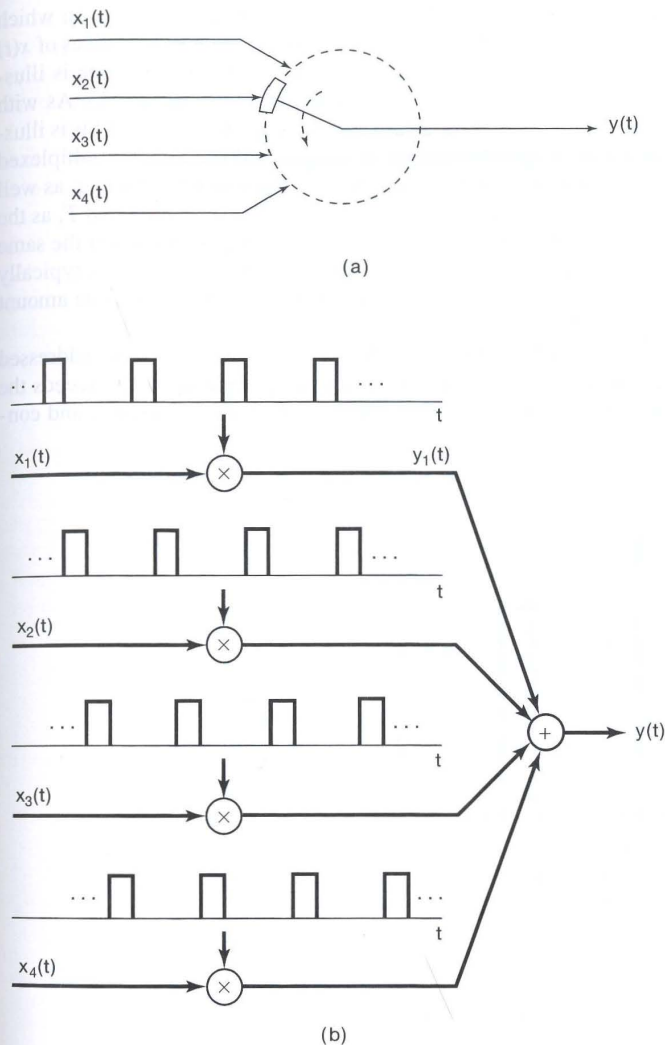


Figure 8.25 Time-division multiplexing.

in order to ensure an alias-free reconstruction of $x(t)$. That is, in principle, we need only transmit the samples $x(nT)$ of the signal $x(t)$.

In fact, in modern communication systems, sampled values of the information-bearing signal $x(t)$, rather than time slices are more typically transmitted. For practical reasons, there are limitations on the maximum amplitude that can be transmitted over a communication channel, so that transmitting impulse-sampled versions of $x(t)$ is not practical. Instead, the samples $x(nT)$ are used to modulate the amplitude of a sequence of pulses, resulting in what is referred to as a *pulse-amplitude modulation* (PAM) system.

The use of rectangular pulses corresponds to a sample-and-hold strategy in which pulses of duration Δ and amplitude proportional to the instantaneous sample values of $x(t)$ are transmitted. The resulting waveform for a single PAM channel of this type is illustrated in Figure 8.26. In the figure, the dotted curve represents the signal $x(t)$. As with the modulation scheme in Section 8.5, PAM signals can be time multiplexed. This is illustrated in Figure 8.27, which depicts the transmitted waveform with three time-multiplexed channels. The pulses associated with each channel are distinguished by shading, as well as by the channel number above each pulse. For a given pulse-repetition period T , as the pulse width decreases, more time-multiplexed channels can be transmitted over the same communication channel or medium. However, as the pulse width decreases, it is typically necessary to increase the amplitude of the transmitted pulses so that a reasonable amount of energy is transmitted in each pulse.

In addition to energy considerations, a number of other issues must be addressed in designing a PAM signal. In particular, as long as the sampling frequency exceeds the Nyquist rate, we know that $x(t)$ can be reconstructed exactly from its samples, and con-

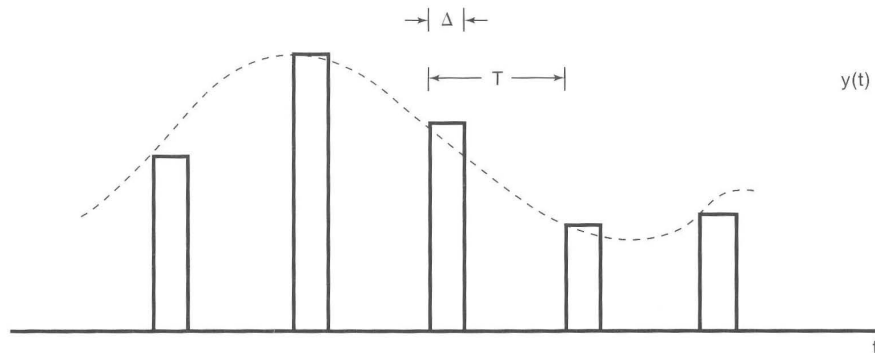


Figure 8.26 Transmitted waveform for a single PAM channel. The dotted curve represents the signal $x(t)$.

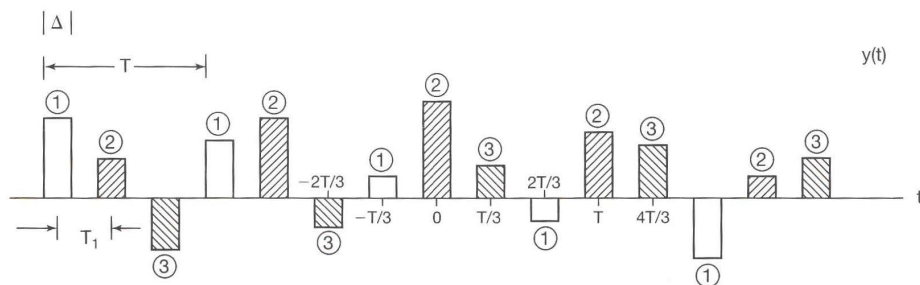


Figure 8.27 Transmitted waveform with three time-multiplexed PAM channels. The pulses associated with each channel are distinguished by shading, as well as by the channel number above each pulse. Here, the intersymbol spacing is $T_1 = T/3$.

sequently we can use these samples to modulate the amplitude of a sequence of pulses of any shape. The choice of pulse shape is dictated by considerations such as the frequency selectivity of the communication medium being used and the problem of intersymbol interference, which we discuss next.

8.6.2 Intersymbol Interference in PAM Systems

In the TDM pulse-amplitude modulation system just described, the receiver can, in principle, separate the channels by sampling the time-multiplexed waveform at appropriate times. For example, consider the time-multiplexed signal in Figure 8.27, which consists of pulse-amplitude-modulated versions of three signals $x_1(t)$, $x_2(t)$, and $x_3(t)$. If we sample $y(t)$ at appropriate times, corresponding, for example, to the midpoints of each pulse, we can separate the samples of the three signals. That is,

$$\begin{aligned} y(t) &= Ax_1(t), & t &= 0, \pm 3T_1, \pm 6T_1, \dots, \\ y(t) &= Ax_2(t), & t &= T_1, T_1 \pm 3T_1, T_1 \pm 6T_1, \dots, \\ y(t) &= Ax_3(t), & t &= 2T_1, 2T_1 \pm 3T_1, T_1 \pm 6T_1, \dots, \end{aligned} \quad (8.27)$$

where T_1 is the intersymbol spacing, here equal to $T/3$, and where A is the appropriate proportionality constant. In other words, samples of $x_1(t)$, $x_2(t)$, and $x_3(t)$ can be obtained by appropriate sampling of the received time-multiplexed PAM signal.

The strategy indicated in the preceding paragraph assumes that the transmitted pulses remain distinct as they propagate over the communication channel. In transmission through any realistic channel, however, the pulses can be expected to be distorted through effects such as additive noise and filtering. Additive noise in the channel will, of course, introduce amplitude errors at the sampling times. Filtering due to the nonideal frequency response of a channel causes a smearing of the individual pulses that can cause the received pulses to overlap in time. This interference is illustrated in Figure 8.28 and is referred to as *intersymbol interference*.

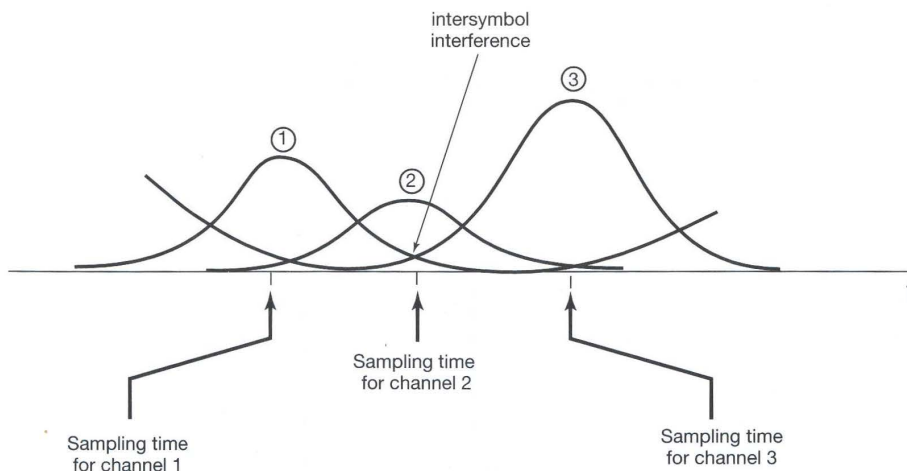


Figure 8.28 Intersymbol interference.

The smearing over time of the idealized pulses in Figure 8.27 can result from the bandwidth constraints of the channel or from phase dispersion caused by nonconstant group delay, as was discussed in Section 6.2.2. (See in particular, Example 6.1.) If the intersymbol interference is due only to the limited bandwidth of the channel, an approach is to use a pulse shape $p(t)$ that is itself band limited and therefore not affected (or only minimally affected) by the restricted bandwidth of the channel. In particular, if the channel has a frequency response $H(j\omega)$ that has no distortion over a specified frequency band (e.g., if $H(j\omega) = 1$ for $|\omega| < W$), then if the pulse that is used is band limited (i.e., if $P(j\omega) = 0$ for $|\omega| \geq W$), each PAM signal will be received without distortion. On the other hand, by using such a pulse, we no longer have pulses without overlap as in Figure 8.27. Nevertheless, intersymbol interference can be avoided in the time domain, even with a band-limited pulse, if the pulse shape is constrained to have zero-crossings at the other sampling times [so that eq. (8.27) continues to hold]. For example, consider the sinc pulse

$$p(t) = \frac{T_1 \sin(\pi t/T_1)}{\pi t}$$

and its corresponding spectrum displayed in Figure 8.29. Since the pulse is zero at integer multiples of the symbol spacing T_1 , as indicated in Figure 8.30, there will be no intersymbol interference at these instants. That is, if we sample the received signal at $t = kT_1$, then the contributions to this sampled value from all of the other pulses, i.e., from $p(t - mT_1)$ for $m \neq k$, will be identically zero. Of course, avoiding interference from adjacent symbols

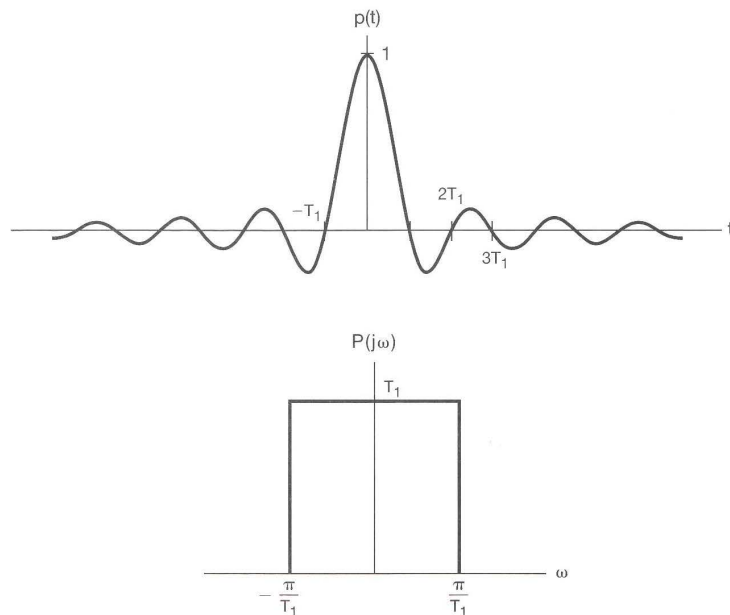


Figure 8.29 A sinc pulse and its corresponding spectrum.

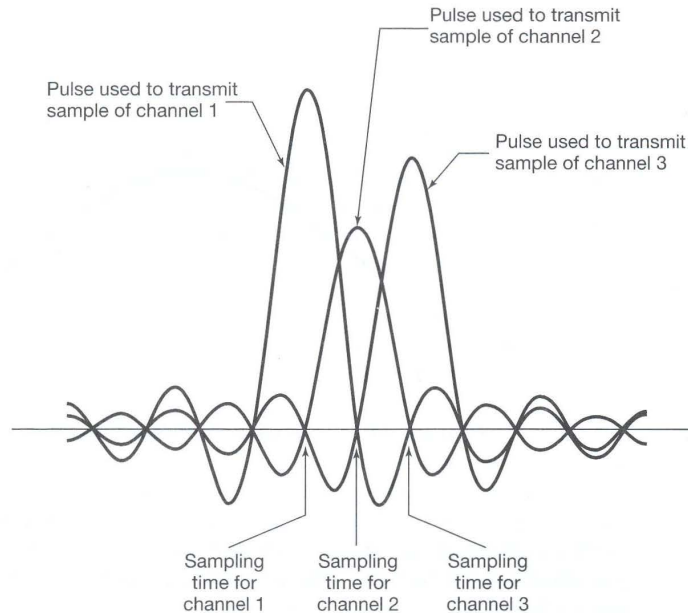


Figure 8.30 Absence of intersymbol interference when sinc pulses with correctly chosen zero-crossings are used.

requires high accuracy in the sampling times, so that sampling occurs at the zero-crossings of the adjacent symbols.

The sinc pulse is only one of many band-limited pulses with time-domain zero-crossings at $\pm T_1$, $\pm 2T_1$, etc. More generally, consider a pulse $p(t)$ with spectrum of the form

$$P(j\omega) = \begin{cases} 1 + P_1(j\omega), & |\omega| \leq \frac{\pi}{T_1}, \\ P_1(j\omega), & \frac{\pi}{T_1} < |\omega| \leq \frac{2\pi}{T_1}, \\ 0, & \text{otherwise} \end{cases} \quad (8.28)$$

and with $P_1(j\omega)$ having odd symmetry around π/T_1 , so that

$$P_1\left(-j\omega + j\frac{\pi}{T_1}\right) = -P_1\left(j\omega + j\frac{\pi}{T_1}\right) \quad 0 \leq \omega \leq \frac{\pi}{T_1}, \quad (8.29)$$

as illustrated in Figure 8.31. If $P_1(j\omega) = 0$, $p(t)$ is the sinc pulse itself. More generally, as explored in Problem 8.42, for any $P(j\omega)$ satisfying the conditions in eqs. (8.28) and (8.29), $p(t)$ will have zero-crossing at $\pm T_1$, $\pm 2T_1$, ...

While signals satisfying eqs. (8.28) and (8.29) allow us to overcome the problem of limited channel bandwidth, other channel distortions may occur that require a different choice of pulse waveform or some additional processing of the received signal prior to the separation of the different TDM signals. In particular, if $|H(j\omega)|$ is not constant over the passband, there may be a need to perform *channel equalization*—i.e., filtering of

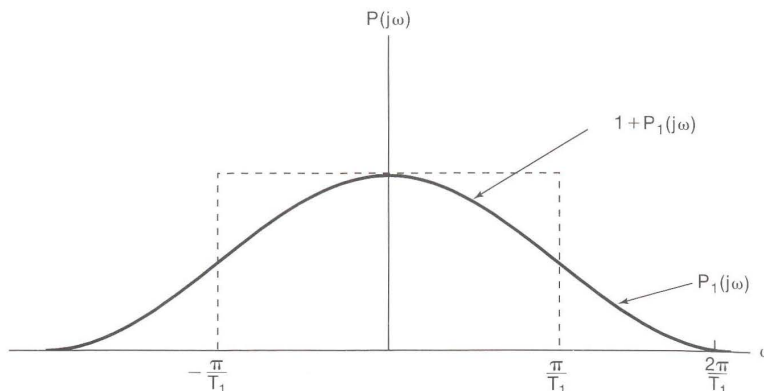


Figure 8.31 Odd symmetry around π/T_1 as defined in eq. (8.29).

the received signal to correct for the nonconstant channel gain. Also, if the channel has nonlinear phase, distortion can result that leads to intersymbol interference, unless compensating signal processing is performed. Problems 8.43 and 8.44 provide illustrations of these effects.

8.6.3 Digital Pulse-Amplitude and Pulse-Code Modulation

The PAM system described in the preceding subsections involves the use of a discrete set of samples to modulate a sequence of pulses. This set of samples can be thought of as a discrete-time signal $x[n]$, and in many applications $x[n]$ is in fact stored in or generated by a digital system. In such cases, the limited word length of a digital system implies that $x[n]$ can take on only a finite, quantized set of values, resulting in only a finite set of possible amplitudes for the modulated pulses.

In fact, in many cases this quantized form of digital PAM is reduced to a system using only a few—typically, only two—amplitude values. In particular, if each sample of $x[n]$ is represented as a binary number (i.e., a finite string of 0's and 1's), then a pulse with one of two possible values (one value corresponding to a 0 and one value to a 1) can be set for each binary digit, or *bit*, in the string. More generally, in order to protect against transmission errors or provide secure communication, the sequence of binary digits representing $x[n]$ might first be transformed or *encoded* into another sequence of 0's and 1's before transmission. For example, a very simple error detection mechanism is to transmit one additional modulated pulse for each sample of $x[n]$, representing a *parity* check. That is, this additional bit would be set to 1 if the binary representation of $x[n]$ has an odd number of 1's in it and to 0 if there is an even number of 1's. The receiver can then check the received parity bit against the other received bits in order to detect inconsistencies. More complex coding and error correction schemes can certainly be employed, and the design of codes with particular desirable properties is an important component of communication system design. For obvious reasons, a PAM system modulated by an encoded sequence of 0's and 1's is referred to as a *pulse-code modulation* (PCM) system.

8.7 SINUSOIDAL FREQUENCY MODULATION

In the preceding sections, we discussed a number of specific amplitude modulation systems in which the modulating signal was used to vary the amplitude of a sinusoidal or a pulse carrier. As we have seen, such systems are amenable to detailed analysis using the frequency-domain techniques we developed in preceding chapters. In another very important class of modulation techniques referred to as *frequency modulation (FM)*, the modulating signal is used to control the frequency of a sinusoidal carrier. Modulation systems of this type have a number of advantages over amplitude modulation systems. As suggested by Figure 8.10, with sinusoidal amplitude modulation the peak amplitude of the envelope of the carrier is directly dependent on the amplitude of the modulating signal $x(t)$, which can have a large dynamic range—i.e., can vary significantly. With frequency modulation, the envelope of the carrier is constant. Consequently, an FM transmitter can always operate at peak power. In addition, in FM systems, amplitude variations introduced over a transmission channel due to additive disturbances or fading can, to a large extent, be eliminated at the receiver. For this reason, in public broadcasting and a variety of other contexts, FM reception is typically better than AM reception. On the other hand, as we will see, frequency modulation generally requires greater bandwidth than does sinusoidal amplitude modulation.

Frequency modulation systems are highly nonlinear and, consequently, are not as straightforward to analyze as are the amplitude modulation systems discussed in the preceding sections. However, the methods we have developed in earlier chapters do allow us to gain some understanding of the nature and operation of these systems.

We begin by introducing the general notion of *angle modulation*. Consider a sinusoidal carrier expressed in the form

$$c(t) = A \cos(\omega_c t + \theta_c) = A \cos \theta(t), \quad (8.30)$$

where $\theta(t) = \omega_c t + \theta_c$ and where ω_c is the frequency and θ_c the phase of the carrier. Angle modulation, in general, corresponds to using the modulating signal to change or vary the angle $\theta(t)$. One form that this sometimes takes is to use the modulating signal $x(t)$ to vary the phase θ_c so that the modulated signal takes the form

$$y(t) = A \cos[\omega_c t + \theta_c(t)], \quad (8.31)$$

where θ_c is now a function of time, specifically of the form

$$\theta_c(t) = \theta_0 + k_p x(t). \quad (8.32)$$

If $x(t)$ is, for example, constant, the phase of $y(t)$ will be constant and proportional to the amplitude of $x(t)$. Angle modulation of the form of eq. (8.31) is referred to as *phase modulation*. Another form of angle modulation corresponds to varying the *derivative* of the angle proportionally with the modulating signal; that is,

$$y(t) = A \cos \theta(t), \quad (8.33)$$

where

$$\frac{d\theta(t)}{dt} = \omega_c + k_f x(t). \quad (8.34)$$

For $x(t)$ constant, $y(t)$ is sinusoidal with a frequency that is offset from the carrier frequency ω_c by an amount proportional to the amplitude of $x(t)$. For that reason, angle modulation of the form of eqs. (8.33) and (8.34) is commonly referred to as *frequency modulation*.

Although phase modulation and frequency modulation are different forms of angle modulation, they can be easily related. From eqs. (8.31) and (8.32), for phase modulation,

$$\frac{d\theta(t)}{dt} = \omega_c + k_p \frac{dx(t)}{dt}, \quad (8.35)$$

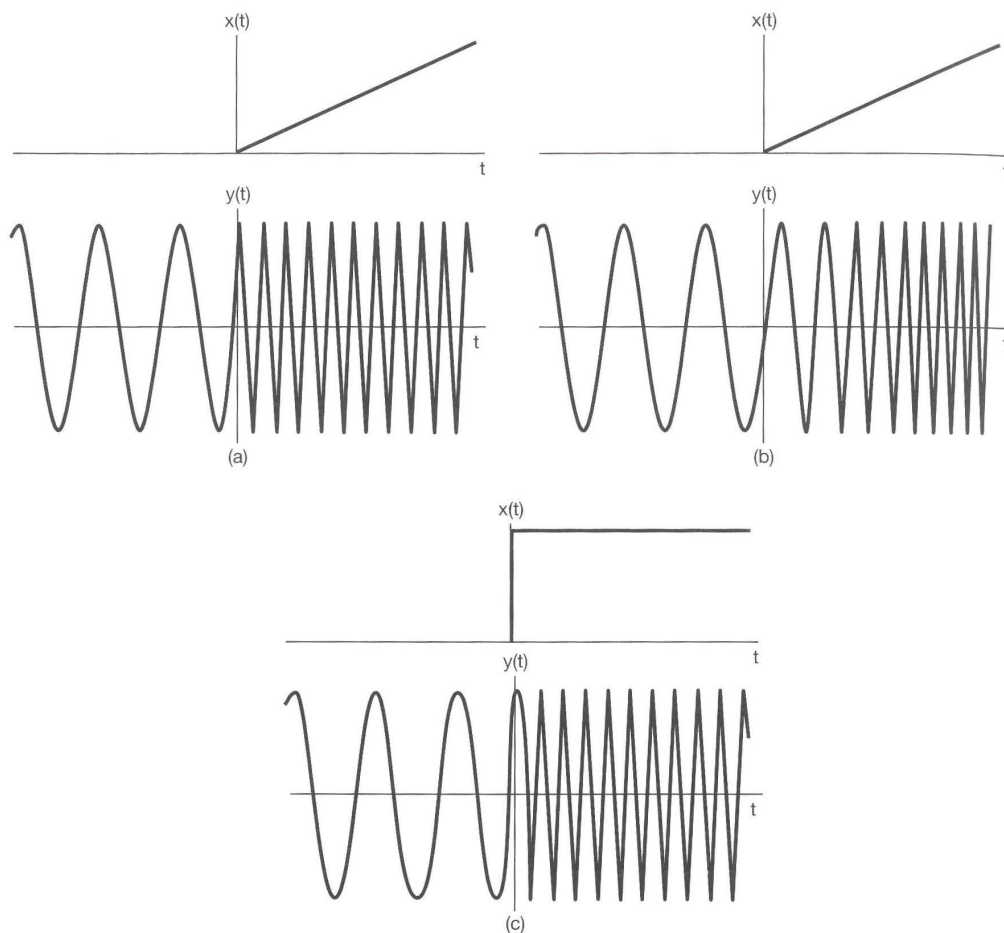


Figure 8.32 Phase modulation, frequency modulation, and their relationship: (a) phase modulation with a ramp as the modulating signal; (b) frequency modulation with a ramp as the modulating signal; (c) frequency modulation with a step (the derivative of a ramp) as the modulating signal.

and thus, comparing eqs. (8.34) and (8.35), we see that phase modulating with $x(t)$ is identical to frequency modulating with the derivative of $x(t)$. Likewise, frequency modulating with $x(t)$ is identical to phase modulating with the integral of $x(t)$. An illustration of phase modulation and frequency modulation is shown in Figures 8.32(a) and (b). In both cases, the modulating signal is $x(t) = tu(t)$ (i.e., a ramp signal increasing linearly with time for $t > 0$). In Figure 8.32(c), an example of frequency modulation is shown with a step (the derivative of a ramp) as the modulating signal [i.e., $x(t) = u(t)$]. The correspondence between Figures 8.32(a) and (c) should be evident.

Frequency modulation with a step corresponds to the frequency of the sinusoidal carrier changing instantaneously from one value to another when $x(t)$ changes value at $t = 0$, much as the frequency of a sinusoidal oscillator changes when the frequency setting is switched instantaneously. When the frequency modulation is a ramp, as in Figure 8.32(b), the frequency changes linearly with time. This notion of a time-varying frequency is often best expressed in terms of the concept of *instantaneous frequency*. For

$$y(t) = A \cos \theta(t), \quad (8.36)$$

the instantaneous frequency of the sinusoid is defined as

$$\omega_i(t) = \frac{d\theta(t)}{dt}. \quad (8.37)$$

Thus, for $y(t)$ truly sinusoidal [i.e., $\theta(t) = (\omega_c t + \theta_0)$], the instantaneous frequency is ω_c , as we would expect. For phase modulation as expressed in eqs. (8.31) and (8.32), the instantaneous frequency is $\omega_c + k_p(dx(t)/dt)$, and for frequency modulation as expressed in eqs. (8.33) and (8.34), the instantaneous frequency is $\omega_c + k_f x(t)$.

Since frequency modulation and phase modulation are easily related, we will phrase the remaining discussion in terms of frequency modulation alone. To gain some insight into how the spectrum of the frequency-modulated signal is affected by the modulating signal $x(t)$, it is useful to consider two cases in which the modulating signal is sufficiently simple so that some of the essential properties of frequency modulation become evident.

8.7.1 Narrowband Frequency Modulation

Consider the case of frequency modulation with

$$x(t) = A \cos \omega_m t. \quad (8.38)$$

From eqs. (8.34) and (8.37), the instantaneous frequency is

$$\omega_i(t) = \omega_c + k_f A \cos \omega_m t, \quad (8.39)$$

which varies sinusoidally between $\omega_c + k_f A$ and $\omega_c - k_f A$. With

$$\Delta\omega = k_f A,$$

we have

$$\omega_i(t) = \omega_c + \Delta\omega \cos \omega_m t,$$

and

$$\begin{aligned} y(t) &= \cos\left[\omega_c t + \int x(t) dt\right] \\ &= \cos\left(\omega_c t + \frac{\Delta\omega}{\omega_m} \sin \omega_m t + \theta_0\right), \end{aligned} \quad (8.40)$$

where θ_0 is a constant of integration. For convenience we will choose $\theta_0 = 0$, so that

$$y(t) = \cos\left[\omega_c t + \frac{\Delta\omega}{\omega_m} \sin \omega_m t\right]. \quad (8.41)$$

The factor $\Delta\omega/\omega_m$, which we denote by m , is defined as the *modulation index* for frequency modulation. The properties of FM systems tend to be different, depending on whether the modulation index m is small or large. The case in which m is small is referred to as *narrowband FM*. In general, we can rewrite eq. (8.41) as

$$y(t) = \cos(\omega_c t + m \sin \omega_m t) \quad (8.42)$$

or

$$y(t) = \cos \omega_c t \cos(m \sin \omega_m t) - \sin \omega_c t \sin(m \sin \omega_m t). \quad (8.43)$$

When m is sufficiently small ($\ll \pi/2$), we can make the approximations

$$\cos(m \sin \omega_m t) \simeq 1, \quad (8.44)$$

$$\sin(m \sin \omega_m t) \simeq m \sin \omega_m t, \quad (8.45)$$

so that eq. (8.42) becomes

$$y(t) \simeq \cos \omega_c t - m(\sin \omega_m t)(\sin \omega_c t). \quad (8.46)$$

The spectrum of $y(t)$ based on this approximation is shown in Figure 8.33. We note that it has a similarity to AM-DSB/WC in that the carrier frequency is present in the spectrum and there are sidebands representing the spectrum of the modulating signal in eq. (8.38). However, in AM-DSB/WC the additional carrier injected is in phase with the modulated carrier, whereas, as we see in eq. (8.46) for the case of the narrowband FM, the carrier signal has a phase difference of $\pi/2$ in relation to the amplitude-modulated carrier. The waveforms corresponding to AM-DSB/WC and FM are also very different. Figure 8.34(a) illustrates the narrowband FM waveform corresponding to eq. (8.46). For comparison, Figure 8.34(b) shows the AM-DSB/WC signal

$$y_2(t) = \cos \omega_c t + m(\cos \omega_m t)(\cos \omega_c t). \quad (8.47)$$

For the narrowband FM signal of eq. (8.46), the bandwidth of the sidebands is equal to the bandwidth of the modulating signal, and in particular, although the approximation in the equation is based on assuming that $m \ll \pi/2$, the bandwidth of the sidebands is otherwise independent of the modulation index m (i.e., it depends only on the bandwidth of the modulating signal, not on its amplitude). A similar statement applies for narrowband FM with a more general modulating signal.

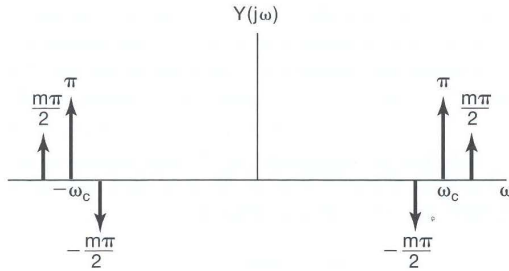


Figure 8.33 Approximate spectrum for narrowband FM.

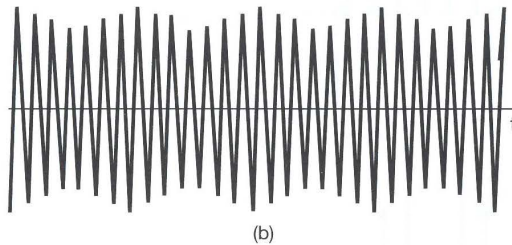
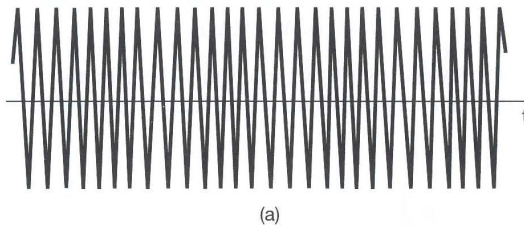


Figure 8.34 Comparison of narrowband FM and AM-DSB/WC: (a) narrowband FM; (b) AM-DSB/WC.

8.7.2 Wideband Frequency Modulation

When m is large, the approximation leading to eq. (8.46) no longer applies, and the spectrum of $y(t)$ depends on both the amplitude and the spectrum of the modulating signal $x(t)$. With $y(t)$ expressed in the form of eq. (8.43), we note that the terms $\cos[m \sin \omega_m t]$ and $\sin[m \sin \omega_m t]$ are periodic signals with fundamental frequency ω_m . Thus, the Fourier transform of each of these signals is an impulse train with impulses at integer multiples of ω_m and amplitudes proportional to the Fourier series coefficients. The coefficients for these two periodic signals involve a class of functions referred to as Bessel functions of the first kind. The first term in eq. (8.43) corresponds to a sinusoidal carrier of the form $\cos \omega_c t$ amplitude modulated by the periodic signal $\cos[m \sin \omega_m t]$ and the second term to a sinusoidal carrier $\sin \omega_c t$ amplitude modulated by the periodic signal $\sin[m \sin \omega_m t]$. Multiplication by the carrier signals has the effect in the frequency domain of translating the spectrum of eq. (8.43) to the carrier frequency, so that it is centered at plus and minus

ω_c . In Figures 8.35(a) and (b) we illustrate, for $\omega > 0$, the magnitude of the spectra of the two individual terms in eq. (8.43), and in Figure 8.35(c) the magnitude of the combined spectrum representing the modulated signal $y(t)$. The spectrum of $y(t)$ consists of impulses at frequencies $\pm\omega_c + n\omega_m$, $n = 0, \pm 1, \pm 2, \dots$, and is not, strictly speaking, band limited around $\pm\omega_c$. However, the behavior of the Fourier series coefficients of $\cos[m \sin \omega_m t]$ and $\sin[m \sin \omega_m t]$ are such that the amplitude of the n th harmonic for $|n| > m$ can be considered negligible, and thus, the total bandwidth B of each sideband centered around $+\omega_c$ and $-\omega_c$ is effectively limited to $2m\omega_m$. That is,

$$B \approx 2m\omega_m, \quad (8.48)$$

or, since $m = k_f A / \omega_m = \Delta\omega / \omega_m$,

$$B \approx 2k_f A = 2\Delta\omega. \quad (8.49)$$

Comparing eqs. (8.39) and (8.49), we note that the effective bandwidth of each sideband is equal to the total excursion of the instantaneous frequency around the carrier

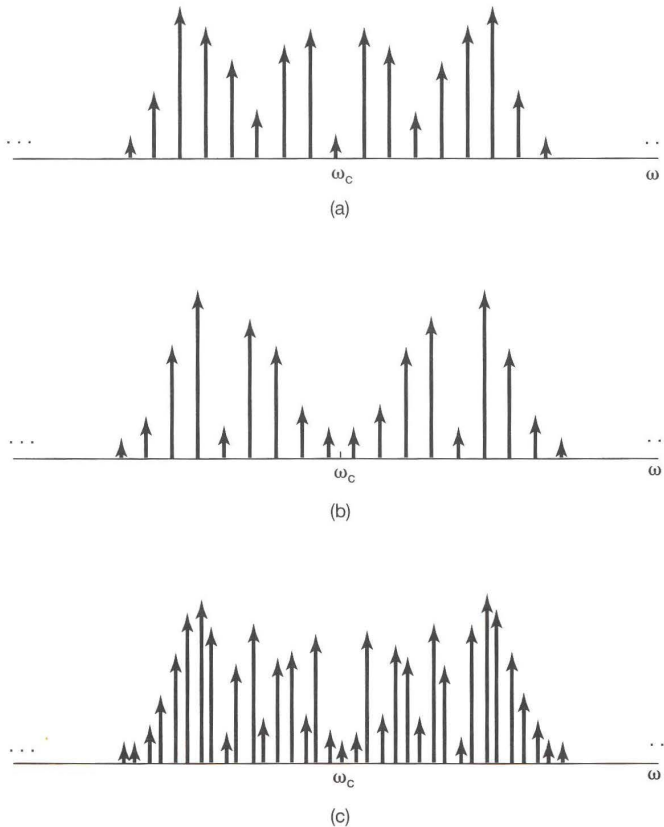


Figure 8.35 Magnitude of spectrum of wideband frequency modulation with $m = 12$: (a) magnitude of spectrum of $\cos \omega_c t \cos[m \sin \omega_m t]$; (b) magnitude of spectrum of $\sin \omega_c t \sin[m \sin \omega_m t]$; (c) combined spectral magnitude of $\cos[\omega_c t + m \sin \omega_m t]$.

frequency. Therefore, for wideband FM, since we assume that m is large, the bandwidth of the modulated signal is much larger than the bandwidth of the modulating signal, and in contrast to the narrowband case, the bandwidth of the transmitted signal in wideband FM is directly proportional to amplitude A of the modulating signal and the gain factor k_f .

8.7.3 Periodic Square-Wave Modulating Signal

Another example that lends insight into the properties of frequency modulation is that of a modulating signal which is a periodic square wave. Referring to eq. (8.39), let $k_f = 1$ so that $\Delta\omega = A$, and let $x(t)$ be given by Figure 8.36. The modulated signal $y(t)$ is illustrated in Figure 8.37. The instantaneous frequency is $\omega_c + \Delta\omega$ when $x(t)$ is positive and $\omega_c - \Delta\omega$ when $x(t)$ is negative. Thus, $y(t)$ can also be written as

$$y(t) = r(t) \cos[(\omega_c + \Delta\omega)t] + r\left(t - \frac{T}{2}\right) \cos[(\omega_c - \Delta\omega)t], \quad (8.50)$$

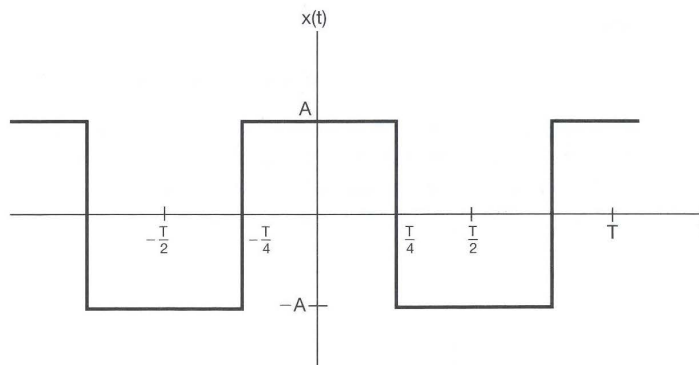


Figure 8.36 Symmetric periodic square wave.

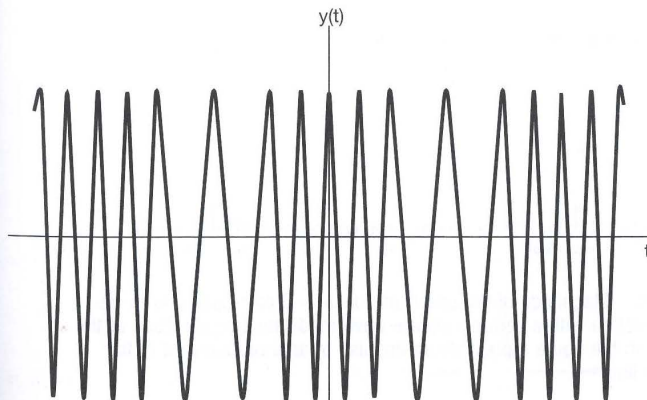


Figure 8.37 Frequency modulation with a periodic square-wave modulating signal.

where $r(t)$ is the symmetric square wave shown in Figure 8.38. Thus, for this particular modulating signal, we are able to recast the problem of determining the spectrum of the FM signal $y(t)$ as the determination of the spectrum of the sum of the two AM signals in eq. (8.50). Specifically,

$$Y(j\omega) = \frac{1}{2}[R(j\omega + j\omega_c + j\Delta\omega) + R(j\omega - j\omega_c - j\Delta\omega)] + \frac{1}{2}[R_T(j\omega + j\omega_c - j\Delta\omega) + R_T(j\omega - j\omega_c + j\Delta\omega)], \quad (8.51)$$

where $R(j\omega)$ is the Fourier transform of the periodic square wave $r(t)$ in Figure 8.38 and $R_T(j\omega)$ is the Fourier transform of $r(t - T/2)$. From Example 4.6, with $T = 4T_1$,

$$R(j\omega) = \sum_{k=-\infty}^{\infty} \frac{2}{2k+1} (-1)^k \delta\left[\omega - \frac{2\pi(2k+1)}{T}\right] + \pi\delta(\omega) \quad (8.52)$$

and

$$R_T(j\omega) = R(j\omega)e^{-j\omega T/2}. \quad (8.53)$$

The magnitude of the spectrum of $Y(j\omega)$ is illustrated in Figure 8.39. As with wideband FM, the spectrum has the general appearance of two sidebands, centered around $\omega_c \pm \Delta\omega$, that decay for $\omega < \omega_c - \Delta\omega$ and $\omega > \omega_c + \Delta\omega$.

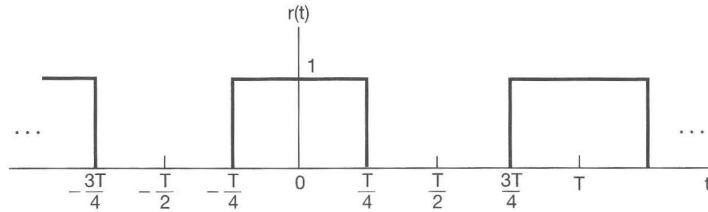


Figure 8.38 Symmetric square wave $r(t)$ in eq. (8.50).

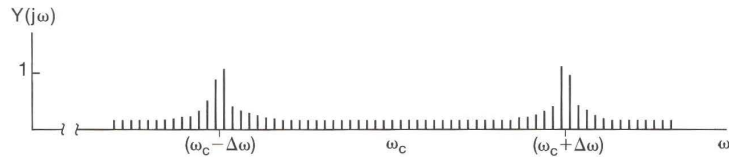


Figure 8.39 Magnitude of the spectrum for $\omega > 0$ corresponding to frequency modulation with a periodic square-wave modulating signal. Each of the vertical lines in the figure represents an impulse of area proportional to the height of the line.

Systems for the demodulation of FM signals typically are of two types. One type of demodulation system corresponds to converting the FM signal to an AM signal through differentiation, while demodulation systems of the second type directly track the phase or frequency of the modulated signal. The foregoing discussion provides only a brief introduction to the characteristics of frequency modulation, and we have again seen how the basic techniques developed in the earlier chapters can be exploited to analyze and gain an insight into an important class of systems.

8.8 DISCRETE-TIME MODULATION

8.8.1 Discrete-Time Sinusoidal Amplitude Modulation

A discrete-time amplitude modulation system is depicted in Figure 8.40, in which $c[n]$ is the carrier and $x[n]$ the modulating signal. The basis for our analysis of continuous-time amplitude modulation was the multiplication property for Fourier transforms—specifically, the fact that multiplication in the time domain corresponds to convolution in the frequency domain. As we discussed in Section 5.5, there is a corresponding property for discrete-time signals which we can use to analyze discrete-time amplitude modulation. Specifically, consider

$$y[n] = x[n]c[n].$$

With $X(e^{j\omega})$, $Y(e^{j\omega})$, and $C(e^{j\omega})$ denoting the Fourier transforms of $x[n]$, $y[n]$, and $c[n]$, respectively, $Y(e^{j\omega})$ is proportional to the periodic convolution of $X(e^{j\omega})$ and $C(e^{j\omega})$; that is,

$$Y(e^{j\omega}) = \frac{1}{2\pi} \int_{2\pi} X(e^{j\theta}) C(e^{j(\omega-\theta)}) d\theta. \quad (8.54)$$

Since $X(e^{j\omega})$ and $C(e^{j\omega})$ are periodic with a period of 2π , the integration can be performed over any frequency interval of length 2π .

Let us first consider sinusoidal amplitude modulation with a complex exponential carrier, so that

$$c[n] = e^{j\omega_c n}. \quad (8.55)$$

As we saw in Section 5.2, the Fourier transform of $c[n]$ is a periodic impulse train; that is,

$$C(e^{j\omega}) = \sum_{k=-\infty}^{+\infty} 2\pi \delta(\omega - \omega_c + k2\pi), \quad (8.56)$$

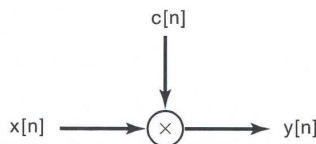


Figure 8.40 Discrete-time amplitude modulation.

which is sketched in Figure 8.41(b). With $X(e^{j\omega})$ as illustrated in Figure 8.41(a), the spectrum of the modulated signal is that shown in Figure 8.41(c). In particular, we note that $Y(e^{j\omega}) = X(e^{j(\omega - \omega_c)})$. This is the discrete-time counterpart to Figure 8.1, and here again, with $x[n]$ real, the modulated signal will be complex. Demodulation is accomplished by multiplying by $e^{-j\omega_c n}$ to translate the spectrum back to its original location on the frequency axis, so that

$$x[n] = y[n]e^{-j\omega_c n}. \quad (8.57)$$

As explored in Problem 6.43, if $\omega_c = \pi$ so that $c[n] = (-1)^n$, the result of modulation in the time domain is to change the algebraic sign of $x[n]$ for odd values of n , while in the frequency domain the consequence is the interchanging of high and low frequencies. Problem 6.44 explores the use of this type of modulation in utilizing a lowpass filter to achieve highpass filtering and vice versa.

As an alternative to a complex exponential carrier, we can use a sinusoidal carrier, in which case, with $x[n]$ real, the modulated signal $y[n]$ will also be real. With $c[n] = \cos \omega_c n$, the spectrum of the carrier consists of periodically repeated pairs of impulses at

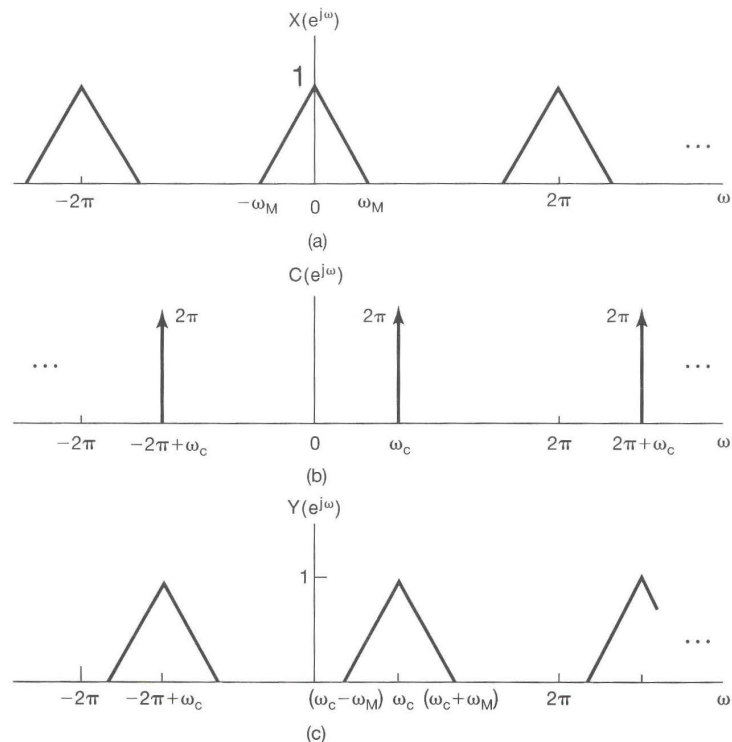


Figure 8.41 (a) Spectrum of $x[n]$; (b) spectrum of $c[n] = e^{j\omega_c n}$; (c) spectrum of $y[n] = x[n]c[n]$.

$\omega = \pm\omega_c + k2\pi$, as illustrated in Figure 8.42(b). With $X(e^{j\omega})$ as shown in Figure 8.42(a), the resulting spectrum for the modulated signal is shown in Figure 8.42(c) and corresponds to replicating $X(e^{j\omega})$ at the frequencies $\omega = \pm\omega_c + k2\pi$. In order that the individual replications of $X(e^{j\omega})$ do not overlap, we require that

$$\omega_c > \omega_M \quad (8.58)$$

and

$$2\pi - \omega_c - \omega_M > \omega_c + \omega_M$$

or, equivalently,

$$\omega_c < \pi - \omega_M. \quad (8.59)$$

The first condition is identical to that in Section 8.2 for continuous-time sinusoidal amplitude modulation, while the second results from the inherent periodicity of discrete-time

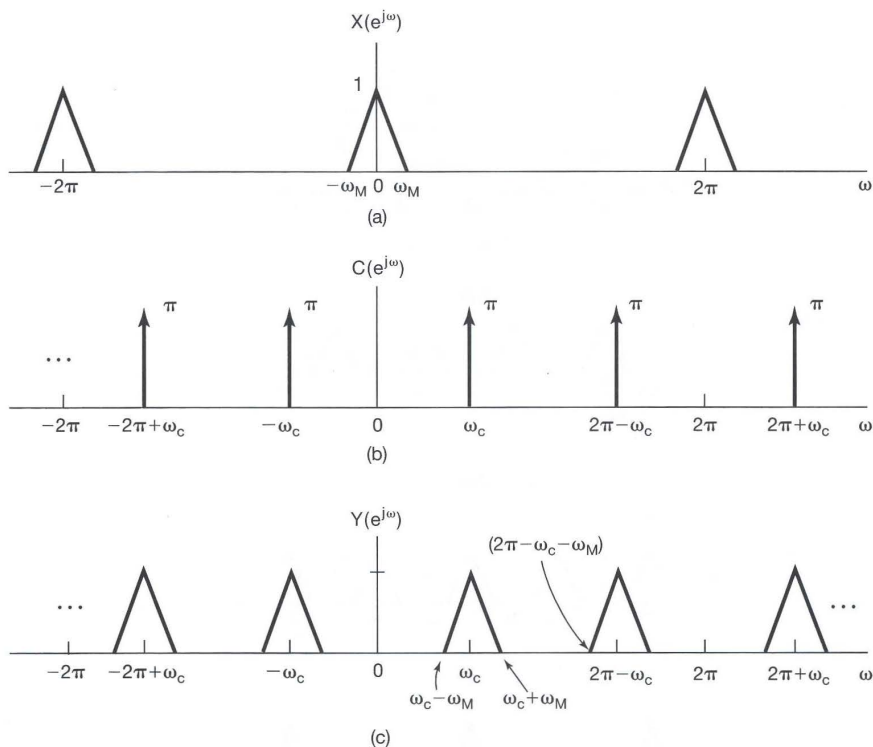


Figure 8.42 Spectra associated with discrete-time modulation using a sinusoidal carrier: (a) spectrum of a bandlimited-signal $x[n]$; (b) spectrum of a sinusoidal carrier signal $c[n] = \cos \omega_c n$; (c) spectrum of the modulated signal $y[n] = x[n]c[n]$.

spectra. Combining eqs. (8.58) and (8.59), we see that for amplitude modulation with a sinusoidal carrier, we must restrict ω_c so that

$$\omega_M < \omega_c < \pi - \omega_M. \quad (8.60)$$

Demodulation can be accomplished in a manner similar to that employed in continuous time. As illustrated in Figure 8.43, multiplication of $y[n]$ with the same carrier used in the modulator results in several replications of the spectrum of the original signal, one of which is centered about $\omega = 0$. By lowpass filtering to eliminate the unwanted replications of $X(e^{j\omega})$, the demodulated signal is obtained.

As should be evident from the foregoing discussion, analysis of discrete-time amplitude modulation proceeds in a manner similar to that of continuous-time amplitude modulation, with only slight differences. For example, as explored in Problem 8.47, in the synchronous modulation and demodulation system, the effect of a phase difference or a frequency difference between the sinusoidal carriers in the modulator and demodulator is identical in both discrete and continuous time. In addition, just as in continuous time, we can use discrete-time sinusoidal AM as the basis for frequency-division multiplexing in

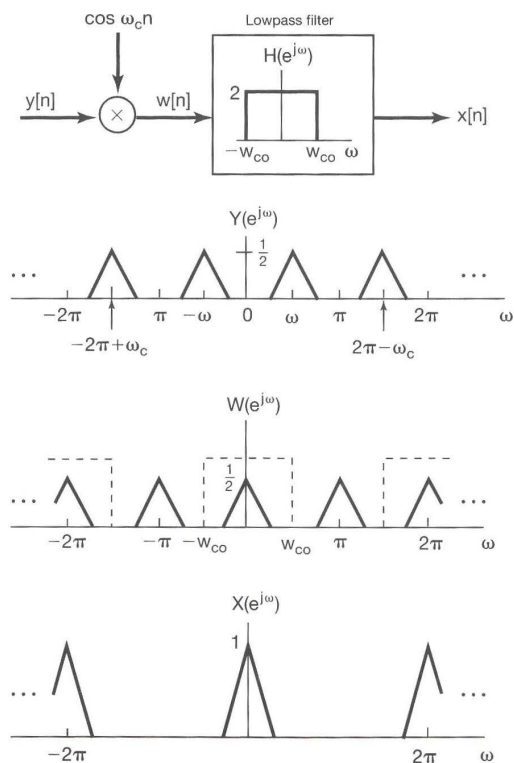


Figure 8.43 System and associated spectra for discrete-time synchronous demodulation.

discrete time. Furthermore, as explored in Problem 8.48, we can also consider using a discrete-time signal to modulate a pulse train, leading to time-division multiplexing of discrete-time signals.

The implementation of discrete-time multiplexing systems provides an excellent example of the flexibility of discrete-time processing in general and the importance of the operation of upsampling (see Section 7.5.2) in particular. Consider a discrete-time FDM system with M sequences that we wish to frequency-division multiplex. With M channels, it is required that the spectral energy for each input channel $x_i[n]$ be band limited; that is,

$$X_i(\omega) = 0, \quad \frac{\pi}{M} < |\omega| < \pi. \quad (8.61)$$

If the sequences originally occupied the entire frequency band corresponding, for example, to having sampled a set of continuous-time signals at the Nyquist rate, then they would first have to be converted to a higher sampling rate (i.e., upsampled) before frequency-division multiplexing. This idea is explored further in Problem 8.33.

8.8.2 Discrete-Time Transmodulation

One context in which discrete-time modulation is widely used, together with the operations of decimation, upsampling, and interpolation introduced in Chapter 7, is digital communication systems. Typically, in such systems continuous-time signals are transmitted over communication channels in the form of discrete-time signals, obtained by sampling. The continuous-time signals are often in the form of time-division-multiplexed (TDM) or frequency-division-multiplexed (FDM) signals. The signals are then converted to discrete-time sequences whose values are represented digitally, for storage or long-distance transmission. In some systems, because of different constraints or requirements at the transmitting end and the receiving end, or because sets of signals that have been individually multiplexed by different methods are then multiplexed together, there is often the requirement for converting from sequences representing TDM signals to sequences representing FDM signals or vice versa. This conversion from one modulation or multiplexing scheme to another is referred to as *transmodulation* or *transmultiplexing*. In the context of digital communication systems, one obvious way of implementing transmultiplexing is to convert back to continuous-time signals, demultiplex and demodulate, and then modulate and multiplex as required. However, if the new signal is then to be converted back to a discrete-time signal, it is clearly more efficient for the entire process to be carried out directly in the discrete-time domain. Figure 8.44 shows, in block diagram form, the steps involved in converting a discrete-time TDM signal to a discrete-time FDM signal. Note that, after demultiplexing the TDM signal, each channel must be upsampled in preparation for frequency-division multiplexing.

8.9 SUMMARY

In this chapter, we have examined a number of the basic concepts associated with communication systems. In particular, we have examined the concept of modulation, in which a signal we wish to communicate is used to modulate a second signal referred to as the

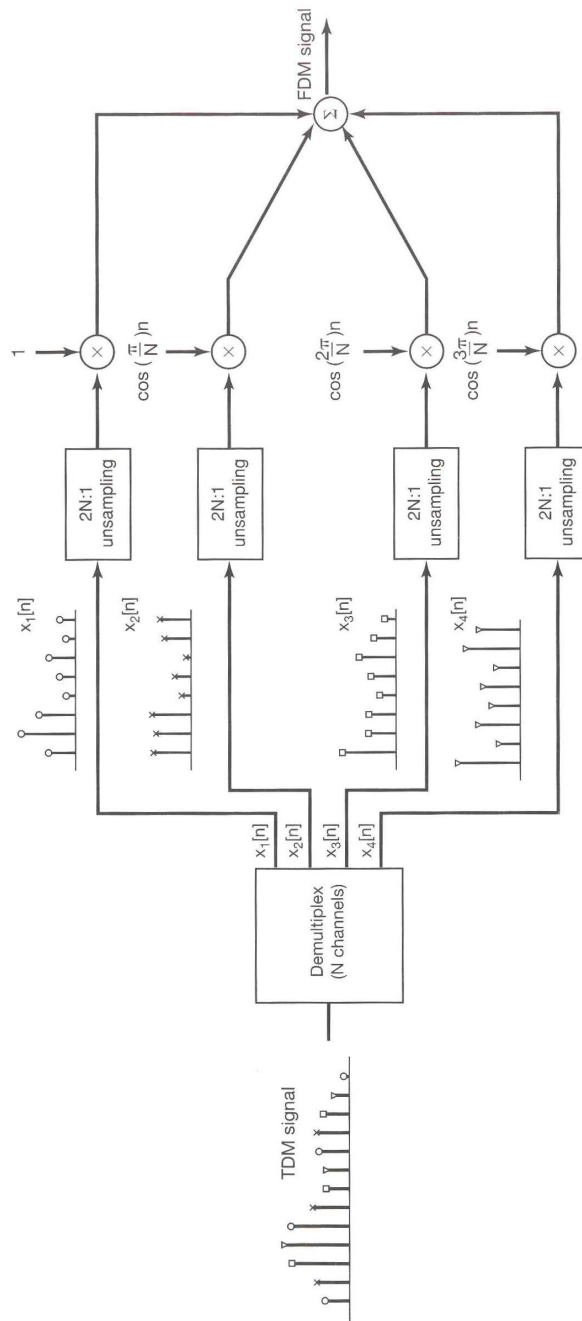


Figure 8.44 Block diagram for TDM-to-FDM transmultiplexing.

carrier, and we have looked in detail at the concept of amplitude modulation. The properties of amplitude modulation are most easily interpreted in the frequency domain through the multiplication property of the Fourier transform. Amplitude modulation with a complex exponential or sinusoidal carrier is typically used to shift the spectrum of a signal in frequency and is applied, for example, in communication systems to place the spectrum in a frequency range suitable for transmission and to permit frequency-division multiplexing. Variations of sinusoidal amplitude modulation, such as the insertion of a carrier signal for asynchronous systems and single- and double-sideband systems, were discussed.

We also examined several other forms of modulation-based communication. In this regard, we briefly introduced the concepts of frequency and phase modulation. Although these forms of modulation are more difficult to analyze in detail, it is possible to gain significant insight into their characteristics through the frequency domain.

We further examined in some detail amplitude modulation of a pulsed signal, which led us to the concepts of time-division multiplexing and pulse-amplitude modulation, in which successive samples of a discrete-time signal are used to modulate the amplitude of a sequence of pulses. This led in turn to an examination of discrete-time modulation and digital communication, in which the flexibility of discrete-time processing facilitates the design and implementation of more sophisticated communication systems involving concepts such as pulse-code modulation and transmodulation.

Chapter 8 Problems

The first section of problems belongs to the basic category, and the answers are provided in the back of the book. The remaining two sections contain problems belonging to the basic and advanced categories, respectively.

BASIC PROBLEMS WITH ANSWERS

- 8.1. Let $x(t)$ be a signal for which $X(j\omega) = 0$ when $|\omega| > \omega_M$. Another signal $y(t)$ is specified as having the Fourier transform $Y(j\omega) = 2X(j(\omega - \omega_c))$. Determine a signal $m(t)$ such that

$$x(t) = y(t)m(t).$$

- 8.2. Let $x(t)$ be a real-valued signal for which $X(j\omega) = 0$ when $|\omega| > 1,000\pi$. Supposing that $y(t) = e^{j\omega_c t} x(t)$, answer the following questions:
- (a) What constraint should be placed on ω_c to ensure that $x(t)$ is recoverable from $y(t)$?
 - (b) What constraint should be placed on ω_c to ensure that $x(t)$ is recoverable from $\Re\{y(t)\}$?