FUTURE VISION BIE

One Stop for All Study Materials

& Lab Programs

Future Vision

By K B Hemanth Raj

Scan the QR Code to Visit the Web Page

Or

Visit : https://hemanthrajhemu.github.io

Gain Access to All Study Materials according to VTU,
CSE – Computer Science Engineering,
ISE – Information Science Engineering,
ECE - Electronics and Communication Engineering
& MORE...

Join Telegram to get Instant Updates: https://bit.ly/VTU_TELEGRAM

Contact: MAIL: futurevisionbie@gmail.com

INSTAGRAM: www.instagram.com/hemanthraj_hemu/

INSTAGRAM: www.instagram.com/futurevisionbie/

WHATSAPP SHARE: https://bit.ly/FVBIESHARE

DIGITAL COMMUNICATION SYSTEMS

Simon Haykin
McMaster University

WILEY

Contents

CHAPTER
2

Fourier Analysis of Signals and Systems

2.1 Introduction

The study of communication systems involves:

- the processing of a modulated message signal generated at the transmitter output so as to facilitate its transportation across a physical channel and
- subsequent processing of the received signal in the receiver so as to deliver an estimate of the original message signal to a user at the receiver output.

In this study, the *representation of signals and systems* features prominently. More specifically, the *Fourier transform* plays a key role in this representation.

The Fourier transform provides the mathematical link between the time-domain representation (i.e., waveform) of a signal and its frequency-domain description (i.e., spectrum). Most importantly, we can go back and forth between these two descriptions of the signal with no loss of information. Indeed, we may invoke a similar transformation in the representation of linear systems. In this latter case, the time-domain and frequency-domain descriptions of a linear time-invariant system are defined in terms of its impulse response and frequency response, respectively.

In light of this background, it is in order that we begin a mathematical study of communication systems by presenting a review of Fourier analysis. This review, in turn, paves the way for the formulation of simplified representations of band-pass signals and systems to which we resort in subsequent chapters. We begin the study by developing the transition from the Fourier series representation of a periodic signal to the Fourier transform representation of a nonperiodic signal; this we do in the next two sections.

2.2 The Fourier Series

Let $g_{T_0}(t)$ denote a *periodic signal,* where the subscript T_0 denotes the duration of periodicity. By using a *Fourier series expansion* of this signal, we are able to resolve it into an infinite sum of sine and cosine terms, as shown by

$$g_{T_0}(t) = a_0 + 2\sum_{n=1}^{\infty} [a_n \cos(2\pi n f_0 t) + b_n \sin(2\pi n f_0 t)] \tag{2.1}$$

where

$$f_0 = \frac{1}{T_0} \tag{2.2}$$

is the *fundamental frequency*. The coefficients a_n and b_n represent the amplitudes of the cosine and sine terms, respectively. The quantity nf_0 represents the nth harmonic of the fundamental frequency f_0. Each of the terms $\cos(2\pi nf_0t)$ and $\sin(2\pi nf_0t)$ is called a *basis function*. These basis functions form an *orthogonal set* over the interval T_0, in that they satisfy three conditions:

$$\int_{-T_0/2}^{T_0/2} \cos(2\pi mf_0t)\cos(2\pi nf_0t)\,\mathrm{d}t = \begin{cases} T_0/2, & m = n \\ 0, & m \neq n \end{cases} \tag{2.3}$$

$$\int_{-T_0/2}^{T_0/2} \cos(2\pi mf_0t)\sin(2\pi nf_0t)\,\mathrm{d}t = 0, \qquad \text{for all } m \text{ and } n \tag{2.4}$$

$$\int_{-T_0/2}^{T_0/2} \sin(2\pi mf_0t)\sin(2\pi nf_0t)\,\mathrm{d}t = \begin{cases} T_0/2, & m = n \\ 0, & m \neq n \end{cases} \tag{2.5}$$

To determine the coefficient a_0, we integrate both sides of (2.1) over a complete period. We thus find that a_0 is the *mean value* of the periodic signal $g_{T_0}(t)$ over one period, as shown by the *time average*

$$a_0 = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} g_{T_0}(t)\,\mathrm{d}t \tag{2.6}$$

To determine the coefficient a_n, we multiply both sides of (2.1) by $\cos(2\pi nf_0t)$ and integrate over the interval $-T_0/2$ to $T_0/2$. Then, using (2.3) and (2.4), we find that

$$a_n = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} g_{T_0}(t)\cos(2\pi nf_0t)\,\mathrm{d}t, \qquad n = 1, 2, \ldots \tag{2.7}$$

Similarly, we find that

$$b_n = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} g_{T_0}(t)\sin(2\pi nf_0t)\,\mathrm{d}t, \qquad n = 1, 2, \ldots \tag{2.8}$$

A basic question that arises at this point is the following:

> Given a periodic signal $g_{T_0}(t)$ of period T_0, how do we know that the Fourier series expansion of (2.1) is *convergent* in that the infinite sum of terms in this expansion is exactly equal to $g_{T_0}(t)$?

To resolve this fundamental issue, we have to show that, for the coefficients a_0, a_n, and b_n calculated in accordance with (2.6) to (2.8), this series will indeed converge to $g_{T_0}(t)$. In general, for a periodic signal $g_{T_0}(t)$ of arbitrary waveform, there is no guarantee that the series of (2.1) will converge to $g_{T_0}(t)$ or that the coefficients a_0, a_n, and b_n will even exist. In a rigorous sense, we may say that a periodic signal $g_{T_0}(t)$ can be expanded in a Fourier

series if the signal $g_{T_0}(t)$ satisfies the *Dirichlet conditions*:[1]

1. The function $g_{T_0}(t)$ is single valued within the interval T_0.
2. The function $g_{T_0}(t)$ has at most a finite number of discontinuities in the interval T_0.
3. The function $g_{T_0}(t)$ has a finite number of maxima and minima in the interval T_0.
4. The function $g_{T_0}(t)$ is absolutely integrable; that is,

$$\int_{-T_0/2}^{T_0/2} \left| g_{T_0}(t) \right| \mathrm{d}t < \infty$$

From an engineering perspective, however, it suffices to say that the Dirichlet conditions are satisfied by the periodic signals encountered in communication systems.

Complex Exponential Fourier Series

The Fourier series of (2.1) can be put into a much simpler and more elegant form with the use of complex exponentials. We do this by substituting into (2.1) the exponential forms for the cosine and sine, namely:

$$\cos(2\pi n f_0 t) = \frac{1}{2}[\exp(\mathrm{j}2\pi n f_0 t) + \exp(-\mathrm{j}2\pi n f_0 t)]$$

$$\sin(2\pi n f_0 t) = \frac{1}{2\mathrm{j}}[\exp(\mathrm{j}2\pi n f_0 t) - \exp(-\mathrm{j}2\pi n f_0 t)]$$

where $\mathrm{j} = \sqrt{-1}$. We thus obtain

$$g_{T_0}(t) = a_0 + \sum_{n=1}^{\infty} [(a_n - \mathrm{j}b_n)\exp(\mathrm{j}2\pi n f_0 t) + (a_n + \mathrm{j}b_n)\exp(-\mathrm{j}2\pi n f_0 t)] \tag{2.9}$$

Let c_n denote a complex coefficient related to a_n and b_n by

$$c_n = \begin{cases} a_n - \mathrm{j}b_n, & n > 0 \\ a_0, & n = 0 \\ a_n + \mathrm{j}b_n, & n < 0 \end{cases} \tag{2.10}$$

Then, we may simplify (2.9) into

$$g_{T_0}(t) = \sum_{n=-\infty}^{\infty} c_n \exp(\mathrm{j}2\pi n f_0 t) \tag{2.11}$$

where

$$c_n = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} g_{T_0}(t)\exp(-\mathrm{j}2\pi n f_0 t)\mathrm{d}t, \qquad n = 0, \pm 1, \pm 2, \ldots \tag{2.12}$$

The series expansion of (2.11) is referred to as the *complex exponential Fourier series*. The c_n themselves are called the *complex Fourier coefficients*.

Given a periodic signal $g_{T_0}(t)$, (2.12) states that we may determine the complete set of complex Fourier coefficients. On the other hand, (2.11) states that, given this set of coefficients, we may reconstruct the original periodic signal $g_{T_0}(t)$ exactly.

The integral on the right-hand side of (2.12) is said to be an *inner product* of the signal $g_{T_0}(t)$ with the *basis functions* $\exp(-j2\pi nf_0t)$, by whose linear combination all square integrable functions can be expressed as in (2.11).

According to this representation, a periodic signal contains all frequencies (both positive and negative) that are harmonically related to the fundamental frequency f_0. The presence of negative frequencies is simply a result of the fact that the mathematical model of the signal as described by (2.11) requires the use of negative frequencies. Indeed, this representation also requires the use of complex-valued basis functions, namely $\exp(j2\pi nf_0t)$, which have no physical meaning either. The reason for using complex-valued basis functions and negative frequency components is merely to provide a compact mathematical description of a periodic signal, which is well-suited for both theoretical and practical work.

2.3 The Fourier Transform

In the previous section, we used the Fourier series to represent a periodic signal. We now wish to develop a similar representation for a signal $g(t)$ that is nonperiodic. In order to do this, we first construct a periodic function $g_{T_0}(t)$ of period T_0 in such a way that $g(t)$ defines exactly one cycle of this periodic function, as illustrated in Figure 2.1. In the limit, we let the period T_0 become infinitely large, so that we may express $g(t)$ as

$$g(t) = \lim_{T_0 \to \infty} g_{T_0}(t) \tag{2.13}$$

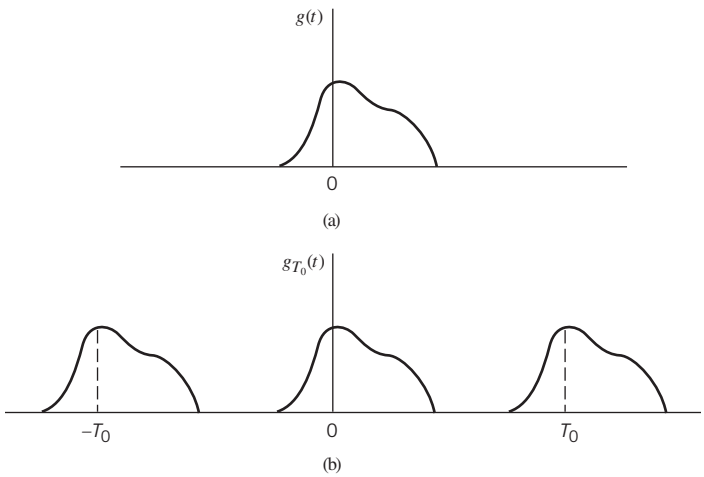


Figure 2.1 Illustrating the use of an arbitrarily defined function of time to construct a periodic waveform. (a) Arbitrarily defined function of time $g(t)$. (b) Periodic waveform $g_{T_0}(t)$ based on $g(t)$.

Representing the periodic function $g_{T_0}(t)$ in terms of the complex exponential form of the Fourier series, we write

$$g_{T_0}(t) = \sum_{n=-\infty}^{\infty} c_n \exp\left(\frac{\mathrm{j}2\pi nt}{T_0}\right)$$

where

$$c_n = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} g_{T_0}(t)\exp\left(-\frac{\mathrm{j}2\pi nt}{T_0}\right)\mathrm{d}t$$

Here, we have purposely replaced f_0 with $1/T_0$ in the exponents. Define

$$\Delta f = \frac{1}{T_0}$$

$$f_n = \frac{n}{T_0}$$

and

$$G(f_n) = c_n T_0$$

We may then go on to modify the original Fourier series representation of $g_{T_0}(t)$ given in (2.11) into a new form described by

$$g_{T_0}(t) = \sum_{n=-\infty}^{\infty} G(f_n)\exp(\mathrm{j}2\pi f_n t)\Delta f \tag{2.14}$$

where

$$G(f_n) = \int_{-T_0/2}^{T_0/2} g_{T_0}(t)\exp(-\mathrm{j}2\pi f_n t)\,\mathrm{d}t \tag{2.15}$$

Equations (2.14) and (2.15) apply to a periodic signal $g_{T_0}(t)$. What we would like to do next is to go one step further and develop a corresponding pair of formulas that apply to a nonperiodic signal $g(t)$. To do this transition, we use the defining equation (2.13). Specifically, two things happen:

1. The discrete frequency f_n in (2.14) and (2.15) approaches the continuous frequency variable f.
2. The discrete sum of (2.14) becomes an integral defining the area under the function $G(f)\exp(\mathrm{j}2\pi ft)$, integrated with respect to time t.

Accordingly, piecing these points together, we may respectively rewrite the limiting forms of (2.15) and (2.14) as

$$G(f) = \int_{-\infty}^{\infty} g(t)\exp(-\mathrm{j}2\pi ft)\mathrm{d}t \tag{2.16}$$

and

$$g(t) = \int_{-\infty}^{\infty} G(f)\exp(\mathrm{j}2\pi ft)\mathrm{d}f \tag{2.17}$$

In words, we may say:

- the *Fourier transform* of the nonperiodic signal $g(t)$ is defined by (2.16);
- given the Fourier transform $G(f)$, the original signal $g(t)$ is recovered exactly from the inverse *Fourier transform* of (2.17).

Figure 2.2 illustrates the interplay between these two formulas, where we see that the frequency-domain description based on (2.16) plays the role of *analysis* and the time-domain description based on (2.17) plays the role of *synthesis*.

From a notational point of view, note that in (2.16) and (2.17) we have used a lowercase letter to denote the time function and an uppercase letter to denote the corresponding frequency function. Note also that these two equations are of identical mathematical form, except for changes in the algebraic signs of the exponents.

For the Fourier transform of a signal $g(t)$ to exist, it is sufficient but not necessary that the nonperiodic signal $g(t)$ satisfies three *Dirichlet's conditions* of its own:

1. The function $g(t)$ is single valued, with a finite number of maxima and minima in any finite time interval.
2. The function $g(t)$ has a finite number of discontinuities in any finite time interval.
3. The function $g(t)$ is absolutely integrable; that is,

$$\int_{-\infty}^{\infty} |g(t)|\,dt < \infty$$

In practice, we may safely ignore the question of the existence of the Fourier transform of a time function $g(t)$ when it is an accurately specified description of a physically realizable signal. In other words, physical realizability is a sufficient condition for the existence of a Fourier transform. Indeed, we may go one step further and state:

> All energy signals are Fourier transformable.

A signal $g(t)$ is said to be an *energy signal* if the condition

$$\int_{-\infty}^{\infty} |g(t)|^2\,dt < \infty \tag{2.18}$$

holds.[2]

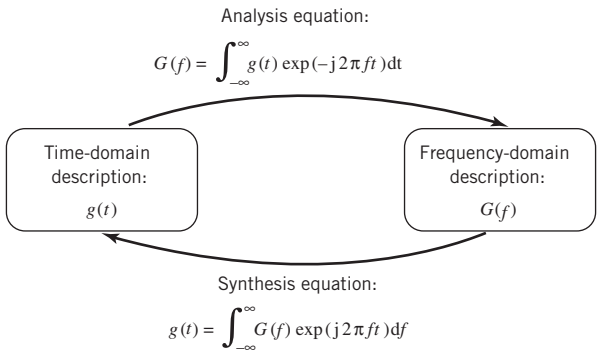


Figure 2.2 Sketch of the interplay between the synthesis and analysis equations embodied in Fourier transformation.

The Fourier transform provides the mathematical tool for measuring the frequency content, or spectrum, of a signal. For this reason, the terms *Fourier transform* and *spectrum* are used interchangeably. Thus, given a signal $g(t)$ with Fourier transform $G(f)$, we may refer to $G(f)$ as the spectrum of the signal $g(t)$. By the same token, we refer to $|G(f)|$ as the *magnitude spectrum* of the signal $g(t)$, and refer to $\arg[G(f)]$ as its *phase spectrum*.

If the signal $g(t)$ is real valued, then the magnitude spectrum of the signal is an even function of frequency f, while the phase spectrum is an odd function of f. In such a case, knowledge of the spectrum of the signal for positive frequencies uniquely defines the spectrum for negative frequencies.

Notations

For convenience of presentation, it is customary to express (2.17) in the short-hand form

$$G(f) = \mathbf{F}[g(t)]$$

where $\mathbf{F}$ plays the role of an *operator*. In a corresponding way, (2.18) is expressed in the short-hand form

$$g(t) = \mathbf{F}^{-1}[G(f)]$$

where $\mathbf{F}^{-1}$ plays the role of an *inverse operator*.

The time function $g(t)$ and the corresponding frequency function $G(f)$ are said to constitute a *Fourier-transform pair*. To emphasize this point, we write

$$g(t) \rightleftharpoons G(f)$$

where the top arrow indicates the forward transformation from $g(t)$ to $G(f)$ and the bottom arrow indicates the inverse transformation. One other notation: the asterisk is used to denote complex conjugation.

Tables of Fourier Tranformations

To assist the user of this book, two tables of Fourier transformations are included:

1. Table 2.1 on page 23 summarizes the properties of Fourier transforms; proofs of them are presented as end-of-chapter problems.
2. Table 2.2 on page 24 presents a list of Fourier-transform pairs, where the items listed on the left-hand side of the table are time functions and those in the center column are their Fourier transforms.

EXAMPLE 1

Binary Sequence for Energy Calculations

Consider the five-digit binary sequence 10010. This sequence is represented by two different waveforms, one based on the rectangular function $\text{rect}(t)$, and the other based on the sinc function $\text{sinc}(t)$. Despite this difference, both waveforms are denoted by $g(t)$, which implies they both have exactly the same total energy, to be demonstrated next.

Case 1: $\text{rect}(t)$ *as the basis function.*

Let binary symbol 1 be represented by $+\text{rect}(t)$ and binary symbol 0 be represented by $-\text{rect}(t)$. Accordingly, the binary sequence 10010 is represented by the waveform

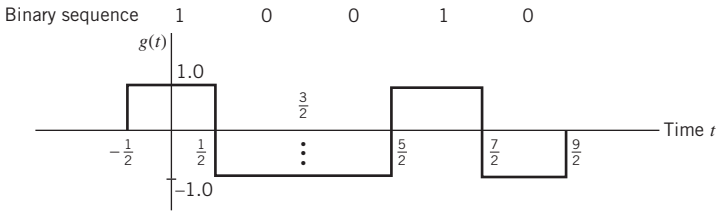


Figure 2.3 Waveform of binary sequence 10010, using rect(t) for symbol 1 and –rect(t) for symbol 0. See Table 2.2 for the definition of rect(t).

shown in Figure 2.3. From this figure, we readily see that, regardless of the representation ±rect(t), each symbol contributes a single unit of energy; hence the total energy for Case 1 is five units.

Case 2: sinc(t) *as the basis function.*

Consider next the representation of symbol 1 by +sinc(t) and the representation of symbol 0 by −sinc(t), which do not interfere with each other in constructing the waveform for the binary sequence 10010. Unfortunately, this time around, it is difficult to calculate the total waveform energy in the time domain. To overcome this difficulty, we do the calculation in the frequency domain.

To this end, in parts a and b of Figure 2.4, we display the waveform of the sinc function in the time domain and its Fourier transform, respectively. On this basis, Figure 2.5 displays the frequency-domain representation of the binary sequence 10010, with part a of the figure displaying the magnitude response $|G(f)|$, and part b displaying the corresponding phrase response $\arg[G(f)]$ expressed in radians. Then, applying Rayleigh's energy theorem, described in Property 14 in Table 2.2, to part a of Figure 2.5, we readily find that the energy of the pulse, ±sinc(t), is equal to one unit, regardless of its amplitude. The total energy of the sinc-based waveform representing the given binary sequence is also exactly five units, confirming what was said at the beginning of this example.

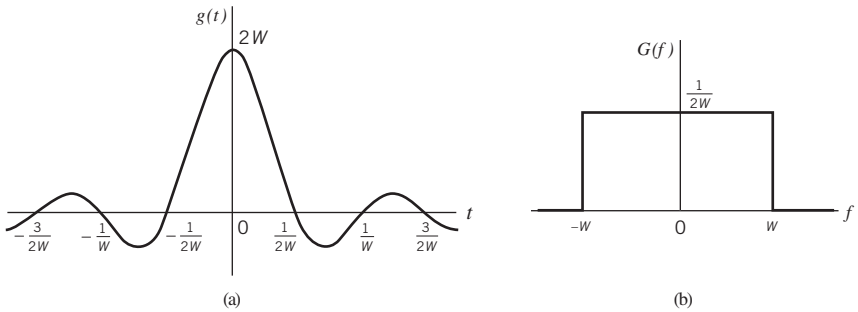


Figure 2.4 (a) Sinc pulse $g(t)$. (b) Fourier transform $G(f)$.

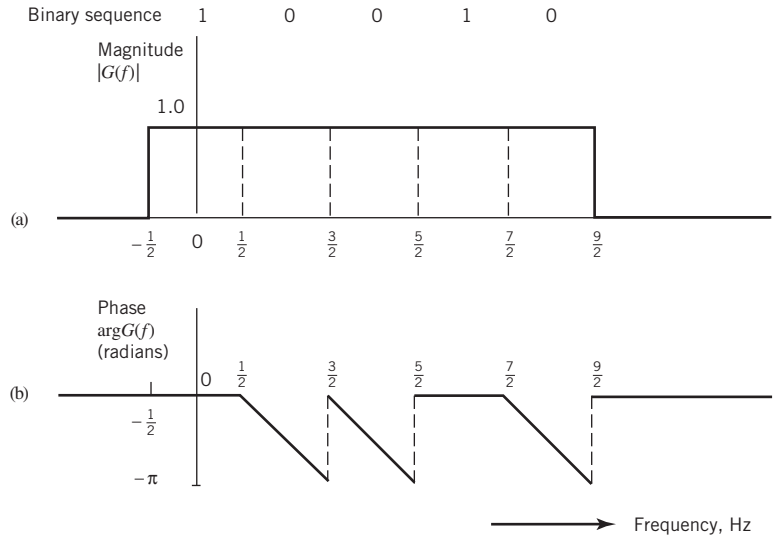


Figure 2.5 (a) Magnitude spectrum of the sequence 10010. (b) Phase spectrum of the sequence.

Observations

1. The dual basis functions, rect(t) and sinc(t), are *dilated* to their simplest forms, each of which has an energy of one unit, hence the equality of the results presented under Cases 1 and 2.
2. Examining the waveform $g(t)$ in Figure 2.3, we clearly see the discrimination between binary symbols 1 and 0. On the other hand, it is the phase response $\arg[G(f)]$ in part b of Figure 2.5 that shows the discrimination between binary symbols 1 and 0.

EXAMPLE 2

Unit Gaussian Pulse

Typically, a pulse signal $g(t)$ and its Fourier transform $G(f)$ have different mathematical forms. This observation is illustrated by the Fourier-transform pair studied in Example 1. In this second example, we consider an exception to this observation. In particular, we use the differentiation property of the Fourier transform to derive the particular form of a *pulse signal that has the same mathematical form as its own Fourier transform*.

Let $g(t)$ denote the pulse signal expressed as a function of time t and $G(f)$ denote its Fourier transform. Differentiating the Fourier transform formula of (2.6) with respect to frequency f yields

$$-\mathrm{j}2\pi t g(t) \rightleftharpoons \frac{\mathrm{d}}{\mathrm{d}f}G(f)$$

or, equivalently,

$$2\pi t g(t) \rightleftharpoons \mathrm{j}\frac{\mathrm{d}}{\mathrm{d}f}G(f) \tag{2.19}$$

Use of the Fourier-transform property on differentiation in the time domain listed in Table 2.1 yields

$$\frac{\mathrm{d}}{\mathrm{d}t}g(t) \rightleftharpoons \mathrm{j}2\pi f G(f) \tag{2.20}$$

Suppose we now impose the equality condition on the left-hand sides of (2.19) and (2.20):

$$\frac{\mathrm{d}}{\mathrm{d}t}g(t) = 2\pi t g(t) \tag{2.21}$$

Then, in a corresponding way, it follows that the right-hand sides of these two equations must (after canceling the common multiplying factor j) satisfy the condition

$$\frac{\mathrm{d}}{\mathrm{d}f}G(f) = 2\pi f G(f) \tag{2.22}$$

Equations (2.21) and (2.22) show that the pulse signal $g(t)$ and its Fourier transform $G(f)$ have exactly the same mathematical form. In other words, provided that the pulse signal $g(t)$ satisfies the differential equation (2.21), then $G(f) = g(f)$, where $g(f)$ is obtained from $g(t)$ simply by substituting f for t. Solving (2.21) for $g(t)$, we obtain

$$g(t) = \exp(-\pi t^2) \tag{2.23}$$

which has a bell-shaped waveform, as illustrated in Figure 2.6. Such a pulse is called a *Gaussian pulse*, the name of which follows from the similarity of the function $g(t)$ to the Gaussian probability density function of probability theory, to be discussed in Chapter 3. By applying the Fourier-transform property on the area under $g(t)$ listed in Table 2.1, we have

$$\int_{-\infty}^{\infty} \exp(-\pi t^2)\,\mathrm{d}t = 1 \tag{2.24}$$

When the central ordinate and the area under the curve of a pulse are both unity, as in (2.23) and (2.24), we say that the Gaussian pulse is a *unit pulse*. Therefore, we may state that the unit Gaussian pulse is its own Fourier transform, as shown by

$$\exp(-\pi t^2) \rightleftharpoons \exp(-\pi f^2) \tag{2.25}$$

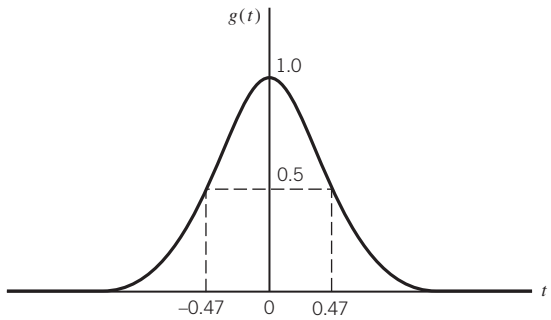


Figure 2.6 Gaussian pulse.

Table 2.1 **Fourier-transform theorems**

Property	Mathematical description
1. Linearity	$ag_1(t) + bg_2(t) \rightleftharpoons aG_1(f) + bG_2(f)$ where a and b are constants
2. Dilation	$g(at) \rightleftharpoons \frac{1}{\lvert a\rvert} G\left(\frac{f}{a}\right)$ where a is a constant
3. Duality	If $g(t) \rightleftharpoons G(f)$, then $G(t) \rightleftharpoons g(-f)$
4. Time shifting	$g(t-t_0) \rightleftharpoons G(f)\exp(-\mathrm{j}2\pi f t_0)$
5. Frequency shifting	$g(t)\exp(-\mathrm{j}2\pi f_0 t) \rightleftharpoons G(f-f_0)$
6. Area under $g(t)$	$\int_{-\infty}^{\infty} g(t)\mathrm{d}t = G(0)$
7. Area under $G(f)$	$g(0) = \int_{-\infty}^{\infty} G(f)\mathrm{d}f$
8. Differentiation in the time domain	$\frac{\mathrm{d}}{\mathrm{d}t}g(t) \rightleftharpoons \mathrm{j}2\pi f G(f)$
9. Integration in the time domain	$\int_{-\infty}^{t} g(\tau)\mathrm{d}\tau \rightleftharpoons \frac{1}{\mathrm{j}2\pi f}G(f) + \frac{G(0)}{2}\delta(f)$
10. Conjugate functions	If $g(t) \rightleftharpoons G(f)$, then $g^*(t) \rightleftharpoons G^*(-f)$
11. Multiplication in the time domain	$g_1(t)g_2(t) \rightleftharpoons \int_{-\infty}^{\infty} G_1(\lambda)G_2(f-\lambda)\mathrm{d}\lambda$
12. Convolution in the time domain	$\int_{-\infty}^{t} g_1(\tau)g_2(t-\tau)\mathrm{d}\tau \rightleftharpoons G_1(f)G_2(f)$
13. Correlation theorem	$\int_{-\infty}^{\infty} g_1(t)g_2^*(t-\tau)\mathrm{d}\tau \rightleftharpoons G_1(f)G_2^*(f)$
14. Rayleigh's energy theorem	$\int_{-\infty}^{\infty} \lvert g(t)\rvert^2\mathrm{d}t = \int_{-\infty}^{\infty} \lvert G(f)\rvert^2\mathrm{d}f$
15. Parseval's power theorem for periodic signal of period T_0	$\frac{1}{T_0}\int_{-T_0/2}^{T_0/2} \lvert g(t)\rvert^2\mathrm{d}t = \sum_{n=-\infty}^{\infty} \lvert G(f_n)\rvert^2, \quad f_n = n/T_0$

Table 2.2 Fourier-transform pairs and commonly used time functions

	Time function	Fourier transform
1.	$\text{rect}\left(\frac{t}{T}\right)$	$T\,\text{sinc}(fT)$
2.	$\text{sinc}(2Wt)$	$\frac{1}{2W}\text{rect}\left(\frac{f}{2Wf}\right)$
3.	$\exp(-at)u(t), \quad a>0$	$\frac{1}{a+\text{j}2\pi f}$
4.	$\exp(-a\lvert t\rvert), \quad a>0$	$\frac{2a}{a^2+(2\pi f)^2}$
5.	$\exp(-\pi t^2)$	$\exp(-\pi f^2)$
6.	$\begin{cases} 1-\frac{\lvert t\rvert}{T}, & \lvert t\rvert < T \\ 0, & \lvert t\rvert \geq T \end{cases}$	$T\,\text{sinc}^2(fT)$
7.	$\delta(t)$	1
8.	1	$\delta(f)$
9.	$\delta(t-t_0)$	$\exp(-\text{j}2\pi f t_0)$
10.	$\exp(\text{j}2\pi f_c t)$	$\delta(f-f_c)$
11.	$\cos(2\pi f_c t)$	$\frac{1}{2}[\delta(f-f_c)+\delta(f+f_c)]$
12.	$\sin(2\pi f_c t)$	$\frac{1}{2}[\delta(f-f_c)-\delta(f+f_c)]$
13.	$\text{sgn}(t)$	$\frac{1}{\text{j}\pi f}$
14.	$\frac{1}{\pi t}$	$-\text{j}\,\text{sgn}(f)$
15.	$u(t)$	$\frac{1}{2}\delta(f)+\frac{1}{\text{j}2\pi f}$
16.	$\sum_{i=-\infty}^{\infty}\delta(t-iT_0)$	$f_0\sum_{n=-\infty}^{\infty}\delta(f-nf_0), \quad f_0=\frac{1}{T_0}$

Definitions

Unit step function:

$$u(t) = \begin{cases} 1, & t>0 \\ \frac{1}{2}, & t=0 \\ 0, & t<0 \end{cases}$$

Dirac delta function:

$\delta(t) = 0$ for $t \neq 0$ and

$$\int_{-\infty}^{\infty}\delta(t)\text{d}t = 1$$

Rectangular function:

$$\text{rect}(t) = \begin{cases} 1, & -\frac{1}{2} < t \leq \frac{1}{2} \\ 0, & \text{otherwise} \end{cases}$$

Signum function:

$$\text{sgn}(t) = \begin{cases} +1, & t>0 \\ 0, & t=0 \\ -1, & t<0 \end{cases}$$

Sinc function:

$$\text{sinc}(t) = \frac{\sin(\pi t)}{\pi t}$$

Gaussian function:

$$g(t) = \exp(-\pi t^2)$$

2.4 The Inverse Relationship between Time-Domain and Frequency-Domain Representations

The time-domain and frequency-domain descriptions of a signal are *inversely* related. In this context, we may make four important statements:

1. If the time-domain description of a signal is changed, the frequency-domain description of the signal is changed in an *inverse* manner, and vice versa. This inverse relationship prevents arbitrary specifications of a signal in both domains. In other words:

   We may specify an arbitrary function of time or an arbitrary spectrum, but we cannot specify them both together.

2. If a signal is strictly limited in frequency, then the time-domain description of the signal will trail on indefinitely, even though its amplitude may assume a progressively smaller value. To be specific, we say:

   A signal is strictly limited in frequency (i.e., strictly band limited) if its Fourier transform is exactly zero outside a finite band of frequencies.

   Consider, for example, the band-limited sinc pulse defined by

   $$\text{sinc}(t) = \frac{\sin(\pi t)}{\pi t}$$

   whose waveform and spectrum are respectively shown in Figure 2.4: part a shows that the sinc pulse is *asymptotically limited in time* and part b of the figure shows that the sinc pulse is indeed *strictly band limited*, thereby confirming statement 2.

3. In a dual manner to statement 2, we say:

   If a signal is strictly limited in time (i.e., the signal is exactly zero outside a finite time interval), then the spectrum of the signal is infinite in extent, even though the magnitude spectrum may assume a progressively smaller value.

   This third statement is exemplified by a *rectangular pulse*, the waveform and spectrum of which are defined in accordance with item 1 in Table 2.2.

4. In light of the duality described under statements 2 and 3, we now make the final statement:

   A signal cannot be strictly limited in both time and frequency.

The Bandwidth Dilemma

The statements we have just made have an important bearing on the *bandwidth* of a signal, which provides a measure of the *extent of significant spectral content of the signal for positive frequencies*. When the signal is strictly band limited, the bandwidth is well defined. For example, the sinc pulse $\text{sinc}(2Wt)$ has a bandwidth equal to W. However, when the signal is not strictly band limited, as is often the case, we encounter difficulty in defining the bandwidth of the signal. The difficulty arises because the meaning of "significant" attached to the spectral content of the signal is mathematically imprecise. Consequently, there is no universally accepted definition of bandwidth. It is in this sense that we speak of the "bandwidth dilemma."

Nevertheless, there are some commonly used definitions for bandwidth, as discussed next. When the spectrum of a signal is symmetric with a main lobe bounded by well-defined nulls (i.e., frequencies at which the spectrum is zero), we may use the main lobe as the basis for defining the bandwidth of the signal. Specifically:

> If a signal is low-pass (i.e., its spectral content is centered around the origin $f = 0$), the bandwidth is defined as one-half the total width of the main spectral lobe, since only one-half of this lobe lies inside the positive frequency region.

For example, a rectangular pulse of duration T seconds has a main spectral lobe of total width $(2/T)$ hertz centered at the origin. Accordingly, we may define the bandwidth of this rectangular pulse as $(1/T)$ hertz.

If, on the other hand, the signal is *band-pass* with main spectral lobes centered around $\pm f_c$, where f_c is large enough, the bandwidth is defined as the width of the main lobe for positive frequencies. This definition of bandwidth is called the *null-to-null bandwidth*. Consider, for example, a radio-frequency (RF) pulse of duration T seconds and frequency f_c, shown in Figure 2.7. The spectrum of this pulse has main spectral lobes of width $(2/T)$ hertz centered around $\pm f_c$, where it is assumed that f_c is large compared with $(1/T)$. Hence, we define the null-to-null bandwidth of the RF pulse of Figure 2.7 as $(2/T)$ hertz.

On the basis of the definitions presented here, we may state that shifting the spectral content of a low-pass signal by a sufficiently large frequency has the effect of doubling the bandwidth of the signal; this frequency translation is attained by using the process of modulation. Basically, the modulation moves the spectral content of the signal for negative frequencies into the positive frequency region, whereupon the negative frequencies become physically measurable.

Another popular definition of bandwidth is the *3 dB bandwidth*. Specifically, if the signal is low-pass, we say:

> The 3 dB bandwidth of a low-pass signal is defined as the separation between zero frequency, where the magnitude spectrum attains its peak value, and the positive frequency at which the amplitude spectrum drops to $1/\sqrt{2}$ of its peak value.

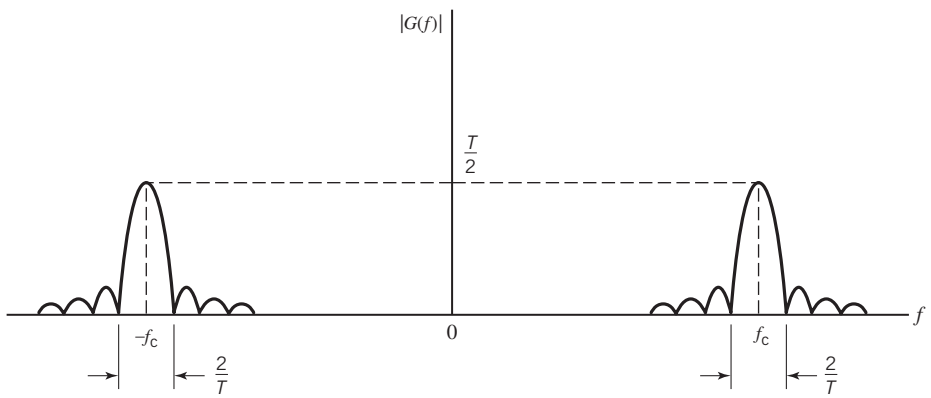


Figure 2.7 Magnitude spectrum of the RF pulse, showing the null-to-null bandwidth to be $2/T$, centered on the mid-band frequency f_c.

For example, the decaying exponential function $\exp(-at)$ has a 3 dB bandwidth of $(a/2\pi)$ hertz.

If, on the other hand, the signal is of a band-pass kind, centered at $\pm f_c$, the 3 dB bandwidth is defined as the separation (along the positive frequency axis) between the two frequencies at which the magnitude spectrum of the signal drops to $1/\sqrt{2}$ of its peak value at f_c.

Regardless of whether we have a low-pass or band-pass signal, the 3 dB bandwidth has the advantage that it can be read directly from a plot of the magnitude spectrum. However, it has the disadvantage that it may be misleading if the magnitude spectrum has slowly decreasing tails.

Time–Bandwidth Product

For any family of pulse signals that differ by a time-scaling factor, the product of the signal's duration and its bandwidth is always a constant, as shown by

duration × bandwidth = constant

This product is called the *time–bandwidth product*. The constancy of the time–bandwidth product is another manifestation of the inverse relationship that exists between the time-domain and frequency-domain descriptions of a signal. In particular, if the duration of a pulse signal is decreased by reducing the time scale by a factor a, the frequency scale of the signal's spectrum, and therefore the bandwidth of the signal is increased by the same factor a. This statement follows from the *dilation property* of the Fourier transform (defined in Property 2 of Table 2.1). The time–bandwidth product of the signal is therefore maintained constant. For example, a rectangular pulse of duration T seconds has a bandwidth (defined on the basis of the positive-frequency part of the main lobe) equal to $(1/T)$ hertz; in this example, the time–bandwidth product of the pulse equals unity.

The important point to take from this discussion is that whatever definitions we use for the bandwidth and duration of a signal, the time–bandwidth product remains constant over certain classes of pulse signals; the choice of particular definitions for bandwidth and duration merely change the value of the constant.

Root-Mean-Square Definitions of Bandwidth and Duration

To put matters pertaining to the bandwidth and duration of a signal on a firm mathematical basis, we first introduce the following definition for bandwidth:

The root-mean-square (rms) bandwidth is defined as the square root of the second moment of a normalized form of the squared magnitude spectrum of the signal about a suitably chosen frequency.

To be specific, we assume that the signal $g(t)$ is of a low-pass kind, in which case the second moment is taken about the origin $f = 0$. The squared magnitude spectrum of the signal is denoted by $|G(f)|^2$. To formulate a nonnegative function, the total area under whose curve is unity, we use the normalizing function

$$\int_{-\infty}^{\infty} |G(f)|^2 \mathrm{d}f$$

We thus mathematically define the rms bandwidth of a low-pass signal $g(t)$ with Fourier transform $G(f)$ as

$$W_{\text{rms}} = \left(\frac{\int_{-\infty}^{\infty} f^2 |G(f)|^2 \, df}{\int_{-\infty}^{\infty} |G(f)|^2 \, df} \right)^{1/2} \tag{2.26}$$

which describes the dispersion of the spectrum $G(f)$ around $f = 0$. An attractive feature of the rms bandwidth W_{rms} is that it lends itself readily to mathematical evaluation. But, it is not as easily measurable in the laboratory.

In a manner corresponding to the rms bandwidth, the *rms duration* of the signal $g(t)$ is mathematically defined by

$$T_{\text{rms}} = \left(\frac{\int_{-\infty}^{\infty} t^2 |g(t)|^2 \, dt}{\int_{-\infty}^{\infty} |g(t)|^2 \, dt} \right)^{1/2} \tag{2.27}$$

where it is assumed that the signal $g(t)$ is centered around the origin $t = 0$. In Problem 2.7, it is shown that, using the rms definitions of (2.26) and (2.27), the time–bandwidth product takes the form

$$T_{\text{rms}} W_{\text{rms}} \geq \frac{1}{4\pi} \tag{2.28}$$

In Problem 2.7, it is also shown that the Gaussian pulse $\exp(-\pi t^2)$ satisfies this condition exactly with the equality sign.

2.5 The Dirac Delta Function

Strictly speaking, the theory of the Fourier transform, presented in Section 2.3, is applicable only to time functions that satisfy the Dirichlet conditions. As mentioned previously, such functions naturally include energy signals. However, it would be highly desirable to extend this theory in two ways:

1. To combine the Fourier series and Fourier transform into a unified theory, so that the Fourier series may be treated as a special case of the Fourier transform.
2. To include power signals in the list of signals to which we may apply the Fourier transform. A signal $g(t)$ is said to be a *power signal* if the condition

$$\frac{1}{T} \int_{-T/2}^{T/2} |g(t)|^2 \, dt < \infty$$

holds, where T is the observation interval.

It turns out that both of these objectives can be met through the "proper use" of the *Dirac delta function*, or *unit impulse*.

The Dirac delta function[3] or just delta function, denoted by $\delta(t)$, is defined as having zero amplitude everywhere except at $t = 0$, where it is infinitely large in such a way that it contains unit area under its curve; that is,

$$\delta(t) = 0, \qquad t \neq 0 \tag{2.29}$$

and

$$\int_{-\infty}^{\infty} \delta(t)\,\mathrm{d}t = 1 \tag{2.30}$$

An implication of this pair of relations is that the delta function $\delta(t)$ is an even function of time t, centered at the origin $t = 0$. Perhaps, the simplest way of describing the Dirac delta function is to view it as the rectangular pulse

$$g(t) = \frac{1}{T}\,\mathrm{rect}\left(\frac{t}{T}\right)$$

whose duration is T and amplitude is $1/T$, as illustrated in Figure 2.8. As T approaches zero, the rectangular pulse $g(t)$ approaches the Dirac delta function $\delta(t)$ in the limit.

For the delta function to have meaning, however, it has to appear as a factor in the integrand of an integral with respect to time, and then, strictly speaking, only when the other factor in the integrand is a continuous function of time. Let $g(t)$ be such a function, and consider the product of $g(t)$ and the time-shifted delta function $\delta(t - t_0)$. In light of the two defining equations (2.29) and (2.30), we may express the integral of this product as

$$\int_{-\infty}^{\infty} g(t)\,\delta(t - t_0)\mathrm{d}t = g(t_0) \tag{2.31}$$

The operation indicated on the left-hand side of this equation sifts out the value $g(t_0)$ of the function $g(t)$ at time $t = t_0$, where $-\infty < t < \infty$. Accordingly, (2.31) is referred to as the *sifting property* of the delta function. This property is sometimes used as the defining equation of a delta function; in effect, it incorporates (2.29) and (2.30) into a single relation.

Noting that the delta function $\delta(t)$ is an even function of t, we may rewrite (2.31) so as to emphasize its resemblance to the convolution integral, as shown by

$$\int_{-\infty}^{\infty} g(\tau)\,\delta(t - \tau)\,\mathrm{d}\tau = g(t) \tag{2.32}$$

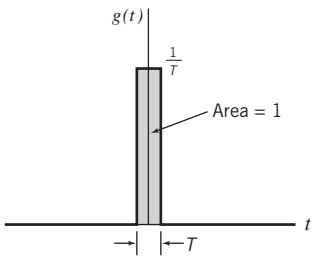


Figure 2.8 Illustrative example of the Dirac delta function as the limiting form of rectangular pulse $\frac{1}{T}\,\mathrm{rect}\left(\frac{t}{T}\right)$ as T approaches zero.

In words, the convolution of any function with the delta function leaves that function unchanged. We refer to this statement as the *replication property* of the delta function.

It is important to realize that no function in the ordinary sense has the two properties of (2.29) and (2.30) or the equivalent sifting property of (2.31). However, we can imagine a sequence of functions that have progressively taller and thinner peaks at $t = 0$, with the area under the curve consistently remaining equal to unity; as this progression is being performed, the value of the function tends to zero at every point except $t = 0$, where it tends to infinity, as illustrated in Figure 2.8, for example. We may therefore say:

> The delta function may be viewed as the limiting form of a pulse of unit area as the duration of the pulse approaches zero.

It is immaterial what sort of pulse shape is used, so long as it is symmetric with respect to the origin; this symmetry is needed to maintain the "even" function property of the delta function.

Two other points are noteworthy:

1. Applicability of the delta function is not confined to the time domain. Rather, it can equally well be applied in the frequency domain; all that we have to do is to replace time t by frequency f in the defining equations (2.29) and (2.30).
2. The area covered by the delta function defines its "strength." As such, the units, in terms of which the strength is measured, are determined by the specifications of the two coordinates that define the delta function.

EXAMPLE 3

The Sinc Function as a Limiting Form of the Delta Function in the Time Domain

As another illustrative example, consider the scaled sinc function $2W\text{sinc}(2Wt)$, whose waveform covers an area equal to unity for all W.

Figure 2.9 displays the evolution of this time function toward the delta function as the parameter W is varied in three stages: $W = 1$, $W = 2$, and $W = 5$. Referring back to Figure 2.4, we may infer that as the parameter W characterizing the sinc pulse is increased, the amplitude of the pulse at time $t = 0$ increases linearly, while at the same time the duration of the main lobe of the pulse decreases inversely. With this objective in mind, as the parameter W is progressively increased, Figure 2.9 teaches us two important things:

1. The scaled sinc function becomes more like a delta function.
2. The constancy of the function's spectrum is maintained at unity across an increasingly wider frequency band, in accordance with the constraint that the area under the function is to remain constant at unity; see Property 6 of Table 2.1 for a validation of this point.

Based on the trend exhibited in Figure 2.9, we may write

$$\delta(t) = \lim_{W \to \infty} 2W\,\text{sinc}(2Wt) \tag{2.33}$$

which, in addition to the rectangular pulse considered in Figure 2.8, is another way of realizing a delta function in the time domain.

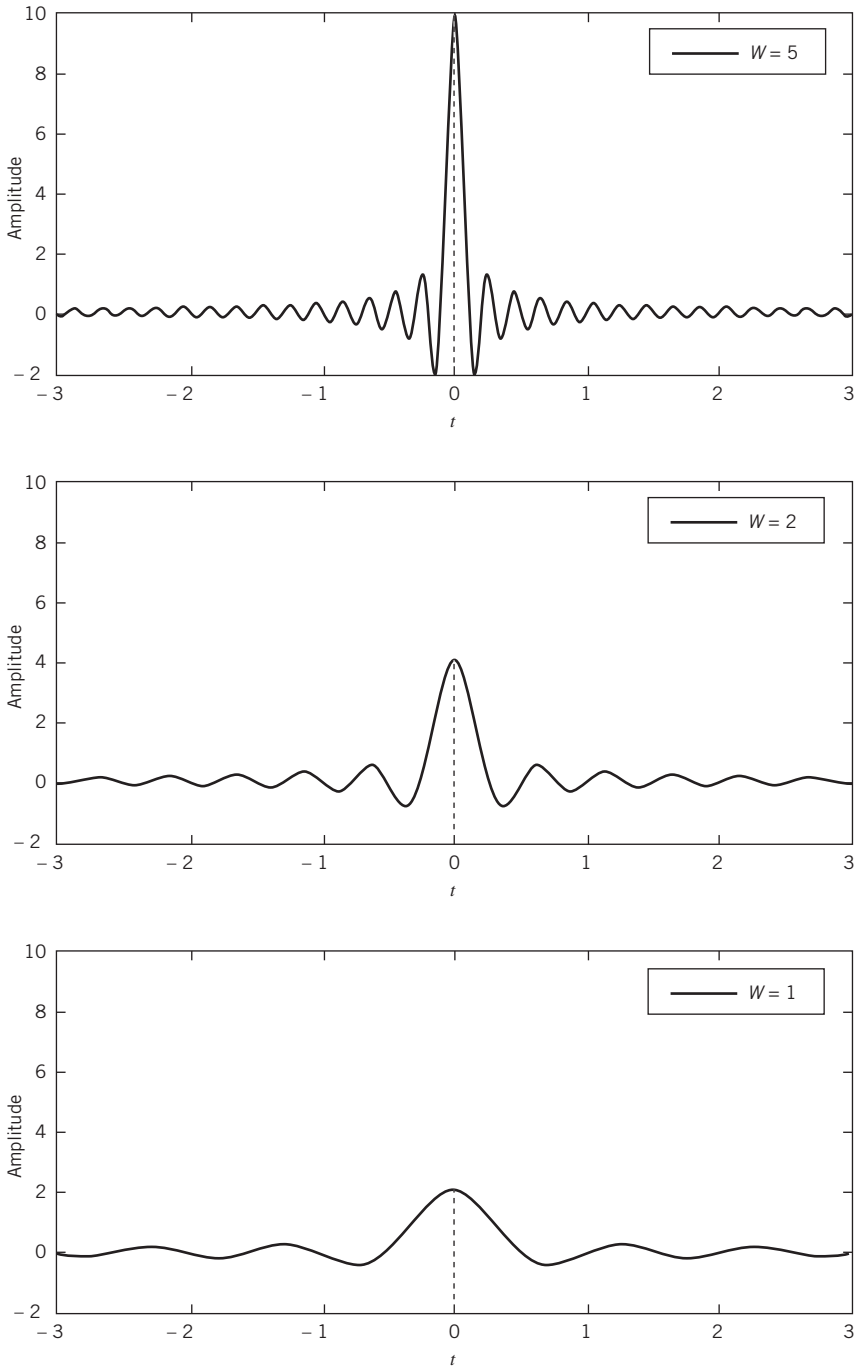


Figure 2.9 Evolution of the sinc function $2W\,\mathrm{sinc}(2Wt)$ toward the delta function as the parameter W progressively increases.

EXAMPLE 4

Evolution of the Sum of Complex Exponentials toward the Delta Function in the Frequency Domain

For yet another entirely different example, consider the infinite summation term $\sum_{m=-\infty}^{\infty} \exp(j2\pi mf)$ over the interval $-1/2 \le f < 1/2$. Using *Euler's formula*

$$\exp(j2\pi mf) = \cos(2\pi mf) + j\sin(2\pi mf)$$

we may express the given summation as

$$\sum_{m=-\infty}^{\infty} \exp(j2\pi mf) = \sum_{m=-\infty}^{\infty} \cos(2\pi mf) + j \sum_{m=-\infty}^{\infty} \sin(2\pi mf)$$

The imaginary part of the summation is zero for two reasons. First, $\sin(2\pi mf)$ is zero for $m = 0$. Second, since $\sin(-2\pi mf) = -\sin(2\pi mf)$, the remaining imaginary terms cancel each other. Therefore,

$$\sum_{m=-\infty}^{\infty} \exp(j2\pi mf) = \sum_{m=-\infty}^{\infty} \cos(2\pi mf)$$

Figure 2.10 plots this real-valued summation versus frequency f over the interval $-1/2 \le f < 1/2$ for three ranges of m:

1. $-5 \le m \le 5$
2. $-10 \le m \le 10$
3. $-20 \le m \le 20$

Building on the results exhibited in Figure 2.10, we may go on to say

$$\delta(f) = \sum_{m=-\infty}^{\infty} \cos(2\pi mf), \qquad -\frac{1}{2} \le f < \frac{1}{2} \tag{2.34}$$

which is one way of realizing a delta function in the frequency domain. Note that the area under the summation term on the right-hand side of (2.34) is equal to unity; we say so because

$$\int_{-1/2}^{1/2} \sum_{m=-\infty}^{\infty} \cos(2\pi mf)\, df = \sum_{m=-\infty}^{\infty} \int_{-1/2}^{1/2} \cos(2\pi mf)\, df$$

$$= \sum_{m=-\infty}^{\infty} \left[\frac{\sin(2\pi mf)}{2\pi m} \right]_{f=-1/2}^{1/2}$$

$$= \sum_{m=-\infty}^{\infty} \left[\frac{\sin(\pi m)}{\pi m} \right]$$

$$= \begin{cases} 1 & \text{for } m = 0 \\ 0 & \text{otherwise} \end{cases}$$

This result, formulated in the frequency domain, confirms (2.34) as one way of defining the delta function $\delta(f)$.

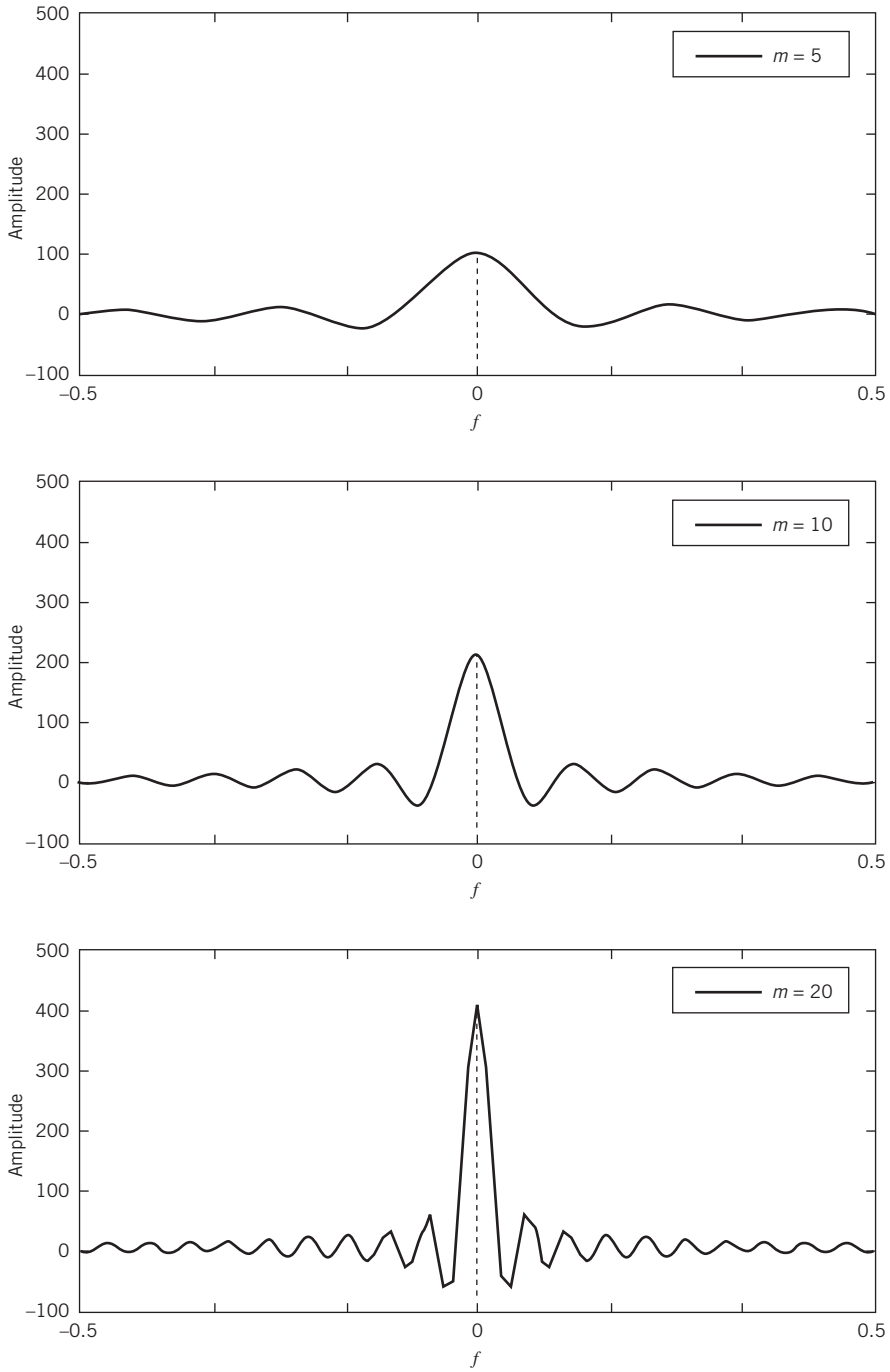


Figure 2.10 Evolution of the sum of m complex exponentials toward a delta function in the frequency domain as m becomes increasingly larger.

2.6 Fourier Transforms of Periodic Signals

We began the study of Fourier analysis by reviewing the Fourier series expansion of periodic signals, which, in turn, paved the way for the formulation of the Fourier transform. Now that we have equipped ourselves with the Dirac delta function, we would like to revisit the Fourier series and show that it can indeed be treated as a special case of the Fourier transform.

To this end, let $g(t)$ be a pulse-like function, which equals a periodic signal $g_{T_0}(t)$ over one period T_0 of the signal and is zero elsewhere, as shown by

$$g(t) = \begin{cases} g_{T_0}(t), & -\frac{T_0}{2} < t \le \frac{T_0}{2} \\ 0, & \text{elsewhere} \end{cases} \tag{2.35}$$

The periodic signal $g_{T_0}(t)$ itself may be expressed in terms of the function $g(t)$ as an infinite summation, as shown by

$$g_{T_0}(t) = \sum_{m=-\infty}^{\infty} g(t - mT_0) \tag{2.36}$$

In light of the definition of the pulselike function $g(t)$ in (2.35), we may view this function as a *generating function*, so called as it generates the periodic signal $g_{T_0}(t)$ in accordance with (2.36).

Clearly, the generating function $g(t)$ is Fourier transformable; let $G(f)$ denote its Fourier transform. Correspondingly, let $G_{T_0}(f)$ denote the Fourier transform of the periodic signal $g_{T_0}(t)$. Hence, taking the Fourier transforms of both sides of (2.36) and applying the time-shifting property of the Fourier transform (Property 4 of Table 2.1), we may write

$$G_{T_0}(f) = G(f) \sum_{m=-\infty}^{\infty} \exp(-j2\pi mfT_0), \qquad -\infty < f < \infty \tag{2.37}$$

where we have taken $G(f)$ outside the summation because it is independent of m.

In Example 4, we showed that

$$\sum_{m=-\infty}^{\infty} \exp(j2\pi mf) = \sum_{m=-\infty}^{\infty} \cos(j2\pi mf) = \delta(f), \qquad -\frac{1}{2} \le f < \frac{1}{2}$$

Let this result be expanded to cover the entire frequency range, as shown by

$$\sum_{m=-\infty}^{\infty} \exp(j2\pi mf) = \sum_{n=-\infty}^{\infty} \delta(f - n), \qquad -\infty < f < \infty \tag{2.38}$$

Equation (2.38) (see Problem 2.8c) represents a *Dirac comb*, consisting of an infinite sequence of uniformly spaced delta functions, as depicted in Figure 2.11.

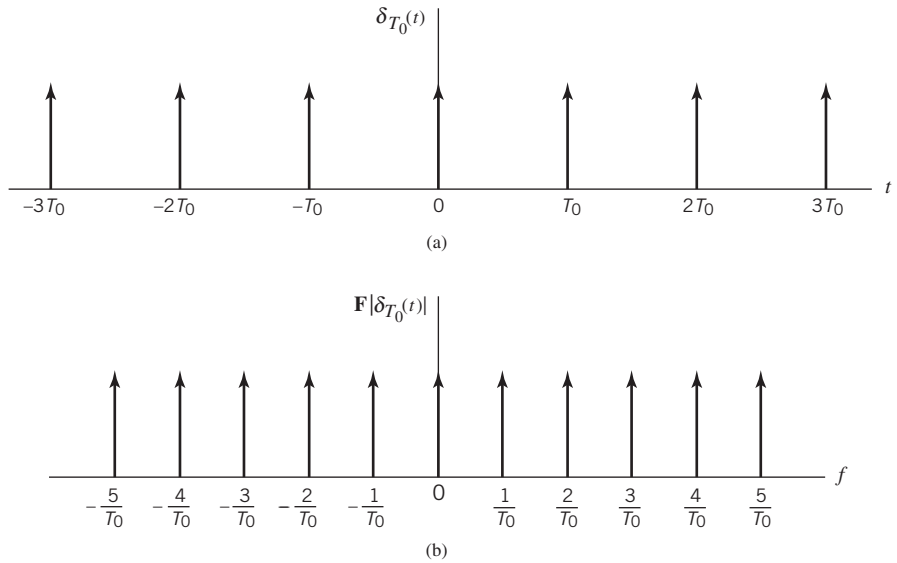


Figure 2.11 (a) Dirac comb. (b) Spectrum of the Dirac comb.

Next, introducing the frequency-scaling factor $f_0 = 1/T_0$ into (2.38), we correspondingly write

$$\sum_{m=-\infty}^{\infty} \exp(j2\pi mfT_0) = f_0 \sum_{n=-\infty}^{\infty} \delta(f - nf_0), \qquad -\infty < f < \infty \tag{2.39}$$

Hence, substituting (2.39) into the right-hand side of (2.37), we get

$$\begin{aligned} G_{T_0}(f) &= f_0 G(f) \sum_{n=-\infty}^{\infty} \delta(f - nf_0) \\ &= f_0 \sum_{n=-\infty}^{\infty} G(f_n)\delta(f - f_n), \qquad -\infty < f < \infty \end{aligned} \tag{2.40}$$

where $f_n = nf_0$.

What we have to show next is that the inverse Fourier transform of $G_{T_0}(f)$ defined in (2.40) is exactly the same as in the Fourier series formula of (2.14). Specifically, substituting (2.40) into the inverse Fourier transform formula of (2.17), we get

$$g_{T_0}(t) = f_0 \int_{-\infty}^{\infty} \left[\sum_{n=-\infty}^{\infty} G(f_n)\delta(f - f_n) \right] \exp(j2\pi ft)\, df$$

Interchanging the order of summation and integration, and then invoking the sifting property of the Dirac delta function (this time in the frequency domain), we may go on to write

$$g_{T_0}(t) = f_0 \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} G(f_n)\exp(j2\pi ft)\,\delta(f-f_n)\,df$$

$$= f_0 \sum_{n=-\infty}^{\infty} G(f_n)\exp(j2\pi f_n t)$$

which is an exact rewrite of (2.14) with $f_0 = \Delta f$. Equivalently, in light of (2.36), we may formulate the Fourier transform pair

$$\sum_{m=-\infty}^{\infty} g(t-mT_0) = f_0 \sum_{n=-\infty}^{\infty} G(f_n)\exp(j2\pi f_n t) \tag{2.41}$$

The result derived in (2.41) is one form of *Poisson's sum formula*.

We have thus demonstrated that the Fourier series representation of a periodic signal is embodied in the Fourier transformation of (2.16) and (2.17), provided, of course, we permit the use of the Dirac delta function. In so doing, we have closed the "circle" by going from the Fourier series to the Fourier transform, and then back to the Fourier series.

Consequences of Ideal Sampling

Consider a Fourier transformable pulselike signal $g(t)$ with its Fourier transform denoted by $G(f)$. Setting $f_n = nf_0$ in (2.41) and using (2.38), we may express Poisson's sum formula

$$\sum_{m=-\infty}^{\infty} g(t-mT_0) \rightleftharpoons f_0 \sum_{n=-\infty}^{\infty} G(nf_0)\,\delta(f-nf_0) \tag{2.42}$$

where $f_0 = 1/T_0$. The summation on the left-hand side of this Fourier-transform pair is a periodic signal with period T_0. The summation on the right-hand side of the pair is a uniformly sampled version of the spectrum $G(f)$. We may therefore make the following statement:

> Uniform sampling of the spectrum $G(f)$ in the frequency domain introduces periodicity of the function $g(t)$ in the time domain.

Applying the duality property of the Fourier transform (Property 3 of Table 2.1) to (2.42), we may also write

$$T_0 \sum_{m=-\infty}^{\infty} g(mT_0)\,\delta(t-mT_0) \rightleftharpoons \sum_{n=-\infty}^{\infty} G(f-nf_0) \tag{2.43}$$

in light of which we may make the following dual statement:

> Uniform sampling of the Fourier transformable function $g(t)$ in the time domain introduces periodicity of the spectrum $G(f)$ in the frequency domain.

2.7 Transmission of Signals through Linear Time-Invariant Systems

A *system* refers to any physical entity that produces an output signal in response to an input signal. It is customary to refer to the input signal as the *excitation* and to the output signal as the *response*. In a linear system, the *principle of superposition* holds; that is, the response of a linear system to a number of excitations applied simultaneously is equal to the sum of the responses of the system when each excitation is applied individually.

In the time domain, a linear system is usually described in terms of its *impulse response*, which is formally defined as follows:

> The impulse response of a linear system is the response of the system (with zero initial conditions) to a unit impulse or delta function $\delta(t)$ applied to the input of the system at time $t = 0$.

If the system is also *time invariant*, then the shape of the impulse response is the same no matter when the unit impulse is applied to the system. Thus, with the unit impulse or delta function applied to the system at time $t = 0$, the impulse response of a linear time-invariant system is denoted by $h(t)$.

Suppose that a system described by the impulse response $h(t)$ is subjected to an arbitrary excitation $x(t)$, as depicted in Figure 2.12. The resulting response of the system $y(t)$, is defined in terms of the impulse response $h(t)$ by

$$y(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)\,\mathrm{d}\tau \tag{2.44}$$

which is called the *convolution integral*. Equivalently, we may write

$$y(t) = \int_{-\infty}^{\infty} h(\tau)x(t-\tau)\,\mathrm{d}\tau \tag{2.45}$$

Equations (2.44) and (2.45) state that convolution is *commutative*.

Examining the convolution integral of (2.44), we see that three different time scales are involved: *excitation time* τ, *response time* t, and *system-memory time* $t - \tau$. This relation is the basis of time-domain analysis of linear time-invariant systems. According to (2.44), the present value of the response of a linear time-invariant system is an integral over the past history of the input signal, weighted according to the impulse response of the system. Thus, the impulse response acts as a *memory function* of the system.

Causality and Stability

A linear system with impulse response $h(t)$ is said to be *causal* if its impulse response $h(t)$ satisfies the condition

$$h(t) = 0 \quad \text{for} \quad t < 0$$

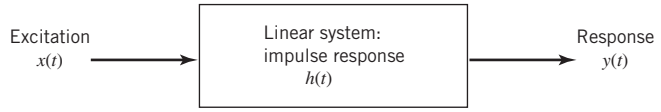


Figure 2.12 Illustrating the roles of excitation $x(t)$, impulse response $h(t)$, and response $y(t)$ in the context of a linear time-invariant system.

The essence of causality is that no response can appear at the output of the system before an excitation is applied to its input. Causality is a necessary requirement for on-line operation of the system. In other words, for a system operating in *real time* to be physically realizable, it has to be causal.

Another important property of a linear system is *stability*. A necessary and sufficient condition for the system to be stable is that its impulse response $h(t)$ must satisfy the inequality

$$\int_{-\infty}^{\infty} |h(t)|\,\mathrm{d}t < \infty$$

This requirement follows from the commonly used criterion of *bounded input–bounded output*. Basically, for the system to be stable, its impulse response must be *absolutely integrable*.

Frequency Response

Let $X(f)$, $H(f)$, and $Y(f)$ denote the Fourier transforms of the excitation $x(t)$, impulse response $h(t)$, and response $y(t)$, respectively. Then, applying Property 12 of the Fourier transform in Table 2.1 to the convolution integral, be it written in the form of (2.44) or (2.45), we get

$$Y(f) = H(f)X(f) \tag{2.46}$$

Equivalently, we may write

$$H(f) = \frac{Y(f)}{X(f)} \tag{2.47}$$

The new frequency function $H(f)$ is called the *transfer function* or *frequency response* of the system; these two terms are used interchangeably. Based on (2.47), we may now formally say:

> The frequency response of a linear time-invariant system is defined as the ratio of the Fourier transform of the response of the system to the Fourier transform of the excitation applied to the system.

In general, the frequency response $H(f)$ is a complex quantity, so we may express it in the form

$$H(f) = |H(f)|\exp[\mathrm{j}\beta(f)] \tag{2.48}$$

where $|H(f)|$ is called the *magnitude response*, and $\beta(f)$ is the *phase response*, or simply *phase*. When the impulse response of the system is real valued, the frequency response exhibits conjugate symmetry, which means that

$$|H(f)| = |H(-f)|$$

and

$$\beta(f) = -\beta(-f)$$

That is, the magnitude response $|H(f)|$ of a linear system with real-valued impulse response is an even function of frequency, whereas the phase $\beta(f)$ is an odd function of frequency.

In some applications it is preferable to work with the logarithm of $H(f)$ expressed in polar form, rather than with $H(f)$ itself. Using ln to denote the natural logarithm, let

$$\ln H(f) = \alpha(f) + \mathrm{j}\beta(f) \tag{2.49}$$

where

$$\alpha(f) = \ln|H(f)| \tag{2.50}$$

The function $\alpha(f)$ is called the *gain* of the system; it is measured in *nepers*. The phase $\beta(f)$ is measured in *radians*. Equation (2.49) indicates that the gain $\alpha(f)$ and phase $\beta(f)$ are, respectively, the real and imaginary parts of the (natural) logarithm of the transfer function $H(f)$. The gain may also be expressed in *decibels* (dB) by using the definition

$$\alpha'(f) = 20\log_{10}|H(f)|$$

The two gain functions $\alpha(f)$ and $\alpha'(f)$ are related by

$$\alpha'(f) = 8.69\,\alpha(f)$$

That is, 1 neper is equal to 8.69 dB.

As a means of specifying the constancy of the magnitude response $|H(f)|$ or gain $\alpha(f)$ of a system, we use the notion of *bandwidth*. In the case of a low-pass system, the bandwidth is customarily defined as the frequency at which the magnitude response $|H(f)|$ is $1/\sqrt{2}$ times its value at zero frequency or, equivalently, the frequency at which the gain $\alpha'(f)$ drops by 3 dB below its value at zero frequency, as illustrated in Figure 2.13a. In the case of a band-pass system, the bandwidth is defined as the range of frequencies over which the magnitude response $|H(f)|$ remains within $1/\sqrt{2}$ times its value at the mid-band frequency, as illustrated in Figure 2.13b.

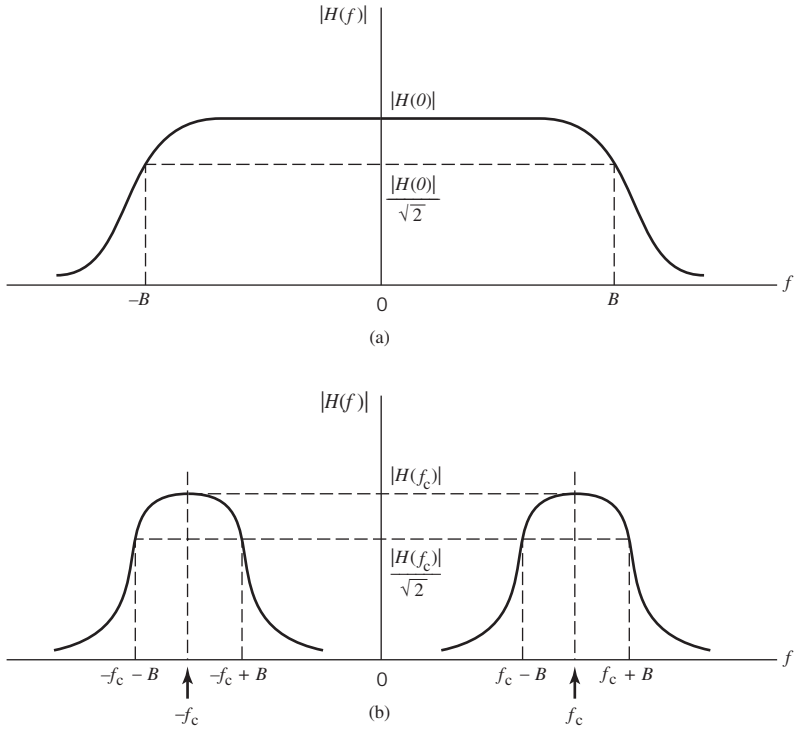


Figure 2.13 Illustrating the definition of system bandwidth. (a) Low-pass system. (b) Band-pass system.

Paley–Wiener Criterion: Another Way of Assessing Causality

A necessary and sufficient condition for a function $\alpha(f)$ to be the gain of a causal filter is the convergence of the integral

$$\int_{-\infty}^{\infty} \frac{|\alpha(f)|}{1+f^2}\,\mathrm{d}f < \infty \tag{2.51}$$

This condition is known as the *Paley–Wiener criterion*.[4] The criterion states that provided the gain $\alpha(f)$ satisfies the condition of (2.51), then we may associate with this gain a suitable phase $\beta(f)$, such that the resulting filter has a causal impulse response that is zero for negative time. In other words, the Paley–Wiener criterion is the frequency-domain equivalent of the causality requirement. A system with a realizable gain characteristic may have infinite attenuation for a discrete set of frequencies, but it cannot have infinite attenuation over a band of frequencies; otherwise, the Paley–Wiener criterion is violated.

Finite-Duration Impulse Response (FIR) Filters

Consider next a linear time-invariant filter with impulse response $h(t)$. We make two assumptions:

1. *Causality*, which means that the impulse response $h(t)$ is zero for $t < 0$.
2. *Finite support*, which means that the impulse response of the filter is of some finite duration T_f, so that we may write $h(t) = 0$ for $t \geq T_f$.

Under these two assumptions, we may express the filter output $y(t)$ produced in response to the input $x(t)$ as

$$y(t) = \int_0^{T_f} h(\tau)x(t-\tau)\,\mathrm{d}\tau \tag{2.52}$$

Let the input $x(t)$, impulse response $h(t)$, and output $y(t)$ be *uniformly sampled* at the rate $(1/\Delta\tau)$ samples per second, so that we may put

$$t = n\Delta\tau$$

and

$$\tau = k\Delta\tau$$

where k and n are integers and $\Delta\tau$ is the *sampling period*. Assuming that $\Delta\tau$ is small enough for the product $h(\tau)x(t-\tau)$ to remain essentially constant for $k\Delta\tau \leq \tau \leq (k+1)\Delta\tau$ for all values of k and τ, we may approximate (2.52) by the *convolution sum*

$$y(n\Delta\tau) = \sum_{k=0}^{N-1} h(k\Delta\tau)x(n\Delta\tau - k\Delta\tau)\Delta\tau$$

where $N\Delta\tau = T_f$. To simplify the notations used in this summation formula, we introduce three definitions:

$$w_k = h(k\Delta\tau)\Delta\tau$$
$$x(n\Delta\tau) = x_n$$
$$y(n\Delta\tau) = y_n$$

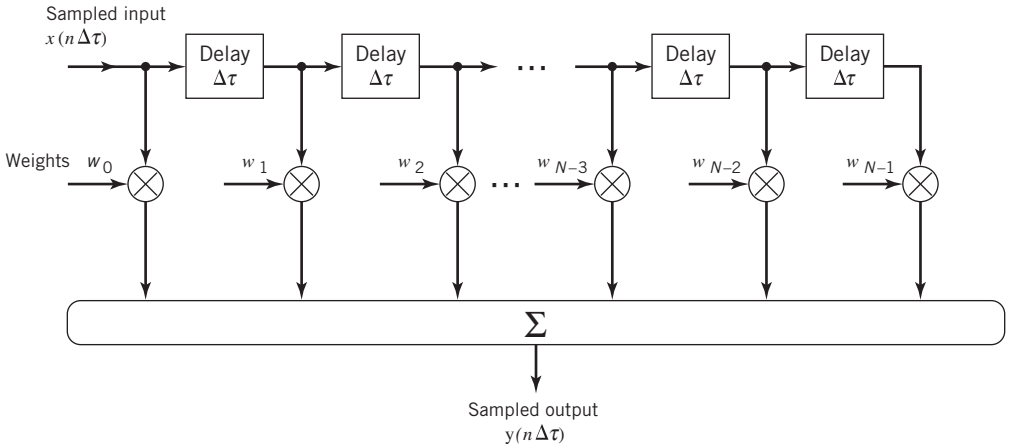


Figure 2.14 Tapped-delay-line (TDL) filter; also referred to as FIR filter.

We may then rewrite the formula for $y(n\Delta\tau)$ in the compact form

$$y_n = \sum_{k=0}^{N-1} w_k x_{n-k}, \qquad n = 0, \pm 1, \pm 2, \ldots \tag{2.53}$$

Equation (2.53) may be realized using the structure shown in Figure 2.14, which consists of a set of *delay elements* (each producing a delay of $\Delta\tau$ seconds), a set of *multipliers* connected to the *delay-line taps*, a corresponding set of *weights* supplied to the multipliers, and a *summer* for adding the multiplier outputs. The sequences x_n and y_n, for integer values of n as described in (2.53), are referred to as the *input* and *output sequences,* respectively.

In the digital signal-processing literature, the structure of Figure 2.14 is known as a *finite-duration impulse response (FIR) filter.* This filter offers some highly desirable practical features:

1. The filter is inherently *stable*, in the sense that a bounded input sequence produces a bounded output sequence.
2. Depending on how the weights $\{w_k\}_{k=0}^{N-1}$ are designated, the filter can perform the function of a low-pass filter or band-pass filter. Moreover, the phase response of the filter can be configured to be a linear function of frequency, which means that there will be no delay distortion.
3. In a digital realization of the filter, the filter assumes a *programmable* form whereby the application of the filter can be changed merely by making appropriate changes to the weights, leaving the structure of the filter completely unchanged; this kind of flexibility is not available with analog filters.

We will have more to say on the FIR filter in subsequent chapters of the book.

2.8 Hilbert Transform

The Fourier transform is particularly useful for evaluating the frequency content of an energy signal or, in a limiting sense, that of a power signal. As such, it provides the mathematical basis for analyzing and designing *frequency-selective filters* for the separation of signals on the basis of their frequency content. Another method of separating signals is based on *phase selectivity*, which uses phase shifts between the pertinent signals to achieve the desired separation. A phase shift of special interest in this context is that of ±90°. In particular, when the phase angles of all components of a given signal are shifted by ±90°, the resulting function of time is known as the *Hilbert transform* of the signal. The Hilbert transform is called a *quadrature filter*; it is so called to emphasize its distinct property of providing a ±90° phase shift.

To be specific, consider a Fourier transformable signal $g(t)$ with its Fourier transform denoted by $G(f)$. The *Hilbert transform* of $g(t)$, which we denote by $\hat{g}(t)$, is defined by[5]

$$\hat{g}(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{g(\tau)}{t-\tau}\,d\tau \tag{2.54}$$

Table 2.3 Hilbert-transform pairs*

Time function	Hilbert transform
1. $m(t)\cos(2\pi f_c t)$	$m(t)\sin(2\pi f_c t)$
2. $m(t)\sin(2\pi f_c t)$	$-m(t)\cos(2\pi f_c t)$
3. $\cos(2\pi f_c t)$	$\sin(2\pi f_c t)$
4. $\sin(2\pi f_c t)$	$-\cos(2\pi f_c t)$
5. $\dfrac{\sin t}{t}$	$\dfrac{1-\cos t}{t}$
6. $\text{rect}(t)$	$-\dfrac{1}{\pi}\ln\left\lvert\dfrac{t-1/2}{t+1/2}\right\rvert$
7. $\delta(t)$	$\dfrac{1}{\pi t}$
8. $\dfrac{1}{1+t^2}$	$\dfrac{t}{1+t^2}$
9. $\dfrac{1}{t}$	$-\pi\delta(t)$

Notes: $\delta(t)$ denotes Dirac delta function; rect(t) denotes rectangular function; ln denotes natural logarithm.
* In the first two pairs, it is assumed that $m(t)$ is band limited to the interval $-W \le f \le W$, where $W < f_c$.

Clearly, Hilbert transformation is a linear operation. The *inverse Hilbert transform*, by means of which the original signal $g(t)$ is linearly recovered from $\hat{g}(t)$, is defined by

$$g(t) = -\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\hat{g}(\tau)}{t-\tau}\,\mathrm{d}\tau \tag{2.55}$$

The functions $g(t)$ and $\hat{g}(t)$ are said to constitute a *Hilbert-transform pair*. A short table of Hilbert-transform pairs is given in Table 2.3 on page 42.

The definition of the Hilbert transform $\hat{g}(t)$ given in (2.54) may be interpreted as the convolution of $g(t)$ with the time function $1/(\pi t)$. We know from the convolution theorem listed in Table 2.1 that the convolution of two functions in the time domain is transformed into the multiplication of their Fourier transforms in the frequency domain.

For the time function $1/(\pi t)$, we have the Fourier-transform pair (see Property 14 in Table 2.2)

$$\frac{1}{\pi t} \rightleftharpoons -\mathrm{j}\,\mathrm{sgn}(f)$$

where $\mathrm{sgn}(f)$ is the *signum function*, defined in the frequency domain as

$$\mathrm{sgn}(f) = \begin{cases} 1, & f>0 \\ 0, & f=0 \\ -1, & f<0 \end{cases} \tag{2.56}$$

It follows, therefore, that the Fourier transform $\hat{G}(f)$ of $\hat{g}(t)$ is given by

$$\hat{G}(f) = -\mathrm{j}\,\mathrm{sgn}(f)G(f) \tag{2.57}$$

Equation (2.57) states that given a Fourier transformable signal $g(t)$, we may obtain the Fourier transform of its Hilbert transform $\hat{g}(t)$ by passing $g(t)$ through a linear time-invariant system whose frequency response is equal to $-\mathrm{j}\,\mathrm{sgn}(f)$. This system may be considered as one that produces a phase shift of –90° for all positive frequencies of the input signal and +90° degrees for all negative frequencies, as in Figure 2.15. The amplitudes of all frequency components in the signal, however, are unaffected by transmission through the device. Such an ideal system is referred to as a *Hilbert transformer*, or *quadrature filter*.

Properties of the Hilbert Transform

The Hilbert transform differs from the Fourier transform in that it operates exclusively in the time domain. It has a number of useful properties of its own, some of which are listed next. The signal $g(t)$ is assumed to be real valued, which is the usual domain of application of the Hilbert transform. For this class of signals, the Hilbert transform has the following properties.

PROPERTY 1 *A signal $g(t)$ and its Hilbert transform $\hat{g}(t)$ have the same magnitude spectrum.*

That is to say,

$$|G(f)| = |\hat{G}(f)|$$

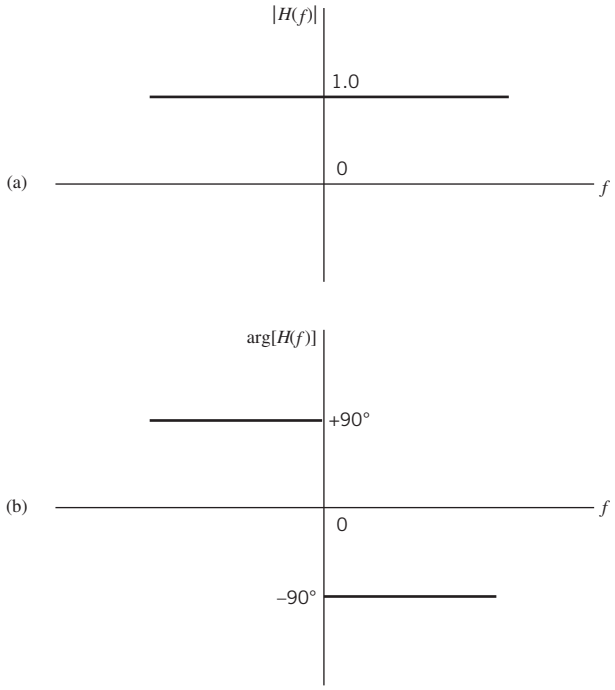


Figure 2.15
(a) Magnitude response and
(b) phase response of Hilbert transform.

PROPERTY 2 *If $\hat{g}(t)$ is the Hilbert transform of $g(t)$, then the Hilbert transform of $\hat{g}(t)$ is $-g(t)$.*

Another way of stating this property is to write

$$\arg[G(f)] = -\arg\{\hat{G}(f)\}$$

PROPERTY 3 *A signal $g(t)$ and its Hilbert transform $\hat{g}(t)$ are orthogonal over the entire time interval $(-\infty, \infty)$.*

In mathematical terms, the orthogonality of $g(t)$ and $\hat{g}(t)$ is described by

$$\int_{-\infty}^{\infty} g(t)\hat{g}(t)\mathrm{d}t = 0$$

Proofs of these properties follow from (2.54), (2.55), and (2.57).

EXAMPLE 5 **Hilbert Transform of Low-Pass Signal**

Consider Figure 2.16a that depicts the Fourier transform of a low-pass signal $g(t)$, whose frequency content extends from $-W$ to W. Applying the Hilbert transform to this signal yields a new signal $\hat{g}(t)$ whose Fourier transform, $\hat{G}(f)$, is depicted in Figure 2.16b. This figure illustrates that the frequency content of a Fourier transformable signal can be radically changed as a result of Hilbert transformation.

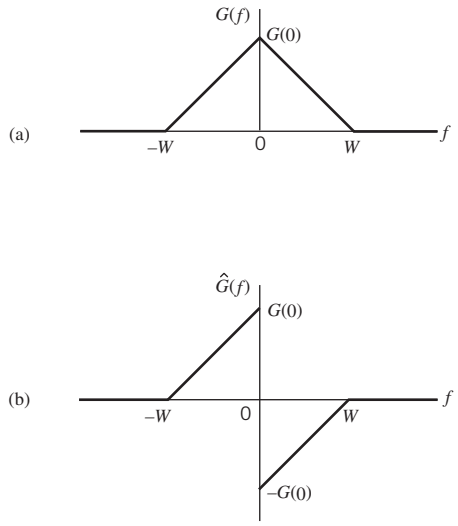


Figure 2.16 Illustrating application of the Hilbert transform to a low-pass signal: (a) Spectrum of the signal $g(t)$; (b) Spectrum of the Hilbert transform $\hat{g}(t)$.

2.9 Pre-envelopes

The Hilbert transform of a signal is defined for both positive and negative frequencies. In light of the spectrum shaping illustrated in Example 5, a question that begs itself is:

> How can we modify the frequency content of a real-valued signal $g(t)$ such that all negative frequency components are completely eliminated?

The answer to this fundamental question lies in the idea of a complex-valued signal called the *pre-envelope*[6] of $g(t)$, formally defined as

$$g_+(t) = g(t) + j\hat{g}(t) \tag{2.58}$$

where $\hat{g}(t)$ is the Hilbert transform of $g(t)$. According to this definition, the given signal $g(t)$ is the real part of the pre-envelope $g_+(t)$, and the Hilbert transform $\hat{g}(t)$ is the imaginary part of the pre-envelope. An important feature of the pre-envelope $g_+(t)$ is the behavior of its Fourier transform. Let $G_+(f)$ denote the Fourier transform of $g_+(t)$. Then, using (2.57) and (2.58) we may write

$$G_+(f) = G(f) + \text{sgn}(f)G(f) \tag{2.59}$$

Next, invoking the definition of the signum function given in (2.56), we may rewrite (2.59) in the equivalent form

$$G_+(f) = \begin{cases} 2G(f), & f > 0 \\ G(0), & f = 0 \\ 0, & f < 0 \end{cases} \tag{2.60}$$

where $G(0)$ is the value of $G(f)$ at the origin $f = 0$. Equation (2.60) clearly shows that the pre-envelope of the signal $g(t)$ has no frequency content (i.e., its Fourier transform vanishes) for all negative frequencies, and the question that was posed earlier has indeed been answered. Note, however, in order to do this, we had to introduce the complex-valued version of a real-valued signal as described in (2.58).

From the foregoing analysis it is apparent that for a given signal $g(t)$ we may determine its pre-envelope $g_+(t)$ in one of two equivalent procedures.

1. *Time-domain procedure*. Given the signal $g(t)$, we use (2.58) to compute the pre-envelope $g_+(t)$.
2. *Frequency-domain procedure*. We first determine the Fourier transform $G(f)$ of the signal $g(t)$, then use (2.60) to determine $G_+(f)$, and finally evaluate the inverse Fourier transform of $G_+(f)$ to obtain

$$g_+(t) = 2\int_0^\infty G(f)\exp(j2\pi ft)\,df \tag{2.61}$$

Depending on the description of the signal, procedure 1 may be easier than procedure 2, or vice versa.

Equation (2.58) defines the pre-envelope $g_+(t)$ for positive frequencies. Symmetrically, we may define the pre-envelope for *negative frequencies* as

$$g_-(t) = g(t) - j\hat{g}(t) \tag{2.62}$$

The two pre-envelopes $g_+(t)$ and $g_-(t)$ are simply the complex conjugate of each other, as shown by

$$g_-(t) = g_+^*(t) \tag{2.63}$$

where the asterisk denotes complex conjugation. The spectrum of the pre-envelope $g_+(t)$ is nonzero only for *positive* frequencies; hence the use of a plus sign as the subscript. On the other hand, the use of a minus sign as the subscript is intended to indicate that the spectrum of the other pre-envelope $g_-(t)$ is nonzero only for *negative* frequencies, as shown by the Fourier transform

$$G_-(f) = \begin{cases} 0, & f > 0 \\ G(0), & f = 0 \\ 2G(f), & f < 0 \end{cases} \tag{2.64}$$

Thus, the pre-envelope $g_+(t)$ and $g_-(t)$ constitute a complementary pair of complex-valued signals. Note also that the sum of $g_+(t)$ and $g_-(t)$ is exactly twice the original signal $g(t)$.

Given a real-valued signal, (2.60) teaches us that the pre-envelope $g_+(t)$ is uniquely defined by the spectral content of the signal for positive frequencies. By the same token, (2.64) teaches us that the other pre-envelope $g_-(t)$ is uniquely defined by the spectral content of the signal for negative frequencies. Since $g_-(t)$ is simply the complex conjugate of $g_+(t)$ as indicated in (2.63), we may now make the following statement:

> The spectral content of a Fourier transformable real-valued signal for positive frequencies uniquely defines that signal.

In other words, given the spectral content of such a signal for positive frequencies, we may uniquely define the spectral content of the signal for negative frequencies. Here then is the mathematical justification for basing the bandwidth of a Fourier transformable signal on its spectral content exclusively for positive frequencies, which is exactly what we did in Section 2.4, dealing with bandwidth.

EXAMPLE 6 **Pre-envelopes of Low-Pass Signal**

Continuing with the low-pass signal $g(t)$ considered in Example 5, Figure 2.17a and b depict the corresponding spectra of the pre-envelope $g_+(t)$ and the second pre-envelope $g_-(t)$, both of which belong to $g(t)$. Whereas the spectrum of $g(t)$ is defined for $-W \le f \le W$ as in Figure 2.16a, we clearly see from Figure 2.17 that the spectral content of $g_+(t)$ is confined entirely to $0 \le f \le W$, and the spectral content of $g_-(t)$ is confined entirely to $-W \le f \le 0$.

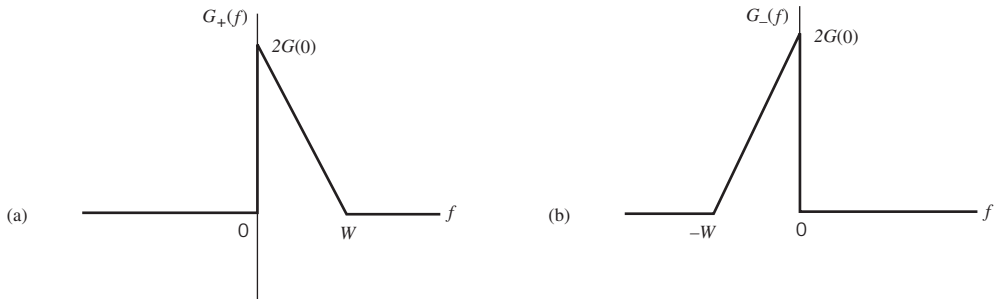


Figure 2.17 Another illustrative application of the Hilbert transform to a low-pass signal: (a) Spectrum of the pre-envelope $g_+(t)$; (b) Spectrum of the other pre-envelope $g_-(t)$.

Practical Importance of the Hilbert Transformation

An astute reader may see an analogy between the use of *phasors* and that of *pre-envelopes*. In particular, just as the use of phasors simplifies the manipulations of alternating currents and voltages in the study of circuit theory, so we find the pre-envelope simplifies the analysis of band-pass signals and band-pass systems in signal theory.

More specifically, by applying the concept of pre-envelope to a band-pass signal, the signal is transformed into an equivalent low-pass representation. In a corresponding way, a band-pass filter is transformed into its own equivalent low-pass representation. Both transformations, rooted in the Hilbert transform, play a key role in the formulation of modulated signals and their demodulation, as demonstrated in what follows in this and subsequent chapters.

2.10 Complex Envelopes of Band-Pass Signals

The idea of pre-envelopes introduced in Section 2.9 applies to any real-valued signal, be it of a low-pass or band-pass kind; the only requirement is that the signal be Fourier transformable. From this point on and for the rest of the chapter, we will restrict attention to band-pass signals. Such signals are exemplified by signals modulated onto a sinusoidal

carrier. In a corresponding way, when it comes to systems we restrict attention to band-pass systems. The primary reason for these restrictions is that the material so presented is directly applicable to analog modulation theory, to be covered in Section 2.14, as well as other digital modulation schemes covered in subsequent chapters of the book. With this objective in mind and the desire to make a consistent use of notation with respect to material to be presented in subsequent chapters, henceforth we will use $s(t)$ to denote a modulated signal. When such a signal is applied to the input of a band-pass system, such as a communication channel, we will use $x(t)$ to denote the resulting system (e.g., channel) output. However, as before, we will use $h(t)$ as the impulse response of the system.

To proceed then, let the band-pass signal of interest be denoted by $s(t)$ and its Fourier transform be denoted by $S(f)$. We assume that the Fourier transform $S(f)$ is essentially confined to a band of frequencies of total extent $2W$, centered about some frequency $\pm f_c$, as illustrated in Figure 2.18a. We refer to f_c as the *carrier frequency*; this terminology is

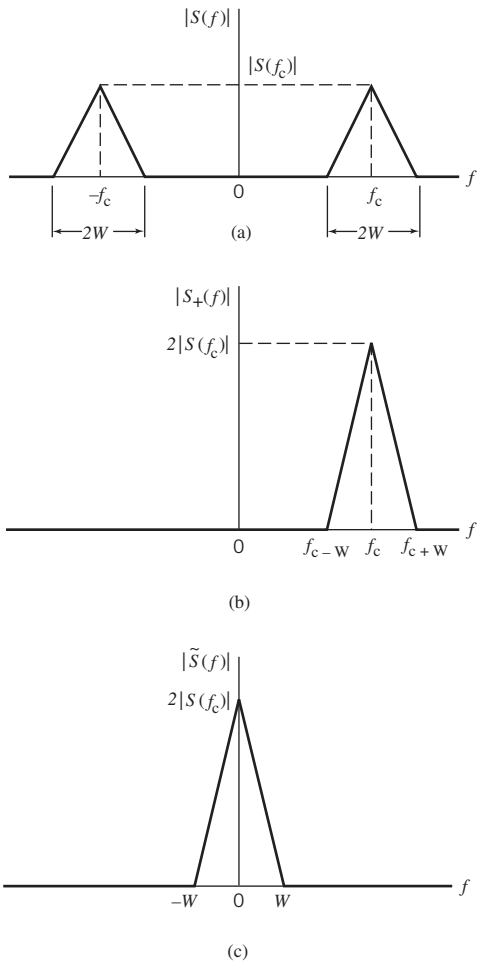


Figure 2.18 (a) Magnitude spectrum of band-pass signal $s(t)$; (b) Magnitude spectrum of pre-envelope $s_+(t)$; (c) Magnitude spectrum of complex envelope $\tilde{s}(t)$.

borrowed from modulation theory. In the majority of communication signals encountered in practice, we find that the bandwidth $2W$ is small compared with f_c, so we may refer to the signal $s(t)$ as a *narrowband signal*. However, a precise statement about how small the bandwidth must be for the signal to be considered narrowband is not necessary for our present discussion. Hereafter, the terms band-pass and narrowband are used interchangeably.

Let the pre-envelope of the narrowband signal $s(t)$ be expressed in the form

$$s_+(t) = \tilde{s}(t)\exp(j2\pi f_c t) \tag{2.65}$$

We refer to $\tilde{s}(t)$ as the *complex envelope* of the band-pass signal $s(t)$. Equation (2.65) may be viewed as the basis of a definition for the complex envelope $\tilde{s}(t)$ in terms of the pre-envelope $s_+(t)$. In light of the narrowband assumption imposed on the spectrum of the band-pass signal $s(t)$, we find that the spectrum of the pre-envelope $s_+(t)$ is limited to the positive frequency band $f_c - W \le f \le f_c + W$, as illustrated in Figure 2.18b. Therefore, applying the frequency-shifting property of the Fourier transform to (2.65), we find that the spectrum of the complex envelope $\tilde{s}(t)$ is correspondingly limited to the band $-W \le f \le W$ and centered at the origin $f = 0$, as illustrated in Figure 2.18c. In other words, the complex envelope $\tilde{s}(t)$ of the band-pass signal $s(t)$ is a *complex low-pass signal*. The essence of the *mapping* from the band-pass signal $s(t)$ to the complex low-pass signal $\tilde{s}(t)$ is summarized in the following threefold statement:

- The information content of a modulated signal $s(t)$ is fully preserved in the complex envelope $\tilde{s}(t)$.
- Analysis of the band-pass signal $s(t)$ is complicated by the presence of the carrier frequency f_c; in contrast, the complex envelope $\tilde{s}(t)$ dispenses with f_c, making its analysis simpler to deal with.
- The use of $\tilde{s}(t)$ requires having to handle complex notations.

2.11 Canonical Representation of Band-Pass Signals

By definition, the real part of the pre-envelope $s_+(t)$ is equal to the original band-pass signal $s(t)$. We may therefore express the band-pass signal $s(t)$ in terms of its corresponding complex envelope $\tilde{s}(t)$ as

$$s(t) = \mathrm{Re}[\tilde{s}(t)\exp(j2\pi f_c t)] \tag{2.66}$$

where the operator Re[.] denotes the real part of the quantity enclosed inside the square brackets. Since, in general, $\tilde{s}(t)$ is a complex-valued quantity, we emphasize this property by expressing it in the Cartesian form

$$\tilde{s}(t) = s_I(t) + js_Q(t) \tag{2.67}$$

where $s_I(t)$ and $s_Q(t)$ are both real-valued low-pass functions; their low-pass property is inherited from the complex envelope $\tilde{s}(t)$. We may therefore use (2.67) in (2.66) to express the original band-pass signal $s(t)$ in the *canonical* or *standard* form

$$s(t) = s_I(t)\cos(2\pi f_c t) - s_Q(t)\sin(2\pi f_c t) \tag{2.68}$$

We refer to $s_I(t)$ as the *in-phase component* of the band-pass signal $s(t)$ and refer to $s_Q(t)$ as the *quadrature-phase component* or simply the *quadrature component* of the signal $s(t)$.

This nomenclature follows from the following observation: if $\cos(2\pi f_c t)$, the multiplying factor of $s_I(t)$, is viewed as the reference sinusoidal carrier, then $\sin(2\pi f_c t)$, the multiplying factor of $s_Q(t)$, is in phase quadrature with respect to $\cos(2\pi f_c t)$.

According to (2.66), the complex envelope $\tilde{s}(t)$ may be pictured as a *time-varying phasor* positioned at the origin of the (s_I, s_Q)-plane, as indicated in Figure 2.19a. With time t varying continuously, the end of the phasor moves about in the plane. Figure 2.19b depicts the phasor representation of the complex exponential $\exp(2\pi f_c t)$. In the definition given in (2.66), the complex envelope $\tilde{s}(t)$ is multiplied by the complex exponential $\exp(j2\pi f_c t)$. The angles of these two phasors, therefore, add and their lengths multiply, as shown in Figure 2.19c. Moreover, in this latter figure, we show the (s_I, s_Q)-phase rotating with an angular velocity equal to $2\pi f_c$ radians per second. Thus, in the picture portrayed in the figure, the phasor representing the complex envelope $\tilde{s}(t)$ moves in the (s_I, s_Q)-plane, while at the very same time the plane itself rotates about the origin. The original band-pass signal $s(t)$ is the projection of this time-varying phasor on a *fixed line* representing the real axis, as indicated in Figure 2.19c.

Since both $s_I(t)$ and $s_Q(t)$ are low-pass signals limited to the band $-W \le f \le W$, they may be extracted from the band-pass signal $s(t)$ using the scheme shown in Figure 2.20a. Both low-pass filters in this figure are designed identically, each with a bandwidth equal to W.

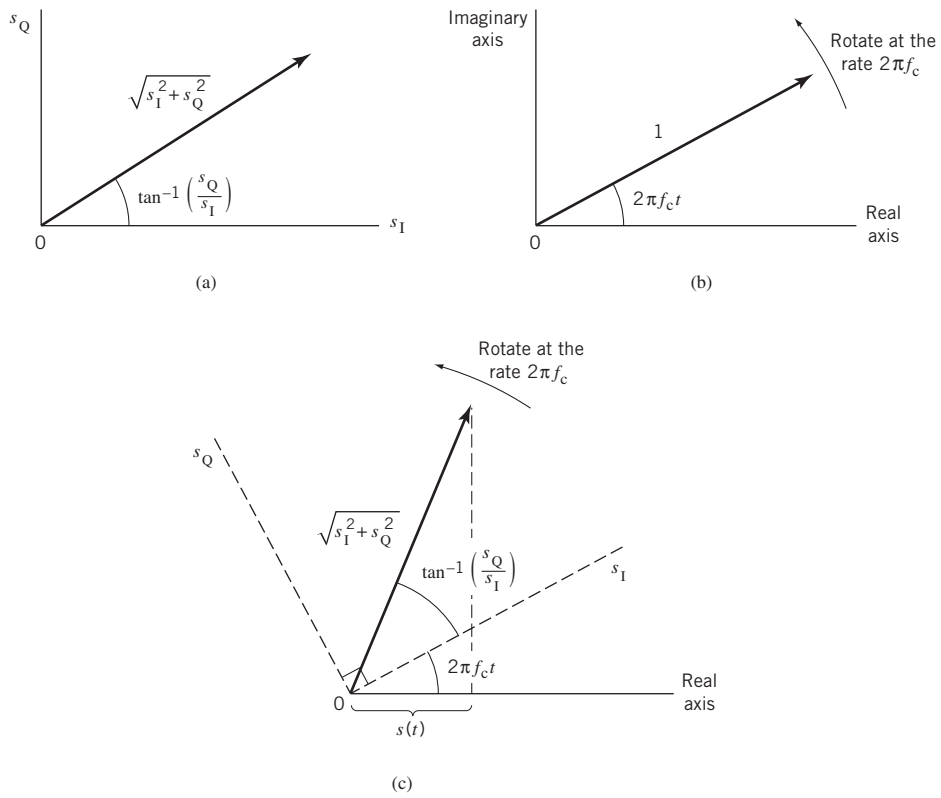


Figure 2.19 Illustrating an interpretation of the complex envelope $\tilde{s}(t)$ and its multiplication by $\exp(j2\pi f_c t)$.

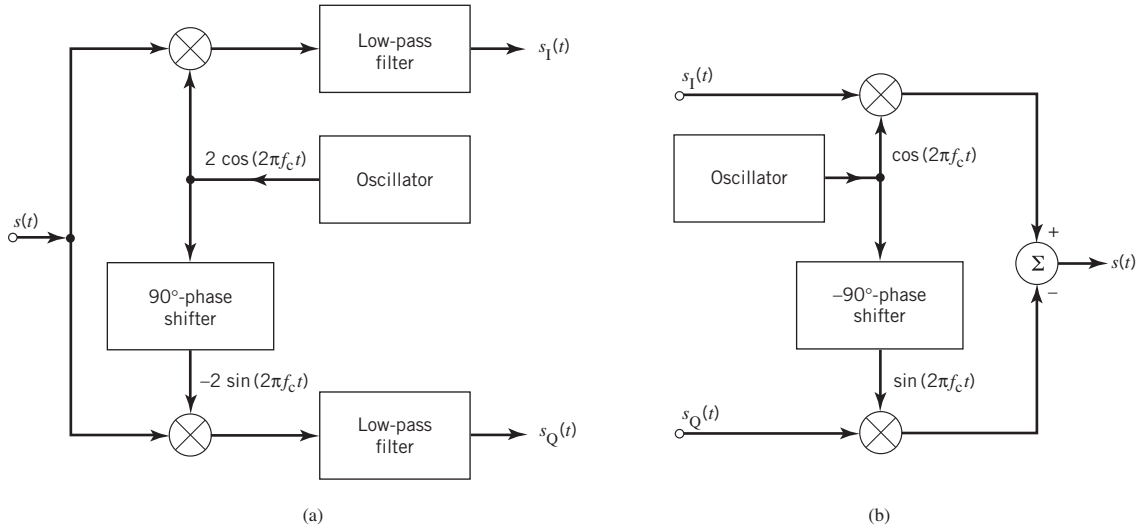


Figure 2.20 (a) Scheme for deriving the in-phase and quadrature components of a band-pass signal $g(t)$. (b) Scheme for reconstructing the band-pass signal from its in-phase and quadrature components.

To reconstruct $s(t)$ from its in-phase and quadrature components, we may use the scheme shown in Figure 2.20b. In light of these statements, we may refer to the scheme in Figure 2.20a as an *analyzer*, in the sense that it extracts the in-phase and quadrature components, $s_I(t)$ and $s_Q(t)$, from the band-pass signal $s(t)$. By the same token, we may refer to the second scheme in Figure 2.20b as a *synthesizer*, in the sense it reconstructs the band-pass signal $s(t)$ from its in-phase and quadrature components, $s_I(t)$ and $s_Q(t)$.

The two schemes shown in Figure 2.20 are basic to the study of *linear modulation schemes, be they of an analog or digital kind*. Multiplication of the low-pass in-phase component $s_I(t)$ by $\cos(2\pi f_c t)$ and multiplication of the quadrature component $s_Q(t)$ by $\sin(2\pi f_c t)$ represent linear forms of modulation. Provided that the carrier frequency f_c is larger than the low-pass bandwidth W, the resulting band-pass function $s(t)$ defined in (2.68) is referred to as a *passband signal waveform*. Correspondingly, the mapping from $s_I(t)$ and $s_Q(t)$ combined into $s(t)$ is known as *passband modulation*.

Polar Representation of Band-Pass Signals

Equation (2.67) is the Cartesian form of defining the complex envelope $\tilde{s}(t)$ of the band-pass signal $s(t)$. Alternatively, we may define $\tilde{s}(t)$ in the *polar form* as

$$\tilde{s}(t) = a(t)\exp[j\phi(t)] \tag{2.69}$$

where $a(t)$ and $\phi(t)$ are both real-valued low-pass functions. Based on the polar representation of (2.69), the original band-pass signal $s(t)$ is itself defined by

$$\tilde{s}(t) = a(t)\cos[2\pi f_c t + \phi(t)] \tag{2.70}$$

We refer to $a(t)$ as the *natural envelope* or simply the *envelope* of the band-pass signal $s(t)$ and refer to $\phi(t)$ as the *phase* of the signal. We now see why the term "pre-envelope" was used in referring to (2.58), the formulation of which *preceded* that of (2.70).

Relationship Between Cartesian and Polar Representations of Band-Pass Signal

The envelope $a(t)$ and phase $\phi(t)$ of a band-pass signal $s(t)$ are respectively related to the in-phase and quadrature components $s_I(t)$ and $s_Q(t)$ as follows (see the time-varying phasor representation of Figure 2.19a):

$$a(t) = \sqrt{s_I^2(t) + s_Q^2(t)} \tag{2.71}$$

and

$$\phi(t) = \tan^{-1}\left(\frac{s_Q(t)}{s_I(t)}\right) \tag{2.72}$$

Conversely, we may write

$$s_I(t) = a(t)\cos[\phi(t)] \tag{2.73}$$

and

$$s_Q(t) = a(t)\sin[\phi(t)] \tag{2.74}$$

Thus, both the in-phase and quadrature components of a band-pass signal contain amplitude and phase information, both of which are uniquely defined for a prescribed phase $\phi(t)$, modulo 2π.

2.12 Complex Low-Pass Representations of Band-Pass Systems

Now that we know how to handle the complex low-pass representation of band-pass signals, it is logical that we develop a corresponding procedure for handling the representation of linear time-invariant band-pass systems. Specifically, we wish to show that the analysis of band-pass systems is greatly simplified by establishing an *analogy*, more precisely an *isomorphism*, between band-pass and low-pass systems. For example, this analogy would help us to facilitate the computer simulation of a wireless communication channel driven by a sinusoidally modulated signal, which otherwise could be a difficult proposition.

Consider a narrowband signal $s(t)$, with its Fourier transform denoted by $S(f)$. We assume that the spectrum of the signal $s(t)$ is limited to frequencies within $\pm W$ hertz of the carrier frequency f_c. We also assume that $W < f_c$. Let the signal $s(t)$ be applied to a linear time-invariant band-pass system with impulse response $h(t)$ and frequency response $H(f)$. We assume that the frequency response of the system is limited to frequencies within $\pm B$ of the carrier frequency f_c. The *system bandwidth* $2B$ is usually narrower than or equal to the input *signal bandwidth* $2W$. We wish to represent the band-pass impulse response $h(t)$ in terms of two quadrature components, denoted by $h_I(t)$ and $h_Q(t)$. In particular, by analogy to the representation of band-pass signals, we express $h(t)$ in the form

$$h(t) = h_I(t)\cos(2\pi f_c t) - h_Q(t)\sin(2\pi f_c t) \tag{2.75}$$

Correspondingly, we define the *complex impulse response* of the band-pass system as

$$\tilde{h}(t) = h_I(t) + jh_Q(t) \tag{2.76}$$

Hence, following (2.66), we may express $h(t)$ in terms of $\tilde{h}(t)$ as

$$h(t) = \mathrm{Re}[\tilde{h}(t)\exp(\mathrm{j}2\pi f_c t)] \tag{2.77}$$

Note that $h_I(t)$, $h_Q(t)$, and $\tilde{h}(t)$ are all low-pass functions, limited to the frequency band $-B \le f \le B$.

We may determine the complex impulse response $\tilde{h}(t)$ in terms of the in-phase and quadrature components $h_I(t)$ and $h_Q(t)$ of the band-pass impulse response $h(t)$ by building on (2.76). Alternatively, we may determine it from the band-pass frequency response $H(f)$ in the following way. We first use (2.77) to write

$$2h(t) = \tilde{h}(t)\exp(\mathrm{j}2\pi f_c t) + \tilde{h}^*(t)\exp(-\mathrm{j}2\pi f_c t) \tag{2.78}$$

where $\tilde{h}^*(t)$ is the *complex conjugate* of $\tilde{h}(t)$; the rationale for introducing the factor of 2 on the left-hand side of (2.78) follows from the fact that if we add a complex signal and its complex conjugate, the sum adds up to twice the real part and the imaginary parts cancel. Applying the Fourier transform to both sides of (2.78) and using the complex-conjugation property of the Fourier transform, we get

$$2H(f) = \tilde{H}(f - f_c) + \tilde{H}^*(-f - f_c) \tag{2.79}$$

where $H(f) \rightleftharpoons h(t)$ and $\tilde{H}(f) \rightleftharpoons \tilde{h}(t)$. Equation (2.79) satisfies the requirement that $H^*(f) = H(-f)$ for a real-valued impulse response $h(t)$. Since $\tilde{H}(f)$ represents a low-pass frequency response limited to $|f| \le B$ with $B < f_c$, we infer from (2.79) that

$$\tilde{H}(f - f_c) = 2H(f), \qquad f > 0 \tag{2.80}$$

Equation (2.80) states:

> For a specified band-pass frequency response $H(f)$, we may determine the corresponding complex low-pass frequency response $\tilde{H}(f)$ by taking the part of $H(f)$ defined for positive frequencies, shifting it to the origin, and scaling it by the factor 2.

Having determined the complex frequency response $\tilde{H}(f)$, we decompose it into its in-phase and quadrature components, as shown by

$$\tilde{H}(f) = \tilde{H}_I(f) + \mathrm{j}\tilde{H}_Q(f) \tag{2.81}$$

where the *in-phase component* is defined by

$$\tilde{H}_I(f) = \frac{1}{2}[\tilde{H}(f) + \tilde{H}^*(-f)] \tag{2.82}$$

and the *quadrature component* is defined by

$$\tilde{H}_Q(f) = \frac{1}{2\mathrm{j}}[\tilde{H}(f) - \mathrm{j}\tilde{H}^*(-f)] \tag{2.83}$$

Finally, to determine the complex impulse response $\tilde{h}(t)$ of the band-pass system, we take the inverse Fourier transform of $\tilde{H}(f)$, obtaining

$$\tilde{h}(t) = \int_{-\infty}^{\infty} \tilde{H}(f)\exp(\mathrm{j}2\pi f t)\, \mathrm{d}f \tag{2.84}$$

which is the formula we have been seeking.

2.13 Putting the Complex Representations of Band-Pass Signals and Systems All Together

Examining (2.66) and (2.77), we immediately see that these two equations share a common multiplying factor: the exponential $\exp(j2\pi f_c t)$. In practical terms, the inclusion of this factor accounts for a sinusoidal carrier of frequency f_c, which facilitates transmission of the modulated (band-pass) signal $s(t)$ across a band-pass channel of midband frequency f_c. In analytic terms, however, the presence of this exponential factor in both (2.66) and (2.77) complicates the analysis of the band-pass system driven by the modulated signal $s(t)$. This analysis can be simplified through the combined use of complex low-pass equivalent representations of both the modulated signal $s(t)$ and the band-pass system characterized by the impulse response $h(t)$. The simplification can be carried out in the time domain or frequency domain, as discussed next.

The Time-Domain Procedure

Equipped with the complex representations of band-pass signals and systems, we are ready to derive an analytically efficient method for determining the output of a band-pass system driven by a corresponding band-pass signal. To proceed with the derivation, assume that $S(f)$, denoting the spectrum of the input signal $s(t)$, and $H(f)$, denoting the frequency response of the system, are both centered around the same frequency f_c. In practice, there is no need to consider a situation in which the carrier frequency of the input signal is not aligned with the midband frequency of the band-pass system, since we have considerable freedom in choosing the carrier or midband frequency. Thus, changing the carrier frequency of the input signal by an amount Δf_c, for example, simply corresponds to absorbing (or removing) the factor $\exp(\pm j2\pi\Delta f_c t)$ in the complex envelope of the input signal or the complex impulse response of the band-pass system. We are therefore justified in proceeding on the assumption that $S(f)$ and $H(f)$ are both centered around the same carrier frequency f_c.

Let $x(t)$ denote the output signal of the band-pass system produced in response to the incoming band-pass signal $s(t)$. Clearly, $x(t)$ is also a band-pass signal, so we may represent it in terms of its own low-pass complex envelope $\tilde{x}(t)$ as

$$x(t) = \mathrm{Re}[\tilde{x}(t)\exp(j2\pi f_c t)] \tag{2.85}$$

The output signal $x(t)$ is related to the input signal $s(t)$ and impulse response $h(t)$ of the system in the usual way by the convolution integral

$$x(t) = \int_{-\infty}^{\infty} h(\tau)s(t-\tau)\,d\tau \tag{2.86}$$

In terms of pre-envelopes, we have $h(t) = \mathrm{Re}[h_+(t)]$ and $s(t) = \mathrm{Re}[s_+(t)]$. We may therefore rewrite (2.86) in terms of the pre-envelopes $s_+(t)$ and $h_+(t)$ as

$$x(t) = \int_{-\infty}^{\infty} \mathrm{Re}[h_+(\tau)]\mathrm{Re}[s_+(t-\tau)]\,d\tau \tag{2.87}$$

To proceed further, we make use of a basic property of pre-envelopes that is described by the following relation:

$$\int_{-\infty}^{\infty} \mathrm{Re}[h_+(\tau)]\mathrm{Re}[s_+(\tau)]\,\mathrm{d}\tau = \frac{1}{2}\mathrm{Re}\left[\int_{-\infty}^{\infty} h_+(\tau)s_+^*(\tau)\,\mathrm{d}\tau\right] \tag{2.88}$$

where we have used τ as the integration variable to be consistent with that in (2.87); details of (2.88) are presented in Problem 2.20. Next, from Fourier-transform theory we note that using $s(-\tau)$ in place of $s(\tau)$ has the effect of removing the complex conjugation on the right-hand side of (2.88). Hence, bearing in mind the algebraic difference between the argument of $s_+(\tau)$ in (2.88) and that of $s_+(t-\tau)$ in (2.87), and using the relationship between the pre-envelope and complex envelope of a band-pass signal, we may express (2.87) in the equivalent form

$$\begin{aligned} x(t) &= \frac{1}{2}\mathrm{Re}\left[\int_{-\infty}^{\infty} h_+(\tau)s_+(t-\tau)\,\mathrm{d}\tau\right] \\ &= \frac{1}{2}\mathrm{Re}\left\{\int_{-\infty}^{\infty} \tilde{h}(\tau)\exp(\mathrm{j}2\pi f_c\tau)\tilde{s}(t-\tau)\exp[\mathrm{j}2\pi f_c(t-\tau)]\,\mathrm{d}\tau\right\} \\ &= \frac{1}{2}\mathrm{Re}\left[\exp(\mathrm{j}2\pi f_c t)\int_{-\infty}^{\infty} \tilde{h}(\tau)\tilde{s}(t-\tau)\,\mathrm{d}\tau\right] \end{aligned} \tag{2.89}$$

Thus, comparing the right-hand sides of (2.85) and (2.89), we readily find that for a large enough carrier frequency f_c, the complex envelope $\tilde{x}(t)$ of the output signal is simply defined in terms of the complex envelope $\tilde{s}(t)$ of the input signal and the complex impulse response $\tilde{h}(t)$ of the band-pass system as follows:

$$\tilde{x}(t) = \frac{1}{2}\int_{-\infty}^{\infty} \tilde{h}(t)\tilde{s}(t-\tau)\,\mathrm{d}\tau \tag{2.90}$$

This important relationship is the result of the *isomorphism* between a band-pass function and the corresponding complex low-pass function, in light of which we may now make the following summarizing statement:

> Except for the scaling factor 1/2, the complex envelope $\tilde{x}(t)$ of the output signal of a band-pass system is obtained by convolving the complex impulse response $\tilde{h}(t)$ of the system with the complex envelope $\tilde{s}(t)$ of the input band-pass signal.

In computational terms, the significance of this statement is profound. Specifically, in dealing with band-pass signals and systems, we need only concern ourselves with the functions $\tilde{s}(t)$, $\tilde{x}(t)$, and $\tilde{h}(t)$, representing the complex low-pass equivalents of the excitation applied to the input of the system, the response produced at the output of the system, and the impulse response of the system respectively, as illustrated in Figure 2.21. The essence of the filtering process performed in the original system of Figure 2.21a is completely retained in the complex low-pass equivalent representation depicted in Figure 2.21b.

The complex envelope $\tilde{s}(t)$ of the input band-pass signal and the complex impulse response $\tilde{h}(t)$ of the band-pass system are defined in terms of their respective in-phase

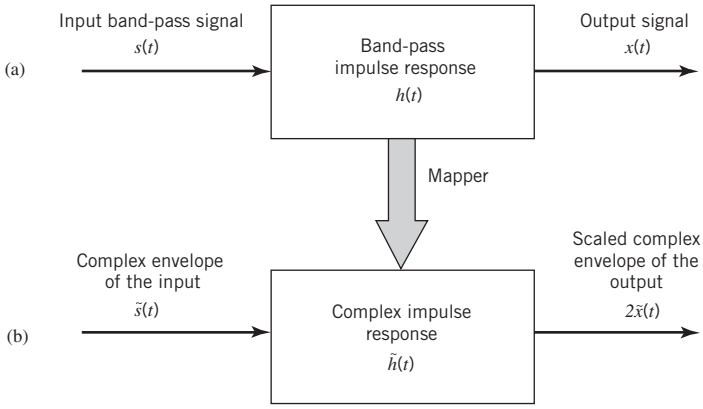


Figure 2.21
(a) Input–output description of a band-pass system; (b) Complex low-pass equivalent model of the band-pass system.

and quadrature components by (2.67) and (2.76), respectively. Substituting these relations into (2.90), we get

$$\begin{aligned} 2\tilde{x}(t) &= \tilde{h}(t)\bigstar\tilde{s}(t) \\ &= [h_{\mathrm{I}}(t) + jh_{\mathrm{Q}}(t)]\bigstar[\mathrm{s}_{\mathrm{I}}(t) + \mathrm{j}s_{\mathrm{Q}}(t)] \end{aligned} \tag{2.91}$$

where the symbol $\bigstar$ denotes convolution. Because convolution is *distributive*, we may rewrite (2.91) in the equivalent form

$$2\tilde{x}(t) = [h_{\mathrm{I}}(t)\bigstar s_{\mathrm{I}}(t) - h_{\mathrm{Q}}(t)\bigstar s_{\mathrm{Q}}(t)] + \mathrm{j}[h_{\mathrm{Q}}(t)\bigstar s_{\mathrm{I}}(t) + h_{\mathrm{I}}\bigstar s_{\mathrm{Q}}(t)] \tag{2.92}$$

Let the complex envelope $\tilde{x}(t)$ of the response be defined in terms of its in-phase and quadrature components as

$$\tilde{x}(t) = x_{\mathrm{I}}(t) + \mathrm{j}x_{\mathrm{Q}}(t) \tag{2.93}$$

Then, comparing the real and imaginary parts in (2.92) and (2.93), we find that the in-phase component $x_{\mathrm{I}}(t)$ is defined by the relation

$$2x_{\mathrm{I}}(t) = h_{\mathrm{I}}(t)\bigstar s_{\mathrm{I}}(t) - h_{\mathrm{Q}}(t)\bigstar s_{\mathrm{Q}}(t) \tag{2.94}$$

and its quadrature component $x_{\mathrm{Q}}(t)$ is defined by the relation

$$2x_{\mathrm{Q}}(t) = h_{\mathrm{Q}}(t)\bigstar s_{\mathrm{I}}(t) + h_{\mathrm{I}}(t)\bigstar s_{\mathrm{Q}}(t) \tag{2.95}$$

Thus, for the purpose of evaluating the in-phase and quadrature components of the complex envelope $\tilde{x}(t)$ of the system output, we may use the *low-pass equivalent model* shown in Figure 2.22. All the signals and impulse responses shown in this model are real-valued low-pass functions; hence a time-domain procedure for simplifying the analysis of band-pass systems driven by band-pass signals.

The Frequency-Domain Procedure

Alternatively, Fourier-transforming the convolution integral of (2.90) and recognizing that convolution in the time domain is changed into multiplication in the frequency domain, we get

$$\tilde{X}(f) = \frac{1}{2}\tilde{H}(f)\tilde{S}(f) \tag{2.96}$$

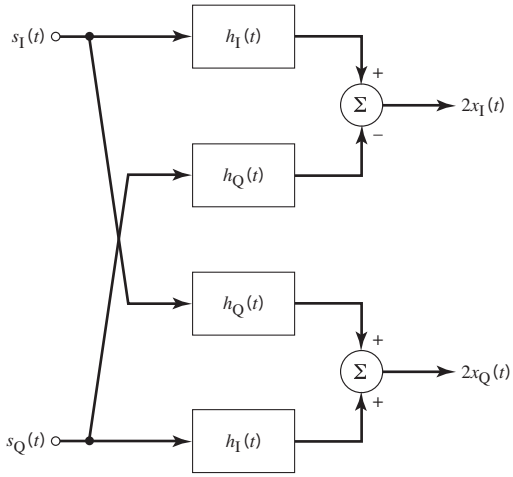


Figure 2.22 Block diagram illustrating the relationship between the in-phase and quadrature components of the response of a band-pass filter and those of the input signal.

where $\tilde{s}(t) \rightleftharpoons \tilde{S}(f)$, $\tilde{h}(t) \rightleftharpoons \tilde{H}(f)$, and $\tilde{x}(t) \rightleftharpoons \tilde{X}(f)$. The $\tilde{H}(f)$ is itself related to the frequency response $H(f)$ of the band-pass system by (2.80). Thus, assuming that $H(f)$ is known, we may use the frequency-domain procedure summarized in Table 2.4 for computing the system output $x(t)$ in response to the system input $s(t)$.

In actual fact, the procedure of Table 2.4 is the frequency-domain representation of the low-pass equivalent to the band-pass system, depicted in Figure 2.21b. In computational terms, this procedure is of profound practical significance. We say so because its use alleviates the analytic and computational difficulty encountered in having to include the carrier frequency f_c in the pertinent calculations.

As discussed earlier in the chapter, the theoretical formulation of the low-pass equivalent in Figure 2.21b is rooted in the Hilbert transformation, the evaluation of which poses a practical problem of its own, because of the wideband 90^o-phase shifter involved in its theory. Fortunately, however, we do not need to invoke the Hilbert transform in constructing the low-pass equivalent. This is indeed so, when a message signal modulated onto a sinusoidal carrier is processed by a band-pass filter, as explained here:

1. Typically, the message signal is band limited for all practical purposes. Moreover, the carrier frequency is larger than the highest frequency component of the signal; the modulated signal is therefore a band-pass signal with a well-defined passband. Hence, the in-phase and quadrature components of the modulated signal $s(t)$, represented respectively by $s_I(t)$ and $s_Q(t)$, are readily obtained from the canonical representation of $s(t)$, described in (2.68).
2. Given the well-defined frequency response $H(f)$ of the band-pass system, we may readily evaluate the corresponding complex low-pass frequency response $\tilde{H}(f)$; see (2.80). Hence, we may compute the system output $x(t)$ produced in response to the carrier-modulated input $s(t)$ without invoking the Hilbert transform.

Table 2.4 Procedure for the computational analysis of a band-pass system driven by a band-pass signal

Given the frequency response $H(f)$ of a band-pass system, computation of the output signal $x(t)$ of the system in response to an input band-pass signal $s(t)$ is summarized as follows:

1. Use (2.80), namely $\tilde{H}(f-f_c) = 2H(f)$, for $f > 0$ to determine $\tilde{H}(f)$.
2. Expressing the input band-pass signal $s(t)$ in the canonical form of (2.68), evaluate the complex envelope $\tilde{s}(t) = s_I(t) + js_Q(t)$ where $s_I(t)$ is the in-phase component of $s(t)$ and $s_Q(t)$ is its quadrature component. Hence, compute the Fourier transform $\tilde{S}(f) = \mathbf{F}[\tilde{s}(t)]$
3. Using (2.96), compute $\tilde{X}(f) = \frac{1}{2}\tilde{H}(f)\tilde{S}(f)$, which defines the Fourier transform of the complex envelope $\tilde{x}(t)$ of the output signal $x(t)$.
4. Compute the inverse Fourier transform of $\tilde{X}(f)$, yielding $\tilde{x}(t) = \mathbf{F}^{-1}[\tilde{X}(f)]$
5. Use (2.85) to compute the desired output signal $x(t) = \mathrm{Re}[\tilde{x}(t)\exp(j2\pi f_c t)]$

Procedure for Efficient Simulation of Communication Systems

To summarize, the frequency-domain procedure described in Table 2.4 is well suited for the efficient simulation of communication systems on a computer for two reasons:

1. The low-pass equivalents of the incoming band-pass signal and the band-pass system work by eliminating the exponential factor $\exp(j2\pi f_c t)$ from the computation without loss of information.
2. The *fast Fourier transform (FFT) algorithm*, discussed later in the chapter, is used for numerical computation of the Fourier transform. This algorithm is used twice in Table 2.4, once in step 2 to perform Fourier transformation, and then again in step 4 to perform inverse Fourier transformation.

The procedure of this table, rooted largely in the frequency domain, assumes availability of the band-pass system's frequency response $H(f)$. If, however, it is the system's impulse response $h(t)$ that is known, then all we need is an additional step to Fourier transform $h(t)$ into $H(f)$ before initiating the procedure of Table 2.4.

2.14 Linear Modulation Theory

The material presented in Sections 2.8–2.13 on the complex low-pass representation of band-pass signals and systems is of profound importance in the study of communication theory. In particular, we may use the canonical formula of (2.68) as the mathematical basis for a unified treatment of linear modulation theory, which is the subject matter of this section.

We start this treatment with a formal definition:

> Modulation is a process by means of which one or more parameters of a sinusoidal carrier are varied in accordance with a message signal so as to facilitate transmission of that signal over a communication channel.

The message signal (e.g., voice, video, data sequence) is referred to as the *modulating signal*, and the result of the modulation process is referred to as the *modulated signal.* Naturally, in a communication system, modulation is performed in the transmitter. The reverse of modulation, aimed at recovery of the original message signal in the receiver, is called *demodulation*.

Consider the block diagram of Figure 2.23, depicting a modulator, where $m(t)$ is the message signal, $\cos(2\pi f_c t)$ is the carrier, and $s(t)$ is the modulated signal. To apply (2.68) to this modulator, the in-phase component $s_I(t)$ in that equation is treated simply as a scaled version of the message signal denoted by $m(t)$. As for the quadrature component $s_Q(t)$, it is defined by a *spectrally shaped* version of $m(t)$ that is performed linearly. In such a scenario, it follows that a modulated signal $s(t)$ defined by (2.68) is a *linear function* of the message signal $m(t)$; hence the reference to this equation as the mathematical basis of *linear modulation theory.*

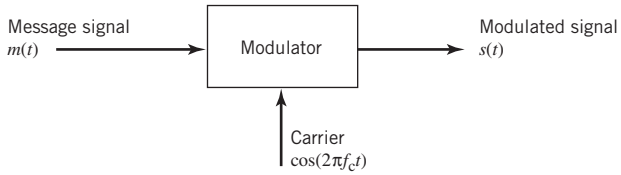


Figure 2.23 Block diagram of a modulator.

To recover the original message signal $m(t)$ from the modulated signal $s(t)$, we may use a demodulator, the block diagram of which is depicted in Figure 2.24. An elegant feature of linear modulation theory is that demodulation of $s(t)$ is also achieved using linear operations. However, for linear demodulation of $s(t)$ to be feasible, the locally generated carrier in the demodulator of Figure 2.24 has to be synchronous with the original sinusoidal carrier used in the modulator of Figure 2.23. Accordingly, we speak of *synchronous demodulation* or *coherent detection*.

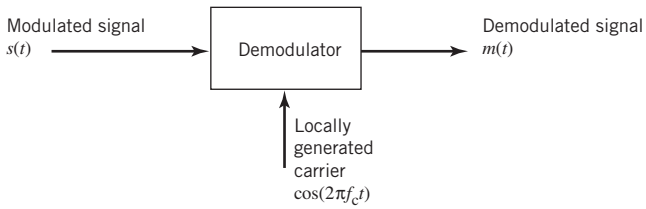


Figure 2.24 Block diagram of a demodulator.

Depending on the spectral composition of the modulated signal, we have three kinds of linear modulation in analog communications:

- double sideband-suppressed carrier (DSB-SC) modulation;
- vestigial sideband (VSB) modulation;
- single sideband (SSB) modulation.

These three methods of modulation are discussed in what follows and in this order.

DSB-SC Modulation

DSB-SC modulation is the simplest form of linear modulation, which is obtained by setting

$$s_I(t) = m(t)$$

and

$$s_Q(t) = 0$$

Accordingly, (2.68) is reduced to

$$s(t) = m(t)\cos(2\pi f_c t) \tag{2.97}$$

the implementation of which simply requires a *product modulator* that multiplies the message signal $m(t)$ by the carrier $\cos(2nf_c t)$, assumed to be of unit amplitude.

For a frequency-domain description of the DSB-SC-modulated signal defined in (2.97), suppose that the message signal $m(t)$ occupies the frequency band $-W \le f \le W$, as depicted in Figure 2.25a; hereafter, W is referred to as the *message bandwidth*. Then, provided that the carrier frequency satisfies the condition $f_c > W$, we find that the spectrum of the DSB-SC-modulated signal consists of an *upper sideband* and *lower sideband*, as depicted in Figure 2.25b. Comparing the two parts of this figure, we immediately see that the *channel bandwidth*, B, required to support the transmission of the DSB-SC-modulated signal from the transmitter to the receiver is twice the message bandwidth.

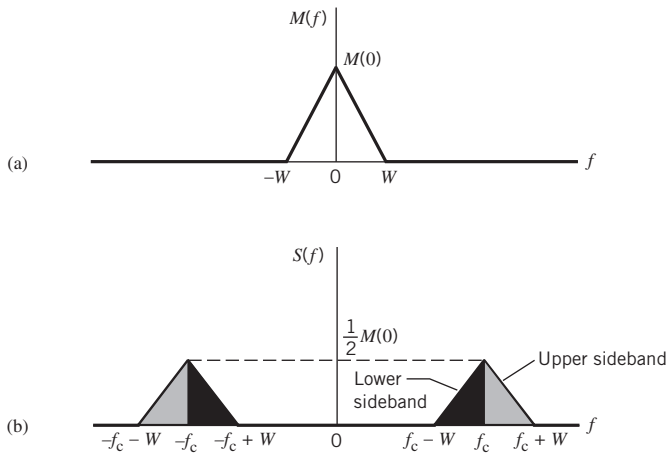


Figure 2.25 (a) Message spectrum. (b) Spectrum of DSB-SC modulated wave $s(t)$, assuming $f_c > W$.

One other interesting point apparent from Figure 2.25b is that the spectrum of the DSB-SC modulated signal is entirely void of delta functions. This statement is further testimony to the fact that the carrier is suppressed from the generation of the modulated signal $s(t)$ of (2.97).

Summarizing the useful features of DSB-SC modulation:

- *suppression of the carrier*, which results in saving of transmitted power;
- *desirable spectral characteristics*, which make it applicable to the modulation of band-limited message signals;
- ease of synchronizing the receiver to the transmitter for coherent detection.

On the downside, DSB-SC modulation is wasteful of channel bandwidth. We say so for the following reason. The two sidebands, constituting the spectral composition of the modulated signal $s(t)$, are actually the *image* of each other with respect to the carrier frequency f_c; hence, the transmission of either sideband is sufficient for transporting $s(t)$ across the channel.

VSB Modulation

In VSB modulation, one sideband is partially suppressed and a *vestige* of the other sideband is configured in such a way to compensate for the partial sideband suppression by exploiting the fact that the two sidebands in DSB-SC modulation are the image of each other. A popular method of achieving this design objective is to use the *frequency discrimination method*. Specifically, a DSB-SC-modulated signal is first generated using a product modulator, followed by a band-pass filter, as shown in Figure 2.26. The desired spectral shaping is thereby realized through the appropriate design of the band-pass filter.

Suppose that a vestige of the lower sideband is to be transmitted. Then, the frequency response of the band-pass filter, $H(f)$, takes the form shown in Figure 2.27; to simplify matters, only the frequency response for positive frequencies is shown in the figure. Examination of this figure reveals two characteristics of the band-pass filter:

1. *Normalization* of the frequency response, which means that

$$H(f) = \begin{cases} 1 & \text{for } f_c + f_v \le |f| < f_c + W \\ \frac{1}{2} & \text{for } |f| = f_c \end{cases} \tag{2.98}$$

   where f_v is the *vestigial bandwidth* and the other parameters are as previously defined.

2. *Odd symmetry of the cutoff portion inside the transition interval* $f_c - f_v \le |f| \le f_c + f_v$, which means that values of the frequency response $H(f)$ at any two frequencies equally spaced above and below the carrier frequency add up to unity.

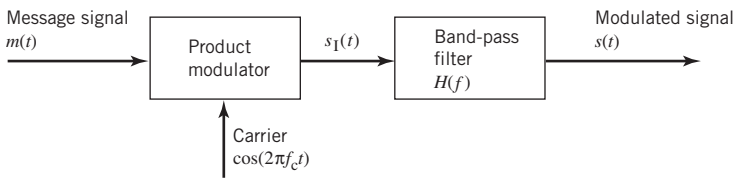


Figure 2.26
Frequency-discrimination method for producing VSB modulation where the intermediate signal $s_I(t)$ is DSB-SC modulated.

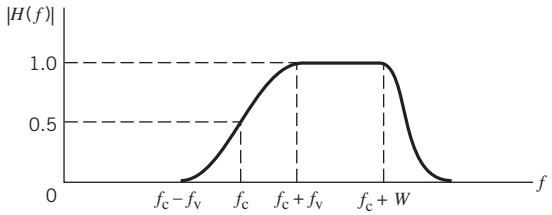


Figure 2.27 Magnitude response of VSB filter; only the positive-frequency portion is shown

Consequently, we find that shifted versions of the frequency response $H(f)$ satisfy the condition

$$H(f-f_c) + H(f+f_c) = 1 \qquad \text{for } -W \le |f| \le W \tag{2.99}$$

Outside the frequency band of interest defined by $|f| \ge f_c + W$, the frequency response $H(f)$ can assume arbitrary values. We may thus express the channel bandwidth required for the transmission of VSB-modulated signals as

$$B = W + f_v \tag{2.100}$$

With this background, we now address the issue of how to specify $H(f)$. We first use the canonical formula of (2.68) to express the VSB-modulated signal $s_1(t)$, containing a vestige of the lower sideband, as

$$s_1(t) = \frac{1}{2}m(t)\cos(2\pi f_c t) - \frac{1}{2}m_Q(t)\sin(2\pi f_c t) \tag{2.101}$$

where $m(t)$ is the message signal, as before, and $m_Q(t)$ is the spectrally shaped version of $m(t)$; the reason for the factor 1/2 will become apparent later. Note that if $m_Q(t)$ is set equal to zero, (2.101) reduces to DSB-SC modulation. It is therefore in the *quadrature signal* $m_Q(t)$ that VSB modulation distinguishes itself from DSB-SC modulation. In particular, the role of $m_Q(t)$ is to interfere with the message signal $m(t)$ in such a way that power in one of the sidebands of the VSB-modulated signal $s(t)$ (e.g., the lower sideband in Figure 2.27) is appropriately reduced.

To determine $m_Q(t)$, we examine two different procedures:

1. *Phase-discrimination,* which is rooted in the time-domain description of (2.101); transforming this equation into the frequency domain, we obtain

$$S_1(f) = \frac{1}{4}[M(f-f_c) + M(f+f_c)] - \frac{1}{4j}[M_Q(f-f_c) - M_Q(f+f_c)] \tag{2.102}$$

where

$$M(f) = \mathbf{F}[m(t)] \qquad \text{and} \qquad M_Q(f) = \mathbf{F}[m_Q(t)]$$

2. *Frequency-discrimination,* which is structured in the manner described in Figure 2.26; passing the DSB-SC-modulated signal (i.e., the intermediate signal $s_I(t)$ in Figure 2.26) through the band-pass filter, we write

$$S_1(f) = \frac{1}{2}[M(f-f_c) + M(f+f_c)]H(f) \tag{2.103}$$

In both (2.102) and (2.103), the spectrum $S_1(f)$ is defined in the frequency interval

$$f_c - W \le |f| \le f_c + W$$

Equating the right-hand sides of these two equations, we get (after canceling common terms)

$$\begin{aligned}\frac{1}{2}[M(f-f_c) + M(f+f_c)] - \frac{1}{2j}[M_Q(f-f_c) - M_Q(f+f_c)] \\ = [M(f-f_c) + M(f+f_c)]H(f)\end{aligned} \tag{2.104}$$

Shifting both sides of (2.104) to the left by the amount f_c, we get (after canceling common terms)

$$\frac{1}{2}M(f) - \frac{1}{2j}M_Q(f) = M(f)H(f+f_c), \qquad -W \le |f| \le W \tag{2.105}$$

where the terms $M(f+2f_c)$ and $M_Q(f+2f_c)$ are ignored as they both lie outside the interval $-W \le |f| \le W$. Next, shifting both sides of (2.104) by the amount f_c, but this time to the *right*, we get (after canceling common terms)

$$\frac{1}{2}M(f) + \frac{1}{2j}M_Q(f) = M(f)H(f-f_c), \qquad -W \le |f| \le W \tag{2.106}$$

where, this time, the terms $M(f-2f_c)$ and $M_Q(f-2f_c)$ are ignored as they both lie outside the interval $-W \le |f| \le W$.

Given (2.105) and (2.106), all that remains to be done now is to follow two simple steps:

1. Adding these two equations and then factoring out the common term $M(f)$, we get the condition of (2.99) previously imposed on $H(f)$; indeed, it is with this condition in mind that we introduced the scaling factor 1/2 in (2.101).
2. Subtracting (2.105) from (2.106) and rearranging terms, we get the desired relationship between $M_Q(f)$ and $M(f)$:

$$M_Q(f) = j[H(f-f_c) - H(f+f_c)]M(f), \qquad -W \le |f| \le W \tag{2.107}$$

Let $H_Q(f)$ denote the frequency response of a *quadrature filter* that operates on the message spectrum $M(f)$ to produce $M_Q(f)$. In light of (2.107), we may readily define $H_Q(f)$ in terms of $H(f)$ as

$$\begin{aligned}H_Q(f) &= \frac{M_Q(f)}{M(f)} \\ &= j[H(f-f_c) - H(f+f_c)], \qquad -W \le |f| \le W\end{aligned} \tag{2.108}$$

Equation (2.108) provides the frequency-domain basis for the *phase-discrimination method* for generating the VSB-modulated signal $s_1(t)$, where only a vestige of the lower sideband is retained. With this equation at hand, it is instructive to plot the frequency response $H_Q(f)$. For the frequency interval $-W \le f \le W$, the term $H(f-f_c)$ is defined by the response $H(f)$ for negative frequencies shifted to the right by f_c, whereas the term $H(f+f_c)$ is defined by the response $H(f)$ for positive frequencies shifted to the left by f_c. Accordingly, building on the positive frequency response plotted in Figure 2.27, we find that the corresponding plot of $H_Q(f)$ is shaped as shown in Figure 2.28.

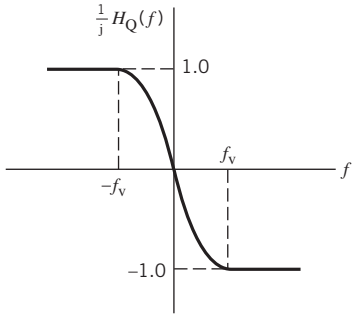


Figure 2.28 Frequency response of the quadrature filter for producing the quadrature component of the VSB wave.

The discussion on VSB modulation has thus far focused on the case where a vestige of the lower sideband is transmitted. For the alternative case when a vestige of the upper sideband is transmitted, we find that the corresponding VSB-modulated wave is described by

$$s_2(t) = \frac{1}{2}m(t)\cos(2\pi f_c t) + \frac{1}{2}m_Q(t)\sin(2\pi f_c t) \tag{2.109}$$

where the quadrature signal $m_Q(t)$ is constructed from the message signal $m(t)$ in exactly the same way as before.

Equations (2.101) and (2.109) are of the same mathematical form, except for an algebraic difference; they may, therefore, be combined into the single formula

$$s(t) = \frac{1}{2}m(t)\cos(2\pi f_c t) \mp \frac{1}{2}m_Q(t)\sin(2\pi f_c t) \tag{2.110}$$

where the minus sign applies to a VSB-modulated signal containing a vestige of the lower sideband and the plus sign applies to the alternative case when the modulated signal contains a vestige of the upper sideband.

The formula of (2.110) for VSB modulation includes DSB-SC modulation as a special case. Specifically, setting $m_Q(t) = 0$, this formula reduces to that of (2.97) for DSB-SC modulation, except for the trivial scaling factor of 1/2.

SSB Modulation

Next, considering *SSB modulation*, we may identify two choices:

1. The carrier and the lower sideband are both suppressed, leaving the upper sideband for transmission in its full spectral content; this first SSB-modulated signal is denoted by $s_{USB}(t)$.
2. The carrier and the upper sideband are both suppressed, leaving the lower sideband for transmission in its full spectral content; this second SSB-modulated signal is denoted by $s_{LSB}(t)$.

The Fourier transforms of these two modulated signals are the *image* of each other with respect to the carrier frequency f_c, which, as mentioned previously, emphasizes that the transmission of either sideband is actually sufficient for transporting the message signal $m(t)$ over the communication channel. In practical terms, both $s_{USB}(t)$ and $s_{LSB}(t)$ require

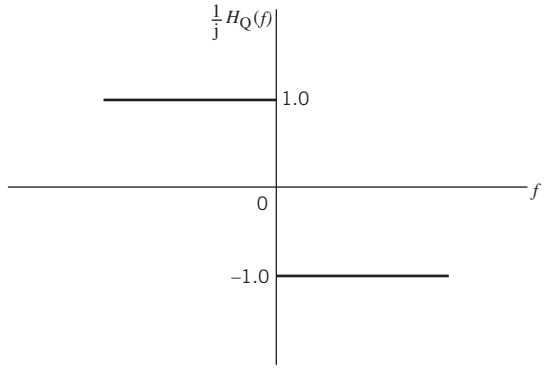


Figure 2.29
Frequency response of the quadrature filter in SSB modulation.

the smallest feasible channel bandwidth, $B = W$, without compromising the perfect recovery of the message signal under noiseless conditions. It is for these reasons that we say SSB modulation is the *optimum form of linear modulation* for analog communications, preserving both the transmitted power and channel bandwidth in the best manner possible.

SSB modulation may be viewed as a special case of VSB modulation. Specifically, setting the vestigial bandwidth $f_v = 0$, we find that the frequency response of the quadrature filter plotted in Figure 2.28 takes the limiting form of the *signum function* shown in Figure 2.29. In light of the material presented in (2.60) on Hilbert transformation, we therefore find that for $f_v = 0$ the quadrature component $m_Q(t)$ becomes the Hilbert transform of the message signal $m(t)$, denoted by $\hat{m}(t)$. Accordingly, using $\hat{m}(t)$ in place of $m_Q(t)$ in (2.110) yields the SSB formula

$$s(t) = \frac{1}{2}m(t)\cos(2\pi f_c t) \mp \frac{1}{2}\hat{m}(t)\sin(2\pi f_c t) \tag{2.111}$$

where the minus sign applies to the SSB-modulated signal $s_{USB}(t)$ and the plus sign applies to the alternative SSB-modulated signal $s_{LSB}(t)$.

Unlike DSB-SC and VSB methods of modulation, SSB modulation is of limited applicability. Specifically, we say:

> For SSB modulation to be feasible in practical terms, the spectral content of the message signal $m(t)$ must have an energy gap centered on the origin.

This requirement, illustrated in Figure 2.30, is imposed on the message signal $m(t)$ so that the band-pass filter in the frequency-discrimination method of Figure 2.26 has a *finite transition* band for the filter to be physically realizable. With the transition band separating the pass-band from the stop-band, it is only when the transition band is finite that the undesired sideband can be suppressed. An example of message signals for which the energy-gap requirement is satisfied is voice signals; for such signals, the energy gap is about 600 Hz, extending from –300 to +300 Hz.

In contrast, the spectral contents of television signals and wideband data extend practically to a few hertz, thereby ruling out the applicability of SSB modulation to this second class of message signals. It is for this reason that VSB modulation is preferred over SSB modulation for the transmission of wideband signals.

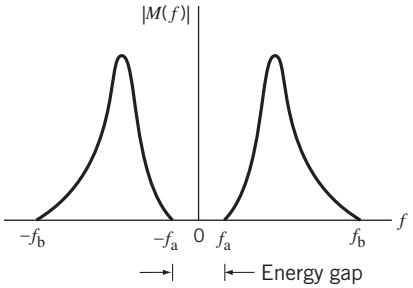


Figure 2.30
Spectrum of a message signal $m(t)$ with an energy gap centered around the origin.

Summary of Linear Modulation Methods

Equation (2.97) for DSB-SC modulation, (2.110) for VSB modulation, and (2.111) for SSB modulation are summarized in Table 2.5 as special cases of the canonical formula of (2.68). Correspondingly, we may treat the time-domain generations of these three linearly modulated signals as special cases of the "synthesizer" depicted in Figure 2.20b.

Table 2.5 Summary of linear modulation methods viewed as special cases of the canonical formula $s(t) = s_I(t)\cos(2\pi f_c t) - s_Q(t)\sin(2\pi f_c t)$

Type of modulation	In-phase component, $s_I(t)$	Quadrature component, $s_Q(t)$	Comments
DSB-SC	$m(t)$	zero	$m(t)$ = message signal
VSB	$\frac{1}{2}m(t)$	$\pm\frac{1}{2}m_Q(t)$	Plus sign applies to using vestige of lower sideband and minus sign applies to using vestige of upper sideband
SSB	$\frac{1}{2}m(t)$	$\pm\frac{1}{2}\hat{m}(t)$	Plus sign applies to transmission of upper sideband and minus sign applies to transmission of lower sideband

2.15 Phase and Group Delays

A discussion of signal transmission through linear time-invariant systems is incomplete without considering the phase and group delays involved in the signal transmission process.

Whenever a signal is transmitted through a dispersive system, exemplified by a communication channel (or band-pass filter), some *delay* is introduced into the output signal, the delay being measured with respect to the input signal. In an ideal channel, the phase response varies *linearly* with frequency inside the passband of the channel, in which case the filter introduces a constant delay equal to t_0, where the parameter t_0 controls the slope of the linear phase response of the channel. Now, what if the phase response of the channel is a nonlinear function of frequency, which is frequently the case in practice? The purpose of this section is to address this practical issue.

To begin the discussion, suppose that a steady sinusoidal signal at frequency f_c is transmitted through a dispersive channel that has a phase-shift of $\beta(f_c)$ radians at that frequency. By using two phasors to represent the input signal and the received signal, we see that the received signal phasor lags the input signal phasor by $\beta(f_c)$ radians. The time taken by the received signal phasor to sweep out this phase lag is simply equal to the ratio $\beta(f_c)/(2\pi f_c)$ seconds. This time is called the *phase delay* of the channel.

It is important to realize, however, that the phase delay is not necessarily the true signal delay. This follows from the fact that a steady sinusoidal signal does *not* carry information, so it would be incorrect to deduce from the above reasoning that the phase delay is the true signal delay. To substantiate this statement, suppose that a slowly varying signal, over the interval $-(T/2) \le t \le (T/2)$, is multiplied by the carrier, so that the resulting modulated signal consists of a narrow group of frequencies centered around the carrier frequency; the DSB-SC waveform of Figure 2.31 illustrates such a modulated signal. When this modulated signal is transmitted through a communication channel, we find that there is indeed a delay between the envelope of the input signal and that of the received signal. This delay, called the *envelope* or *group delay* of the channel, represents the true signal delay insofar as the information-bearing signal is concerned.

Assume that the dispersive channel is described by the transfer function

$$H(f) = K\exp[j\beta(f)] \tag{2.112}$$

where the amplitude K is a constant scaling factor and the phase $\beta(f)$ is a nonlinear function of frequency f; it is the nonlinearity of $\beta(f)$ that is responsible for the dispersive

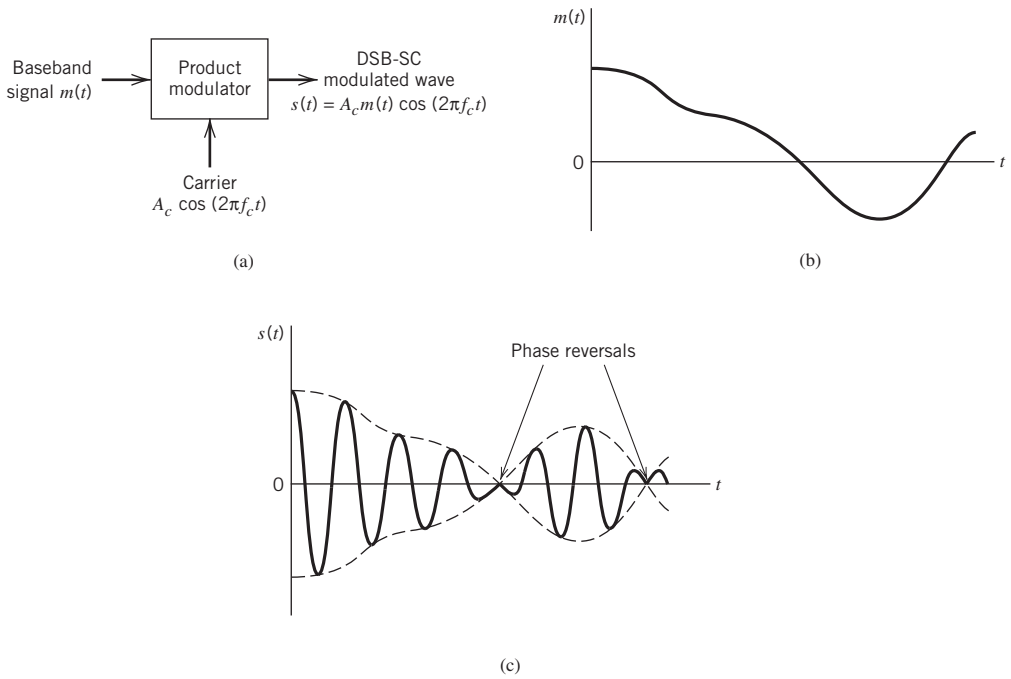


Figure 2.31 (a) Block diagram of product modulator; (b) Baseband signal; (c) DSB-SC modulated wave.

nature of the channel. The input signal $s(t)$ is assumed to be of the kind displayed in Figure 2.31; that is, the DSB-SC-modulated signal

$$s(t) = m(t)\cos(2\pi f_c t) \tag{2.113}$$

where $m(t)$ is the message signal, assumed to be of a low-pass kind and limited to the frequency interval $|f| \leq W$. Moreover, we assume that the carrier frequency $f_c > W$. By expanding the phase $\beta(f)$ in a *Taylor series* about the point $f = f_c$ and retaining only the first two terms, we may approximate $\beta(f)$ as

$$\beta(f) \approx \beta(f_c) + (f - f_c)\frac{\partial \beta(f)}{\partial f}\bigg|_{f=f_c} \tag{2.114}$$

Define two new terms:

$$\tau_p = -\frac{\beta(f_c)}{2\pi f_c} \tag{2.115}$$

and

$$\tau_g = -\frac{1}{2\pi}\frac{\partial \beta(f)}{\partial f}\bigg|_{f=f_c} \tag{2.116}$$

Then, we may rewrite (2.114) in the equivalent form

$$\beta(f) \approx -2\pi f_c \tau_p - 2\pi(f - f_c)\tau_g \tag{2.117}$$

Correspondingly, the transfer function of the channel takes the approximate form

$$H(f) \approx K\exp[-j2\pi f_c \tau_p - j2\pi(f - f_c)\tau_g] \tag{2.118}$$

Following the band-pass-to-low-pass transformation described in Section 2.12, in particular using (2.80), we may replace the band-pass channel described by $H(f)$ by an equivalent low-pass filter whose transfer function is approximately given by

$$\tilde{H}(f) \approx 2K\exp(-j2\pi f_c \tau_p - j2\pi f \tau_g), \qquad f > f_c \tag{2.119}$$

Correspondingly, using (2.67) we may replace the modulated signal $s(t)$ of (2.113) by its low-pass complex envelope, which, for the DSB-SC example at hand, is simply defined by

$$\tilde{s}(t) = m(t) \tag{2.120}$$

Transforming $\tilde{s}(t)$ into the frequency domain, we may write

$$\tilde{S}(f) = M(f) \tag{2.121}$$

Therefore, in light of (2.96), the Fourier transform of the complex envelope of the signal received at the channel output is given by

$$\begin{aligned}\tilde{X}(f) &= \frac{1}{2}\tilde{H}(f)\tilde{S}(f) \\ &\approx K\exp(-j2\pi f_c \tau_p)\exp(-j2\pi f_c \tau_g)M(f)\end{aligned} \tag{2.122}$$

We note that the multiplying factor $K\exp(-j2\pi f_c \tau_p)$ is a constant for fixed values of f_c and τ_p. We also note from the time-shifting property of the Fourier transform that the term $\exp(-j2\pi f_c \tau_g)M(f)$ represents the Fourier transform of the delayed signal $m(t - \tau_g)$. Accordingly, the complex envelope of the channel output is

$$\tilde{x}(t) = K\exp(-j2\pi f_c \tau_p)m(t - \tau_g) \tag{2.123}$$

Finally, using (2.66) we find that the actual channel output is itself given by

$$\begin{aligned} x(t) &= \mathrm{Re}[\tilde{x}(t)\exp(\mathrm{j}2\pi f_c t)] \\ &= Km(t-\tau_g)\cos[2\pi f_c(t-\tau_p)] \end{aligned} \tag{2.124}$$

Equation (2.124) reveals that, as a result of transmitting the modulated signal $s(t)$ through the dispersive channel, two different delay effects occur at the channel output:

1. The sinusoidal carrier wave $\cos(2\pi f_c t)$ is delayed by τ_p seconds; hence, τ_p represents the *phase delay*; sometimes τ_p is referred to as the *carrier delay.*
2. The envelope $m(t)$ is delayed by τ_g seconds; hence, τ_g represents the *envelope* or *group delay.*

Note that τ_g is related to the slope of the phase $\beta(f)$, measured at $f = f_c$. Note also that when the phase response $\beta(f)$ varies linearly with frequency f and $\beta(f_c)$ is zero, the phase delay and group delay assume a common value. It is only then that we can think of these two delays being equal.

2.16 Numerical Computation of the Fourier Transform

The material presented in this chapter clearly testifies to the importance of the Fourier transform as a theoretical tool for the representation of deterministic signals and linear time-invariant systems, be they of the low-pass or band-pass kind. The importance of the Fourier transform is further enhanced by the fact that there exists a class of algorithms called FFT algorithms[6] for numerical computation of the Fourier transform in an efficient manner.

The FFT algorithm is derived from the discrete Fourier transform (DFT) in which, as the name implies, both time and frequency are represented in discrete form. The DFT provides an *approximation* to the Fourier transform. In order to properly represent the information content of the original signal, we have to take special care in performing the sampling operations involved in defining the DFT. A detailed treatment of the sampling process is presented in Chapter 6. For the present, it suffices to say that, given a band-limited signal, the sampling rate should be greater than twice the highest frequency component of the input signal. Moreover, if the samples are uniformly spaced by T_s seconds, the spectrum of the signal becomes periodic, repeating every $f_s = (1/T_s)$ hz in accordance with (2.43). Let N denote the number of frequency samples contained in the interval f_s. Hence, the *frequency resolution* involved in numerical computation of the Fourier transform is defined by

$$\Delta f = \frac{f_s}{N} = \frac{1}{NT_s} = \frac{1}{T} \tag{2.125}$$

where T is the total duration of the signal.

Consider then a *finite data sequence* $\{g_0, g_1, \ldots, g_{N-1}\}$. For brevity, we refer to this sequence as g_n, in which the subscript is the *time index* $n = 0, 1, \ldots, N-1$. Such a sequence may represent the result of sampling an analog signal $g(t)$ at times $t = 0, T_s, \ldots, (N-1)T_s$, where T_s is the sampling interval. The ordering of the data sequence defines the sample

time in that $g_0, g_1, ..., g_{N-1}$ denote samples of $g(t)$ taken at times $0, T_s, ..., (N-1)T_s$, respectively. Thus we have

$$g_n = g(nT_s) \tag{2.126}$$

We formally define the DFT of g_n as

$$G_k = \sum_{n=0}^{N-1} g_n \exp\left(-\frac{j2\pi}{N}kn\right) \qquad k = 0, 1, \ldots, N-1 \tag{2.127}$$

The sequence $\{G_0, G_1, ..., G_{N-1}\}$ is called the *transform sequence*. For brevity, we refer to this second sequence simply as G_k, in which the subscript is the *frequency index* $k = 0, 1, ..., N-1$.

Correspondingly, we define the *inverse discrete Fourier transform* (IDFT) of G_k as

$$g_n = \frac{1}{N}\sum_{k=0}^{N-1} G_k \exp\left(\frac{j2\pi}{N}kn\right) \qquad n = 0, 1, \ldots, N-1 \tag{2.128}$$

The DFT and the IDFT form a discrete transform pair. Specifically, given a data sequence g_n, we may use the DFT to compute the transform sequence G_k; and given the transform sequence G_k, we may use the IDFT to recover the original data sequence g_n. A distinctive feature of the DFT is that, for the finite summations defined in (2.127) and (2.128), there is no question of convergence.

When discussing the DFT (and algorithms for its computation), the words "sample" and "point" are used interchangeably to refer to a sequence value. Also, it is common practice to refer to a sequence of length N as an *N-point sequence* and to refer to the DFT of a data sequence of length N as an *N-point DFT*.

Interpretation of the DFT and the IDFT

We may visualize the DFT process described in (2.127) as a collection of N *complex heterodyning* and *averaging* operations, as shown in Figure 2.32a. We say that the heterodyning is complex in that samples of the data sequence are multiplied by *complex exponential sequences*. There is a total of N complex exponential sequences to be considered, corresponding to the frequency index $k = 0, 1, ..., N-1$. Their periods have been selected in such a way that each complex exponential sequence has precisely an integer number of cycles in the total interval 0 to $N-1$. The zero-frequency response, corresponding to $k = 0$, is the only exception.

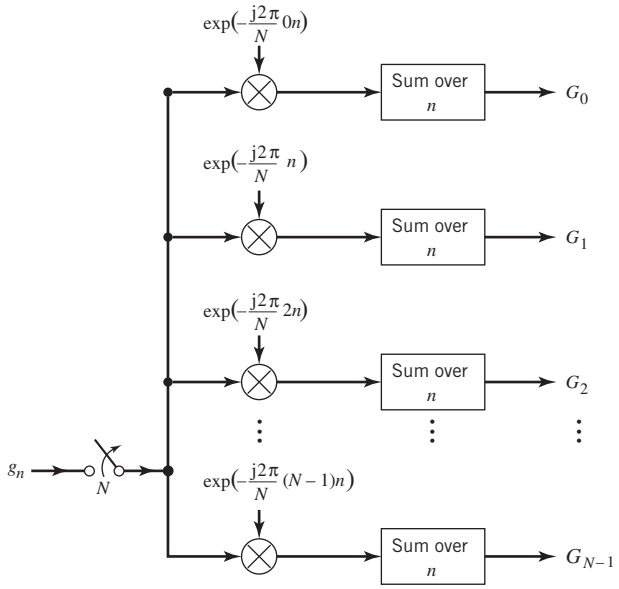
For the interpretation of the IDFT process, described in (2.128), we may use the scheme shown in Figure 2.32b. Here we have a collection of N *complex signal generators*, each of which produces the complex exponential sequence

$$\begin{aligned} \exp\left(\frac{j2\pi}{N}kn\right) &= \cos\left(\frac{2\pi}{N}kn\right) + j\sin\left(\frac{2\pi}{N}kn\right) \\ &= \left\{\cos\left(\frac{2\pi}{N}kn\right), \sin\left(\frac{2\pi}{N}kn\right)\right\}_{k=0}^{N-1} \end{aligned} \tag{2.129}$$

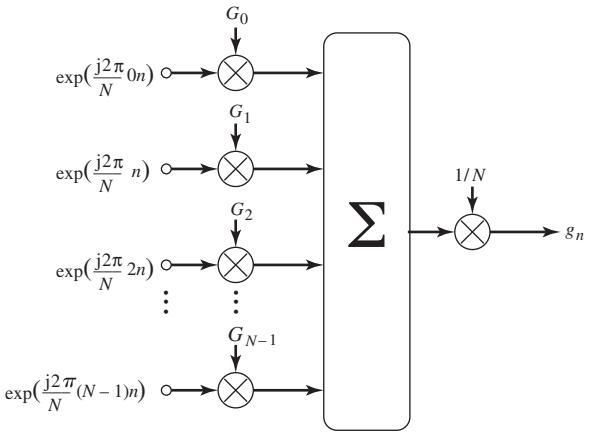
Thus, in reality, each complex signal generator consists of a pair of generators that output a cosinusoidal and a sinusoidal sequence of k cycles per observation interval. The output

of each complex signal generator is weighted by the complex Fourier coefficient G_k. At each time index n, an output is formed by summing the weighted complex generator outputs.

It is noteworthy that although the DFT and the IDFT are similar in their mathematical formulations, as described in (2.127) and (2.128), their interpretations as depicted in Figure 2.32a and b are so completely different.



(a)



(b)

Figure 2.32 Interpretations of (a) the DFT and (b) the IDFT.

Also, the addition of harmonically related periodic signals, involved in these two parts of the figure, suggests that their outputs G_k and g_n must be both periodic. Moreover, the processors shown in Figure 2.32 are linear, suggesting that the DFT and IDFT are both linear operations. This important property is also obvious from the defining equations (2.127) and (2.128).

FFT Algorithms

In the DFT both the input and the output consist of sequences of numbers defined at uniformly spaced points in time and frequency, respectively. This feature makes the DFT ideally suited for direct numerical evaluation on a computer. Moreover, the computation can be implemented most efficiently using a class of algorithms, collectively called *FFT algorithms*. An algorithm refers to a "recipe" that can be written in the form of a computer program.

FFT algorithms are efficient because they use a greatly reduced number of arithmetic operations as compared with the brute force (i.e., direct) computation of the DFT. Basically, an FFT algorithm attains its computational efficiency by following the engineering strategy of "divide and conquer," whereby the original DFT computation is decomposed successively into smaller DFT computations. In this section, we describe one version of a popular FFT algorithm, the development of which is based on such a strategy.

To proceed with the development, we first rewrite (2.127), defining the DFT of g_n, in the convenient mathematical form

$$G_k = \sum_{n=0}^{N-1} g_n W^{kn}, \qquad k = 0, 1, \ldots, N-1 \tag{2.130}$$

where we have introduced the complex parameter

$$W = \exp\left(-\frac{j2\pi}{N}\right) \tag{2.131}$$

From this definition, we readily see that

$$\begin{aligned} W^N &= 1 \\ W^{N/2} &= -1 \\ W^{(l+lN)(n+mN)} &= W^{kn}, \qquad (m, l) = 0, \pm 1, \pm 2, \ldots \end{aligned}$$

That is, W^{kn} is periodic with period N. The periodicity of W^{kn} is a key feature in the development of FFT algorithms.

Let N, the number of points in the data sequence, be an integer power of two, as shown by

$$N = 2^L$$

where L is an integer; the rationale for this choice is explained later. Since N is an even integer, $N/2$ is an integer, and so we may divide the data sequence into the first half and last half of the points.

Thus, we may rewrite (2.130) as

$$\begin{aligned} G_k &= \sum_{n=0}^{(N/2)-1} g_n W^{kn} + \sum_{n=N/2}^{N-1} g_n W^{kn} \\ &= \sum_{n=0}^{(N/2)-1} g_n W^{kn} + \sum_{n=0}^{(N/2)-1} g_{n+N/2} W^{k(n+N/2)} \\ &= \sum_{n=0}^{(N/2)-1} (g_n + g_{n+N/2} W^{kN/2}) W^{kn} \qquad k = 0, 1, \ldots, N-1 \end{aligned} \tag{2.132}$$

Since $W^{N/2} = -1$, we have

$$W^{kN/2} = (-1)^k$$

Accordingly, the factor $W^{kN/2}$ in (2.132) takes on only one of two possible values, namely +1 or –1, depending on whether the frequency index k is even or odd, respectively. These two cases are considered in what follows.

First, let k be even, so that $W^{kN/2} = 1$. Also let

$$k = 2l, \qquad l = 0, 1, \ldots, \frac{N}{2} - 1$$

and define

$$x_n = g_n + g_{n+N/2} \tag{2.133}$$

Then, we may put (2.132) into the new form

$$\begin{aligned} G_{2l} &= \sum_{n=0}^{(N/2)-1} x_n W^{2ln} \\ &= \sum_{n=0}^{(N/2)-1} x_n (W^2)^{ln} \qquad l = 0, 1, \ldots, \frac{N}{2} - 1 \end{aligned} \tag{2.134}$$

From the definition of W given in (2.131), we readily see that

$$\begin{aligned} W^2 &= \exp\left(-\frac{j4\pi}{N}\right) \\ &= \exp\left(-\frac{j2\pi}{N/2}\right) \end{aligned}$$

Hence, we recognize the sum on the right-hand side of (2.134) as the $(N/2)$-point DFT of the sequence x_n.

Next, let k be odd so that $W^{kN/2} = -1$. Also, let

$$k = 2l + 1, \qquad l = 0, 1, \ldots, \frac{N}{2} - 1$$

and define

$$y_n = g_n - g_{n+N/2} \tag{2.135}$$

Then, we may put (2.132) into the corresponding form

$$\begin{aligned} G^{2l+1} &= \sum_{n=0}^{(N/2)-1} y_n W^{(2l+1)n} \\ &= \sum_{n=0}^{(N/2)-1} [y_n W^n](W^2)^{ln} \qquad l = 0, 1, \ldots, \frac{N}{2} - 1 \end{aligned} \tag{2.136}$$

We recognize the sum on the right-hand side of (2.136) as the $(N/2)$-point DFT of the sequence $y_n W^n$. The parameter W^n associated with y_n is called the *twiddle factor*.

Equations (2.134) and (2.136) show that the even- and odd-valued samples of the transform sequence G_k can be obtained from the $(N/2)$-point DFTs of the sequences x_n and $y_n W^n$, respectively. The sequences x_n and y_n are themselves related to the original data sequence g_n by (2.133) and (2.135), respectively. Thus, the problem of computing an N-point DFT is reduced to that of computing two $(N/2)$-point DFTs. The procedure just described is repeated a second time, whereby an $(N/2)$-point DFT is decomposed into two $(N/4)$-point DFTs. The decomposition procedure is continued in this fashion until (after $L = \log_2 N$ stages) we reach the trivial case of N single-point DFTs.

Figure 2.33 illustrates the computations involved in applying the formulas of (2.134) and (2.136) to an eight-point data sequence; that is, $N = 8$. In constructing left-hand portions of the figure, we have used signal-flow graph notation. A *signal-flow graph* consists of an interconnection of *nodes* and *branches*. The *direction* of signal transmission along a branch is indicated by an arrow. A branch multiplies the variable at a node (to which it is connected) by the branch *transmittance*. A node sums the outputs of all incoming branches. The convention used for branch transmittances in Figure 2.33 is as follows. When no coefficient is indicated on a branch, the transmittance of that branch is assumed to be unity. For other branches, the transmittance of a branch is indicated by –1 or an integer power of W, placed alongside the arrow on the branch.

Thus, in Figure 2.33a the computation of an eight-point DFT is reduced to that of two four-point DFTs. The procedure for the eight-point DFT may be mimicked to simplify the computation of the four-point DFT. This is illustrated in Figure 2.33b, where the computation of a four-point DFT is reduced to that of two two-point DFTs. Finally, the computation of a two-point DFT is shown in Figure 2.33c.

Combining the ideas described in Figure 2.33, we obtain the complete signal-flow graph of Figure 2.34 for the computation of the eight-point DFT. A repetitive structure, called the *butterfly* with two inputs and two outputs, can be discerned in the FFT algorithm of Figure 2.34. Examples of butterflies (for the three stages of the algorithm) are shown by the bold-faced lines in Figure 2.34.

For the general case of $N = 2^L$, the algorithm requires $L = \log_2 N$ stages of computation. Each stage requires $(N/2)$ butterflies. Each butterfly involves one complex multiplication and two complex additions (to be precise, one addition and one subtraction). Accordingly, the FFT structure described here requires $(N/2)\log_2 N$ complex multiplications and $N\log_2 N$

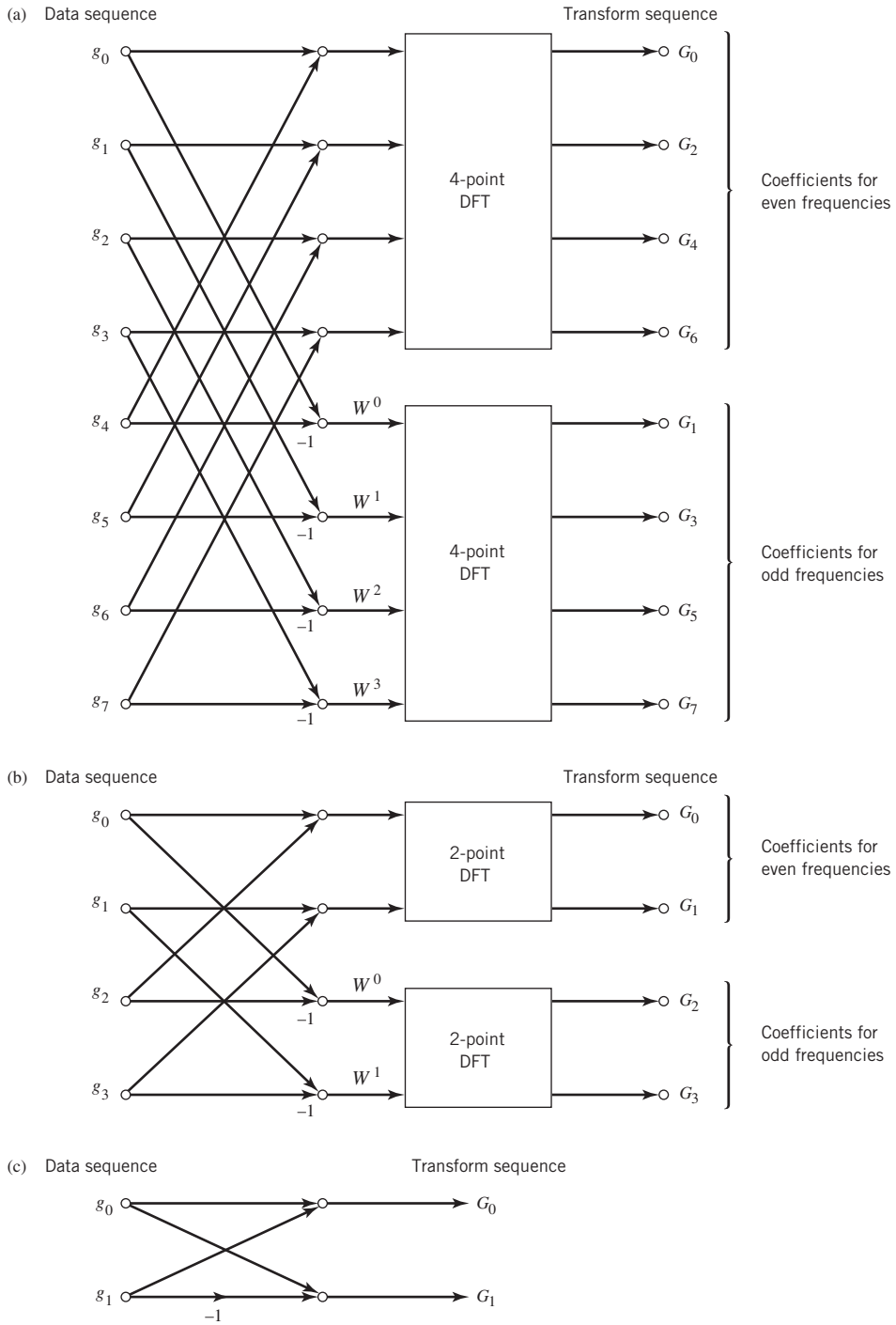


Figure 2.33 (a) Reduction of eight-point DFT into two four-point DFTs. (b) Reduction of four-point DFT into two two-point DFTs. (c) Trivial case of two-point DFT.

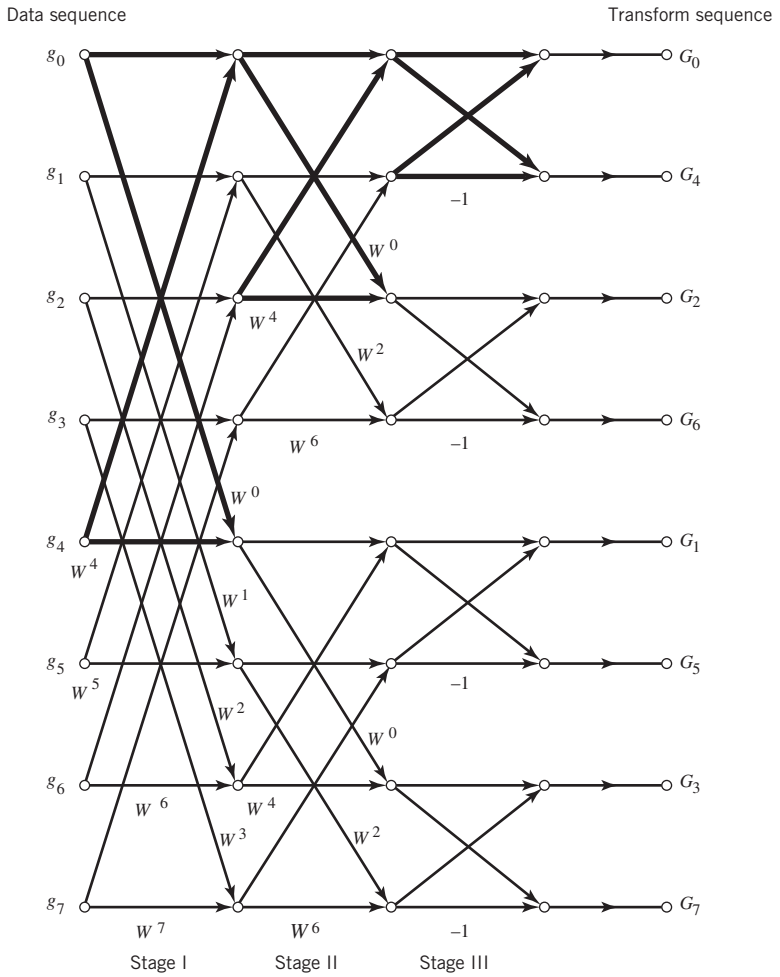


Figure 2.34
Decimation-in-frequency FFT algorithm.

complex additions; actually, the number of multiplications quoted is pessimistic, because we may omit all twiddle factors $W^0 = 1$ and $W^{N/2} = -1$, $W^{N/4} = j$, $W^{3N/4} = -j$. This computational complexity is significantly smaller than that of the N^2 complex multiplications and $N(N-1)$ complex additions required for *direct* computation of the DFT. The computational savings made possible by the FFT algorithm become more substantial as we increase the data length N. For example, for $N = 8192 = 2^{11}$, the direct approach requires approximately 630 times as many arithmetic operations as the FFT algorithm, hence the popular use of the FFT algorithm in computing the DFT.

We may establish two other important features of the FFT algorithm by carefully examining the signal-flow graph shown in Figure 2.34:

1. At each stage of the computation, the new set of N complex numbers resulting from the computation can be stored in the same memory locations used to store the previous set. This kind of computation is referred to as *in-place computation*.

2. The samples of the transform sequence G_k are stored in a bit-reversed order. To illustrate the meaning of this terminology, consider Table 2.6 constructed for the case of $N = 8$. At the left of the table, we show the eight possible values of the frequency index k (in their natural order) and their 3-bit binary representations. At the right of the table, we show the corresponding bit-reversed binary representations and indices. We observe that the bit-reversed indices in the rightmost column of Table 2.6 appear in the same order as the indices at the output of the FFT algorithm in Figure 2.34.

Table 2.6 Illustrating bit reversal

Frequency index, k	Binary representation	Bit-reversed binary representation	Bit-reversed index
0	000	000	0
1	001	100	4
2	010	010	2
3	011	110	6
4	100	001	1
5	101	101	5
6	110	011	3
7	111	111	7

The FFT algorithm depicted in Figure 2.34 is referred to as a *decimation-in-frequency algorithm*, because the transform (frequency) sequence G_k is divided successively into smaller subsequences. In another popular FFT algorithm, called a *decimation-in-time algorithm*, the data (time) sequence g_n is divided successively into smaller subsequences. Both algorithms have the same computational complexity. They differ from each other in two respects. First, for decimation-in-frequency, the input is in natural order, whereas the output is in bit-reversed order; the reverse is true for decimation-in-time. Second, the butterfly for decimation-in-time is slightly different from that for decimation-in-frequency. The reader is invited to derive the details of the decimation-in-time algorithm using the divide-and-conquer strategy that led to the development of the algorithm described in Figure 2.34.

In devising the FFT algorithm presented herein, we placed the factor $1/N$ in the formula for the forward DFT, as shown in (2.128). In some other FFT algorithms, location of the factor $1/N$ is reversed. In yet other formulations, the factor $1/\sqrt{N}$ is placed in the formulas for both the forward and inverse DFTs for the sake of symmetry.

Computation of the IDFT

The IDFT of the transform G_k is defined by (2.128). We may rewrite this equation in terms of the complex parameter W as

$$g_n = \frac{1}{N}\sum_{k=0}^{N-1} G_k W^{-kn}, \qquad n = 0, 1, \ldots, N-1 \tag{2.137}$$

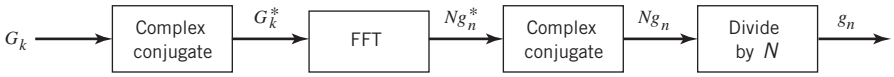


Figure 2.35 Use of the FFT algorithm for computing the IDFT.

Taking the complex conjugate of (2.137) and multiplying by N, we get

$$Ng_n^* = \sum_{k=0}^{N-1} G_k^* W^{-kn}, \qquad n = 0, 1, \ldots, N-1 \tag{2.138}$$

The right-hand side of (2.138) is recognized as the N-point DFT of the complex-conjugated sequence G_k^*. Accordingly, (2.138) suggests that we may compute the desired sequence g_n using the scheme shown in Figure 2.35, based on an N-point FFT algorithm. Thus, the same FFT algorithm can be used to handle the computation of both the IDFT and the DFT.

2.17 Summary and Discussion

In this chapter we have described the Fourier transform as a fundamental tool for relating the time-domain and frequency-domain descriptions of a deterministic signal. The signal of interest may be an energy signal or a power signal. The Fourier transform includes the exponential Fourier series as a special case, provided that we permit the use of the Dirac delta function.

An inverse relationship exists between the time-domain and frequency-domain descriptions of a signal. Whenever an operation is performed on the waveform of a signal in the time domain, a corresponding modification is applied to the spectrum of the signal in the frequency domain. An important consequence of this inverse relationship is the fact that the time–bandwidth product of an energy signal is a constant; the definitions of signal duration and bandwidth merely affect the value of the constant.

An important signal-processing operation frequently encountered in communication systems is that of linear filtering. This operation involves the convolution of the input signal with the impulse response of the filter or, equivalently, the multiplication of the Fourier transform of the input signal by the transfer function (i.e., Fourier transform of the impulse response) of the filter. Low-pass and band-pass filters represent two commonly used types of filters. Band-pass filtering is usually more complicated than low-pass filtering. However, through the combined use of a complex envelope for the representation of an input band-pass signal and the complex impulse response for the representation of a band-pass filter, we may formulate a complex low-pass equivalent for the band-pass filtering problem and thereby replace a difficult problem with a much simpler one. It is also important to note that there is no loss of information in establishing this equivalence. A rigorous treatment of the concepts of complex envelope and complex impulse response as presented in this chapter is rooted in Hilbert transformation.

The material on Fourier analysis, as presented in this chapter, deals with signals whose waveforms can be nonperiodic or periodic, and whose spectra can be continuous or discrete functions of frequency. In this sense, the material has general appeal.

Building on the canonical representation of a band-pass signal involving the in-phase and quadrature components of the signal, we showed that this representation provides an elegant way of describing the three basic forms of linear modulation, namely DSB-SC, VSB, and SSB.

With the Fourier transform playing such a pervasive role in the study of signals and linear systems, we finally described the FFT algorithm as an efficient tool for numerical computation of the DFT that represents the uniformly sampled versions of the forward and inverse forms of the ordinary Fourier transform.

Problems

The Fourier Transform

2.1 Prove the dilation property of the Fourier transform, listed as Property 2 in Table 2.1.

2.2 a. Prove the duality property of the Fourier transform, listed as Property 3 in Table 2.1.

b. Prove the time-shifting property, listed as Property 4; and then use the duality property to prove the frequency-shifting property, listed as Property 5 in the table.

c. Using the frequency-shifting property, determine the Fourier transform of the radio frequency RF pulse

$$g(t) = A\,\mathrm{rect}\left(\frac{t}{T}\right)\cos(2\pi f_c t)$$

assuming that f_c is larger than $(1/T)$.

2.3 a. Prove the multiplication-in-the-time-domain property of the Fourier transform, listed as Property 11 in Table 2.1.

b. Prove the convolution in the time-domain property, listed as Property 12.

c. Using the result obtained in part b, prove the correlation theorem, listed as Property 13.

2.4 Prove Rayleigh's energy theorem listed as Property 14 in Table 2.1.

2.5 The following expression may be viewed as an approximate representation of a pulse with finite rise time:

$$g(t) = \frac{1}{\tau}\int_{t-T}^{t+T} \exp\left(-\frac{\pi u^2}{\tau^2}\right) \mathrm{d}u$$

where it is assumed that $T >> \tau$. Determine the Fourier transform of $g(t)$. What happens to this transform when we allow τ to become zero? *Hint:* Express $g(t)$ as the superposition of two signals, one corresponding to integration from $t - T$ to 0, and the other from 0 to $t + T$.

2.6 The Fourier transform of a signal $g(t)$ is denoted by $G(f)$. Prove the following properties of the Fourier transform:

a. If a real signal $g(t)$ is an even function of time t, the Fourier transform $G(f)$ is purely real. If a real signal $g(t)$ is an odd function of time t, the Fourier transform $G(f)$ is purely imaginary.

b.

$$t^n g(t) \rightleftharpoons \left(\frac{\mathrm{j}}{2\pi}\right)^n G^{(n)}(f)$$

where $G^{(n)}(f)$ is the nth derivative of $G(f)$ with respect to f.

c.

$$\int_{-\infty}^{\infty} t^n g(t)\,\mathrm{d}t = \left(\frac{\mathrm{j}}{2\pi}\right)^n G^{(n)}(0)$$

d. Assuming that both $g_1(t)$ and $g_2(t)$ are complex signals, show that:

$$g_1(t)g_2^*(t) \rightleftharpoons \int_{-\infty}^{\infty} G_1(\lambda)G_2^*(\lambda - f)\,d\lambda$$

and

$$\int_{-\infty}^{\infty} g_1(t)g_2^*(t)dt = \int_{-\infty}^{\infty} G_1(f)G_2^*(f)\,df$$

2.7 a. The *root mean-square (rms) bandwidth* of a low-pass signal $g(t)$ of finite energy is defined by

$$W_{\text{rms}} = \left[\frac{\int_{-\infty}^{\infty} f^2|G(f)|^2 df}{\int_{-\infty}^{\infty} |G(f)|^2 df}\right]^{1/2}$$

where $|G(f)|^2$ is the energy spectral density of the signal. Correspondingly, the *root mean-square (rms) duration* of the signal is defined by

$$T_{\text{rms}} = \left[\frac{\int_{-\infty}^{\infty} t^2|g(t)|^2 dt}{\int_{-\infty}^{\infty} |g(t)|^2 dt}\right]^{1/2}$$

Using these definitions, show that

$$T_{\text{rms}}W_{\text{rms}} \geq \frac{1}{4\pi}$$

Assume that $|g(t)| \to 0$ faster than $1/\sqrt{|t|}$ as $|t| \to \infty$.

b. Consider a Gaussian pulse defined by

$$g(t) = \exp(-\pi t^2)$$

Show that for this signal the equality

$$T_{\text{rms}}W_{\text{rms}} = \frac{1}{4\pi}$$

is satisfied.

Hint: Use Schwarz's inequality

$$\left(\int_{-\infty}^{\infty} [g_1^*(t)g_2(t) + g_1(t)g_2{}^*(t)]dt\right)^2 \leq 4\int_{-\infty}^{\infty} |g_1(t)|^2 dt \int_{-\infty}^{\infty} |g_2(t)|^2 dt$$

in which we set

$$g_1(t) = tg(t)$$

and

$$g_2(t) = \frac{dg(t)}{dt}$$

2.8 The *Dirac comb*, formulated in the time domain, is defined by

$$\delta_{T_0}(t) = \sum_{m=-\infty}^{\infty} \delta(t - mT_0)$$

where T_0 is the period.

a. Show that the Dirac comb is its own Fourier transform. That is, the Fourier transform of $\delta_{T_0}(t)$ is also an infinitely long periodic train of delta functions, weighted by the factor $f_0 = (1/T_0)$ and *regularly spaced by* f_0 along the frequency axis.

b. Hence, prove the pair of dual relations:

$$\sum_{m=-\infty}^{\infty} \delta(t - mT_0) = f_0 \sum_{n=-\infty}^{\infty} \exp(j2\pi n f_0 t)$$

$$T_0 \sum_{m=-\infty}^{\infty} \exp(j2\pi m f T_0) = \sum_{n=-\infty}^{\infty} \delta(f - nf_0)$$

c. Finally, prove the validity of (2.38).

Signal Transmission through Linear Time-invariant Systems

2.9 The periodic signal

$$x(t) = \sum_{m=-\infty}^{\infty} x(nT_0)\,\delta(t - nT_0)$$

is applied to a linear system of impulse response $h(t)$. Show that the average power of the signal $y(t)$ produced at the system output is defined by

$$P_{\text{av},y} = \sum_{n=-\infty}^{\infty} \left|x(nT_0)\right|^2 \left|H(nf_0)\right|^2$$

where $H(f)$ is the frequency response of the system, and $f_0 = 1/T_0$.

2.10 According to the bounded input–bounded output stability criterion, the impulse response $h(t)$ of a linear-invariant system must be absolutely integrable; that is,

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

Prove that this condition is both necessary and sufficient for stability of the system.

Hilbert Transform and Pre-envelopes

2.11 Prove the three properties of the Hilbert transform itemized on pages 43 and 44.

2.12 Let $\hat{g}(t)$ denote the Hilbert transform of $g(t)$. Derive the set of Hilbert-transform pairs listed as items 5 to 8 in Table 2.3.

2.13 Evaluate the inverse Fourier transform $g(t)$ of the one-sided frequency function:

$$G(f) = \begin{cases} \exp(-f), & f > 0 \\ \frac{1}{2}, & f = 0 \\ 0, & f < 0 \end{cases}$$

Show that $g(t)$ is complex, and that its real and imaginary parts constitute a Hilbert-transform pair.

2.14 Let $\hat{g}(t)$ denote the Hilbert transform of a Fourier transformable signal $g(t)$. Show that $\frac{d}{dt}\hat{g}(t)$ is equal to the Hilbert transform of $\frac{d}{dt}g(t)$.

2.15 In this problem, we revisit Problem 2.14, except that this time we use integration rather than differentiation. Doing so, we find that, in general, the integral $\int_{-\infty}^{\infty} \hat{g}(t)\,dt$ is *not* equal to the Hilbert transform of the integral $\int_{-\infty}^{\infty} g(t)\,dt$.

a. Justify this statement.

b. Find the condition for which exact equality holds.

2.16 Determine the pre-envelope $g_+(t)$ corresponding to each of the following two signals:

a. $g(t) = \text{sinc}(t)$

b. $g(t) = [1 + k\cos(2\pi f_m t)]\cos(2\pi f_c t)$

Complex Envelope

2.17 Show that the complex envelope of the sum of two narrowband signals (with the same carrier frequency) is equal to the sum of their individual complex envelopes.

2.18 The definition of the complex envelope $\tilde{s}(t)$ of a band-pass signal given in (2.65) is based on the pre-envelope $s_+(t)$ for positive frequencies. How is the complex envelope defined in terms of the pre-envelope $s\_(t)$ for negative frequencies? Justify your answer.

2.19 Consider the signal

$$s(t) = c(t)m(t)$$

whose $m(t)$ is a low-pass signal whose Fourier transform $M(f)$ vanishes for $|f| > W$, and $c(t)$ is a high-pass signal whose Fourier transform $C(f)$ vanishes for $|f| < W$. Show that the Hilbert transform of $s(t)$ is $\hat{s}(t) = \hat{c}(t)m(t)$, where $\hat{c}(t)$ is the Hilbert transform of $c(t)$.

2.20 a. Consider two real-valued signals $s_1(t)$ and $s_2(t)$ whose pre-envelopes are denoted by $s_{1+}(t)$ and $s_{2+}(t)$, respectively. Show that

$$\int_{-\infty}^{\infty} \text{Re}[s_{1+}(t)]\text{Re}[s_{2+}(t)]dt = \frac{1}{2}\text{Re}\left[\int_{-\infty}^{\infty} s_{1+}(t)s_{2+}^{*}(t)\,dt\right]$$

b. Suppose that $s_2(t)$ is replaced by $s_2(-t)$. Show that this modification has the effect of removing the complex conjugation in the right-hand side of the formula given in part a.

c. Assuming that $s(t)$ is a narrowband signal with complex envelope $\tilde{s}(t)$ and carrier frequency f_c, use the result of part a to show that

$$\int_{-\infty}^{\infty} s^2(t)\,dt = \frac{1}{2}\int_{-\infty}^{\infty} |\tilde{s}(t)|^2\,dt$$

2.21 Let a narrow-band signal $s(t)$ be expressed in the form

$$s(t) = s_I(t)\cos(2\pi f_c t) - s_Q(t)\sin(2\pi f_c t)$$

Using $S_+(f)$ to denote the Fourier transform of the pre-envelope of $s_+(t)$, show that the Fourier transforms of the in-phase component $s_I(t)$ and quadrature component $s_Q(t)$ are given by

$$S_I(f) = \frac{1}{2}[S_+(f+f_c) + S_+^{*}(-f+f_c)]$$

$$S_Q(f) = \frac{1}{2j}[S_+(f+f_c) - S_+^{*}(-f+f_c)]$$

respectively, where the asterisk denotes complex conjugation.

2.22 The block diagram of Figure 2.20a illustrates a method for extracting the in-phase component $s_I(t)$ and quadrature component $s_Q(t)$ of a narrowband signal $s(t)$. Given that the spectrum of $s(t)$ is limited to the interval $f_c - W \le |f| f_c + W$, demonstrate the validity of this method. Hence, show that

$$S_I(f) = \begin{cases} S(f-f_c) + S(f+f_c), & -W \le f \le W \\ 0, & \text{elsewhere} \end{cases}$$

and

$$S_Q(f) = \begin{cases} j[S(f-f_c) - S(f+f_c)], & -W \le f \le W \\ 0, & \text{elsewhere} \end{cases}$$

where $S_I(f)$, $S_Q(f)$, and $S(f)$ are the Fourier transforms of $s_I(t)$, $s_Q(t)$, and $s(t)$, respectively.

Low-Pass Equivalent Models of Band-Pass Systems

2.23 Equations (2.82) and (2.83) define the in-phase component $\tilde{H}_I(f)$ and the quadrature component $\tilde{H}_Q(f)$ of the frequency response $\tilde{H}(f)$ of the complex low-pass equivalent model of a band-pass system of impulse response $h(t)$. Prove the validity of these two equations.

2.24 Explain what happens to the low-pass equivalent model of Figure 2.21b when the amplitude response of the corresponding bandpass filter has even symmetry and the phase response has odd symmetry with respect to the mid-band frequency f_c.

2.25 The rectangular RF pulse

$$x(t) = \begin{cases} A\cos(2\pi f_c t), & 0 \le t \le T \\ 0, & \text{elsewhere} \end{cases}$$

is applied to a linear filter with impulse response

$$h(t) = x(T-t)$$

Assume that the frequency f_c equals a large integer multiple of $1/T$. Determine the response of the filter and sketch it.

2.26 Figure P2.26 depicts the frequency response of an idealized band-pass filter in the receiver of a communication system, namely $H(f)$, which is characterized by a bandwidth of $2B$ centered on the carrier frequency f_c. The signal applied to the band-pass filter is described by the modulated sinc function:

$$x(t) = 4A_c B\,\text{sinc}(2Bt)\cos[2\pi(f_c \pm \Delta f)t]$$

where Δf is *frequency misalignment* introduced due to the receiver's imperfections, measured with respect to the carrier $A_c\cos(2\pi f_c t)$.

a. Find the complex low-pass equivalent models of the signal $x(t)$ and the frequency response $H(f)$.

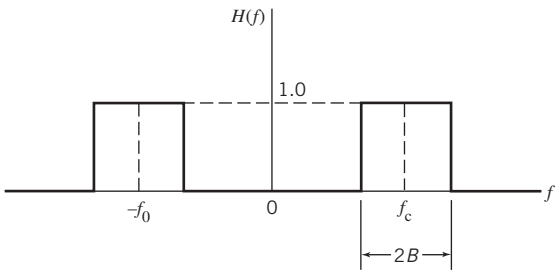


Figure P2.26

b. Then, go on to find the complex low-pass response of the filter output, denoted by $\tilde{y}(t)$, which includes distortion due to $\pm\Delta f$.

c. Building on the formula derived for $\tilde{y}(t)$ obtained in part b, explain how you would mitigate the misalignment distortion in the receiver.

Nonlinear Modulations

2.27 In analog communications, amplitude modulation is defined by

$$s_{\mathrm{AM}}(t) = A_c[1 + k_{\mathrm{a}}m(t)]\cos(2\pi f_c t)$$

where $A_c\cos(2\pi f_c t)$ is the carrier, $m(t)$ is the message signal, and k_{a} is a constant called *amplitude sensitivity* of the modulator. Assume that $|k_{\mathrm{a}}m(t)| < 1$ for all time t.

a. Justify the statement that, in a strict sense, $s_{\mathrm{AM}}(t)$ violates the principle of superposition.

b. Formulate the complex envelope $\tilde{s}_{\mathrm{AM}}(t)$ and its spectrum.

c. Compare the result obtained in part b with the complex envelope of DSB-SC. Hence, comment on the advantages and disadvantages of amplitude modulation.

2.28 Continuing on with analog communications, *frequency modulation* (FM) is defined by

$$s_{\mathrm{FM}}(t) = A_{\mathrm{c}}\left[\cos(2\pi f_{\mathrm{c}}t) + k_{\mathrm{f}}\int_o^t m(\tau)\,\mathrm{d}\tau\right]$$

where $A_{\mathrm{c}}\cos(2\pi f_{\mathrm{c}}t)$ is the carrier, $m(t)$ is the message signal, and k_{f} is a constant called the *frequency sensitivity* of the modulator.

a. Show that frequency modulation is nonlinear in that it violates the principle of superposition.

b. Formulate the complex envelope of the FM signal, namely $\tilde{s}_{\mathrm{FM}}(t)$.

c. Consider the message signal to be in the form of a square wave as shown in Figure P2.28. The modulation frequencies used for the positive and negative amplitudes of the square wave, namely f_1 and f_2, are defined as follows:

$$f_1 + f_2 = \frac{2}{T_{\mathrm{b}}}$$

$$f_1 - f_2 = \frac{1}{T_{\mathrm{b}}}$$

where T_{b} is the duration of each positive or negative amplitude in the square wave. Show that under these conditions the complex envelope $\tilde{s}_{\mathrm{FM}}(t)$ maintains *continuity* for all time t, including the switching times between positive and negative amplitudes.

d. Plot the real and imaginary parts of $\tilde{s}_{\mathrm{FM}}(t)$ for the following values:

$$T_b = \frac{1}{3}\ \mathrm{s}$$

$$f_1 = 4\frac{1}{2}\ \mathrm{Hz}$$

$$f_2 = 1\frac{1}{2}\ \mathrm{Hz}$$

Phase and Group Delays

2.29 The phase response of a band-pass communication channel is defined by.

$$\phi(f) = -\tan^{-1}\left(\frac{f^2 - f_{\mathrm{c}}^2}{ff_{\mathrm{c}}}\right)$$

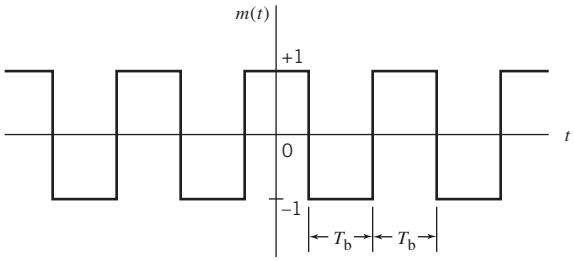


Figure P2.28

A sinusoidally modulated signal defined by

$$s(t) = A_c \cos(2\pi f_m t)\cos(2\pi f_c t)$$

is transmitted through the channel; f_c is the carrier frequency and f_m is the modulation frequency.

a. Determine the phase delay τ_p.

b. Determine the group delay τ_g.

c. Display the waveform produced at the channel output; hence, comment on the results obtained in parts a and b.

Notes

1. For a proof of convergence of the Fourier series, see Kammler (2000).

2. If a time function $g(t)$ is such that the value of the energy $\int_{-\infty}^{\infty} |g(t)|^2 dt$ is defined and finite, then the Fourier transform $G(f)$ of the function $g(t)$ exists and

$$\lim_{A \to \infty}\left[\int_{-\infty}^{\infty}\left|g(t) - \int_{-A}^{A} G(f)\exp(j2\pi ft)\, df\right|^2\right] = 0$$

This result is known as *Plancherel's theorem*. For a proof of this theorem, see Titchmarsh (1950).

3. The notation $\delta(t)$ for a delta function was first introduced into quantum mechanics by Dirac. This notation is now in general use in the signal processing literature. For detailed discussions of the delta function, see Bracewell (1986).

In a rigorous sense, the Dirac delta function is a distribution, not a function; for a rigorous treatment of the subject, see the book by Lighthill (1958).

4. The Paley–Wiener criterion is named in honor of the authors of the paper by Paley and Wiener (1934).

5. The integral in (2.54), defining the Hilbert transform of a signal, is an *improper* integral in that the integrand has a singularity at $\tau = t$. To avoid this singularity, the integration must be carried out in a symmetrical manner about the point $\tau = t$. For this purpose, we use the definition

$$P\int_{-\infty}^{\infty} \frac{g(\tau)}{t-\tau} d\tau = \lim_{g \to 0}\left[\int_{-\infty}^{t=\epsilon} \frac{g(\tau)}{t-\tau} d\tau + \int_{t+\epsilon}^{\infty} \frac{g(\tau)}{t-\tau} d\tau\right]$$

where the symbol P denotes Cauchy's principal value of the integral and ϵ is incrementally small. For notational simplicity, the symbol P has been omitted from (2.54) and (2.55).

6. The complex representation of an arbitrary signal defined in (2.58) was first described by Gabor (1946). Gabor used the term "analytic signal." The term "pre-envelope" was used in Arens (1957) and Dungundji (1958). For a review of the different envelopes, see the paper by Rice (1982).

7. The FFT is *ubiquitous* in that it is applicable to a great variety of unrelated fields. For a detailed mathematical treatment of this widely used tool and its applications, the reader is referred to Brigham (1988).

CHAPTER
6

Conversion of Analog Waveforms into Coded Pulses

6.1 Introduction

In *continuous-wave (CW) modulation*, which was studied briefly in Chapter 2, some parameter of a sinusoidal carrier wave is varied continuously in accordance with the message signal. This is in direct contrast to *pulse modulation*, which we study in this chapter. In pulse modulation, some parameter of a pulse train is varied in accordance with the message signal. On this basis, we may distinguish two families of pulse modulation:

1. *Analog pulse modulation*, in which a periodic pulse train is used as the carrier wave and some characteristic feature of each pulse (e.g., amplitude, duration, or position) is varied in a continuous manner in accordance with the corresponding *sample* value of the message signal. Thus, in analog pulse modulation, information is transmitted basically in analog form but the transmission takes place at discrete times.
2. *Digital pulse modulation*, in which the message signal is represented in a form that is discrete in both time and amplitude, thereby permitting transmission of the message in digital form as a sequence of *coded pulses*; this form of signal transmission has *no* CW counterpart.

The use of coded pulses for the transmission of analog information-bearing signals represents a basic ingredient in digital communications. In this chapter, we focus attention on digital pulse modulation, which, in basic terms, is described as the *conversion of analog waveforms into coded pulses*. As such, the conversion may be viewed as the transition from analog to digital communications.

Three different kinds of digital pulse modulation are studied in the chapter:

1. *Pulse-code modulation* (PCM), which has emerged as the most favored scheme for the digital transmission of analog information-bearing signals (e.g., voice and video signals). The important advantages of PCM are summarized thus:
   - *robustness* to channel noise and interference;
   - efficient *regeneration* of the coded signal along the transmission path;
   - efficient *exchange* of increased channel bandwidth for improved signal-to-quantization noise ratio, obeying an exponential law;
   - a *uniform format* for the transmission of different kinds of baseband signals, hence their integration with other forms of digital data in a common network;

- comparative *ease* with which message sources may be dropped or reinserted in a multiplex system;
- *secure* communication through the use of special modulation schemes or encryption.

These advantages, however, are attained at the cost of increased system complexity and increased transmission bandwidth. Simply stated:

There is no free lunch.

For every gain we make, there is a price to pay.

2. *Differential pulse-code modulation* (DPCM), which exploits the use of *lossy data compression* to remove the redundancy inherent in a message signal, such as voice or video, so as to reduce the bit rate of the transmitted data without serious degradation in overall system response. In effect, increased system complexity is traded off for reduced bit rate, therefore reducing the bandwidth requirement of PCM.
3. *Delta modulation* (DM), which addresses another practical limitation of PCM: the need for simplicity of implementation when it is a necessary requirement. DM satisfies this requirement by intentionally "oversampling" the message signal. In effect, increased transmission bandwidth is traded off for reduced system complexity. DM may therefore be viewed as the dual of DPCM.

Although, indeed, these three methods of analog-to-digital conversion are quite different, they do share two basic signal-processing operations, namely sampling and quantization:

- the process of sampling, followed by
- pulse-amplitude modulation (PAM) and finally
- amplitude quantization

are studied in what follows in this order.

6.2 Sampling Theory

The *sampling process* is usually described in the time domain. As such, it is an operation that is basic to digital signal processing and digital communications. Through use of the sampling process, an analog signal is converted into a corresponding sequence of samples that are usually spaced uniformly in time. Clearly, for such a procedure to have practical utility, it is necessary that we choose the sampling rate properly in relation to the bandwidth of the message signal, so that the sequence of samples uniquely defines the original analog signal. This is the essence of the sampling theorem, which is derived in what follows.

Frequency-Domain Description of Sampling

Consider an arbitrary signal $g(t)$ of finite energy, which is specified for all time t. A segment of the signal $g(t)$ is shown in Figure 6.1a. Suppose that we sample the signal $g(t)$ instantaneously and at a uniform rate, once every T_s seconds. Consequently, we obtain an infinite sequence of samples spaced T_s seconds apart and denoted by $\{g(nT_s)\}$, where n takes on all possible integer values, positive as well as negative. We refer to T_s as the *sampling period*, and to its reciprocal $f_s = 1/T_s$ as the *sampling rate*. For obvious reasons, this ideal form of sampling is called *instantaneous sampling*.

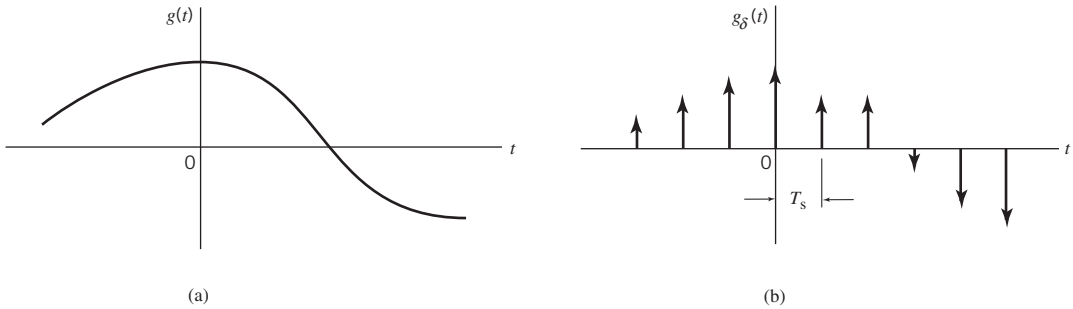


Figure 6.1 The sampling process. (a) Analog signal. (b) Instantaneously sampled version of the analog signal.

Let $g_\delta(t)$ denote the signal obtained by individually weighting the elements of a periodic sequence of delta functions spaced T_s seconds apart by the sequence of numbers $\{g(nT_s)\}$, as shown by (see Figure 6.1b):

$$g_\delta(t) = \sum_{n=-\infty}^{\infty} g(nT_s)\delta(t - nT_s) \tag{6.1}$$

We refer to $g_\delta(t)$ as the *ideal sampled signal*. The term $\delta(t - nT_s)$ represents a delta function positioned at time $t = nT_s$. From the definition of the delta function, we recall from Chapter 2 that such an idealized function has unit area. We may therefore view the multiplying factor $g(nT_s)$ in (6.1) as a "mass" assigned to the delta function $\delta(t - nT_s)$. A delta function weighted in this manner is closely approximated by a rectangular pulse of duration Δt and amplitude $g(nT_s)/\Delta t$; *the smaller we make* Δt *the better the approximation will be*.

Referring to the table of Fourier-transform pairs in Table 2.2, we have

$$g_\delta(t) \rightleftharpoons f_s \sum_{m=-\infty}^{\infty} G(f - mf_s) \tag{6.2}$$

where $G(f)$ is the Fourier transform of the original signal $g(t)$ and f_s is the sampling rate. Equation (6.2) states:

> The process of uniformly sampling a continuous-time signal of finite energy results in a periodic spectrum with a frequency equal to the sampling rate.

Another useful expression for the Fourier transform of the ideal sampled signal $g_\delta(t)$ may be obtained by taking the Fourier transform of both sides of (6.1) and noting that the Fourier transform of the delta function $\delta(t - nT_s)$ is equal to $\exp(-j2\pi nfT_s)$. Letting $G_\delta(f)$ denote the Fourier transform of $g_\delta(t)$, we may write

$$G_\delta(f) = \sum_{n=-\infty}^{\infty} g(nT_s)\exp(-j2\pi nfT_s) \tag{6.3}$$

Equation (6.3) describes the *discrete-time Fourier transform*. It may be viewed as a complex Fourier series representation of the periodic frequency function $G_\delta(f)$, with the sequence of samples $\{g(nT_s)\}$ defining the coefficients of the expansion.

The discussion presented thus far applies to any continuous-time signal $g(t)$ of finite energy and infinite duration. Suppose, however, that the signal $g(t)$ is *strictly band limited*, with no frequency components higher than W hertz. That is, the Fourier transform $G(f)$ of the signal $g(t)$ has the property that $G(f)$ is zero for $|f| \geq W$, as illustrated in Figure 6.2a; the shape of the spectrum shown in this figure is merely intended for the purpose of illustration. Suppose also that we choose the sampling period $T_s = 1/2W$. Then the corresponding spectrum $G_\delta(f)$ of the sampled signal $g_\delta(t)$ is as shown in Figure 6.2b. Putting $T_s = 1/2W$ in (6.3) yields

$$G_\delta(f) = \sum_{n=-\infty}^{\infty} g\left(\frac{n}{2W}\right) \exp\left(-\frac{j\pi nf}{W}\right) \tag{6.4}$$

Isolating the term on the right-hand side of (6.2), corresponding to $m = 0$, we readily see that the Fourier transform of $g_\delta(t)$ may also be expressed as

$$G_\delta(f) = f_s G(f) + f_s \sum_{\substack{m=-\infty \\ m \neq 0}}^{\infty} G(f - mf_s) \tag{6.5}$$

Suppose, now, we impose the following two conditions:

1. $G(f) = 0$ for $|f| \geq W$.
2. $f_s = 2W$.

We may then reduce (6.5) to

$$G(f) = \frac{1}{2W} G_\delta(f), \qquad -W < f < W \tag{6.6}$$

Substituting (6.4) into (6.6), we may also write

$$G(f) = \frac{1}{2W} \sum_{n=-\infty}^{\infty} g\left(\frac{n}{2W}\right) \exp\left(-\frac{j\pi nf}{W}\right), \qquad -W < f < W \tag{6.7}$$

Equation (6.7) is the desired formula for the frequency-domain description of sampling. This formula reveals that if the sample values $g(n/2W)$ of the signal $g(t)$ are specified for all n, then the Fourier transform $G(f)$ of that signal is uniquely determined. Because $g(t)$ is related to $G(f)$ by the inverse Fourier transform, it follows, therefore, that $g(t)$ is itself uniquely determined by the sample values $g(n/2W)$ for $-\infty < n < \infty$. In other words, the sequence $\{g(n/2W)\}$ has all the information contained in the original signal $g(t)$.

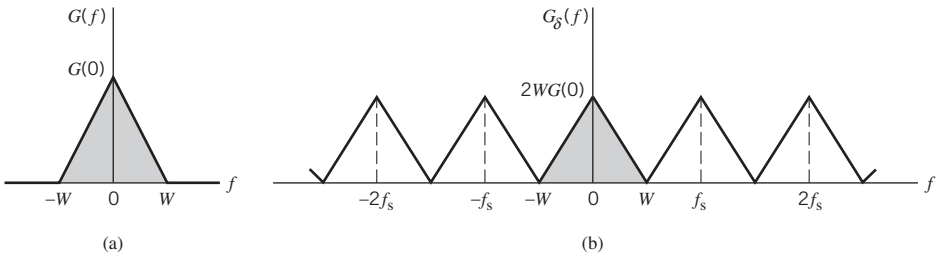


Figure 6.2 (a) Spectrum of a strictly band-limited signal $g(t)$. (b) Spectrum of the sampled version of $g(t)$ for a sampling period $T_s = 1/2W$.

Consider next the problem of reconstructing the signal $g(t)$ from the sequence of sample values $\{g(n/2W)\}$. Substituting (6.7) in the formula for the inverse Fourier transform

$$g(t) = \int_{-\infty}^{\infty} G(f)\exp(j2\pi ft)\,df$$

and interchanging the order of summation and integration, which is permissible because both operations are linear, we may go on to write

$$g(t) = \sum_{n=-\infty}^{\infty} g\left(\frac{n}{2W}\right)\frac{1}{2W}\int_{-W}^{W}\exp\left[j2\pi f\left(t-\frac{n}{2W}\right)\right]df \tag{6.8}$$

The definite integral in (6.8), including the multiplying factor $1/2W$, is readily evaluated in terms of the sinc function, as shown by

$$\begin{aligned}\frac{1}{2W}\int_{-W}^{W}\exp\left[j2\pi f\left(t-\frac{n}{2W}\right)\right]df &= \frac{\sin(2\pi Wt-n\pi)}{2\pi Wt-n\pi}\\ &= \operatorname{sinc}(2Wt-n)\end{aligned}$$

Accordingly, (6.8) reduces to the infinite-series expansion

$$g(t) = \sum_{n=-\infty}^{\infty} g\left(\frac{n}{2W}\right)\operatorname{sinc}(2Wt-n), \qquad -\infty < t < \infty \tag{6.9}$$

Equation (6.9) is the desired *reconstruction formula.* This formula provides the basis for reconstructing the original signal $g(t)$ from the sequence of sample values $\{g(n/2W)\}$, with the sinc function $\operatorname{sinc}(2Wt)$ playing the role of a *basis function* of the expansion. Each sample, $g(n/2W)$, is multiplied by a delayed version of the *basis function*, $\operatorname{sinc}(2Wt-n)$, and all the resulting individual waveforms in the expansion are added to reconstruct the original signal $g(t)$.

The Sampling Theorem

Equipped with the frequency-domain description of sampling given in (6.7) and the reconstruction formula of (6.9), we may now state the *sampling theorem* for strictly band-limited signals of finite energy in two equivalent parts:

1. A band-limited signal of finite energy that has no frequency components higher than W hertz is completely described by specifying the values of the signal instants of time separated by $1/2W$ seconds.
2. A band-limited signal of finite energy that has no frequency components higher than W hertz is completely recovered from a knowledge of its samples taken at the rate of $2W$ samples per second.

Part 1 of the theorem, following from (6.7), is performed in the transmitter. Part 2 of the theorem, following from (6.9), is performed in the receiver. For a signal bandwidth of W hertz, the sampling rate of $2W$ samples per second, for a signal bandwidth of W hertz, is called the *Nyquist rate*; its reciprocal $1/2W$ (measured in seconds) is called the *Nyquist interval*; see the classic paper (Nyquist, 1928b).

Aliasing Phenomenon

Derivation of the sampling theorem just described is based on the assumption that the signal $g(t)$ is strictly band limited. In practice, however, a message signal is *not* strictly band limited, with the result that some degree of undersampling is encountered, as a consequence of which *aliasing* is produced by the sampling process. Aliasing refers to the phenomenon of a high-frequency component in the spectrum of the signal seemingly taking on the identity of a lower frequency in the spectrum of its sampled version, as illustrated in Figure 6.3. The aliased spectrum, shown by the solid curve in Figure 6.3b, pertains to the undersampled version of the message signal represented by the spectrum of Figure 6.3a.

To combat the effects of aliasing in practice, we may use two corrective measures:

1. Prior to sampling, a low-pass *anti-aliasing filter* is used to attenuate those high-frequency components of the signal that are not essential to the information being conveyed by the message signal $g(t)$.
2. The filtered signal is sampled at a rate slightly higher than the Nyquist rate.

The use of a sampling rate higher than the Nyquist rate also has the beneficial effect of easing the design of the *reconstruction filter* used to recover the original signal from its sampled version. Consider the example of a message signal that has been anti-alias (low-pass) filtered, resulting in the spectrum shown in Figure 6.4a. The corresponding spectrum of the instantaneously sampled version of the signal is shown in Figure 6.4b, assuming a sampling rate higher than the Nyquist rate. According to Figure 6.4b, we readily see that design of the reconstruction filter may be specified as follows:

- The reconstruction filter is low-pass with a passband extending from $-W$ to W, which is itself determined by the anti-aliasing filter.
- The reconstruction filter has a transition band extending (for positive frequencies) from W to $(f_s - W)$, where f_s is the sampling rate.

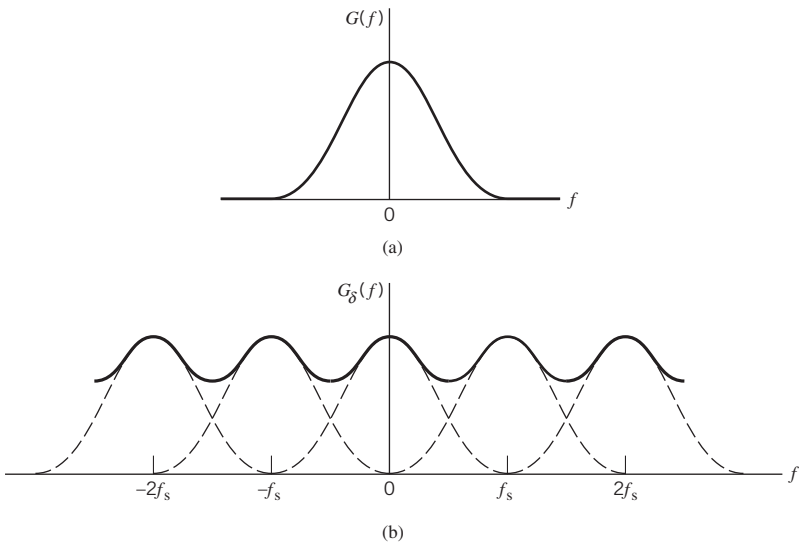


Figure 6.3 (a) Spectrum of a signal. (b) Spectrum of an under-sampled version of the signal exhibiting the aliasing phenomenon.

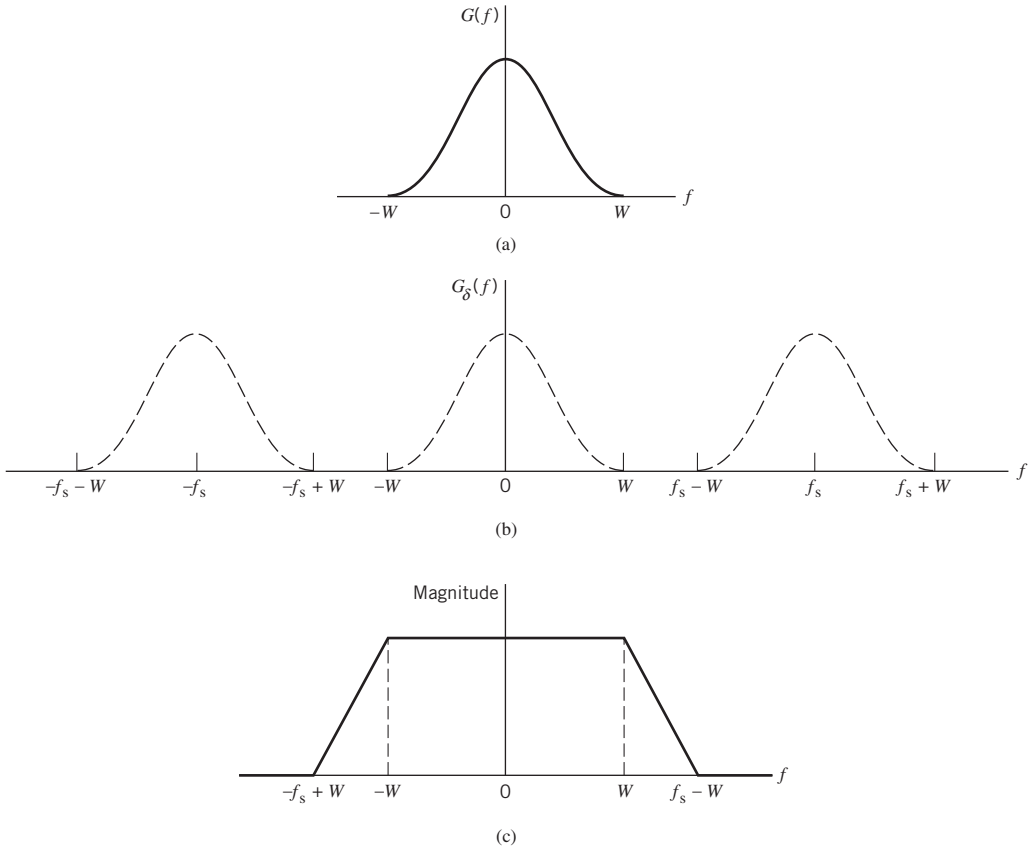


Figure 6.4 (a) Anti-alias filtered spectrum of an information-bearing signal. (b) Spectrum of instantaneously sampled version of the signal, assuming the use of a sampling rate greater than the Nyquist rate. (c) Magnitude response of reconstruction filter.

EXAMPLE 1

Sampling of Voice Signals

As an illustrative example, consider the sampling of voice signals for waveform coding. Typically, the frequency band, extending from 100 Hz to 3.1 kHz, is considered to be adequate for telephonic communication. This limited frequency band is accomplished by passing the voice signal through a low-pass filter with its cutoff frequency set at 3.1 kHz; such a filter may be viewed as an anti-aliasing filter. With such a cutoff frequency, the Nyquist rate is $f_s = 2 \times 3.1 = 6.2$ kHz. The standard sampling rate for the waveform coding of voice signals is 8 kHz. Putting these numbers together, design specifications for the reconstruction (low-pass) filter in the receiver are as follows:

Cutoff frequency	3.1 kHz
Transition band	6.2 to 8 kHz
Transition-band width	1.8 kHz.

6.3 Pulse-Amplitude Modulation

Now that we understand the essence of the sampling process, we are ready to formally define PAM, which is the simplest and most basic form of analog pulse modulation. It is formally defined as follows:

> PAM is a linear modulation process where the amplitudes of regularly spaced pulses are varied in proportion to the corresponding sample values of a continuous message signal.

The pulses themselves can be of a rectangular form or some other appropriate shape.

The waveform of a PAM signal is illustrated in Figure 6.5. The dashed curve in this figure depicts the waveform of a message signal $m(t)$, and the sequence of amplitude-modulated rectangular pulses shown as solid lines represents the corresponding PAM signal $s(t)$. There are two operations involved in the generation of the PAM signal:

1. *Instantaneous sampling* of the message signal $m(t)$ every T_s seconds, where the sampling rate $f_s = 1/T_s$ is chosen in accordance with the sampling theorem.
2. *Lengthening* the duration of each sample so obtained to some constant value T.

In digital circuit technology, these two operations are jointly referred to as "sample and hold." One important reason for intentionally lengthening the duration of each sample is to avoid the use of an excessive channel bandwidth, because bandwidth is inversely proportional to pulse duration. However, care has to be exercised in how long we make the sample duration T, as the following analysis reveals.

Let $s(t)$ denote the sequence of flat-top pulses generated in the manner described in Figure 6.5. We may express the PAM signal as a *discrete convolution sum*:

$$s(t) = \sum_{n=-\infty}^{\infty} m(nT_s)h(t - nT_s) \tag{6.10}$$

where T_s is the *sampling period* and $m(nT_s)$ is the sample value of $m(t)$ obtained at time $t = nT_s$. The $h(t)$ is a Fourier-transformal pulse. With spectral analysis of $s(t)$ in mind, we would like to recast (6.10) in the form of a convolution integral. To this end, we begin by invoking the sifting property of a delta function (discussed in Chapter 2) to express the delayed version of the pulse shape $h(t)$ in (6.10) as

$$h(t - nT_s) = \int_{-\infty}^{\infty} h(t - \tau)\delta(t - nT_s)\, d\tau \tag{6.11}$$

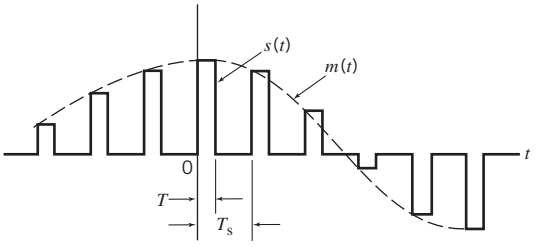


Figure 6.5 Flat-top samples, representing an analog signal.

Hence, substituting (6.11) into (6.10), and interchanging the order of summation and integration, we get

$$s(t) = \int_{-\infty}^{\infty} \left[\sum_{n=-\infty}^{\infty} m(nT_s)\delta(t - nT_s) \right] h(t - \tau)\, d\tau \tag{6.12}$$

Referring to (6.1), we recognize that the expression inside the brackets in (6.12) is simply the instantaneously sampled version of the message signal $m(t)$, as shown by

$$m_\delta(t) = \sum_{n=-\infty}^{\infty} m(nT_s)\delta(t - nT_s) \tag{6.13}$$

Accordingly, substituting (6.13) into (6.12), we may reformulate the PAM signal $s(t)$ in the desired form

$$\begin{aligned} s(t) &= \int_{-\infty}^{\infty} m_\delta(t) h(t - \tau)\, d\tau \\ &= m_\delta(t) \bigstar h(t) \end{aligned} \tag{6.14}$$

which is the convolution of the two time functions; $m_\delta(t)$ and $h(t)$.

The stage is now set for taking the Fourier transform of both sides of (6.14) and recognizing that the convolution of two time functions is transformed into the multiplication of their respective Fourier transforms; we get the simple result

$$S(f) = M_\delta(f)H(f) \tag{6.15}$$

where $S(f) = \mathbf{F}[s(t)]$, $M_\delta(f) = \mathbf{F}[m_\delta(t)]$, and $H(f) = \mathbf{F}[h(t)]$. Adapting (6.2) to the problem at hand, we note that the Fourier transform $M_\delta(f)$ is related to the Fourier transform $M(f)$ of the original message signal $m(t)$ as follows:

$$M_\delta(f) = f_s \sum_{k=-\infty}^{\infty} M(f - kf_s) \tag{6.16}$$

where f_s is the sampling rate. Therefore, the substitution of (6.16) into (6.15) yields the desired formula for the Fourier transform of the PAM signal $s(t)$, as shown by

$$S(f) = f_s \sum_{k=-\infty}^{\infty} M(f - kf_s)H(f) \tag{6.17}$$

Given this formula, how do we recover the original message signal $m(t)$? As a first step in this reconstruction, we may pass $s(t)$ through a low-pass filter whose frequency response is defined in Figure 6.4c; here, it is assumed that the message signal is limited to bandwidth W and the sampling rate f_s is larger than the Nyquist rate $2W$. Then, from (6.17) we find that the spectrum of the resulting filter output is equal to $M(f)H(f)$. This output is equivalent to passing the original message signal $m(t)$ through another low-pass filter of frequency response $H(f)$.

Equation (6.17) applies to any Fourier-transformable pulse shape $h(t)$.

Consider now the special case of a rectangular pulse of unit amplitude and duration T, as shown in Figure 6.6a; specifically:

$$h(t) = \begin{cases} 1, & 0 < t < T \\ \frac{1}{2}, & t = 0, t = T \\ 0, & \text{otherwise} \end{cases} \tag{6.18}$$

Correspondingly, the Fourier transform of $h(t)$ is given by

$$H(f) = T\text{sinc}(fT)\exp(-j\pi fT) \tag{6.19}$$

which is plotted in Figure 6.6b. We therefore find from (6.17) that by using flat-top samples to generate a PAM signal we have introduced *amplitude distortion* as well as a *delay* of $T/2$. This effect is rather similar to the variation in transmission with frequency that is caused by the finite size of the scanning aperture in television. Accordingly, the distortion caused by the use of PAM to transmit an analog information-bearing signal is referred to as the *aperture effect*.

To correct for this distortion, we connect an *equalizer* in cascade with the low-pass reconstruction filter, as shown in Figure 6.7. The equalizer has the effect of decreasing the in-band loss of the reconstruction filter as the frequency increases in such a manner as to

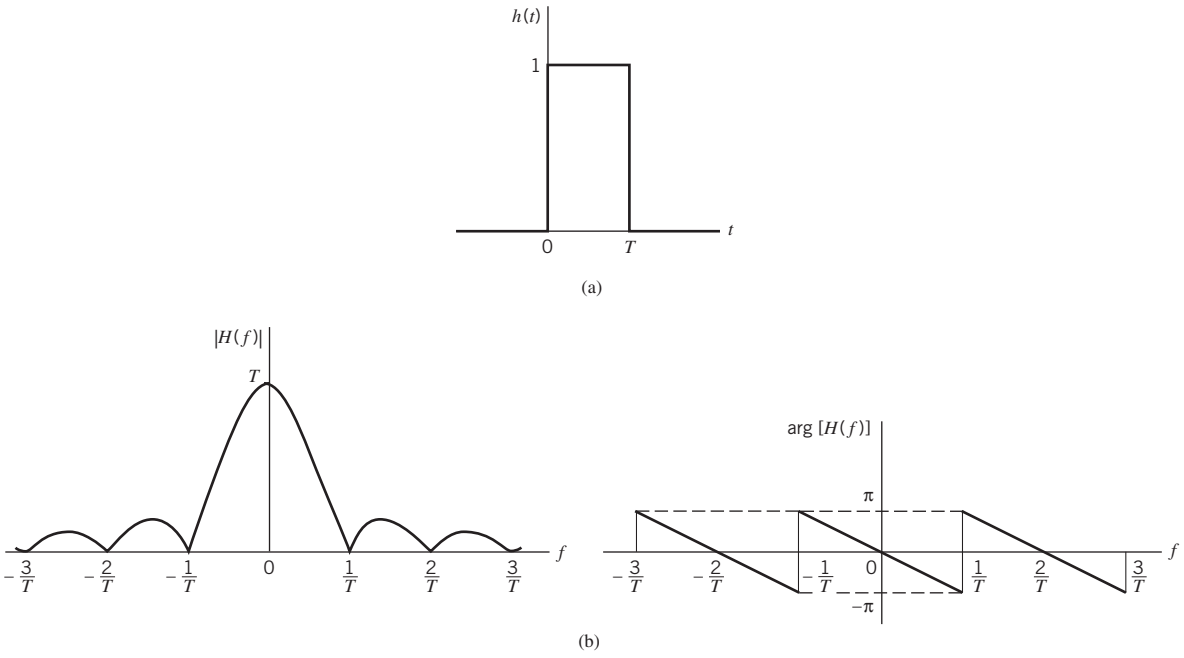


Figure 6.6 (a) Rectangular pulse $h(t)$. (b) Transfer function $H(f)$, made up of the magnitude $|H(f)|$ and phase arg$[H(f)]$.

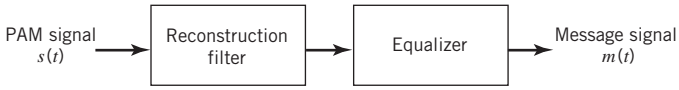


Figure 6.7 System for recovering message signal $m(t)$ from PAM signal $s(t)$.

compensate for the aperture effect. In light of (6.19), the magnitude response of the equalizer should ideally be

$$\frac{1}{|H(f)|} = \frac{1}{T\text{sinc}(fT)} = \frac{\pi f}{\sin(\pi fT)}$$

The amount of equalization needed in practice is usually small. Indeed, for a duty cycle defined by the ratio $T/T_s \leq 0.1$, the amplitude distortion is less than 0.5%. In such a situation, the need for equalization may be omitted altogether.

Practical Considerations

The transmission of a PAM signal imposes rather stringent requirements on the frequency response of the channel, because of the relatively short duration of the transmitted pulses. One other point that should be noted: relying on amplitude as the parameter subject to modulation, the noise performance of a PAM system can never be better than baseband-signal transmission. Accordingly, in practice, we find that for transmission over a communication channel PAM is used only as the preliminary means of message processing, whereafter the PAM signal is changed to some other more appropriate form of pulse modulation.

With analog-to-digital conversion as the aim, what would be the appropriate form of modulation to build on PAM? Basically, there are three potential candidates, each with its own advantages and disadvantages, as summarized here:

1. *PCM*, which, as remarked previously in Section 6.1, is robust but demanding in both transmission bandwidth and computational requirements. Indeed, PCM has established itself as the standard method for the conversion of speech and video signals into digital form.
2. *DPCM*, which provides a method for the reduction in transmission bandwidth but at the expense of increased computational complexity.
3. *DM*, which is relatively simple to implement but requires a significant increase in transmission bandwidth.

Before we go on, a comment on terminology is in order. The term "modulation" used herein is a *misnomer*. In reality, PCM, DM, and DPCM are different forms of source coding, with source coding being understood in the sense described in Chapter 5 on information theory. Nevertheless, the terminologies used to describe them have become embedded in the digital communications literature, so much so that we just have to live with them.

Despite their basic differences, PCM, DPCM and DM do share an important feature: the message signal is represented in discrete form in both time and amplitude. PAM takes care of the discrete-time representation. As for the discrete-amplitude representation, we resort to a process known as quantization, which is discussed next.

6.4 Quantization and its Statistical Characterization

Typically, an analog message signal (e.g., voice) has a continuous range of amplitudes and, therefore, its samples have a continuous amplitude range. In other words, within the finite amplitude range of the signal, we find an infinite number of amplitude levels. In actual fact, however, it is not necessary to transmit the exact amplitudes of the samples for the following reason: any human sense (the ear or the eye) as ultimate receiver can detect only finite intensity differences. This means that the message signal may be *approximated* by a signal constructed of discrete amplitudes selected on a minimum error basis from an available set. The existence of a finite number of discrete amplitude levels is a basic condition of waveform coding exemplified by PCM. Clearly, if we assign the discrete amplitude levels with sufficiently close spacing, then we may make the approximated signal practically indistinguishable from the original message signal. For a formal definition of *amplitude quantization*, or just *quantization* for short, we say:

> Quantization is the process of transforming the sample amplitude $m(nT_s)$ of a message signal $m(t)$ at time $t = nT_s$ into a discrete amplitude $v(nT_s)$ taken from a finite set of possible amplitudes.

This definition assumes that the *quantizer* (i.e., the device performing the quantization process) is *memoryless and instantaneous*, which means that the transformation at time $t = nT_s$ is not affected by earlier or later samples of the message signal $m(t)$. This simple form of scalar quantization, though not optimum, is commonly used in practice.

When dealing with a memoryless quantizer, we may simplify the notation by dropping the time index. Henceforth, the symbol m_k is used in place of $m(kT_s)$, as indicated in the block diagram of a quantizer shown in Figure 6.8a. Then, as shown in Figure 6.8b, the signal amplitude m is specified by the index k if it lies inside the *partition cell*

$$J_k: \{m_k < m \leq m_{k+1}\}, \qquad k = 1, 2, \ldots, L \tag{6.20}$$

where

$$m_k = m(kT_s) \tag{6.21}$$

and L is the total number of amplitude levels used in the quantizer. The discrete amplitudes m_k, $k = 1, 2, \ldots, L$, at the quantizer input are called *decision levels* or *decision thresholds*. At the quantizer output, the index k is transformed into an amplitude v_k that represents all amplitudes of the cell J_k; the discrete amplitudes v_k, $k = 1, 2, \ldots, L$, are called *representation levels* or *reconstruction levels*. The spacing between two adjacent representation levels is called a *quantum* or *step-size*. Thus, given a quantizer denoted by $g(\cdot)$, the quantized output v equals v_k if the input sample m belongs to the interval J_k. In effect, the mapping (see Figure 6.8a)

$$v = g(m) \tag{6.22}$$

defines the *quantizer characteristic*, described by a staircase function.

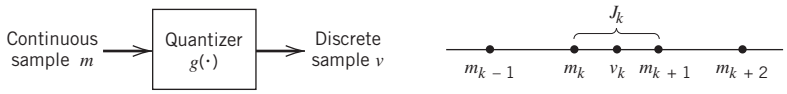


Figure 6.8
Description of a memoryless quantizer.

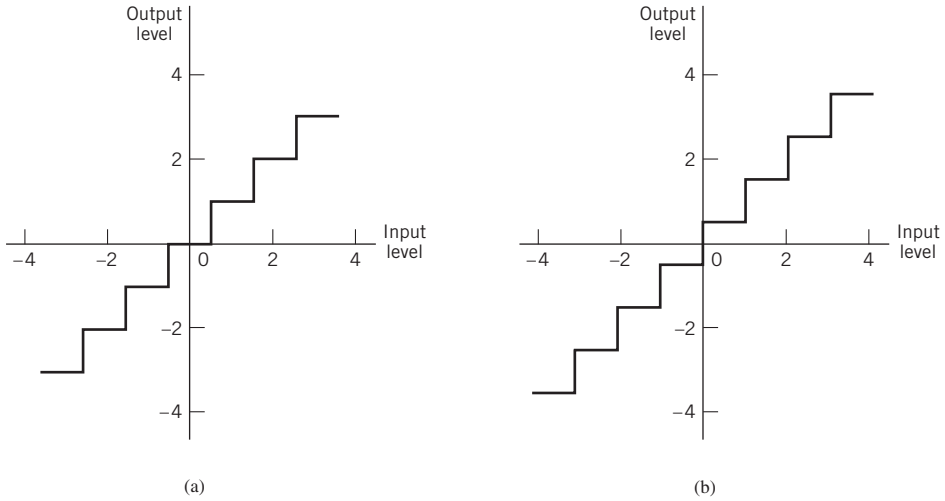


Figure 6.9 Two types of quantization: (a) midtread and (b) midrise.

Quantizers can be of a *uniform* or *nonuniform* type. In a uniform quantizer, the representation levels are uniformly spaced; otherwise, the quantizer is nonuniform. In this section, we consider only uniform quantizers; nonuniform quantizers are considered in Section 6.5. The quantizer characteristic can also be of *midtread* or *midrise type*. Figure 6.9a shows the input–output characteristic of a uniform quantizer of the midtread type, which is so called because the origin lies in the middle of a tread of the staircaselike graph. Figure 6.9b shows the corresponding input–output characteristic of a uniform quantizer of the midrise type, in which the origin lies in the middle of a rising part of the staircaselike graph. Despite their different appearances, both the midtread and midrise types of uniform quantizers illustrated in Figure 6.9 are *symmetric* about the origin.

Quantization Noise

Inevitably, the use of quantization introduces an error defined as the difference between the continuous input sample m and the quantized output sample v. The error is called *quantization noise*.[1] Figure 6.10 illustrates a typical variation of quantization noise as a function of time, assuming the use of a uniform quantizer of the midtread type.

Let the quantizer input m be the sample value of a zero-mean random variable M. (If the input has a nonzero mean, we can always remove it by subtracting the mean from the input and then adding it back after quantization.) A quantizer, denoted by $g(\cdot)$, maps the input random variable M of continuous amplitude into a discrete random variable V; their respective sample values m and v are related by the nonlinear function $g(\cdot)$ in (6.22). Let the quantization error be denoted by the random variable Q of sample value q. We may thus write

$$q = m - v \tag{6.23}$$

or, correspondingly,

$$Q = M - V \tag{6.24}$$

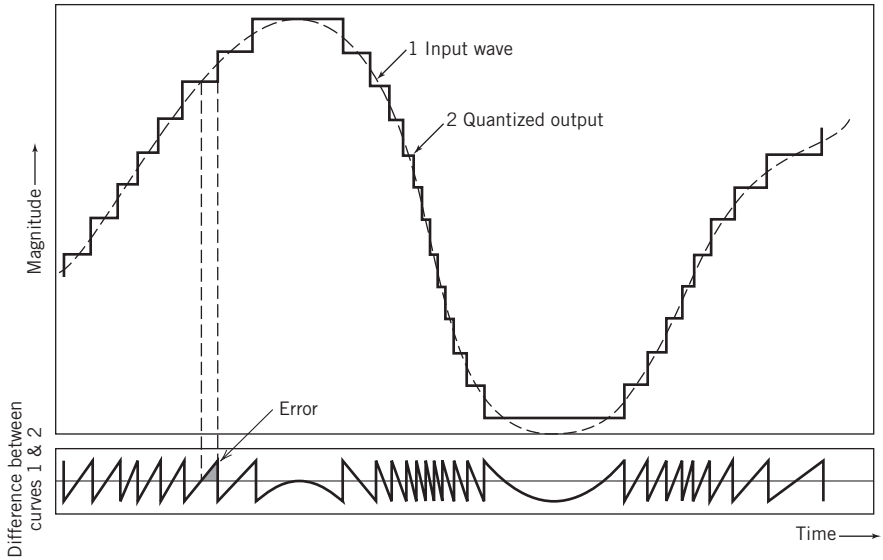


Figure 6.10
Illustration of the quantization process.

With the input M having zero mean and the quantizer assumed to be symmetric as in Figure 6.9, it follows that the quantizer output V and, therefore, the quantization error Q will also have zero mean. Thus, for a partial statistical characterization of the quantizer in terms of output signal-to-(quantization) noise ratio, we need only find the mean-square value of the quantization error Q.

Consider, then, an input m of continuous amplitude, which, symmetrically, occupies the range $[-m_{\max}, m_{\max}]$. Assuming a uniform quantizer of the midrise type illustrated in Figure 6.9b, we find that the step size of the quantizer is given by

$$\Delta = \frac{2m_{\max}}{L} \tag{6.25}$$

where L is the total number of representation levels. For a uniform quantizer, the quantization error Q will have its sample values bounded by $-\Delta/2 \le q \le \Delta/2$. If the step size Δ is sufficiently small (i.e., the number of representation levels L is sufficiently large), it is reasonable to assume that the quantization error Q is a *uniformly distributed* random variable and the interfering effect of the quantization error on the quantizer input is similar to that of thermal noise, hence the reference to quantization error as *quantization noise*. We may thus express the probability density function of the quantization noise as

$$f_Q(q) = \begin{cases} \dfrac{1}{\Delta}, & -\dfrac{\Delta}{2} < q \le \dfrac{\Delta}{2} \\ 0, & \text{otherwise} \end{cases} \tag{6.26}$$

For this to be true, however, we must ensure that the incoming continuous sample does *not* overload the quantizer. Then, with the mean of the quantization noise being zero, its variance σ_Q^2 is the same as the mean-square value; that is,

$$\begin{aligned}\sigma_Q^2 &= \mathbb{E}[Q^2] \\ &= \int_{-\Delta/2}^{\Delta/2} q^2 f_Q(q)\, \mathrm{d}q\end{aligned} \tag{6.27}$$

Substituting (6.26) into (6.27), we get

$$\begin{aligned}\sigma_Q^2 &= \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} q^2\, \mathrm{d}q \\ &= \frac{\Delta^2}{12}\end{aligned} \tag{6.28}$$

Typically, the L-ary number k, denoting the kth representation level of the quantizer, is transmitted to the receiver in binary form. Let R denote the *number of bits per sample* used in the construction of the binary code. We may then write

$$L = 2^R \tag{6.29}$$

or, equivalently,

$$R = \log_2 L \tag{6.30}$$

Hence, substituting (6.29) into (6.25), we get the step size

$$\Delta = \frac{2m_{\max}}{2^R} \tag{6.31}$$

Thus, the use of (6.31) in (6.28) yields

$$\sigma_Q^2 = \frac{1}{3}m_{\max}^2 2^{-2R} \tag{6.32}$$

Let P denote the average power of the original message signal $m(t)$. We may then express the *output signal-to-noise ratio* of a uniform quantizer as

$$\begin{aligned}(\mathrm{SNR})_{\mathrm{O}} &= \frac{P}{\sigma_Q^2} \\ &= \left(\frac{3P}{m_{\max}^2}\right)2^{2R}\end{aligned} \tag{6.33}$$

Equation (6.33) shows that the output signal-to-noise ratio of a uniform quantizer $(\mathrm{SNR})_{\mathrm{O}}$ increases *exponentially* with increasing number of bits per sample R, which is intuitively satisfying.

EXAMPLE 2 **Sinusoidal Modulating Signal**

Consider the special case of a full-load sinusoidal modulating signal of amplitude A_{m}, which utilizes all the representation levels provided. The average signal power is (assuming a load of 1 Ω)

$$P = \frac{A_{\mathrm{m}}^2}{2}$$

The total range of the quantizer input is $2A_m$, because the modulating signal swings between $-A_m$ and A_m. We may, therefore, set $m_{max} = A_m$, in which case the use of (6.32) yields the average power (variance) of the quantization noise as

$$\sigma_Q^2 = \frac{1}{3}A_m^2 2^{-2R}$$

Thus, the output signal-to-noise of a uniform quantizer, for a full-load test tone, is

$$(\text{SNR})_O = \frac{A_m^2/2}{A_m^2 2^{-2R}/3} = \frac{3}{2}(2^{2R}) \tag{6.34}$$

Expressing the signal-to-noise (SNR) in decibels, we get

$$10 \log_{10}(\text{SNR})_O = 1.8 + 6R \tag{6.35}$$

The corresponding values of signal-to-noise ratio for various values of L and R, are given in Table 6.1. For sinusoidal modulation, this table provides a basis for making a quick estimate of the number of bits per sample required for a desired output signal-to-noise ratio.

Table 6.1 Signal-to-(quantization) noise ratio for varying number of representation levels for sinusoidal modulation

No. of representation levels L	No. of bits per sample R	SNR (dB)
32	5	31.8
64	6	37.8
128	7	43.8
256	8	49.8

Conditions of Optimality of Scalar Quantizers

In designing a scalar quantizer, the challenge is how to select the representation levels and surrounding partition cells so as to minimize the average quantization power for a fixed number of representation levels.

To state the problem in mathematical terms: consider a message signal $m(t)$ drawn from a stationary process and whose dynamic range, denoted by $-A \le m \le A$, is partitioned into a set of L cells, as depicted in Figure 6.11. The boundaries of the partition cells are defined by a set of real numbers $m_1, m_2, \ldots, m_{L-1}$ that satisfy the following three conditions:

$$m_1 = -A$$
$$m_{L-1} = A$$
$$m_k \le m_{k-1} \text{ for } k = 1, 2, \ldots, L$$

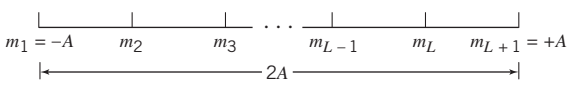


Figure 6.11
Illustrating the partitioning of the dynamic range $-A \le m \le A$ of a message signal $m(t)$ into a set of L cells.

The kth partition cell is defined by (6.20), reproduced here for convenience:

$$J_k : m_k < m < m_{k-1} \text{ for } k = 1, 2, \ldots, L \tag{6.36}$$

Let the representation levels (i.e., quantization values) be denoted by v_k, $k = 1, 2, \ldots, L$. Then, assuming that $d(m,v_k)$ denotes a *distortion measure* for using v_k to represent all those values of the input m that lie inside the partition cell J_k, the goal is to find the two sets $\{v_k\}_{k=1}^{L}$ and $\{J_k\}_{k=1}^{L}$ that minimize the *average distortion*

$$D = \sum_{k=1}^{L} \int_{m \in v_k} d(m,v_k) f_M(m)\, \mathrm{d}m \tag{6.37}$$

where $f_M(m)$ is the probability density function of the random variable M with sample value m.

A commonly used distortion measure is defined by

$$d(m,v_k) = (m - v_k)^2 \tag{6.38}$$

in which case we speak of the *mean-square distortion*. In any event, the optimization problem stated herein is nonlinear, defying an explicit, closed-form solution. To get around this difficulty, we resort to an *algorithmic approach* for solving the problem in an *iterative manner*.

Structurally speaking, the quantizer consists of two components with interrelated design parameters:

- An encoder characterized by the set of partition cells $\{J_k\}_{k=1}^{L}$; this is located in the transmitter.
- A decoder characterized by the set of representation levels $\{v_k\}_{k=1}^{L}$; this is located in the receiver.

Accordingly, we may identify two critically important conditions that provide the mathematical basis for all algorithmic solutions to the optimum quantization problem. One condition assumes that we are given a decoder and the problem is to find the optimum encoder in the transmitter. The other condition assumes that we are given an encoder and the problem is to find the optimum decoder in the receiver. Henceforth, these two conditions are referred to as condition I and II, respectively.

Condition I: Optimality of the Encoder for a Given Decoder

The availability of a decoder means that we have a certain *codebook* in mind. Let the codebook be defined by

$$\mathscr{C} : \{v_k\}_{k=1}^{L} \tag{6.39}$$

Given the codebook $\mathscr{C}$, the problem is to find the set of partition cells $\{J_k\}_{k=1}^{L}$ that minimizes the mean-square distortion D. That is, we wish to find the encoder defined by the nonlinear mapping

$$g(m) = v_k, \qquad k = 1, 2, \ldots, L \tag{6.40}$$

such that we have

$$D = \int_{-A}^{A} d(m, g(m)) f_M(m)\, \mathrm{d}M \geq \sum_{k=1}^{L} \int_{m \in J_k} [\min d(m, v_k)] f_M(m)\, \mathrm{d}m \tag{6.41}$$

For the lower bound specified in (6.41) to be attained, we require that the nonlinear mapping of (6.40) be satisfied only if the condition

$$d(m, v_k) \leq d(m, v_j) \qquad \text{holds for all } j \neq k \tag{6.42}$$

The necessary condition described in (6.42) for optimality of the encoder for a specified codebook $\mathscr{C}$ is recognized as the *nearest-neighbor condition*. In words, the nearest neighbor condition requires that the partition cell J_k should embody all those values of the input m that are closer to v_k than any other element of the codebook $\mathscr{C}$. This optimality condition is indeed intuitively satisfying.

Condition II: Optimality of the Decoder for a Given Encoder

Consider next the reverse situation to that described under condition I, which may be stated as follows: optimize the codebook $\mathscr{C} = \{v_k\}_{k=1}^{L}$ for the decoder, given that the set of partition cells $\{J_k\}_{k=1}^{L}$ characterizing the encoder is fixed. The criterion for optimization is the average (mean-square) distortion:

$$D = \sum_{k=1}^{L} \int_{m \in J_k} (m - v_k)^2 f_M(m)\, \mathrm{d}m \tag{6.43}$$

The probability density function $f_M(m)$ is clearly independent of the codebook $\mathscr{C}$. Hence, differentiating D with respect to the representation level v_k, we readily obtain

$$\frac{\partial D}{\partial v_k} = -2 \sum_{k=1}^{L} \int_{m \in J_k} (m - v_k) f_M(m)\, \mathrm{d}m \tag{6.44}$$

Setting $\partial D / \partial v_k$ equal to zero and then solving for v_k, we obtain the optimum value

$$v_{k,\,\mathrm{opt}} = \frac{\displaystyle\int_{m \in J_k} m f_M(m)\, \mathrm{d}m}{\displaystyle\int_{m \in J_k} f_M(m)\, \mathrm{d}m} \tag{6.45}$$

The denominator in (6.45) is just the probability p_k that the random variable M with sample value m lies in the partition cell J_k, as shown by

$$\begin{aligned} p_k &= \mathbb{P}(m_k < M \leq m_k + 1) \\ &= \int_{m \in J_k} f_M(m)\, \mathrm{d}m \end{aligned} \tag{6.46}$$

Accordingly, we may interpret the optimality condition of (6.45) as choosing the representation level v_k to equal the *conditional mean* of the random variable M, given that M lies in the partition cell J_k. We can thus formally state that the condition for optimality of the decoder for a given encoder as follows:

$$v_{k,\,\mathrm{opt}} = \mathbb{E}[M \,|\, m_k < M \leq m_{k+1}] \tag{6.47}$$

where $\mathbb{E}$ is the expectation operator. Equation (6.47) is also intuitively satisfying.

Note that the nearest neighbor condition (I) for optimality of the encoder for a given decoder was proved for a generic average distortion. However, the conditional mean requirement (condition II) for optimality of the decoder for a given encoder was proved for

the special case of a mean-square distortion. In any event, these two conditions are necessary for optimality of a scalar quantizer. Basically, the algorithm for designing the quantizer consists of alternately optimizing the encoder in accordance with condition I, then optimizing the decoder in accordance with condition II, and continuing in this manner until the average distortion D reaches a minimum. The optimum quantizer designed in this manner is called the *Lloyd–Max quantizer*.[2]

6.5 Pulse-Code Modulation

With the material on sampling, PAM, and quantization presented in the preceding sections, the stage is set for describing PCM, for which we offer the following definition:

> PCM is a discrete-time, discrete-amplitude waveform-coding process, by means of which an analog signal is directly represented by a sequence of coded pulses.

Specifically, the transmitter consists of two components: a *pulse-amplitude modulator* followed by an *analog-to-digital (A/D) converter*. The latter component itself embodies a *quantizer* followed by an *encoder*. The receiver performs the inverse of these two operations: *digital-to-analog (D/A) conversion* followed by *pulse-amplitude demodulation*. The communication channel is responsible for transporting the encoded pulses from the transmitter to the receiver.

Figure 6.12, a block diagram of the PCM, shows the transmitter, the transmission path from the transmitter output to the receiver input, and the receiver.

It is important to realize, however, that once distortion in the form of quantization noise is introduced into the encoded pulses, there is absolutely nothing that can be done at the receiver to compensate for that distortion. The only design precaution that can be taken is to choose a number of representation levels in the receiver that is large enough to ensure that the quantization noise is imperceptible for human use at the receiver output.

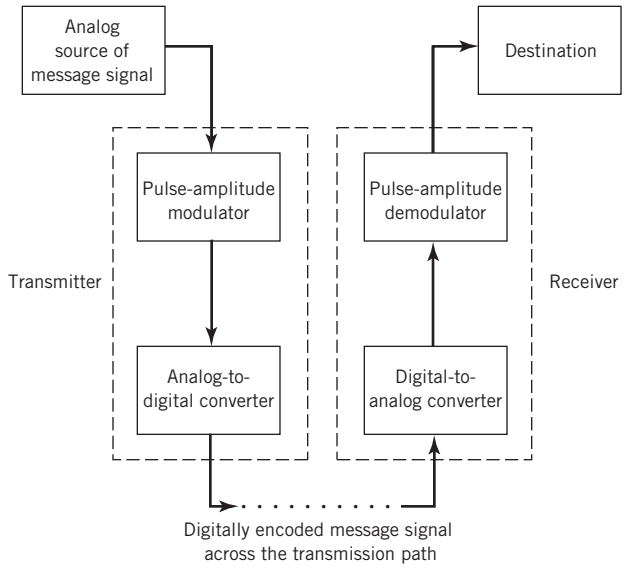


Figure 6.12 Block diagram of PCM system.

Sampling in the Transmitter

The incoming message signal is sampled with a train of rectangular pulses short enough to closely approximate the instantaneous sampling process. To ensure perfect reconstruction of the message signal at the receiver, the sampling rate must be greater than twice the highest frequency component W of the message signal in accordance with the sampling theorem. In practice, a low-pass anti-aliasing filter is used at the front end of the pulse-amplitude modulator to exclude frequencies greater than W before sampling and which are of negligible practical importance. Thus, the application of sampling permits the reduction of the continuously varying message signal to a limited number of discrete values per second.

Quantization in the Transmitter

The PAM representation of the message signal is then quantized in the analog-to-digital converter, thereby providing a new representation of the signal that is discrete in both time and amplitude. The quantization process may follow a uniform law as described in Section 6.4. In telephonic communication, however, it is preferable to use a variable separation between the representation levels for efficient utilization of the communication channel. Consider, for example, the quantization of voice signals. Typically, we find that the range of voltages covered by voice signals, from the peaks of loud talk to the weak passages of weak talk, is on the order of 1000 to 1. By using a *nonuniform quantizer* with the feature that the step size increases as the separation from the origin of the input–output amplitude characteristic of the quantizer is increased, the large end-steps of the quantizer can take care of possible excursions of the voice signal into the large amplitude ranges that occur relatively infrequently. In other words, the weak passages needing more protection are favored at the expense of the loud passages. In this way, a nearly uniform percentage precision is achieved throughout the greater part of the amplitude range of the input signal. The end result is that fewer steps are needed than would be the case if a uniform quantizer were used; hence the improvement in channel utilization.

Assuming memoryless quantization, the use of a nonuniform quantizer is equivalent to passing the message signal through a *compressor* and then applying the compressed signal to a *uniform quantizer*, as illustrated in Figure 6.13a. A particular form of *compression law* that is used in practice is the so-called *μ-law*,[3] which is defined by

$$|v| = \frac{\ln(1 + \mu|m|)}{\ln(1 + \mu)} \tag{6.48}$$

where ln, i.e., $\log_e$, denotes the natural logarithm, m and v are the input and output voltages of the *compressor*, and μ is a positive constant. It is assumed that m and,

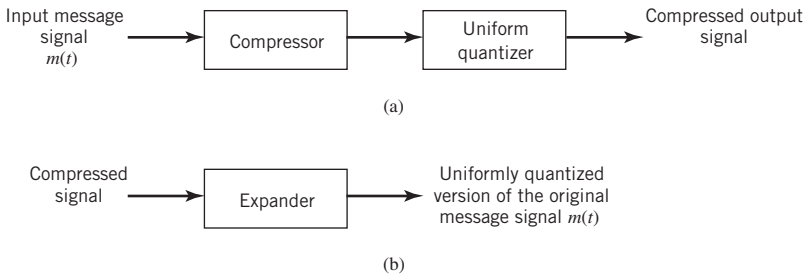


Figure 6.13
(a) Nonuniform quantization of the message signal in the transmitter. (b) Uniform quantization of the original message signal in the receiver.

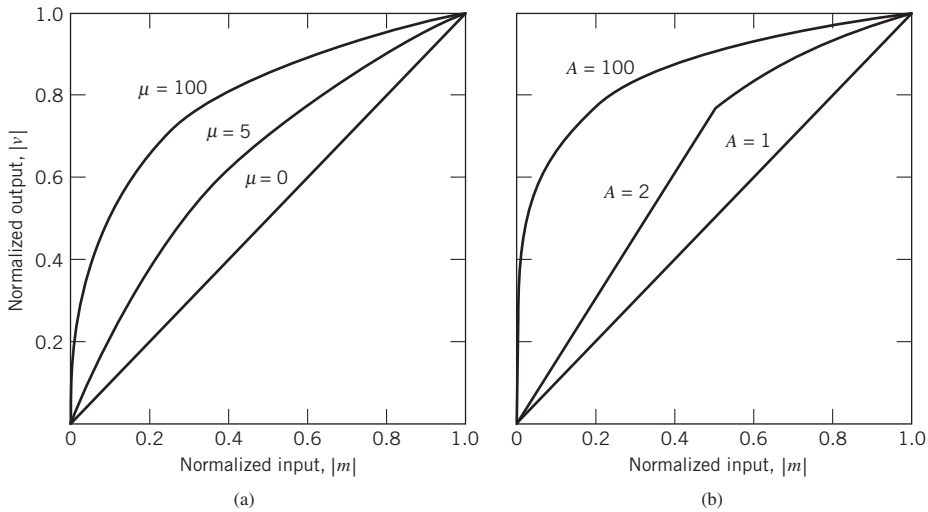


Figure 6.14
Compression laws:
(a) μ-law;
(b) A-law.

therefore, v are scaled so that they both lie inside the interval [–1, 1]. The μ-law is plotted for three different values of μ in Figure 6.14a. The case of uniform quantization corresponds to $\mu = 0$. For a given value of μ, the reciprocal slope of the compression curve that defines the quantum steps is given by the derivative of the absolute value $|m|$ with respect to the corresponding absolute value $|v|$; that is,

$$\frac{d|m|}{d|v|} = \frac{\ln(1+\mu)}{\mu}(1+\mu|m|) \tag{6.49}$$

From (6.49) it is apparent that the μ-law is neither strictly linear nor strictly logarithmic. Rather, it is approximately linear at low input levels corresponding to $\mu|m| \ll 1$ and approximately logarithmic at high input levels corresponding to $\mu|m| \gg 1$.

Another compression law that is used in practice is the so-called *A-law*, defined by

$$|v| = \begin{cases} \dfrac{A|m|}{1+\ln A}, & 0 \le |m| \le \dfrac{1}{A} \\ \dfrac{1+\ln(A|m|)}{1+\ln A}, & \dfrac{1}{A} \le |m| \le 1 \end{cases} \tag{6.50}$$

where A is another positive constant. Equation (6.50) is plotted in Figure 6.14b for varying A. The case of uniform quantization corresponds to $A = 1$. The reciprocal slope of this second compression curve is given by the derivative of $|m|$ with respect to $|v|$, as shown by

$$\frac{d|m|}{d|v|} = \begin{cases} \dfrac{1+\ln A}{A}, & 0 \le |m| \le \dfrac{1}{A} \\ (1+\ln A)|m|, & \dfrac{1}{A} \le |m| \le 1 \end{cases} \tag{6.51}$$

To restore the signal samples to their correct relative level, we must, of course, use a device in the receiver with a characteristic complementary to the compressor. Such a device is called an *expander*. Ideally, the compression and expansion laws are exactly the inverse of each other. With this provision in place, we find that, except for the effect of quantization, the expander output is equal to the compressor input. The cascade combination of a *compressor* and an *expander,* depicted in Figure 6.13, is called a *compander*.

For both the μ-law and A-law, the dynamic range capability of the compander improves with increasing μ and A, respectively. The SNR for low-level signals increases at the expense of the SNR for high-level signals. To accommodate these two conflicting requirements (i.e., a reasonable SNR for both low- and high-level signals), a compromise is usually made in choosing the value of parameter μ for the μ-law and parameter A for the A-law. The typical values used in practice are $\mu = 255$ for the μ–law and $A = 87.6$ for the A-law.[4]

Encoding in the Transmitter

Through the combined use of sampling and quantization, the specification of an analog message signal becomes limited to a discrete set of values, but not in the form best suited to transmission over a telephone line or radio link. To exploit the advantages of sampling and quantizing for the purpose of making the transmitted signal more robust to noise, interference, and other channel impairments, we require the use of an *encoding process* to translate the discrete set of sample values to a more appropriate form of signal. Any plan for representing each of this discrete set of values as a particular arrangement of discrete events constitutes a *code*. Table 6.2 describes the one-to-one correspondence between representation levels and codewords for a binary number system for $R = 4$ bits per sample. Following the terminology of Chapter 5, the two symbols of a binary code are customarily denoted as 0 and 1. In practice, the binary code is the preferred choice for encoding for the following reason:

> The maximum advantage over the effects of noise encountered in a communication system is obtained by using a binary code because a binary symbol withstands a relatively high level of noise and, furthermore, it is easy to regenerate.

The last signal-processing operation in the transmitter is that of *line coding*, the purpose of which is to represent each binary codeword by a sequence of pulses; for example, symbol 1 is represented by the presence of a pulse and symbol 0 is represented by absence of the pulse. Line codes are discussed in Section 6.10. Suppose that, in a binary code, each codeword consists of R bits. Then, using such a code, we may represent a total of 2^R distinct numbers. For example, a sample quantized into one of 256 levels may be represented by an 8-bit codeword.

Inverse Operations in the PCM Receiver

The first operation in the receiver of a PCM system is to *regenerate* (i.e., reshape and clean up) the received pulses. These clean pulses are then regrouped into codewords and decoded (i.e., mapped back) into a quantized pulse-amplitude modulated signal. The *decoding* process involves generating a pulse the amplitude of which is the linear sum of all the pulses in the codeword. Each pulse is weighted by its place value ($2^0, 2^1, 2^2, \ldots, 2^{R-1}$) in the code, where R is the number of bits per sample. Note, however, that whereas the analog-to-digital

Table 6.2 Binary number system for $T = 4$ bits/sample

Ordinal number of representation level	Level number expressed as sum of powers of 2	Binary number
0		0000
1	2^0	0001
2	2^1	0010
3	$2^1 + 2^0$	0011
4	2^2	0100
5	$2^2 + 2^0$	0101
6	$2^2 + 2^1$	0110
7	$2^2 + 2^1 + 2^0$	0111
8	2^3	1000
9	$2^3 + 2^0$	1001
10	$2^3 + 2^1$	1010
11	$2^3 + 2^1 + 2^0$	1011
12	$2^3 + 2^2$	1100
13	$2^3 + 2^2 + 2^0$	1101
14	$2^3 + 2^2 + 2^1$	1110
15	$2^3 + 2^2 + 2^1 + 2^0$	1111

converter in the transmitter involves both quantization and encoding, the digital-to-analog converter in the receiver involves decoding only, as illustrated in Figure 6.12.

The final operation in the receiver is that of *signal reconstruction*. Specifically, an estimate of the original message signal is produced by passing the decoder output through a *low-pass reconstruction filter* whose cutoff frequency is equal to the message bandwidth W. Assuming that the transmission link (connecting the receiver to the transmitter) is error free, the reconstructed message signal includes no noise with the exception of the initial distortion introduced by the quantization process.

PCM Regeneration along the Transmission Path

The most important feature of a PCM systems is its ability to control the effects of distortion and noise produced by transmitting a PCM signal through the channel, connecting the receiver to the transmitter. This capability is accomplished by reconstructing the PCM signal through a chain of *regenerative repeaters*, located at sufficiently close spacing along the transmission path.

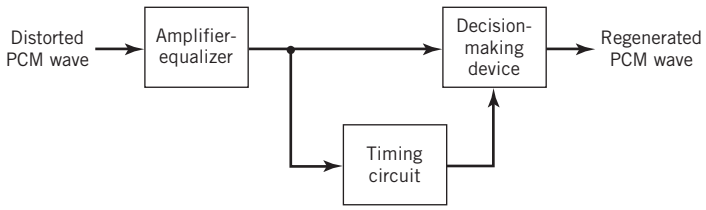


Figure 6.15
Block diagram of regenerative repeater.

As illustrated in Figure 6.15, three basic functions are performed in a regenerative repeater: *equalization*, *timing*, and *decision making*. The equalizer shapes the received pulses so as to compensate for the effects of amplitude and phase distortions produced by the non-ideal transmission characteristics of the channel. The timing circuitry provides a periodic pulse train, derived from the received pulses, for sampling the equalized pulses at the instants of time where the SNR ratio is a maximum. Each sample so extracted is compared with a predetermined *threshold* in the decision-making device. In each bit interval, a decision is then made on whether the received symbol is 1 or 0 by observing whether the threshold is exceeded or not. If the threshold is exceeded, a clean new pulse representing symbol 1 is transmitted to the next repeater; otherwise, another clean new pulse representing symbol 0 is transmitted. In this way, it is possible for the accumulation of distortion and noise in a repeater span to be almost completely removed, provided that the disturbance is not too large to cause an error in the decision-making process. Ideally, except for delay, the regenerated signal is exactly the same as the signal originally transmitted. In practice, however, the regenerated signal departs from the original signal for two main reasons:

1. The unavoidable presence of channel noise and interference causes the repeater to make wrong decisions occasionally, thereby introducing *bit errors* into the regenerated signal.
2. If the spacing between received pulses deviates from its assigned value, a *jitter* is introduced into the regenerated pulse position, thereby causing distortion.

The important point to take from this subsection on PCM is the fact that regeneration along the transmission path is provided across the spacing between individual regenerative repeaters (including the last stage of regeneration at the receiver input) provided that the spacing is short enough. If the transmitted SNR ratio is high enough, then the regenerated PCM data stream is the same as the transmitted PCM data stream, except for a practically negligibly small *bit error rate* (BER). In other words, under these operating conditions, performance degradation in the PCM system is essentially confined to quantization noise in the transmitter.

6.6 Noise Considerations in PCM Systems

The performance of a PCM system is influenced by two major sources of noise:

1. *Channel noise*, which is introduced anywhere between the transmitter output and the receiver input; channel noise is always present, once the equipment is switched on.
2. *Quantization noise*, which is introduced in the transmitter and is carried all the way along to the receiver output; unlike channel noise, quantization noise is *signal dependent,* in the sense that it disappears when the message signal is switched off.

Naturally, these two sources of noise appear simultaneously once the PCM system is in operation. However, the traditional practice is to consider them separately, so that we may develop insight into their individual effects on the system performance.

The main effect of channel noise is to introduce *bit errors* into the received signal. In the case of a binary PCM system, the presence of a bit error causes symbol 1 to be mistaken for symbol 0, or vice versa. Clearly, the more frequently bit errors occur, the more dissimilar the receiver output becomes compared with the original message signal. The fidelity of information transmission by PCM in the presence of channel noise may be measured in terms of the *average probability of symbol error*, which is defined as the probability that the reconstructed symbol at the receiver output differs from the transmitted binary symbol on the average. The average probability of symbol error, also referred to as the BER, assumes that all the bits in the original binary wave are of equal importance. When, however, there is more interest in restructuring the analog waveform of the original message signal, different symbol errors may be *weighted* differently; for example, an error in the most significant bit in a codeword (representing a quantized sample of the message signal) is more harmful than an error in the least significant bit.

To optimize system performance in the presence of channel noise, we need to minimize the average probability of symbol error. For this evaluation, it is customary to model the channel noise as an ideal *additive white Gaussian noise* (AWGN) channel. The effect of channel noise can be made practically negligible by using an adequate signal energy-to-noise density ratio through the provision of short-enough spacing between the regenerative repeaters in the PCM system. In such a situation, the performance of the PCM system is essentially limited by quantization noise acting alone.

From the discussion of quantization noise presented in Section 6.4, we recognize that quantization noise is essentially under the designer's control. It can be made negligibly small through the use of an adequate number of representation levels in the quantizer and the selection of a companding strategy matched to the characteristics of the type of message signal being transmitted. We thus find that the use of PCM offers the possibility of building a communication system that is *rugged* with respect to channel noise on a scale that is beyond the capability of any analog communication system; hence its use as a *standard* against which other waveform coders (e.g., DPCM and DM) are compared.

Error Threshold

The underlying theory of BER calculation in a PCM system is deferred to Chapter 8. For the present, it suffices to say that the average probability of symbol error in a binary encoded PCM receiver due to AWGN depends solely on E_b/N_0, which is defined as *the ratio of the transmitted signal energy per bit* E_b, *to the noise spectral density* N_0. Note that the ratio E_b/N_0 is dimensionless even though the quantities E_b and N_0 have different physical meaning. In Table 6.3, we present a summary of this dependence for the case of a binary PCM system, in which symbols 1 and 0 are represented by rectangular pulses of equal but opposite amplitudes. The results presented in the last column of the table assume a bit rate of 10^5 bits/s.

From Table 6.3 it is clear that there is an *error threshold* (at about 11 dB). For E_b/N_0 below the error threshold the receiver performance involves significant numbers of errors, and above it the effect of channel noise is practically negligible. In other words, provided that the ratio E_b/N_0 exceeds the error threshold, channel noise has virtually no effect on

Table 6.3 Influence of E_b/N_0 on the probability of error

E_b/N_0 (dB)	Probability of error P_e	For a bit rate of 10^5 bits/s, this is about one error every
4.3	10^{-2}	10^{-3} s
8.4	10^{-4}	10^{-1} s
10.6	10^{-6}	10 s
12.0	10^{-8}	20 min
13.0	10^{-10}	1 day
14.0	10^{-12}	3 months

the receiver performance, which is precisely the goal of PCM. When, however, E_b/N_0 drops below the error threshold, there is a sharp increase in the rate at which errors occur in the receiver. Because decision errors result in the construction of incorrect codewords, we find that when the errors are frequent, the reconstructed message at the receiver output bears little resemblance to the original message signal.

An important characteristic of a PCM system is its *ruggedness to interference*, caused by impulsive noise or cross-channel interference. The combined presence of channel noise and interference causes the error threshold necessary for satisfactory operation of the PCM system to increase. If, however, an adequate margin over the error threshold is provided in the first place, the system can withstand the presence of relatively large amounts of interference. In other words, a PCM system is *robust* with respect to channel noise and interference, providing further confirmation to the point made in the previous section that performance degradation in PCM is essentially confined to quantization noise in the transmitter.

PCM Noise Performance Viewed in Light of the Information Capacity Law

Consider now a PCM system that is known to operate above the error threshold, in which case we would be justified to ignore the effect of channel noise. In other words, the noise performance of the PCM system is essentially determined by quantization noise acting alone. Given such a scenario, how does the PCM system fare compared with the information capacity law, derived in Chapter 5?

To address this question of practical importance, suppose that the system uses a codeword consisting of n symbols with each symbol representing one of M possible discrete amplitude levels; hence the reference to the system as an "M-ary" PCM system. For this system to operate above the error threshold, there must be provision for a large enough noise margin.

For the PCM system to operate above the error threshold as proposed, the requirement for a noise margin that is sufficiently large to maintain a negligible error rate due to channel noise. This, in turn, means there must be a certain separation between the M discrete amplitude levels. Call this separation $c\sigma$, where c is a constant and $\sigma^2 = N_0B$ is the

noise variance measured in a channel bandwidth B. The number of amplitude levels M is usually an integer power of 2. The average transmitted power will be least if the amplitude range is symmetrical about zero. Then, the discrete amplitude levels, normalized with respect to the separation $c\sigma$, will have the values $\pm 1/2, \pm 3/2, \ldots, \pm(M-1)/2$. We assume that these M different amplitude levels are equally likely. Accordingly, we find that the average transmitted power is given by

$$\begin{aligned} P &= \frac{2}{M}\left[\left(\frac{1}{2}\right)^2 + \left(\frac{3}{2}\right)^2 + \cdots + \left(\frac{M-1}{2}\right)^2\right](c\sigma)^2 \\ &= c^2\sigma^2\left(\frac{M^2-1}{12}\right) \end{aligned} \tag{6.52}$$

Suppose that the M-ary PCM system described herein is used to transmit a message signal with its highest frequency component equal to W hertz. The signal is sampled at the Nyquist rate of $2W$ samples per second. We assume that the system uses a quantizer of the midrise type, with L equally likely representation levels. Hence, the probability of occurrence of any one of the L representation levels is $1/L$. Correspondingly, the amount of information carried by a single sample of the signal is $\log_2 L$ bits. With a maximum sampling rate of $2W$ samples per second, the maximum rate of information transmission of the PCM system measured in bits per second is given by

$$R_b = 2W \log_2 L \text{ bits/s} \tag{6.53}$$

Since the PCM system uses a codeword consisting of n code elements with each one having M possible discrete amplitude values, we have M^n different possible codewords. For a unique encoding process, therefore, we require

$$L = M^n \tag{6.54}$$

Clearly, the rate of information transmission in the system is unaffected by the use of an encoding process. We may, therefore, eliminate L between (6.53) and (6.54) to obtain

$$R_b = 2Wn \log_2 M \text{ bits/s} \tag{6.55}$$

Equation (6.52) defines the average transmitted power required to maintain an M-ary PCM system operating above the error threshold. Hence, solving this equation for the number of discrete amplitude levels, we may express the number M in terms of the average transmitted power P and channel noise variance $\sigma^2 = N_0 B$ as follows:

$$M = \left(1 + \frac{12P}{c^2 N_0 B}\right)^{1/2} \tag{6.56}$$

Therefore, substituting (6.56) into (6.55), we obtain

$$R_b = Wn \log_2\left(1 + \frac{12P}{c^2 N_0 B}\right) \tag{6.57}$$

The channel bandwidth B required to transmit a rectangular pulse of duration $1/(2nW)$, representing a symbol in the codeword, is given by

$$B = \kappa n W \tag{6.58}$$

where κ is a constant with a value lying between 1 and 2. Using the minimum possible value $\kappa = 1$, we find that the channel bandwidth $B = nW$. We may thus rewrite (6.57) as

$$R_b = B \log_2\left(1 + \frac{12P}{c^2 N_0 B}\right) \text{ bits/s} \tag{6.59}$$

which defines the upper bound on the information capacity realizable by an M-ary PCM system.

From Chapter 5 we recall that, in accordance with Shannon's information capacity law, the *ideal transmission system* is described by the formula

$$C = B \log_2\left(1 + \frac{P}{N_0 B}\right) \text{ bits/s} \tag{6.60}$$

The most interesting point derived from the comparison of (6.59) with (6.60) is the fact that (6.59) is of the right mathematical form in an information-theoretic context. To be more specific, we make the following statement:

> Power and bandwidth in a PCM system are exchanged on a logarithmic basis, and the information capacity of the system is proportional to the channel bandwidth B.

As a corollary, we may go on to state:

> When the SNR ratio is high, the bandwidth-noise trade-off follows an exponential law in PCM.

From the study of noise in analog modulation systems,[5] it is known that the use of frequency modulation provides the best improvement in SNR ratio. To be specific, when the carrier-to-noise ratio is high enough, the bandwidth-noise trade-off follows a *square law* in frequency modulation (FM). Accordingly, in comparing the noise performance of FM with that of PCM we make the concluding statement:

> PCM is more efficient than FM in trading off an increase in bandwidth for improved noise performance.

Indeed, this statement is further testimony for the PCM being viewed as a standard for waveform coding.

6.7 Prediction-Error Filtering for Redundancy Reduction

When a voice or video signal is sampled at a rate slightly higher than the Nyquist rate, as usually done in PCM, the resulting sampled signal is found to exhibit a high degree of *correlation* between adjacent samples. The meaning of this high correlation is that, in an average sense, the signal does not change rapidly from one sample to the next. As a result, the difference between adjacent samples has a variance that is smaller than the variance of the original signal. When these highly correlated samples are encoded, as in the standard PCM system, the resulting encoded signal contains *redundant information*. This kind of signal structure means that symbols that are not absolutely essential to the transmission of

information are generated as a result of the conventional encoding process described in Section 6.5. By reducing this redundancy before encoding, we obtain a *more efficient* coded signal, which is the basic idea behind DPCM. Discussion of this latter form of waveform coding is deferred to the next section. In this section we discuss prediction-error filtering, which provides a method for reduction and, therefore, improved waveform coding.

Theoretical Considerations

To elaborate, consider the block diagram of Figure 6.16a, which includes:

- a direct forward path from the input to the output;
- a predictor in the forward direction as well; and
- a comparator for computing the difference between the input signal and the predictor output.

The difference signal, so computed, is called the *prediction error*. Correspondingly, a filter that operates on the message signal to produce the prediction error, illustrated in Figure 6.16a, is called a *prediction-error filter*.

To simplify the presentation, let

$$m_n = m(nT_s) \tag{6.61}$$

denote a sample of the message signal $m(t)$ taken at time $t = nT_s$. Then, with $\hat{m}_n$ denoting the corresponding predictor output, the prediction error is defined by

$$e_n = m_n - \hat{m}_n \tag{6.62}$$

where e_n is the amount by which the predictor fails to predict the input sample m_n exactly. In any case, the objective is to design the predictor so as to *minimize the variance* of the prediction error e_n. In so doing, we effectively end up using a smaller number of bits to represent e_n than the original message sample m_n; hence, the need for a smaller transmission bandwidth.

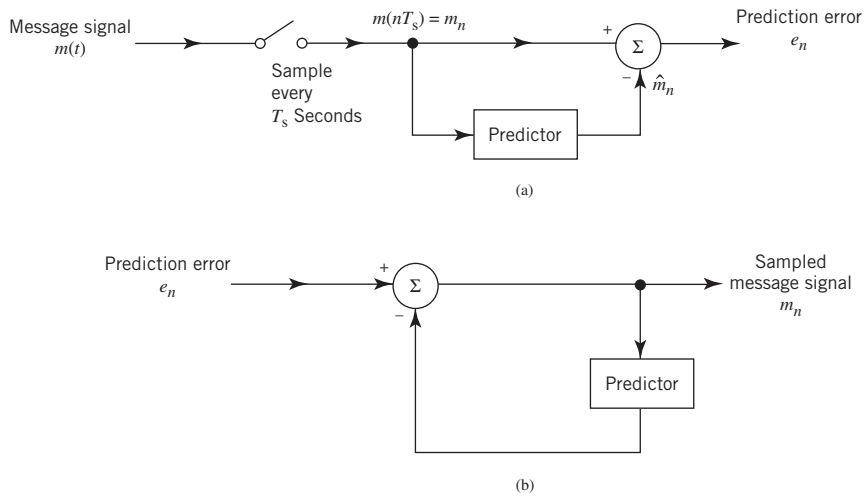


Figure 6.16 Block diagram of (a) prediction-error filter and (b) its inverse.

The prediction-error filter operates on the message signal on a sample-by-sample basis to produce the prediction error. With such an operation performed in the transmitter, how do we recover the original message signal from the prediction error at the receiver? To address this fundamental question in a simple-minded and yet practical way, we invoke the use of *linearity*. Let the *operator* $\mathbf{L}$ denote the action of the predictor, as shown by

$$\hat{m}_n = \mathbf{L}[m_n] \tag{6.63}$$

Accordingly, we may rewrite (6.62) in operator form as follows:

$$\begin{aligned} e_n &= m_n - \mathbf{L}[m_n] \\ &= (1 - \mathbf{L})[m_n] \end{aligned} \tag{6.64}$$

Under the assumption of linearity, we may invert (6.64) to recover the message sample from the prediction error, as shown by

$$m_n = \left(\frac{1}{1 - \mathbf{L}}\right)[e_n] \tag{6.65}$$

Equation (6.65) is immediately recognized as the equation of a *feedback system*, as illustrated in Figure 6.16b. Most importantly, in functional terms, this feedback system may be viewed as the *inverse of prediction-error filtering*.

Discrete-Time Structure for Prediction

To simplify the design of the linear predictor in Figure 6.16, we propose to use a discrete-time structure in the form of a *finite-duration impulse response (FIR) filter*, which is well known in the digital signal-processing literature. The FIR filter was briefly discussed in Chapter 2.

Figure 6.17 depicts an FIR filter, consisting of two functional components:

- a set of p *unit-delay elements*, each of which is represented by z^{-1}; and
- a corresponding set of *adders* used to sum the scaled versions of the delayed inputs,

$$m_{n-1}, m_{n-2}, \ldots, m_{n-p}.$$

The overall linearly predicted output is thus defined by the *convolution sum*

$$\hat{m}_n = \sum_{k=1}^{p} w_k m_{n-k} \tag{6.66}$$

where p is called the *prediction order*. Minimization of the prediction-error variance is achieved by a proper choice of the FIR filter-coefficients as described next.

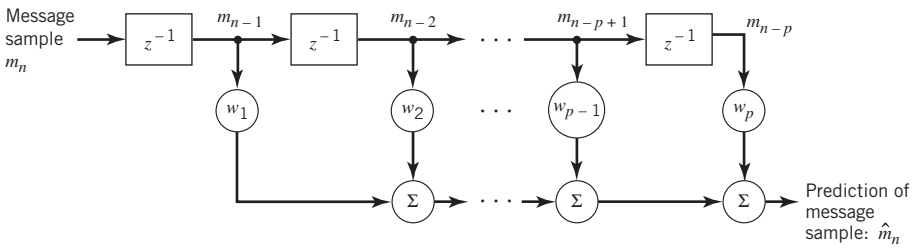


Figure 6.17 Block diagram of an FIR filter of order p.

First, however, we make the following assumption:

> The message signal $m(t)$ is drawn from a stationary stochastic processor $M(t)$ with zero mean.

This assumption may be satisfied by processing the message signal on a block-by-block basis, with each block being just long enough to satisfy the assumption in a *pseudo-stationary manner*. For example, a block duration of 40 ms is considered to be adequate for voice signals.

With the random variable M_n assumed to have zero mean, it follows that the variance of the prediction error e_n is the same as its mean-square value. We may thus define

$$J = \mathbb{E}[e^2(n)] \tag{6.67}$$

as the *index of performance*. Substituting (6.65) and (6.66) into (6.67) and then expanding terms, the index of performance is expressed as follows:

$$J(\mathbf{w}) = \mathbb{E}[m_n^2] - 2\sum_{k=1}^{p} w_k \mathbb{E}[m_n m_{n-k}] + \sum_{j=1}^{p}\sum_{k=1}^{p} w_j w_k \mathbb{E}[m_{n-j} m_{n-k}] \tag{6.68}$$

Moreover, under the above assumption of pseudo-stationarity, we may go on to introduce the following second-order statistical parameters for m_n treated as a sample of the stochastic process $M(t)$ at $t = nT_s$:

1. *Variance*

$$\begin{aligned}\sigma_M^2 &= \mathbb{E}[(m_n - \mathbb{E}[m_n])^2] \\ &= \mathbb{E}[m_n^2] \quad \text{for} \quad \mathbb{E}[m_n] = 0\end{aligned} \tag{6.69}$$

2. *Autocorrelation function*

$$R_{M,k-j} = \mathbb{E}[m_{n-j} m_{n-k}] \tag{6.70}$$

Note that to simplify the notation in (6.67) to (6.70), we have applied the expectation operator $\mathbb{E}$ to samples rather than the corresponding random variables.

In any event, using (6.69) and (6.70), we may reformulate the index of performance of (6.68) in the new form involving statistical parameters:

$$J(\mathbf{w}) = \sigma_M^2 - 2\sum_{k=1}^{p} w_k R_{M,k} + \sum_{j=1}^{p}\sum_{k=1}^{p} w_j w_k R_{M,k-j} \tag{6.71}$$

Differentiating this index of performance with respect to the filter coefficients, setting the resulting expression equal to zero, and then rearranging terms, we obtain the following system of simultaneous equations:

$$\sum_{j=1}^{p} w_{o,j} R_{M,k-j} = R_{M,k}, \qquad k = 1, 2, \ldots, p \tag{6.72}$$

where $w_{o,j}$ is the optimal value of the jth filter coefficient w_j. This optimal set of equations is the discrete-time version of the celebrated *Wiener–Hopf equations* for linear prediction.

With compactness of mathematical exposition in mind, we find it convenient to formulate the Wiener–Hopf equations in matrix form, as shown by

$$\mathbf{R}_{\mathrm{M}}\mathbf{w}_{\mathrm{o}} = \mathbf{r}_{\mathrm{M}} \tag{6.73}$$

where

$$\mathbf{w}_{\mathrm{o}} = [w_{\mathrm{o},1}, w_{\mathrm{o},2}, w_{\mathrm{o},p}]^{\mathrm{T}} \tag{6.74}$$

is the p-by-1 *optimum coefficient vector* of the FIR predictor,

$$\mathbf{r}_{\mathrm{M}} = [R_{\mathrm{M},1}, R_{\mathrm{M},2}, \ldots, R_{\mathrm{M},p}]^{\mathrm{T}} \tag{6.75}$$

is the p-by-1 *autocorrelation vector* of the original message signal, excluding the mean-square value represented by $\mathrm{R}_{\mathrm{M},0}$, and

$$\mathbf{R}_{\mathrm{M}} = \begin{bmatrix} R_{\mathrm{M},0} & R_{\mathrm{M},1} & \cdots & R_{\mathrm{M},p-1} \\ R_{\mathrm{M},1} & R_{\mathrm{M},0} & \cdots & R_{\mathrm{M},p-2} \\ \cdots & \cdots & \cdots & \cdots \\ R_{\mathrm{M},p-1} & R_{\mathrm{M},p-2} & \cdots & R_{\mathrm{M},0} \end{bmatrix} \tag{6.76}$$

is the p-by-y *correlation matrix* of the original message signal, including $R_{M,0}$.[6]

Careful examination of (6.76) reveals the *Toeplitz property* of the autocorrelation matrix $\mathbf{R}_{\mathrm{M}}$, which embodies two distinctive characteristics:

1. All the elements on the main diagonal of the matrix $\mathbf{R}_{\mathrm{M}}$ are equal to the mean-square value or, equivalently under the zero-mean assumption, the variance of the message sample m_n, as shown by
$$R_{\mathrm{M}}(0) = \sigma_{\mathrm{M}}^2$$
2. The matrix is *symmetric* about the main diagonal.

This Toeplitz property is a direct consequence of the assumption that message signal $m(t)$ is the sample function of a stationary stochastic process. From a practical perspective, the Toeplitz property of the autocorrelation matrix $\mathbf{R}_{\mathrm{M}}$ is important in that all of its elements are uniquely defined by the *autocorrelation sequence* $\{R_{\mathrm{M},k}\}_{k=0}^{p-1}$. Moreover, from the defining equation (6.75), it is clear that the autocorrelation vector $\mathbf{r}_{\mathrm{M}}$ is uniquely defined by the autocorrelation sequence $\{R_{\mathrm{M},k}\}_{k=1}^{p}$. We may therefore make the following statement:

> The p filter coefficients of the optimized linear predictor, configured in the form of an FIR filter, are uniquely defined by the variance $\sigma_{\mathrm{M}}^2 = R_{\mathrm{M}}(0)$ and the autocorrelation sequence $\{R_{\mathrm{M},k}\}_{k=0}^{p-1}$, which pertain to the message signal $m(t)$ drawn from a weakly stationary process.

Typically, we have

$$|R_{\mathrm{M},k}| < R_{\mathrm{M}}(0) \qquad \text{for } k = 1, 2, \ldots, p$$

Under this condition, we find that the autocorrelation matrix $\mathbf{R}_M$ is also invertible; that is, the inverse matrix $\mathbf{R}_M^{-1}$ exists. We may therefore solve (6.73) for the unknown value of the optimal coefficient vector $\mathbf{w}_o$ using the formula[7]

$$\mathbf{w}_o = \mathbf{R}_M^{-1}\mathbf{r}_M \tag{6.77}$$

Thus, given the variance σ_M^2 and autocorrelation sequence $\{R_{M,k}\}_{k=1}^{p}$, we may uniquely determine the optimized coefficient vector of the linear predictor, $\mathbf{w}_o$, defining an FIR filter of order p; and with it our design objective is satisfied.

To complete the linear prediction theory presented herein, we need to find the minimum mean-square value of prediction error, resulting from the use of the optimized predictor. We do this by first reformulating (6.71) in the matrix form:

$$J(\mathbf{w}_o) = \sigma_M^2 - 2\mathbf{w}_o^T\mathbf{r}_M + \mathbf{w}_o^T\mathbf{R}_M\mathbf{w}_o \tag{6.78}$$

where the superscript T denotes *matrix transposition*, $\mathbf{w}_o^T\mathbf{r}_M$ is the *inner product* of the p-by-1 vectors $\mathbf{w}_o$ and $\mathbf{r}_M$, and the matrix product $\mathbf{w}_o^T\mathbf{R}_M\mathbf{w}_o$ is a *quadratic form*. Then, substituting the optimum formula of (6.77) into (6.78), we find that the *minimum mean-square value of prediction error* is given by

$$\begin{aligned} J_{\min} &= \sigma_M^2 - 2(\mathbf{R}_M^{-1}\mathbf{r}_M)^T\mathbf{r}_M + (\mathbf{R}_M^{-1}\mathbf{r}_M)^T\mathbf{R}_M(\mathbf{R}_M^{-1}r_M) \\ &= \sigma_M^2 - 2\mathbf{r}_M^T\mathbf{R}_M^{-1}\mathbf{r}_M + \mathbf{r}_M^T\mathbf{R}_M^{-1}r_M \\ &= \sigma_M^2 - \mathbf{r}_M^T\mathbf{R}_M^{-1}\mathbf{r}_M \end{aligned} \tag{6.79}$$

where we have used the property that the autocorrelation matrix of a weakly stationary process is *symmetric*; that is,

$$\mathbf{R}_M^T = \mathbf{R}_M \tag{6.80}$$

By definition, the quadratic form $\mathbf{r}_M^T\mathbf{R}_M^{-1}\mathbf{r}_M$ is always positive. Accordingly, from (6.79) it follows that the minimum value of the mean-square prediction error $J_{\min}$ is always smaller than the variance σ_M^2 of the zero-mean message sample m_n that is being predicted. Through the use of linear prediction as described herein, we have thus satisfied the objective:

> To design a prediction-error filter the output of which has a smaller variance than the variance of the message sample applied to its input, we need to follow the optimum formula of (6.77).

This statement provides the rationale for going on to describe how the bandwidth requirement of the standard PCM can be reduced through redundancy reduction. However, before proceeding to do so, it is instructive that we consider an adaptive implementation of the linear predictor.

Linear Adaptive Prediction

The use of (6.77) for calculating the optimum weight vector of a linear predictor requires knowledge of the autocorrelation function $R_{m,k}$ of the message signal sequence $\{m_k\}_{k=0}^{p}$ where p is the prediction order. What if knowledge of this sequence is not available? In situations of this kind, which occur frequently in practice, we may resort to the use of an *adaptive predictor*.

The predictor is said to be adaptive in the following sense:

- Computation of the tap weights w_k, $k = 1, 2, \ldots, p$, proceeds in an iterative manner, starting from some arbitrary initial values of the tap weights.
- The algorithm used to adjust the tap weights (from one iteration to the next) is "self-designed," operating solely on the basis of available data.

The aim of the algorithm is to find the minimum point of the *bowl-shaped error surface* that describes the dependence of the cost function J on the tap weights. It is, therefore, intuitively reasonable that successive adjustments to the tap weights of the predictor be made in the direction of the steepest descent of the error surface; that is, in a direction opposite to the *gradient vector* whose elements are defined by

$$g_k = \frac{\partial J}{\partial w_k}, \qquad k = 1, 2, \ldots, p \tag{6.81}$$

This is indeed the idea behind the *method of deepest descent*. Let $w_{k,n}$ denote the value of the kth tap weight at iteration n. Then, the updated value of this weight at iteration $n + 1$ is defined by

$$w_{k,n+1} = w_{k,n} - \frac{1}{2}\mu g_k, \qquad k = 1, 2, \ldots, p \tag{6.82}$$

where μ is a *step-size parameter* that controls the speed of adaptation and the factor 1/2 is included for convenience of presentation. Differentiating the cost function J of (6.68) with respect to w_k, we readily find that

$$g_k = -2\mathbb{E}[m_n m_{n-k}] + \sum_{j=1}^{p} w_j \mathbb{E}[m_{n-j} m_{n-k}] \tag{6.83}$$

From a practical perspective, the formula for the gradient g_k in (6.83) could do with further simplification that ignores the expectation operator. In effect, *instantaneous values are used as estimates of autocorrelation functions*. The motivation for this simplification is to permit the adaptive process to proceed forward on a step-by-step basis in a self-organized manner. Clearly, by ignoring the expectation operator in (6.83), the gradient g_k takes on a time-dependent value, denoted by $g_{k,n}$. We may thus write

$$g_{k,n} = -2m_n m_{n-k} + 2m_{n-k} \sum_{j=1}^{p} \hat{w}_{j,n} m_{n,j}, \qquad k = 1, 2, \ldots, p \tag{6.84}$$

where $\hat{w}_{j,n}$ is an estimate of the filter coefficient $w_{j,n}$ at time n.

The stage is now set for substituting (6.84) into (6.82), where in the latter equation $\hat{w}_{k,n}$ is substituted for $w_{k,n}$; this change is made to account for dispensing with the expectation operator:

$$\begin{aligned}\hat{w}_{k,n+1} &= \hat{w}_{k,n} - \frac{1}{2}\mu g_{k,n} \\ &= \hat{w}_{k,n} + \mu\left(m_n m_{n-k} - \sum_{j=1}^{p} \hat{w}_{j,n} m_{n-j} m_{n-k}\right) \\ &= \hat{w}_{k,n} + \mu m_{n-k}\left(m_n - \sum_{j=1}^{p} \hat{w}_{j,n} m_{n-j}\right) \\ &= \hat{w}_{k,n} + \mu m_{n-k} e_n\end{aligned} \tag{6.85}$$

where e_n is the *new prediction error* defined by

$$e_n = m_n - \sum_{j=1}^{p} \hat{w}_{j,n} m_{n-j} \tag{6.86}$$

Note that the current value of the message signal, m_n, plays a role as the *desired response for predicting* the value of m_n given the past values of the message signal: m_{n-1}, m_{n-2}, …, m_{n-p}.

In words, we may express the adaptive filtering algorithm of (6.85) as follows:

$$\begin{pmatrix}\text{Updated value of the } k\text{th} \\ \text{filter coefficient at time } n+1\end{pmatrix} = \begin{pmatrix}\text{Old value of the same} \\ \text{filter coefficient at time } n\end{pmatrix} + \begin{pmatrix}\text{Step-size} \\ \text{parameter}\end{pmatrix} \times \begin{pmatrix}\text{Message signal } m_n \\ \text{delayed by } k \text{ time steps}\end{pmatrix} \begin{pmatrix}\text{Prediction error} \\ \text{computed at time } n\end{pmatrix}$$

The algorithm just described is the popular *least-mean-square (LMS) algorithm*, formulated for the purpose of linear prediction. The reason for popularity of this adaptive filtering algorithm is the simplicity of its implementation. In particular, the computational complexity of the algorithm, measured in terms of the number of additions and multiplications, is *linear* in the prediction order p. Moreover, the algorithm is not only *computationally efficient* but it is also *effective in performance*.

The LMS algorithm is a *stochastic* adaptive filtering algorithm, stochastic in the sense that, starting from the *initial condition* defined by $\{w_{k,0}\}_{k=1}^{p}$, it seeks to find the minimum point of the error surface by following a zig-zag path. However, it never finds this minimum point exactly. Rather, it continues to execute a random motion around the minimum point of the error surface (Haykin, 2013).

6.8 Differential Pulse-Code Modulation

DPCM, the scheme to be considered for *channel-bandwidth conservation*, exploits the idea of linear prediction theory with a practical difference:

> In the transmitter, the linear prediction is performed on a quantized version of the message sample instead of the message sample itself, as illustrated in Figure 6.18.

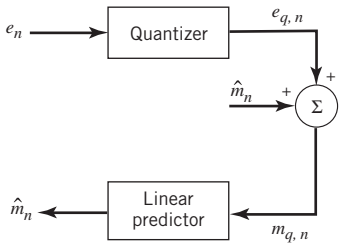


Figure 6.18 Block diagram of a differential quantizer.

The resulting process is referred to as *differential quantization*. The motivation behind the use of differential quantization follows from two practical considerations:

1. Waveform encoding in the transmitter requires the use of quantization.
2. Waveform decoding in the receiver, therefore, has to process a quantized signal.

In order to cater to both requirements in such a way that the *same structure* is used for predictors in both the transmitter and the receiver, the transmitter has to perform prediction-error filtering on the quantized version of the message signal rather than the signal itself, as shown in Figure 6.19a. Then, assuming a noise-free channel, the predictors in the transmitter and receiver operate on exactly the same sequence of quantized message samples.

To demonstrate this highly desirable and distinctive characteristic of differential PCM, we see from Figure 6.19a that

$$e_{q,n} = e_n + q_n \tag{6.87}$$

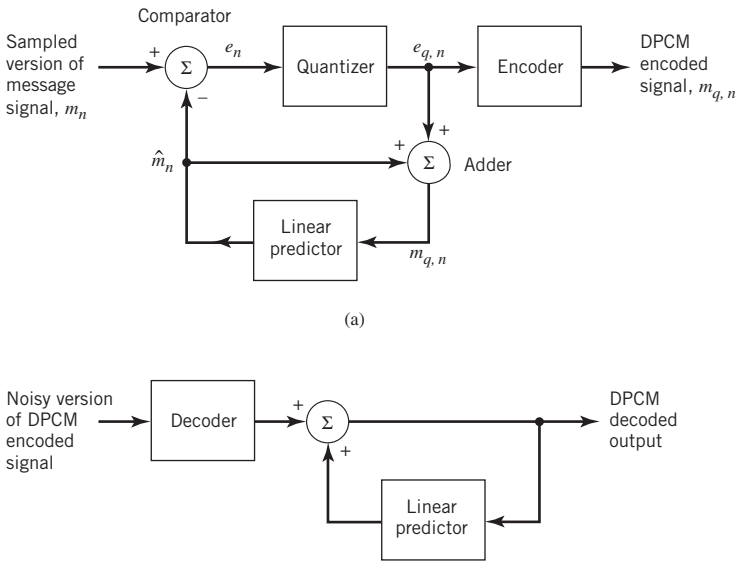


Figure 6.19 DPCM system: (a) transmitter; (b) receiver.

where q_n is the quantization noise produced by the quantizer operating on the prediction error e_n. Moreover, from Figure 6.19a, we readily see that

$$m_{q,n} = \hat{m}_n + e_{q,n} \tag{6.88}$$

where $\hat{m}_n$ is the predicted value of the original message sample m_n; thus, (6.88) is in perfect agreement with Figure 6.18. Hence, the use of (6.87) in (6.88) yields

$$m_{q,n} = \hat{m}_n + e_n + q_n \tag{6.89}$$

We may now invoke (6.88) of linear prediction theory to rewrite (6.89) in the equivalent form:

$$m_{q,n} = m_n + q_n \tag{6.90}$$

which describes a quantized version of the original message sample m_n.

With the differential quantization scheme of Figure 6.19a at hand, we may now expand on the structures of the transmitter and receiver of DPCM.

DPCM Transmitter

Operation of the DPCM transmitter proceeds as follows:

1. Given the predicted message sample $\hat{m}_n$, the comparator at the transmitter input computes the prediction error e_n, which is quantized to produce the quantized version of e_n in accordance with (6.87).
2. With $\hat{m}_n$ and $e_{q,n}$ at hand, the adder in the transmitter produces the quantized version of the original message sample m_n, namely $m_{q,n}$, in accordance with (6.88).
3. The required one-step prediction $\hat{m}_n$ is produced by applying the sequence of quantized samples $\{m_{q,k}\}_{k=1}^{p}$ to a linear FIR predictor of order p.

This multistage operation is clearly *cyclic*, encompassing three steps that are repeated at each time step n. Moreover, at each time step, the encoder operates on the quantized prediction error $e_{q,n}$ to produce the DPCM-encoded version of the original message sample m_n. The DPCM code so produced is a *lossy-compressed* version of the PCM code; it is "lossy" because of the prediction error.

DPCM Receiver

The structure of the receiver is much simpler than that of the transmitter, as depicted in Figure 6.19b. Specifically, first, the decoder reconstructs the quantized version of the prediction error, namely $e_{q,n}$. An estimate of the original message sample m_n is then computed by applying the decoder output to the same predictor used in the transmitter of Figure 6.19a. In the absence of channel noise, the encoded signal at the receiver input is identical to the encoded signal at the transmitter output. Under this ideal condition, we find that the corresponding receiver output is equal to $m_{q,n}$, which differs from the original signal sample m_n only by the quantization error q_n incurred as a result of quantizing the prediction error e_n.

From the foregoing analysis, we thus observe that, in a noise-free environment, the linear predictors in the transmitter and receiver of DPCM operate on the same sequence of samples, $m_{q,n}$. It is with this point in mind that a feedback path is appended to the quantizer in the transmitter of Figure 6.19a.

Processing Gain

The output SNR of the DPCM system, shown in Figure 6.19, is, by definition,

$$(\text{SNR})_{\text{O}} = \frac{\sigma_{\text{M}}^2}{\sigma_{\text{Q}}^2} \tag{6.91}$$

where σ_{M}^2 is the variance of the original signal sample m_n, assumed to be of zero mean, and σ_{Q}^2 is the variance of the quantization error q_n, also of zero mean. We may rewrite (6.91) as the product of two factors, as shown by

$$\begin{aligned}(\text{SNR})_{\text{O}} &= \left(\frac{\sigma_{\text{M}}^2}{\sigma_{\text{E}}^2}\right)\left(\frac{\sigma_{\text{E}}^2}{\sigma_{\text{Q}}^2}\right) \\ &= G_{\text{p}}(\text{SNR})_{\text{Q}}\end{aligned} \tag{6.92}$$

where, in the first line, σ_{E}^2 is the variance of the prediction error e_n. The factor $(\text{SNR})_{\text{Q}}$ introduced in the second line is the *signal-to-quantization noise ratio*, which is itself defined by

$$(\text{SNR})_{\text{Q}} = \frac{\sigma_{\text{E}}^2}{\sigma_{\text{Q}}^2} \tag{6.93}$$

The other factor G_{p} is the *processing gain* produced by the differential quantization scheme; it is formally defined by

$$G_{\text{p}} = \frac{\sigma_{\text{M}}^2}{\sigma_{\text{E}}^2} \tag{6.94}$$

The quantity G_{p}, when it is greater than unity, represents a *gain in signal-to-noise ratio*, which is due to the differential quantization scheme of Figure 6.19. Now, for a given message signal, the variance σ_{M}^2 is fixed, so that G_{p} is maximized by minimizing the variance σ_{M}^2 of the prediction error e_n. Accordingly, the objective in implementing the DPCM should be to design the prediction filter so as to minimize the prediction-error variance, σ_{E}^2.

In the case of voice signals, it is found that the optimum signal-to-quantization noise advantage of the DPCM over the standard PCM is in the neighborhood of 4–11 dB. Based on experimental studies, it appears that the greatest improvement occurs in going from no prediction to first-order prediction, with some additional gain resulting from increasing the order p of the prediction filter up to 4 or 5, after which little additional gain is obtained. Since 6 dB of quantization noise is equivalent to 1 bit per sample by virtue of the results presented in Table 6.1 for sinusoidal modulation, the advantage of DPCM may also be expressed in terms of bit rate. For a constant signal-to-quantization noise ratio, and assuming a sampling rate of 8 kHz, the use of DPCM may provide a saving of about 8–16 kHz (i.e., 1 to 2 bits per sample) compared with the standard PCM.

6.9 Delta Modulation

In choosing DPCM for waveform coding, we are, in effect, economizing on transmission bandwidth by increasing system complexity, compared with standard PCM. In other words, DPCM exploits the *complexity–bandwidth tradeoff*. However, in practice, the need may arise for reduced system complexity compared with the standard PCM. To achieve this other objective, transmission bandwidth is traded off for reduced system complexity, which is precisely the motivation behind DM. Thus, whereas DPCM exploits the *complexity–bandwidth tradeoff*, DM exploits the *bandwidth–complexity tradeoff*. We may, therefore, differentiate between the standard PCM, the DPCM, and the DM along the lines described in Figure 6.20. With the bandwidth–complexity tradeoff being at the heart of DM, the incoming message signal $m(t)$ is *oversampled*, which requires the use of a sampling rate higher than the Nyquist rate. Accordingly, the correlation between adjacent samples of the message signal is purposely increased so as to permit the use of a *simple* quantizing strategy for constructing the encoded signal.

DM Transmitter

In the DM transmitter, system complexity is reduced to the minimum possible by using the combination of two strategies:

1. *Single-bit quantizer*, which is the simplest quantizing strategy; as depicted in Figure 6.21, the quantizer acts as a hard limiter with only two decision levels, namely, $\pm\Delta$.
2. *Single unit-delay element*, which is the most primitive form of a predictor; in other words, the only component retained in the FIR predictor of Figure 6.17 is the front-end block labeled z^{-1}, which acts as an *accumulator*.

Thus, replacing the multilevel quantizer and the FIR predictor in the DPCM transmitter of Figure 6.19a in the manner described under points 1 and 2, respectively, we obtain the block diagram of Figure 6.21a for the DM transmitter.

From this figure, we may express the equations underlying the operation of the DM transmitter by the following set of equations (6.95)–(6.97):

$$\begin{aligned} e_n &= m_n - \hat{m}_n \\ &= m_n - m_{q,\,n-1} \end{aligned} \tag{6.95}$$

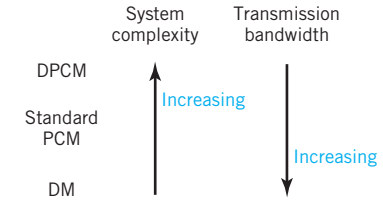


Figure 6.20 Illustrating the tradeoffs between standard PCM, DPCM, and DM.

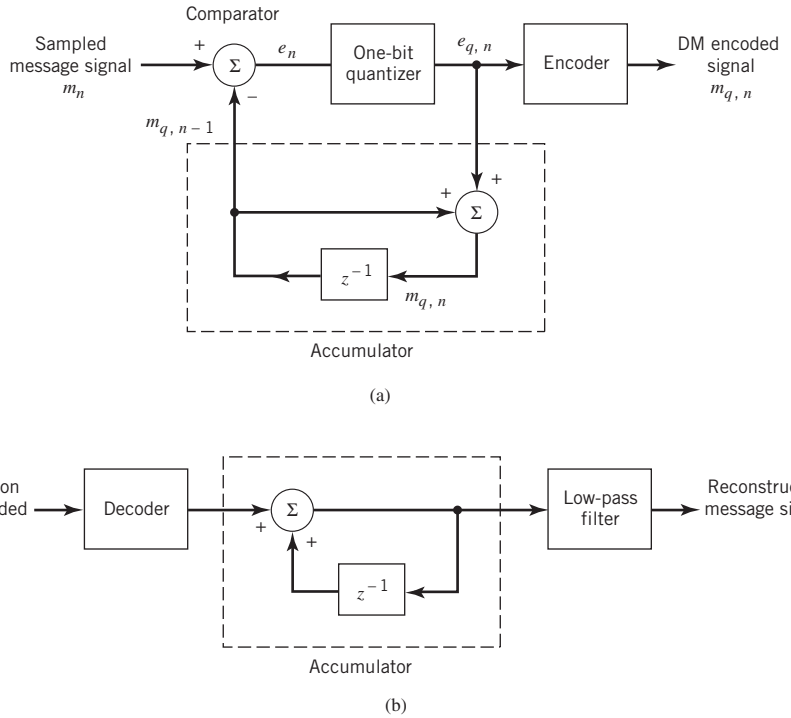


Figure 6.21 DM system: (a) transmitter; (b) receiver.

$$e_{q,n} = \Delta \operatorname{sgn}[e_n]$$

$$= \begin{cases} +\Delta \text{ if } e_n > 0 \\ -\Delta \text{ if } e_n < 0 \end{cases} \tag{6.96}$$

$$m_{q,n} = m_{q,n-1} + e_{q,n} \tag{6.97}$$

According to (6.95) and (6.96), two possibilities may naturally occur:

1. The error signal e_n (i.e., the difference between the message sample m_n and its approximation $\hat{m}_n$) is positive, in which case the approximation $\hat{m}_n = m_{q,n-1}$ is increased by the amount Δ; in this first case, the encoder sends out symbol 1.
2. The error signal e_n is negative, in which case the approximation $\hat{m}_n = m_{q,n-1}$ is reduced by the amount Δ; in this second case, the encoder sends out symbol 0.

From this description it is apparent that the delta modulator produces a staircase approximation to the message signal, as illustrated in Figure 6.22a. Moreover, the rate of data transmission in DM is equal to the sampling rate $f_s = 1/T_s$, as illustrated in the binary sequence of Figure 6.22b.

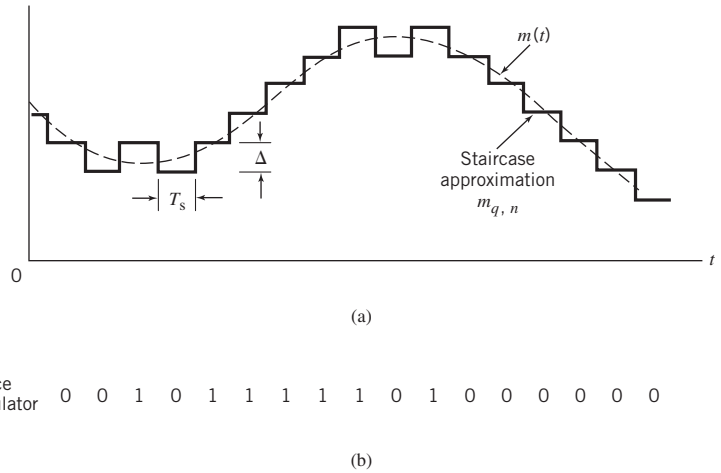


Figure 6.22 Illustration of DM.

DM Receiver

Following a procedure similar to the way in which we constructed the DM transmitter of Figure 6.21a, we may construct the DM receiver of Figure 6.21b as a special case of the DPCM receiver of Figure 6.19b. Working through the operation of the DM receiver, we find that reconstruction of the staircase approximation to the original message signal is achieved by passing the sequence of positive and negative pulses (representing symbols 1 and 0, respectively) through the block labeled "accumulator."

Under the assumption that the channel is distortionless, the accumulated output is the desired $m_{q,n}$ given that the decoded channel output is $e_{q,n}$. The out-of-band quantization noise in the high-frequency staircase waveform in the accumulator output is suppressed by passing it through a low-pass filter with a cutoff frequency equal to the message bandwidth.

Quantization Errors in DM

DM is subject to two types of quantization error: slope overload distortion and granular noise. We will discuss the case of slope overload distortion first.

Starting with (6.97), we observe that this equation is the *digital equivalent of integration*, in the sense that it represents the *accumulation* of positive and negative increments of magnitude Δ. Moreover, denoting the quantization error applied to the message sample m_n by q_n, we may express the quantized message sample as

$$m_{q,n} = m_n + q_n \tag{6.98}$$

With this expression for $m_{q,n}$ at hand, we find from (6.98) that the quantizer input is

$$e_n = m_n - (m_{n-1} + q_{n-1}) \tag{6.99}$$

Thus, except for the delayed quantization error q_{n-1}, the quantizer input is a *first backward difference* of the original message sample. This difference may be viewed as a

digital approximation to the quantizer input or, equivalently, as the *inverse* of the digital integration process carried out in the DM transmitter. If, then, we consider the maximum slope of the original message signal $m(t)$, it is clear that in order for the sequence of samples $\{m_{q,n}\}$ to increase as fast as the sequence of message samples $\{m_n\}$ in a region of maximum slope of $m(t)$, we require that the condition

$$\frac{\Delta}{T_s} \geq \max\left|\frac{dm(t)}{dt}\right| \tag{6.100}$$

be satisfied. Otherwise, we find that the step-size Δ is too small for the staircase approximation $m_q(t)$ to follow a steep segment of the message signal $m(t)$, with the result that $m_q(t)$ falls behind $m(t)$, as illustrated in Figure 6.23. This condition is called *slope overload*, and the resulting quantization error is called *slope-overload distortion* (*noise*). Note that since the maximum slope of the staircase approximation $m_q(t)$ is fixed by the step size Δ, increases and decreases in $m_q(t)$ tend to occur along straight lines. For this reason, a delta modulator using a fixed step size is often referred to as a *linear delta modulator.*

In contrast to slope-overload distortion, *granular noise* occurs when the step size Δ is too large relative to the local slope characteristics of the message signal $m(t)$, thereby causing the staircase approximation $m_q(t)$ to hunt around a relatively flat segment of $m(t)$; this phenomenon is also illustrated in the tail end of Figure 6.23. Granular noise is analogous to quantization noise in a PCM system.

Adaptive DM

From the discussion just presented, it is appropriate that we need to have a large step size to accommodate a wide dynamic range, whereas a small step size is required for the accurate representation of relatively low-level signals. It is clear, therefore, that the choice of the optimum step size that minimizes the mean-square value of the quantization error in a linear delta modulator will be the result of a compromise between slope-overload distortion and granular noise. To satisfy such a requirement, we need to make the delta modulator "adaptive," in the sense that the step size is made to vary in accordance with the input signal. The step size is thereby made variable, such that it is enlarged during intervals when the slope-overload distortion is dominant and reduced in value when the granular (quantization) noise is dominant.

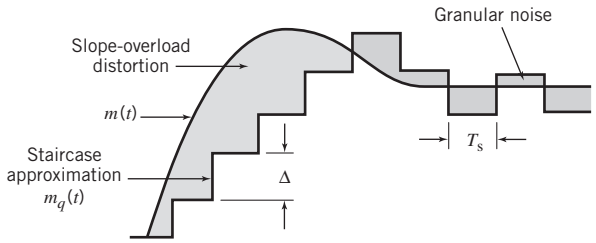


Figure 6.23 Illustration of the two different forms of quantization error in DM.

6.10 Line Codes

In this chapter, we have described three basic waveform-coding schemes: PCM, DPCM, and DM. Naturally, they differ from each other in several ways: transmission–bandwidth requirement, transmitter–receiver structural composition and complexity, and quantization noise. Nevertheless, all three of them have a common need: *line codes* for electrical representation of the encoded binary streams produced by their individual transmitters, so as to facilitate transmission of the binary streams across the communication channel.

Figure 6.24 displays the waveforms of five important line codes for the example data stream 01101001. Figure 6.25 displays their individual power spectra (for positive frequencies) for randomly generated binary data, assuming that first, symbols 0 and 1 are equiprobable, second, the average power is normalized to unity, and third, the frequency f is normalized with respect to the bit rate $1/T_b$. In what follows, we describe the five line codes involved in generating the coded waveforms of Figure 6.24.

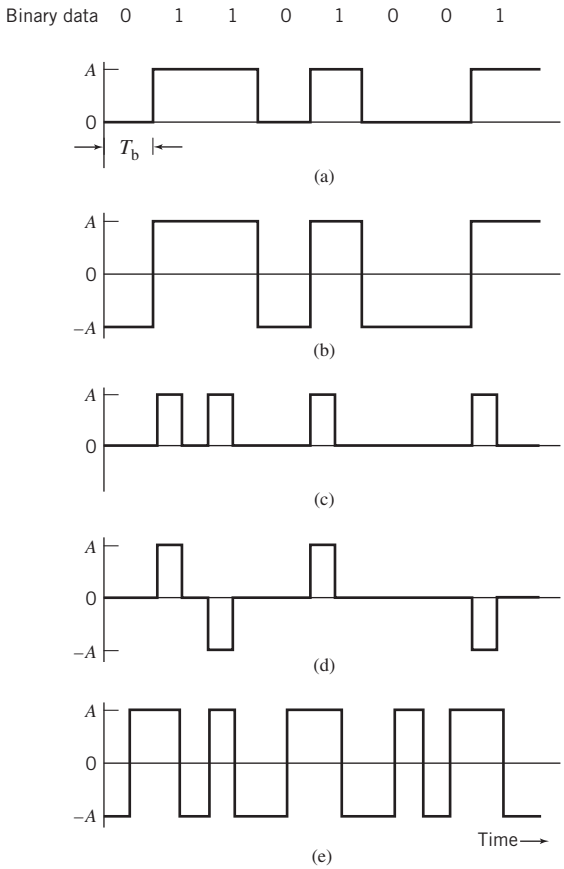


Figure 6.24 Line codes for the electrical representations of binary data: (a) unipolar nonreturn-to-zero (NRZ) signaling; (b) polar NRZ signaling; (c) unipolar return-to-zero (RZ) signaling; (d) bipolar RZ signaling; (e) split-phase or Manchester code.

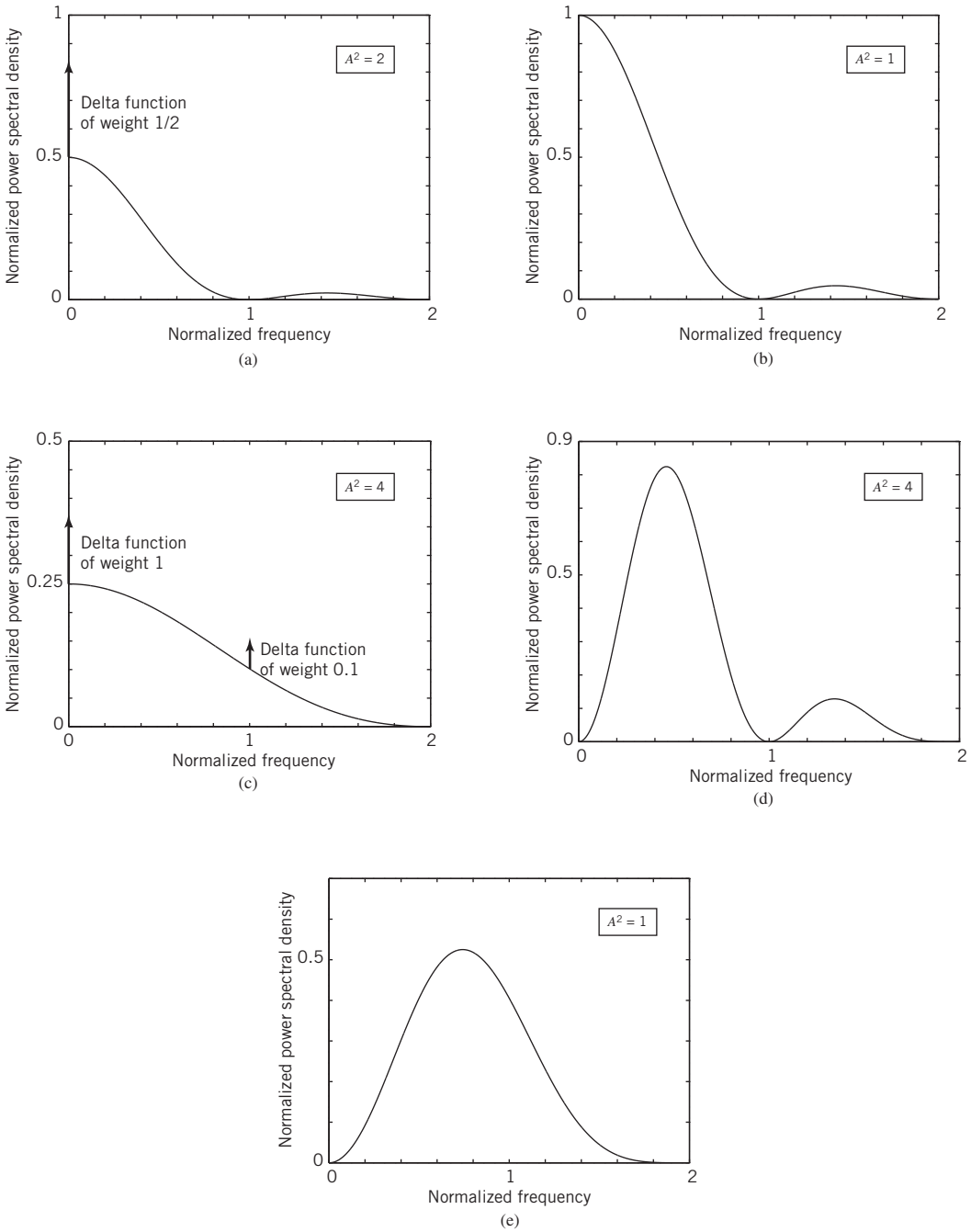


Figure 6.25 Power spectra of line codes: (a) unipolar NRZ signal; (b) polar NRZ signal; (c) unipolar RZ signal; (d) bipolar RZ signal; (e) Manchester-encoded signal. The frequency is normalized with respect to the bit rate $1/T_b$, and the average power is normalized to unity.

Unipolar NRZ Signaling

In this line code, symbol 1 is represented by transmitting a pulse of amplitude A for the duration of the symbol, and symbol 0 is represented by switching off the pulse, as in Figure 6.24a. The unipolar NRZ line code is also referred to as *on–off signaling*. Disadvantages of on–off signaling are the waste of power due to the transmitted DC level and the fact that the power spectrum of the transmitted signal does not approach zero at zero frequency.

Polar NRZ Signaling

In this second line code, symbols 1 and 0 are represented by transmitting pulses of amplitudes $+A$ and $-A$, respectively, as illustrated in Figure 6.24b. The polar NRZ line code is relatively easy to generate, but its disadvantage is that the power spectrum of the signal is large near zero frequency.

Unipolar RZ Signaling

In this third line code, symbol 1 is represented by a rectangular pulse of amplitude A and half-symbol width and symbol 0 is represented by transmitting *no* pulse, as illustrated in Figure 6.24c. An attractive feature of the unipolar RZ line code is the presence of delta functions at $f = 0, \pm 1/T_b$ in the power spectrum of the transmitted signal; the delta functions can be used for *bit-timing recovery* at the receiver. However, its disadvantage is that it requires 3 dB more power than polar RZ signaling for the same probability of symbol error.

Bipolar RZ Signaling

This line code uses three amplitude levels, as indicated in Figure 6.24(d). Specifically, positive and negative pulses of equal amplitude (i.e., $+A$ and $-A$) are used alternately for symbol 1, with each pulse having a half-symbol width; no pulse is always used for symbol 0. A useful property of the bipolar RZ signaling is that the power spectrum of the transmitted signal has no DC component and relatively insignificant low-frequency components for the case when symbols 1 and 0 occur with equal probability. The bipolar RZ line code is also called *alternate mark inversion* (AMI) signaling.

Split-Phase (Manchester Code)

In this final method of signaling, illustrated in Figure 6.24e, symbol 1 is represented by a positive pulse of amplitude A followed by a negative pulse of amplitude $-A$, with both pulses being half-symbol wide. For symbol 0, the polarities of these two pulses are reversed. A unique property of the Manchester code is that it suppresses the DC component and has relatively insignificant low-frequency components, *regardless of the signal statistics*. This property is essential in some applications.

6.11 Summary and Discussion

In this chapter we introduced two fundamental and complementary processes:

- *Sampling*, which operates in the time domain; the sampling process is the link between an analog waveform and its discrete-time representation.
- *Quantization*, which operates in the amplitude domain; the quantization process is the link between an analog waveform and its discrete-amplitude representation.

The sampling process builds on the *sampling theorem*, which states that a strictly band-limited signal with no frequency components higher than W Hz is represented uniquely by a sequence of samples taken at a uniform rate equal to or greater than the Nyquist rate of $2W$ samples per second. The quantization process exploits the fact that any human sense, as ultimate receiver, can only detect finite intensity differences.

The sampling process is basic to the operation of all pulse modulation systems, which may be classified into analog pulse modulation and digital pulse modulation. The distinguishing feature between them is that analog pulse modulation systems maintain a continuous amplitude representation of the message signal, whereas digital pulse modulation systems also employ quantization to provide a representation of the message signal that is discrete in both time and amplitude.

Analog pulse modulation results from varying some parameter of the transmitted pulses, such as amplitude, duration, or position, in which case we speak of PAM, pulse-duration modulation, or pulse-position modulation, respectively. In this chapter we focused on PAM, as it is used in all forms of digital pulse modulation.

Digital pulse modulation systems transmit analog message signals as a sequence of coded pulses, which is made possible through the combined use of sampling and quantization. PCM is an important form of digital pulse modulation that is endowed with some unique system advantages, which, in turn, have made it the standard method of modulation for the transmission of such analog signals as voice and video signals. The advantages of PCM include robustness to noise and interference, efficient regeneration of the coded pulses along the transmission path, and a uniform format for different kinds of baseband signals.

Indeed, it is because of this list of advantages unique to PCM that it has become the method of choice for the construction of public switched telephone networks (PSTNs). In this context, the reader should carefully note that the telephone channel viewed from the PSTN by an Internet service provider, for example, is *nonlinear* due to the use of companding and, most importantly, it is *entirely digital*. This observation has a significant impact on the design of high-speed modems for communications between a computer user and server, which will be discussed in Chapter 8.

DM and DPCM are two other useful forms of digital pulse modulation. The principal advantage of DM is the simplicity of its circuitry, which is achieved at the expense of increased transmission bandwidth. In contrast, DPCM employs increased circuit complexity to reduce channel bandwidth. The improvement is achieved by using the idea of prediction to reduce redundant symbols from an incoming data stream. A further improvement in the operation of DPCM can be made through the use of adaptivity to account for statistical variations in the input data. By so doing, bandwidth requirement may be reduced significantly without serious degradation in system performance.[8]

Problems

Sampling Process

6.1 In natural sampling, an analog signal $g(t)$ is multiplied by a periodic train of rectangular pulses $c(t)$, each of unit area. Given that the pulse repetition frequency of this periodic train is f_s and the duration of each rectangular pulse is T (with $f_s T \ll 1$), do the following:

a. Find the spectrum of the signal $s(t)$ that results from the use of natural sampling; you may assume that time $t = 0$ corresponds to the midpoint of a rectangular pulse in $c(t)$.

b. Show that the original signal $g(t)$ may be recovered exactly from its naturally sampled version, provided that the conditions embodied in the sampling theorem are satisfied.

6.2 Specify the Nyquist rate and the Nyquist interval for each of the following signals:

a. $g(t) = \text{sinc}(200t)$.

b. $g(t) = \text{sinc}^2(200t)$.

c. $g(t) = \text{sinc}(200t) + \text{sinc}^2(200t)$.

6.3 Discussion of the sampling theorem presented in Section 6.2 was confined to the time domain. Describe how the sampling theorem can be applied in the frequency domain.

Pulse-Amplitude Modulation

6.4 Figure P6.4 shows the idealized spectrum of a message signal $m(t)$. The signal is sampled at a rate equal to 1 kHz using flat-top pulses, with each pulse being of unit amplitude and duration 0.1 ms. Determine and sketch the spectrum of the resulting PAM signal.

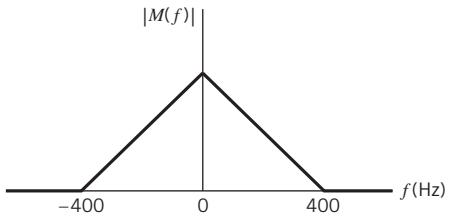


Figure P6.4

6.5 In this problem, we evaluate the equalization needed for the aperture effect in a PAM system. The operating frequency $f = f_s/2$, which corresponds to the highest frequency component of the message signal for a sampling rate equal to the Nyquist rate. Plot $1/\text{sinc}(0.5T/T_s)$ versus T/T_s, and hence find the equalization needed when $T/T_s = 0.1$.

6.6 Consider a PAM wave transmitted through a channel with white Gaussian noise and minimum bandwidth $B_T = 1/2T_s$, where T_s is the sampling period. The noise is of zero mean and power spectral density $N_0/2$. The PAM signal uses a standard pulse $g(t)$ with its Fourier transform defined by

$$G(f) = \begin{cases} \dfrac{1}{2B_T}, & |f| < B_T \\ 0, & |f| > B_T \end{cases}$$

By considering a full-load sinusoidal modulating wave, show that PAM and baseband-signal transmission have equal SNRs for the same average transmitted power.

6.7 Twenty-four voice signals are sampled uniformly and then time-division multiplexed (TDM). The sampling operation uses flat-top samples with 1 μs duration. The multiplexing operation includes

provision for synchronization by adding an extra pulse of sufficient amplitude and also 1 μs duration. The highest frequency component of each voice signal is 3.4 kHz.

a. Assuming a sampling rate of 8 kHz, calculate the spacing between successive pulses of the multiplexed signal.

b. Repeat your calculation assuming the use of Nyquist rate sampling.

6.8 Twelve different message signals, each with a bandwidth of 10 kHz, are to be multiplexed and transmitted. Determine the minimum bandwidth required if the multiplexing/modulation method used is time-division multiplexing (TDM), which was discussed in Chapter 1.

Pulse-Code Modulation

6.9 A speech signal has a total duration of 10 s. It is sampled at the rate of 8 kHz and then encoded. The signal-to-(quantization) noise ratio is required to be 40 dB. Calculate the minimum storage capacity needed to accommodate this digitized speech signal.

6.10 Consider a uniform quantizer characterized by the input-output relation illustrated in Figure 6.9a. Assume that a Gaussian-distributed random variable with zero mean and unit variance is applied to this quantizer input.

a. What is the probability that the amplitude of the input lies outside the range –4 to +4?

b. Using the result of part a, show that the output SNR of the quantizer is given by

$$(\text{SNR})_{\text{O}} = 6R - 7.2 \text{ dB}$$

where R is the number of bits per sample. Specifically, you may assume that the quantizer input extends from –4 to +4. Compare the result of part b with that obtained in Example 2.

6.11 A PCM system uses a uniform quantizer followed by a 7-bit binary encoder. The bit rate of the system is equal to 50×10^6 bits/s.

a. What is the maximum message bandwidth for which the system operates satisfactorily?

b. Determine the output signal-to-(quantization) noise when a full-load sinusoidal modulating wave of frequency 1 MHz is applied to the input.

6.12 Show that with a nonuniform quantizer the mean-square value of the quantization error is approximately equal to $(1/12)\Sigma_i \Delta_i^2 p_i$, where Δ_i is the ith step size and p_i is the probability that the input signal amplitude lies within the ith interval. Assume that the step size Δ_i is small compared with the excursion of the input signal.

6.13 a. A sinusoidal signal with an amplitude of 3.25 V is applied to a uniform quantizer of the midtread type whose output takes on the values 0, ±1, ±2, ±3 V. Sketch the waveform of the resulting quantizer output for one complete cycle of the input.

b. Repeat this evaluation for the case when the quantizer is of the midrise type whose output takes on the values 0.5, ±1.5, ±2.5, ±3.5 V.

6.14 The signal

$$m(t) \text{ (volts)} = 6\sin(2\pi t)$$

is transmitted using a 40-bit binary PCM system. The quantizer is of the midrise type, with a step size of 1 V. Sketch the resulting PCM wave for one complete cycle of the input. Assume a sampling rate of four samples per second, with samples taken at t(s) = ±1/8, ±3/8, ±5/8, …

6.15 Figure P6.15 shows a PCM signal in which the amplitude levels of +1 V and –1 V are used to represent binary symbols 1 and 0, respectively. The codeword used consists of three bits. Find the sampled version of an analog signal from which this PCM signal is derived.

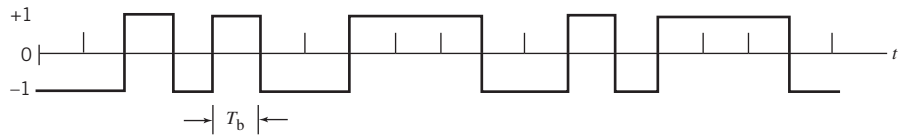


Figure P6.15

6.16 Consider a chain of $(n-1)$ regenerative repeaters, with a total of n sequential decisions made on a binary PCM wave, including the final decision made at the receiver. Assume that any binary symbol transmitted through the system has an independent probability p_1 of being inverted by any repeater. Let p_n represent the probability that a binary symbol is in error after transmission through the complete system.

a. Show that

$$p_n = \frac{1}{2}[1-(1-2p_1)^n]$$

b. If p_1 is very small and n is not too large, what is the corresponding value of p_n?

6.17 Discuss the basic issues involved in the design of a regenerative repeater for PCM.

Linear Prediction

6.18 A one-step linear predictor operates on the sampled version of a sinusoidal signal. The sampling rate is equal to $10f_0$, where f_0 is the frequency of the sinusoid. The predictor has a single coefficient denoted by w_1.

a. Determine the optimum value of w_1 required to minimize the prediction-error variance.

b. Determine the minimum value of the prediction error variance.

6.19 A stationary process $X(t)$ has the following values for its autocorrelation function:

$$R_X(0) = 1$$
$$R_X(0) = 0.8$$
$$R_X(0) = 0.6$$
$$R_X(0) = 0.4$$

a. Calculate the coefficients of an optimum linear predictor involving the use of three unit-time delays.

b. Calculate the variance of the resulting prediction error.

6.20 Repeat the calculations of Problem 6.19, but this time use a linear predictor with two unit-time delays. Compare the performance of this second optimum linear predictor with that considered in Problem 6.19.

Differential Pulse-Code Modulation

6.21 A DPCM system uses a linear predictor with a single tap. The normalized autocorrelation function of the input signal for a lag of one sampling interval is 0.75. The predictor is designed to minimize the prediction-error variance. Determine the processing gain attained by the use of this predictor.

6.22 Calculate the improvement in processing gain of a DPCM system using the optimized three-tap linear predictor. For this calculation, use the autocorrelation function values of the input signal specified in Problem 6.19.

6.23 In this problem, we compare the performance of a DPCM system with that of an ordinary PCM system using companding.

For a sufficiently large number of representation levels, the signal-to-(quantization) noise ratio of PCM systems, in general, is defined by

$$10 \log_{10}(\text{SNR})_O \text{ (dB)} = \alpha + 6n$$

where 2^n is the number of representation levels. For a companded PCM system using the μ-law, the constant α is itself defined by

$$\alpha\,(\text{dB}) \approx 4.77 - 20 \log_{10} \log(1 + \mu)$$

For a DPCM system, on the other hand, the constant α lies in the range $-3 < \alpha < 15$ dBs. The formulas quoted herein apply to telephone-quality speech signals.

Compare the performance of the DPCM system against that of the μ-companded PCM system with $\mu = 255$ for each of the following scenarios:

a. The improvement in $(\text{SNR})_O$ realized by DPCM over companded PCM for the same number of bits per sample.

b. The reduction in the number of bits per sample required by DPCM, compared with the companded PCM for the same $(\text{SNR})_O$.

6.24 In the DPCM system depicted in Figure P6.24, show that in the absence of channel noise, the transmitting and receiving prediction filters operate on slightly different input signals.

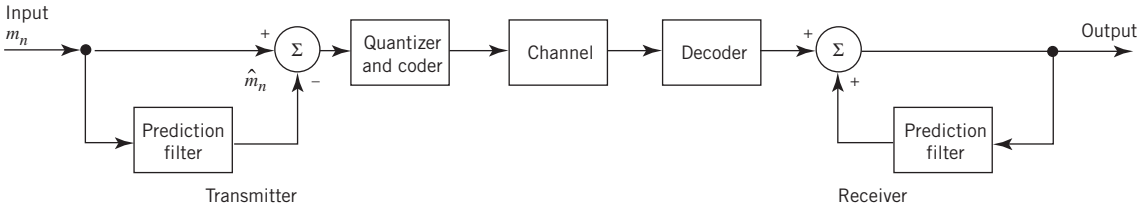


Figure P6.24

6.25 Figure P6.25 depicts the block diagram of adaptive quantization for DPCM. The quantization is of a backward estimation kind because samples of the quantization output and prediction errors are used to continuously derive backward estimates of the variance of the message signal. This estimate computed at time n is denoted by $\hat{\sigma}^2_{m,n}$. Given this estimate, the step size is varied so as to match the actual variance of the message sample m_n, as shown by

$$\Delta_n = \phi \sigma_{m,n}$$

where $\hat{\sigma}^2_{m,n}$ is the estimate of the standard deviation and ϕ is a constant. An attractive feature of the adaptive scheme in Figure P6.25 is that samples of the quantization output and the prediction error are used to compute the predictor's coefficients.

Modify the block diagram of the DPCM transmitter in Figure 6.19a so as to accommodate *adaptive prediction with backward estimation*.

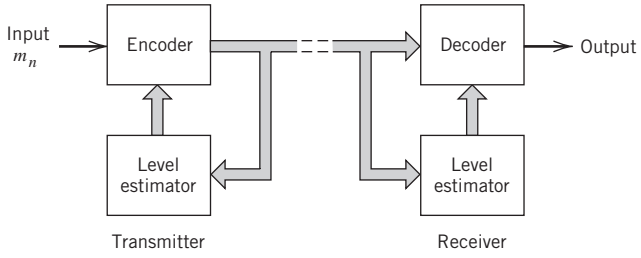


Figure P6.25

Delta Modulation

6.26 Consider a test signal $m(t)$ defined by a hyperbolic tangent function:

$$m(t) = A\tanh(\beta t)$$

where A and β are constants. Determine the minimum step size Δ for DM of this signal, which is required to avoid slope-overload distortion.

6.27 Consider a sine wave of frequency f_m and amplitude A_m, which is applied to a delta modulator of step size Δ. Show that slope-overload distortion will occur if

$$A_m > \frac{\Delta}{2\pi f_m T_s}$$

where T_s is the sampling period. What is the maximum power that may be transmitted without slope-overload distortion?

6.28 A linear delta modulator is designed to operate on speech signals limited to 3.4 kHz. The specifications of the modulator are as follows:

- Sampling rate = $10f_{Nyquist}$, where $f_{Nyquist}$ is the Nyquist rate of the speech signal.
- Step size Δ = 100 mV.

The modulator is tested with a 1 kHz sinusoidal signal. Determine the maximum amplitude of this test signal required to avoid slope-overload distortion.

6.29 In this problem, we derive an empirical formula for the average signal-to-(quantization) noise ratio of a DM system with a sinusoidal signal of amplitude A and frequency f_m as the test signal. Assume that the power spectral density of the granular noise generated by the system is governed by the formula

$$S_N(f) = \frac{\Delta^2}{6f_s}$$

where f_s is the sampling rate and Δ is the step size. (Note that this formula is basically the same as that for the power spectral density of quantization noise in a PCM system with $\Delta/2$ for PCM being replaced by Δ for DM.) The DM system is designed to handle analog message signals limited to bandwidth W.

a. Show that the average quantization noise power produced by the system is

$$N = \frac{4\pi^2 A^2 f_m^2 W}{3f_s^3}$$

where it is assumed that the step size Δ has been chosen in accordance with the formula used in Problem 6.28 so as to avoid slope-overload distortion.

b. Hence, determine the signal-to-(quantization) noise ratio of the DM system for a sinusoidal input.

6.30 Consider a DM system designed to accommodate analog message signals limited to bandwidth W = 5 kHz. A sinusoidal test signal of amplitude A = 1V and frequency f_m = 1 kHz is applied to the system. The sampling rate of the system is 50 kHz.

a. Calculate the step size Δ required to minimize slope overload distortion.

b. Calculate the signal-to-(quantization) noise ratio of the system for the specified sinusoidal test signal.

For these calculations, use the formula derived in Problem 6.29.

6.31 Consider a low-pass signal with a bandwidth of 3 kHz. A linear DM system with step size Δ = 0.1V is used to process this signal at a sampling rate 10 times the Nyquist rate.

a. Evaluate the maximum amplitude of a test sinusoidal signal of frequency 1 kHz, which can be processed by the system without slope-overload distortion.

b. For the specifications given in part a, evaluate the output SNR under (i) prefiltered and (ii) postfiltered conditions.

6.32 In the conventional form of DM, the quantizer input may be viewed as an approximate to the *derivative* of the incoming message signal $m(t)$. This behavior leads to a drawback of DM: transmission disturbances (e.g., noise) result in an accumulation error in the demodulated signal. This drawback can be overcome by *integrating* the message signal $m(t)$ prior to DM, resulting in three beneficial effects:

a. Low frequency content of $m(t)$ is pre-emphasized.

b. Correlation between adjacent samples of $m(t)$ is increased, tending to improve overall system performance by reducing the variance of the error signal at the quantizer input.

c. Design of the receiver is simplified.

Such a DM scheme is called *delta–sigma modulation*.

Construct a block diagram of the delta–sigma modulation system in such a way that it provides an interpretation of the system as a "smoothed" version of 1-bit PCM in the following composite sense:

- smoothness implies that the comparator output is integrated prior to quantization, and
- 1-bit modulation merely restates that the quantizer consists of a hard limiter with only two representation levels.

Explain how the receiver of the delta–sigma modulation system is simplified, compared with conventional DM.

Line Codes

6.33 In this problem, we derive the formulas used to compute the power spectra of Figure 6.25 for the five line codes described in Section 6.10. In the case of each line code, the bit duration is T_b and the pulse amplitude A is conditioned to normalize the average power of the line code to unity as indicated in Figure 6.25. Assume that the data stream is randomly generated and symbols 0 and 1 are equally likely.

Derive the power spectral densities of these line codes as summarized here:

a. Unipolar NRZ signals:

$$S(f) = \frac{A^2 T_b}{4} \operatorname{sinc}^2(fT_b)\left(1 + \frac{1}{T_b}\delta(f)\right)$$

b. Polar NRZ signals:

$$S(f) = A^2 T_b \operatorname{sinc}^2(fT_b)$$

c. Unipolar RZ signals:

$$S(f) = \frac{A^2 T_b}{16} \operatorname{sinc}^2\left(\frac{fT_b}{2}\right)\left[1 + \frac{1}{T_b}\sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T_b}\right)\right]$$

d. Bipolar RZ signals:

$$S(f) = \frac{A^2 T_b}{4} \operatorname{sinc}^2\left(\frac{fT_b}{2}\right)\sin^2(\pi f T_b)$$

e. Manchester-encoded signals:

$$S(f) = \frac{A^2 T_b}{4} \operatorname{sinc}^2\left(\frac{fT_b}{2}\right)\sin^2\left(\frac{\pi f T_b}{2}\right)$$

Hence, confirm the spectral plots displayed in Figure 6.25.

6.34 A randomly generated data stream consists of equiprobable binary symbols 0 and 1. It is encoded into a polar NRZ waveform with each binary symbol being defined as follows:

$$s(t) = \begin{cases} \cos\left(\dfrac{\pi t}{T_b}\right), & -\dfrac{T_b}{2} < t \le \dfrac{T_b}{2} \\ 0, & \text{otherwise} \end{cases}$$

a. Sketch the waveform so generated, assuming that the data stream is 00101110.

b. Derive an expression for the power spectral density of this signal and sketch it.

c. Compare the power spectral density of this random waveform with that defined in part b of Problem 6.33.

6.35 Given the data stream 1110010100, sketch the transmitted sequence of pulses for each of the following line codes:

a. unipolar NRZ

b. polar NRZ

c. unipolar RZ

d. bipolar RZ

e. Manchester code.

Computer Experiments

**6.36 A sinusoidal signal of frequency $f_0 = 10^4/2\pi$ Hz is sampled at the rate of 8 kHz and then applied to a sample-and-hold circuit to produce a flat-topped PAM signal $s(t)$ with pulse duration $T = 500$ μs.

a. Compute the waveform of the PAM signal $s(t)$.

b. Compute $|S(f)|$, denoting the magnitude spectrum of the PAM signal $s(t)$.

c. Compute the envelope of $|S(f)|$. Hence confirm that the frequency at which this envelope goes through zero for the first time is equal to $(1/T) = 20$ kHz.

**6.37 In this problem, we use computer simulation to compare the performance of a companded PCM system using the μ-law against that of the corresponding system using a uniform quantizer. The simulation is to be performed for a sinusoidal input signal of varying amplitude.

With a companded PCM system in mind, Table 6.4 describes the 15-segment *pseudo-linear* characteristic that consists of 15 linear segments configured to approximate the logarithmic μ-law

Table 6.4 **The 15-segment companding characteristic (μ = 255)**

Linear segment number	Step-size	Projections of segment end points onto the horizontal axis
0	2	±31
1*a*, 1*b*	4	±95
2*a*, 2*b*	8	±223
3*a*, 3*b*	16	±479
4*a*, 4*b*	32	±991
5*a*, 5*b*	64	±2015
6*a*, 6*b*	128	±4063
7*a*, 7*b*	256	±8159

of (6.48), with $\mu = 255$. This approximation is constructed in such a way that the segment endpoints in Table 6.4 lie on the compression curve computed from (6.48).

a. Using the μ-law described in Table 6.4, plot the output signal-to-noise ratio as a function of the input signal-to-noise ratio, both ratios being expressed in decibels.

b. Compare the results of your computation in part (a) with a uniform quantizer having 256 representation levels.

**6.38 In this experiment we study the linear adaptive prediction of a signal x_n governed by the following recursion:

$$x_n = 0.8x_{n-1} - 0.1x_{n-2} + 0.1v_n$$

where v_n is drawn from a discrete–time white noise process of zero mean and unit variance. (A process generated in this manner is referred to as an *autoregressive process of order two*.) Specifically, the adaptive prediction is performed using the *normalized LMS algorithm* defined by

$$\hat{x}_n = \sum_{k=1}^{p} w_{k,n} x_{n-k}$$

$$e_n = x_n - \hat{x}_n$$

$$w_{k,n+1} = w_{k,n} + \mu \Big/ \left(\sum_{k=1}^{p} x_{n-k}^2 \right) x_{n-k} e_n \qquad k = 1, 2, \ldots, p$$

where p is the prediction order and μ is the normalized step-size parameter. The important point to note here is that μ is dimensionless and stability of the algorithm is assured by choosing it in accordance with the formula

$$0 < \mu < 2$$

The algorithm is initiated by setting

$$w_{k,0} = 0 \qquad \text{for all } k$$

The *learning curve* of the algorithm is defined as a plot of the mean-square error versus the number of iterations n for specified parameter values, which is obtained by averaging the plot of e_n^2 versus n over a large number of different realizations of the algorithm.

a. Plot the learning curves for the adaptive prediction of x_n for a fixed prediction order $p = 5$ and three different values of step-size parameter: $\mu = 0.0075$, 0.05, and 0.5.

b. What observations can you make from the learning curves of part a?

**6.39 In this problem, we study adaptive delta modulation, the underlying principle of which is two-fold:

1. If successive errors are of opposite polarity, then the delta modulator is operating in the granular mode, in which case the step size Δ is reduced.
2. If, on the other hand, the successive errors are of the same polarity, then the delta modulator is operating in the slope-overload mode, in which case the step size Δ is increased.

Parts a and b of Figure P6.39 depict the block diagrams of the transmitter and receiver of the adaptive delta modulator, respectively, in which the step size, Δ, is increased or decreased by a factor of 50% at each iteration of the adaptive process, as shown by:

$$\Delta_n = \begin{cases} \dfrac{\Delta_{n-1}}{m_{q,n}} (m_{q,n} + 0.5 m_{q,n-1}) & \text{if } \Delta_{n-1} \geq \Delta_{\min} \\ \Delta_{\min} & \text{if } \Delta_{n-1} < \Delta_{\min} \end{cases}$$

where Δ_n is the step size at iteration (time step) n of the adaptation algorithm, and $m_{q,n}$ is the 1-bit quantizer output that equals ± 1.

Specifications: The input signal applied to the transmitter is sinusoidal as shown by

$$m_t = A \sin(2\pi f_m t)$$

where $A = 10$ and $f_m = f_s/100$ where f_s is the sampling frequency; the step size $\Delta_n = 1$ for all n; $\Delta_{\min} = 1/8$.

a. Using the above-described adaptation algorithm, use a computer to plot the resulting waveform for one complete cycle of the sinusoidal modulating signal, and also display the coded modulator output in the transmitter.

b. For the same specifications, repeat the computation using linear modulation.

c. Comment on the results obtained in parts a and b of the problem.

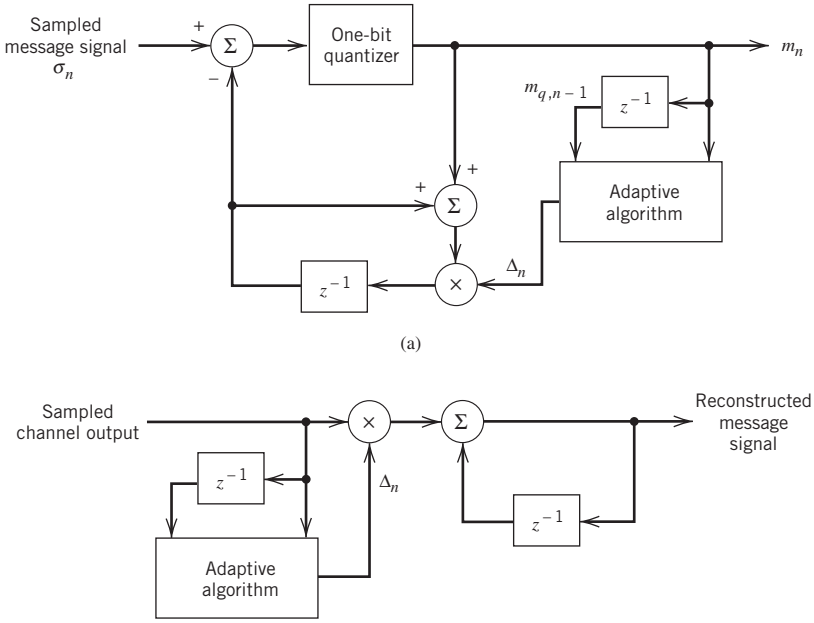


Figure P6.39

Notes

1. For an exhaustive study of quantization noise in signal processing and communications, see Widrow and Kollar (2008).

2. The two necessary conditions of (3.42) and (3.47) for optimality of a scalar quantizer were reported independently by Lloyd (1957) and Max (1960), hence the name "Lloyd–Max quantizer." The derivation of these two optimality conditions presented in this chapter follows the book by Gersho and Gray (1992).

3. The μ-law is used in the USA, Canada, and Japan. On the other hand, in Europe, the A-law is used for signal compression.

4. In actual PCM systems, the companding circuitry does not produce an exact replica of the nonlinear compression curves shown in Figure 6.14. Rather, it provides a *piecewise linear* approximation to the desired curve. By using a large enough number of linear segments, the approximation can approach the true compression curve very closely; for detailed discussion of this issue, see Bellamy (1991).

5. For a discussion of noise in analog modulation systems with particular reference to FM, see Chapter 4 of *Communication Systems* (Haykin, 2001).

6. To simplify notational matters, $\mathbf{R}_{\mathrm{M}}$ is used to denote the autocorrelation matrix in (6.70) rather than $\mathbf{R}_{\mathrm{MM}}$ as in Chapter 4 on Stochastic Processes. To see the rationale for this simplification, the reader is referred to (6.79) for simplicity. For the same reason, henceforth the practice adopted in this chapter will be continued for the rest of the book, dealing with autocorrelation matrices and power spectral density.

7. An optimum predictor that follows (6.77) is said to be a special case of the *Wiener filter.*

8. For a detailed discussion of adaptive DPCM involving the use of adaptive quantization with forward estimation as well as backward estimation, the reader is referred to the classic book (Jayant and Noll, 1984).