CHAPTER FOUR

SPECTRAL DENSITY AND CORRELATION

In this Chapter we complete the characterization of signals and systems by focusing on the *energy* or *power* of a signal. In so doing, we introduce the notion of *spectral density,* which defines the distribution of energy or power per unit bandwidth as a function of frequency. When dealing with energy signals, it is natural to use *energy spectral density* as the parameter of interest. Likewise, when dealing with power signals, *power spectral density* is used to characterize the signal. In this chapter we also introduce another important parameter called *correlation,* which may be viewed as the time-domain counterpart of spectral density. Throughout the chapter, we deal with real-valued energy and power signals. We begin the discussion with energy spectral density.

4.1 ENERGY SPECTRAL DENSITY

Consider an *energy signal* $g(t)$ defined over the interval $-\infty < t < \infty$, and let its Fourier transform or spectrum be denoted by $G(f)$. The signal $g(t)$ is assumed to be *real valued*. The total *energy* of the signal is defined by (see Section 1.2)

$$E = \int_{-\infty}^{\infty} g^2(t)\, dt \tag{4.1}$$

Equation 4.1 is the standard formula for evaluating the energy E. Nevertheless, there is another method based on the amplitude spectrum $|G(f)|$, which may also be used to evaluate the energy E. To develop this alternative method, we start with the relation (see Exercise 10, Chapter 2):

$$\int_{-\infty}^{\infty} g_1(t)g_2(t)\, dt = \int_{-\infty}^{\infty} G_1(f)G_2(-f)\, df \tag{4.2}$$

where $g_1(t)$ and $g_2(t)$ are a pair of energy signals with Fourier transforms $G_1(f)$ and $G_2(f)$, respectively. Let

$$g_1(t) = g_2(t) = g(t)$$

Correspondingly, we may set

$$G_1(f) = G(f)$$

and for real-valued signals,

$$G_2(-f) = G^*(f)$$

Accordingly, we may simplify Eq. 4.2 as

$$\int_{-\infty}^{\infty} g^2(t)\, dt = \int_{-\infty}^{\infty} |G(f)|^2\, df \tag{4.3}$$

where $|G(f)|$ is the amplitude spectrum of the signal $g(t)$. Equation 4.3 is known as the *Rayleigh energy theorem*.

The Rayleigh energy theorem is important not only because it provides a useful method for evaluating energy, but also because it highlights $|G(f)|^2$ as the distribution of energy of the signal $g(t)$ in the frequency domain. It is for this reason that the squared amplitude spectrum $|G(f)|^2$ is called the *energy spectral density* or *energy density spectrum*. Using $\Psi_g(f)$ to denote this new parameter, we may thus write

$$\Psi_g(f) = |G(f)|^2 \tag{4.4}$$

EXAMPLE 1 SINC PULSE

Consider the sinc pulse defined by

$$g(t) = A \operatorname{sinc}(2Wt)$$

The energy of this pulse equals

$$E = A^2 \int_{-\infty}^{\infty} \operatorname{sinc}^2(2Wt)\, dt \tag{4.5}$$

The integral on the right side of Eq. 4.5 is difficult to evaluate. We may obtain the desired result indirectly by applying the Rayleigh energy theorem. We start with the Fourier transform pair (see Example 6, Chapter 2)

$$A \operatorname{sinc}(2Wt) \rightleftharpoons \frac{A}{2W} \operatorname{rect}\left(\frac{f}{2W}\right)$$

Hence, with the Fourier transform

$$G(f) = \frac{A}{2W} \operatorname{rect}\left(\frac{f}{2W}\right)$$

and $\operatorname{rect}^2(f/2W) = \operatorname{rect}(f/2W)$, the energy spectral density of the sinc pulse is given by

$$\Psi_g(f) = \left(\frac{A}{2W}\right)^2 \operatorname{rect}\left(\frac{f}{2W}\right) \tag{4.6}$$

Hence, the application of the Rayleigh energy theorem yields the result

$$E = \left(\frac{A}{2W}\right)^2 \int_{-\infty}^{\infty} \operatorname{rect}\left(\frac{f}{2W}\right) df$$

$$= \left(\frac{A}{2W}\right)^2 \int_{-W}^{W} df$$

$$= \frac{A^2}{2W} \tag{4.7}$$

EXERCISE 1 Show that the total area under the curve of $\text{sinc}^2(t)$ equals 1; that is,

$$\int_{-\infty}^{\infty} \text{sinc}^2(t)\, dt = 1 \tag{4.8}$$

PROPERTIES OF ENERGY SPECTRAL DENSITY

The energy spectral density $\Psi_g(f)$ has several properties that follow from the basic definition given in Eq. 4.4, which are formally described in the sequel.

PROPERTY 1

The energy spectral density of an energy signal $g(t)$ is a nonnegative real-valued function of frequency; that is,

$$\Psi_g(f) \geq 0, \qquad \text{for all } f \tag{4.9}$$

This property follows directly from the fact that the amplitude spectrum $|G(f)|$ of a signal $g(t)$ is a nonnegative real function of the frequency f.

PROPERTY 2

The energy spectral density of a real-valued energy signal $g(t)$ is an even function of frequency, that is,

$$\Psi_g(-f) = \Psi_g(f) \tag{4.10}$$

This property means that the energy spectral density of a real-valued signal is symmetric about zero frequency. It follows directly from the fact that the amplitude spectrum $|G(f)|$ of a real-valued signal $g(t)$ is an even function of the frequency f, as shown by

$$|G(-f)| = |G(f)|$$

PROPERTY 3

The total area under the curve of energy spectral density of an energy signal $g(t)$ equals the signal energy; that is,

$$E = \int_{-\infty}^{\infty} \Psi_g(f)\, df \tag{4.11}$$

Suppose that $g(t)$ denotes the voltage of a source connected across a 1-ohm load resistor. Then, the integral

$$\int_{-\infty}^{\infty} g^2(t)\, dt$$

equals the energy E delivered by the source to the load. From Rayleigh's energy theorem described by Eq. 4.3, the total area under the $\Psi_x(f)$ curve equals the energy E.

EXERCISE 2 Using the energy spectral density for the sinc pulse given by Eq. 4.6, derived in Example 1, demonstrate the validity of Properties 1 through 3.

PROPERTY 4

When an energy signal is transmitted through a linear time-invariant system, the energy spectral density of the output equals the energy spectral density of the input multiplied by the squared amplitude response of the system.

This property follows from the frequency-domain description of a linear time-invariant system given in Eq. 3.23. Specifically, with $X(f)$ denoting the Fourier transform of a signal $x(t)$ applied to the input of a linear time-invariant system of transfer function $H(f)$, the Fourier transform $Y(f)$ of the signal $y(t)$ produced at the output of the system is given by

$$Y(f) = H(f)X(f) \tag{4.12}$$

Taking the squared amplitude of both sides of this equation, we get

$$|Y(f)|^2 = |H(f)|^2\, |X(f)|^2 \tag{4.13}$$

Equivalently, we may write

$$\Psi_y(f) = |H(f)|^2\, \Psi_x(f) \tag{4.14}$$

where $\Psi_y(f) = |Y(f)|^2$ and $\Psi_x(f) = |X(f)|^2$. The quantities $\Psi_y(f)$ and $\Psi_x(f)$ denote the energy spectral densities of the output $y(t)$ and the input $x(t)$, respectively. Equation 4.14 is a mathematical statement of Property 4.

EXAMPLE 2

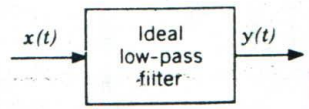
A rectangular pulse of unit amplitude and unit duration is passed through an ideal low-pass filter of bandwidth B, as illustrated in Fig. 4.1*a*. Part *b* of the figure depicts the waveform of the rectangular pulse. The amplitude response of the filter is defined by (see Fig. 4.1*c*)

$$|H(f)| = \begin{cases} 1, & -B \leq f \leq B \\ 0, & \text{otherwise} \end{cases}$$

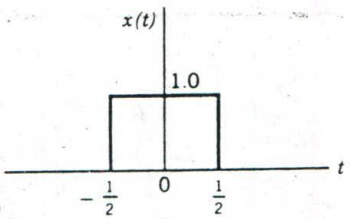
The rectangular pulse constituting the filter input has unit energy. We wish to evaluate the effect of varying the bandwidth B on the energy of the filter output.

We start with the Fourier transform pair:

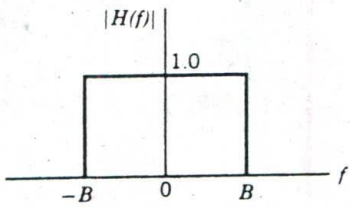
$$\text{rect}(t) \rightleftharpoons \text{sinc}(f)$$



(a)



(b)



(c)

Figure 4.1
(a) *Ideal low-pass filter.* (b) *Filter input.* (c) *Amplitude response of the filter.*

This represents the *normalized* version of the Fourier transform pair given in Eq. 2.33. Hence, with the filter input defined by

$$x(t) = \text{rect}(t)$$

its Fourier transform equals

$$X(f) = \text{sinc}(f)$$

The energy spectral density of the filter input therefore equals

$$\begin{aligned} \Psi_x(f) &= |X(f)|^2 \\ &= \text{sinc}^2(f) \end{aligned} \tag{4.15}$$

This normalized energy spectral density is shown plotted in Fig. 4.2.

To evaluate the energy spectral density $\Psi_y(f)$ of the filter output $y(t)$, we use Eq. 4.14. We thus obtain

$$\begin{aligned} \Psi_y(f) &= |H(f)|^2 \, \Psi_x(f) \\ &= \begin{cases} \Psi_x(f), & -B \leq f \leq B \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{4.16}$$

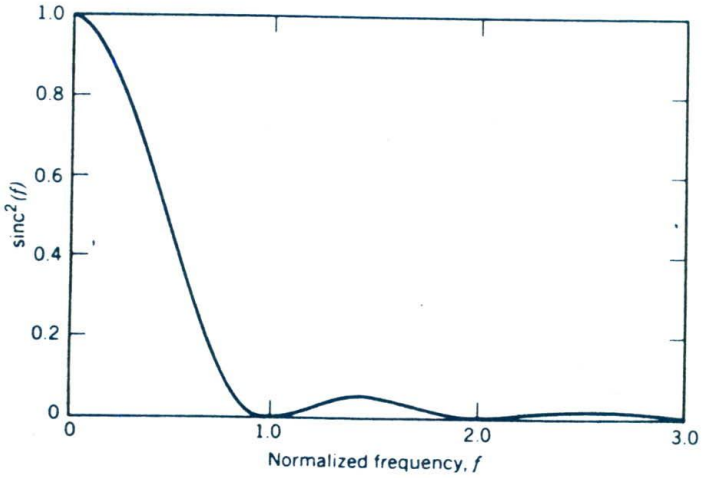


Figure 4.2
Normalized energy spectral density of filter input x(t).

The energy of the filter output therefore equals

$$E_y = \int_{-\infty}^{\infty} \Psi_y(f)\, df$$

$$= \int_{-B}^{B} \Psi_x(f)\, df$$

$$= 2\int_{0}^{B} \Psi_x(f)\, df \tag{4.17}$$

Substituting Eq. 4.15 in 4.17 yields

$$E_y = 2\int_{0}^{B} \text{sinc}^2(f)\, df \tag{4.18}$$

Since the filter input is normalized to have unit energy, we may also view the result given in Eq. 4.18 as *the ratio of the energy of the filter ouput to that of the filter input* for the general case of a rectangular pulse of arbitrary amplitude and arbitrary duration, processed by an ideal band-pass filter of bandwidth B. Accordingly, we may also write

$$\rho = \frac{\text{Energy of filter output}}{\text{Energy of filter input}}$$

$$= 2\int_{0}^{B} \text{sinc}^2(f)\, df \tag{4.19}$$

According to Fig. 4.1*b*, the rectangular pulse applied to the filter input has

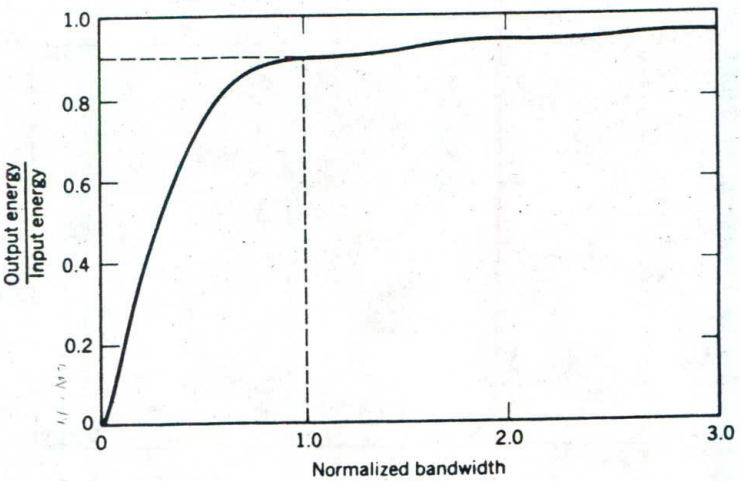


Figure 4.3
Output energy-to-input energy ratio versus normalized bandwidth.

unit duration; hence, the variable f in Eq. 4.19 represents a *normalized frequency*. Equation 4.19 is plotted in Fig. 4.3.

The graph of Fig. 4.3 shows that just over 90% of the total energy of a rectangular pulse lies inside the main spectral lobe of this pulse.

INTERPRETATION OF THE ENERGY SPECTRAL DENSITY

Equation 4.14 is important because it not only relates the output energy spectral density of a linear time-invariant system to the input energy spectral density but it also provides a basis for the physical interpretation of the concept of energy spectral density itself. To be specific, consider the arrangement shown in Fig. 4.4*a*, where an energy signal $x(t)$ is passed

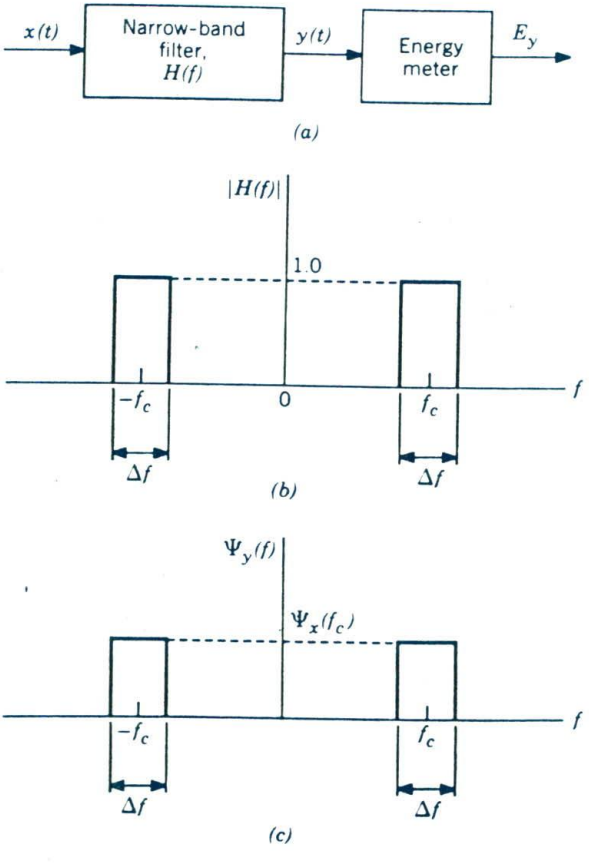


Figure 4.4
(a) *Circuit arrangement for measuring energy spectral density.* (b) *Idealized amplitude response of the filter.* (c) *Energy spectral density of filter output.*

through a narrow-band filter followed by an *energy meter*. Figure 4.4*b* shows the idealized amplitude response of the filter. That is, the amplitude response of the filter is defined by

$$|H(f)| = \begin{cases} 1, & f_c - \frac{\Delta f}{2} \leq |f| \leq f_c + \frac{\Delta f}{2} \\ 0, & \text{otherwise} \end{cases} \tag{4.20}$$

We assume that the filter bandwidth Δf is small enough for the amplitude response of the input signal $x(t)$ to be essentially constant over the frequency interval covered by the passband of the filter. Accordingly, we may express the amplitude spectrum of the filter output by the approximate formula:

$$\begin{aligned} |Y(f)| &= |H(f)|\,|X(f)| \\ &\simeq \begin{cases} |X(f_c)|, & f_c - \frac{\Delta f}{2} \leq |f| \leq f_c + \frac{\Delta f}{2} \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{4.21}$$

Correspondingly, the energy spectral density $\Psi_y(f)$ of the filter output $y(t)$ is approximately related to the energy spectral density $\Psi_x(f)$ of the filter input $x(t)$ as follows

$$\Psi_y(f) \simeq \begin{cases} \Psi_x(f_c) & f_c - \frac{\Delta f}{2} \leq |f| \leq f_c + \frac{\Delta f}{2} \\ 0, & \text{otherwise} \end{cases} \tag{4.22}$$

This relation is illustrated in Fig. 4.4*c*, which shows that only the frequency components of the signal $x(t)$ that lie inside the narrow pass band of the ideal band-pass filter reach the output. From Rayleigh's energy theorem, the energy of the filter output $y(t)$ is given by

$$\begin{aligned} E_y &= \int_{-\infty}^{\infty} \Psi_y(f)\, df \\ &= 2\int_0^{\infty} \Psi_y(f)\, df \\ &\simeq 2\Psi_x(f_c)\,\Delta f \end{aligned} \tag{4.23}$$

where the multiplying factor of 2 accounts for the contributions of negative as well as positive frequency components. We may rewrite Eq. 4.23 as

$$\Psi_x(f_c) \simeq \frac{E_y}{2\,\Delta f} \tag{4.24}$$

Equation 4.24 states that the energy spectral density of the filter input at some frequency f_c equals the energy of the filter output divided by $2\Delta f$, where Δf is the filter bandwidth centered on f_c. We may therefore interpret the energy spectral density of an energy signal for any frequency f *as the energy per unit bandwidth, which is contributed by frequency components of the signal around the frequency* f.

The arrangement shown in the block diagram of Fig. 4.4*a* provides the basis for measuring the energy spectral density of an energy signal. Specifically, by using a *variable* band-pass filter to scan the frequency band of interest, and determining the energy of the filter output for each midband frequency setting of the filter, a plot of the energy spectral density versus frequency is obtained.

4.2 CORRELATION OF ENERGY SIGNALS

The energy spectral density is an important frequency-dependent parameter of an energy signal. With the interplay between time-domain and frequency-domain descriptions of a signal that we have become accustomed to, it is natural for us to seek the time-domain counterpart of energy spectral density. From the defining equation (4.4), we have

$$\Psi_g(f) = G(f)G^*(f) \tag{4.25}$$

where the signal $g(t)$ and its Fourier transform $G(f)$ are related by

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

Equation 4.25 states that the energy spectral density $\Psi_g(f)$ of an energy signal $g(t)$ equals the product of $G(f)$, the Fourier transform of $g(t)$, and its complex conjugate, $G^*(f)$. Hence, given $G(f)$, we need to perform two frequency-domain operations to get $\Psi_g(f)$, namely, complex conjugation and multiplication. This suggests that we may determine the inverse Fourier transform of $\Psi_g(f)$ by making use of two fundamental properties of the Fourier transform (see Section 2.3):

1. The complex conjugation property, according to which time reversal of a real-valued signal translates to complex conjugation of its Fourier transform.
2. The time-domain convolution property, according to which the convolution of two signals translates to the multiplication of their Fourier transforms.

Accordingly, we may formulate the following Fourier transform pair:

$$g(\tau) \star g(-\tau) \rightleftharpoons G(f)\,G^*(f) \tag{4.26}$$

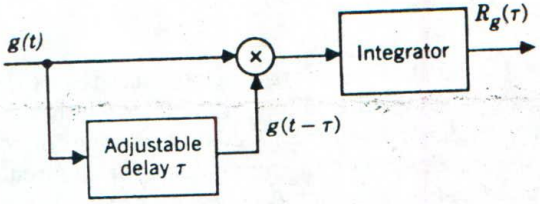


Figure 4.5
Circuit arrangement for measuring the autocorrelation function of an energy signal.

where ☆ denotes convolution. For the convolution on the left side of Eq. 4.26, we have purposely used τ as the time variable of the energy signal of interest, because we wish to reserve the use of time t as the dummy variable of the integral describing the convolution there. Specifically, we define the time-domain convolution on the left side of Eq. 4.26 as

$$\begin{aligned} R_g(\tau) &= g(\tau) \star g(-\tau) \\ &= \int_{-\infty}^{\infty} g(t)g(t-\tau)\,dt \end{aligned} \tag{4.27}$$

The τ-dependent parameter $R_g(\tau)$ is called the *autocorrelation function* of the energy signal $g(t)$. In the defining equation (4.27), the time function $g(t-\tau)$ represents a *delayed version of the signal* $g(t)$, and $R_g(\tau)$ provides a measure of the similarity between the waveforms of the time functions $g(t)$ and $g(t-\tau)$. In particular, the *time lag* or *time delay* τ plays the role of a *scanning* or *searching parameter*. This role is highlighted in the block diagram of Fig. 4.5, which provides the basis for measuring the autocorrelation function $R_g(\tau)$.

EXERCISE 3 Show that the definition for the autocorrelation function $R_g(\tau)$ may also be formulated as

$$R_g(\tau) = \int_{-\infty}^{\infty} g(t+\tau)g(t)\,dt \tag{4.28}$$

PROPERTIES OF THE AUTOCORRELATION FUNCTION OF ENERGY SIGNALS

The autocorrelation function of a real-valued energy signal has several useful properties. They follow directly from the defining equation (4.27) or (4.26), from which the definition of autocorrelation function originated.

PROPERTY 1

The autocorrelation function of a real-valued energy signal $g(t)$ is a real-valued even function, as shown by

$$R_g(-\tau) = R_g(\tau) \tag{4.29}$$

This property follows directly from Eq. 4.27. The implication of this property is that the autocorrelation function exhibits symmetry about the origin.

PROPERTY 2

The value of the autocorrelation function of an energy signal $g(t)$ at the origin is equal to the energy of the signal; that is,

$$R_g(0) = E \tag{4.30}$$

This result is obtained by putting $\tau = 0$ in Eq. 4.27.

PROPERTY 3

The maximum value of the autocorrelation function of an energy signal $g(t)$ occurs at the origin, as shown by

$$|R_g(\tau)| \leq R_g(0), \qquad \text{for all } \tau \tag{4.31}$$

To prove this property, we start with the observation that for any τ,

$$[g(t) \pm g(t - \tau)]^2 \geq 0$$

Equivalently, we may write

$$\pm 2g(t)g(t - \tau) \leq g^2(t) + g^2(t - \tau)$$

Integrating both sides of this relation with respect to time t from $-\infty$ to $+\infty$, and using Eqs. 4.27 and 4.30, we get the result given in Eq. 4.31.

According to Property 3, the degree of similarity between the signal $g(t)$ and its time-delayed version of $g(t - \tau)$ attains its maximum value at $\tau = 0$. This is intuitively satisfying.

PROPERTY 4

For an energy signal $g(t)$, the autocorrelation function and energy spectral density form a Fourier transform pair; that is,

$$R_g(\tau) \rightleftharpoons \Psi_g(f) \tag{4.32}$$

This property follows directly from Eqs. 4.25 through 4.27.

EXERCISE 4 Write the formulas for the Fourier transform of $R_g(\tau)$ and the inverse Fourier transform of $\Psi_g(f)$. Hence, do the following:

(a) Show that the total area under $\Psi_g(f)$ equals the signal energy, by evaluating the formula for the Fourier transform of $R_g(\tau)$ for $\tau = 0$.
(b) Show that the total area under $R_g(\tau)$ equals $\Psi_g(0)$, by evaluating the inverse Fourier transform of $\Psi_g(f)$ for $f = 0$.

EXAMPLE 3 SINC PULSE (CONTINUED)

From Example 1, the energy spectral density of the sinc pulse $A\ \mathrm{sinc}(2Wt)$ is given by (see Eq. 4.6)

$$\Psi_g(f) = \left(\frac{A}{2W}\right)^2 \mathrm{rect}\left(\frac{f}{2W}\right)$$

Taking the inverse Fourier transform of $\Psi_g(f)$, we find that the autocorrelation function of the sinc pulse $A\ \mathrm{sinc}(2Wt)$ is given by

$$R_g(\tau) = \frac{A^2}{2W}\mathrm{sinc}(2W\tau) \tag{4.33}$$

which has a similar waveform to the sinc pulse itself.

CROSS-CORRELATION OF ENERGY SIGNALS

The autocorrelation function provides a measure of the similarity between a signal and its time-delayed version. In a similar way, we may use the *cross-correlation function* as a measure of the similarity between a signal and the time-delayed version of a second signal. Let $g_1(t)$ and $g_2(t)$ denote a pair of real-valued energy signals. The cross-correlation function of this pair of signals is defined by

$$R_{12}(\tau) = \int_{-\infty}^{\infty} g_1(t)g_2(t-\tau)\,dt \tag{4.34}$$

We see that if the two signals $g_1(t)$ and $g_2(t)$ are somewhat similar, then the cross-correlation function $R_{12}(\tau)$ will be finite over some range of τ, thereby providing a quantitative measure of the similarity, or coherence, between them. The energy signals $g_1(t)$ and $g_2(t)$ are said to be *orthogonal* over the entire time interval if $R_{12}(0)$ is zero, that is, if

$$\int_{-\infty}^{\infty} g_1(t)g_2(t)\,dt = 0 \tag{4.35}$$

Equation 4.34 defines one possible value for the cross-correlation function for a specified value of the delay variable τ. We may define a second cross-correlation function for the energy signals $g_1(t)$ and $g_2(t)$ as

$$R_{21}(\tau) = \int_{-\infty}^{\infty} g_2(t)g_1(t - \tau)\, dt \tag{4.36}$$

From the definitions of the cross-correlation functions $R_{12}(\tau)$ and $R_{21}(\tau)$ just given, we obtain the fundamental relationship

$$R_{12}(\tau) = R_{21}(-\tau) \tag{4.37}$$

Equation 4.37 indicates that unlike convolution, correlation is not in general commutative, that is, $R_{12}(\tau) \neq R_{21}(\tau)$.

Another important property of cross-correlation is shown by the Fourier transform pair

$$R_{12}(\tau) \rightleftharpoons G_1(f)\, G_2^*(f) \tag{4.38}$$

This relation is known as the *correlation theorem*. The correlation theorem states that *the cross-correlation of two energy signals corresponds to the multiplication of the Fourier transform of one signal by the complex conjugate of the Fourier transform of the other.*

EXERCISE 5 Prove the property of cross-correlation functions described in Eq. 4.37.

EXERCISE 6 Prove the correlation theorem described by Eq. 4.38.

4.3 POWER SPECTRAL DENSITY

Consider next the case of a *power signal* $g(t)$, which remains finite as time t approaches infinity. We assume $g(t)$ to be real valued. The *average power* of the signal is defined by (see Section 1.2)

$$P = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} g^2(t)\, dt \tag{4.39}$$

To develop a frequency-domain description of power, we need to know the Fourier transform of the signal $g(t)$. However, this may pose a problem, because power signals have infinite energy and may therefore not be Fou-

rier transformable. To overcome the problem, we consider a *truncated* version of the signal $g(t)$. In particular, we define

$$g_T(t) = g(t)\,\text{rect}\left(\frac{t}{2T}\right) = \begin{cases} g(t), & -T \leq t \leq T \\ 0, & \text{otherwise} \end{cases} \tag{4.40}$$

As long as T is finite, the truncated signal $g_T(t)$ has finite energy; hence $g_T(t)$ is Fourier transformable. Let $G_T(f)$ denote the Fourier transform of $g_T(t)$; that is,

$$g_T(t) \rightleftharpoons G_T(f)$$

Using the definition of $g_T(t)$, we may rewrite Eq. 4.39 for the average power P in terms of $g_T(t)$ as

$$P = \lim_{T\to\infty} \frac{1}{2T} \int_{-\infty}^{\infty} g_T^2(t)\,dt \tag{4.41}$$

Since $g_T(t)$ has finite energy, we may use the Rayleigh energy theorem to express the energy of $g_T(t)$ in terms of its Fourier transform $G_T(f)$ as

$$\int_{-\infty}^{\infty} g_T^2(t)\,dt = \int_{-\infty}^{\infty} |G_T(f)|^2\,df \tag{4.42}$$

where $|G_T(f)|$ is the amplitude spectrum of $g_T(t)$. Accordingly, we may rewrite Eq. 4.41 in the equivalent form

$$P = \lim_{T\to\infty} \frac{1}{2T} \int_{-\infty}^{\infty} |G_T(f)|^2\,df \tag{4.43}$$

As T increases, the energy of $g_T(t)$ increases. Correspondingly, the energy spectral density $|G_T(f)|^2$ increases with T. Indeed as T approaches infinity, so will $|G_T(f)|^2$. However, for the average power P to be finite, $|G_T(f)|^2$ must approach infinity at the same rate as T. This requirement ensures the *convergence* of the integral on the right side of Eq. 4.43 in the limit as T approaches infinity. This convergence, in turn, permits us to interchange the order in which the limiting operation and integration in Eq. 4.43 are performed. We may thus rewrite this equation as

$$P = \int_{-\infty}^{\infty} \lim_{T\to\infty} \frac{1}{2T} |G_T(f)|^2\,df \tag{4.44}$$

Let the integrand be denoted by

$$S_g(f) = \lim_{T\to\infty} \frac{1}{2T} |G_T(f)|^2 \tag{4.45}$$

The frequency-dependent function $S_g(f)$ is called the *power spectral density* or *power spectrum* of a power signal, and $|G_T(f)|^2/2T$ is called the *periodogram*[1] of the signal.

EXAMPLE 4 MODULATED WAVE

Consider the *modulated wave*

$$x(t) = g(t)\cos(2\pi f_c t) \tag{4.46}$$

where $g(t)$ is a power signal band-limited to B hertz. We refer to $x(t)$ as a "modulated wave" in the sense that the amplitude of the sinusoidal "carrier" of frequency f_c is varied linearly with the signal $g(t)$. We wish to find the power spectral density of $x(t)$ in terms of that of $g(t)$, given that the frequency f_c is larger than the bandwidth B.

Adapting the formula of Eq. 4.45 to the situation at hand, we may define the power spectral density of the modulated wave $x(t)$ as

$$S_x(f) = \lim_{T\to\infty} \frac{1}{2T} |X_T(f)|^2 \tag{4.47}$$

where $X_T(f)$ is the Fourier transform of $x_T(t)$, the truncated version of $x(t)$. From Eq. 4.46, we have

$$x_T(t) = g_T(t)\cos(2\pi f_c t) \tag{4.48}$$

where the truncated signal $g_T(t)$ is itself defined in Eq. 4.40. Since

$$\cos(2\pi f_c t) = \frac{1}{2}[\exp(j2\pi f_c t) + \exp(-j2\pi f_c t)] \tag{4.49}$$

it follows from the frequency-shifting property of the Fourier transform that

$$X_T(f) = \frac{1}{2}[G_T(f - f_c) + G_T(f + f_c)] \tag{4.50}$$

where $G_T(f)$ is the Fourier transform of $g_T(t)$.

[1]The periodogram is a misnomer since it is a function of frequency not period. Nevertheless, the term has wide usage. It was first used by statisticians to look for periodicities such as seasonal trends in data.

Given that $f_c > B$, we find that $G_T(f - f_c)$ and $G_T(f + f_c)$ represent *nonoverlapping spectra;* their product is therefore zero. Accordingly, using Eq. 4.50 to evaluate the squared amplitude of $X_T(f)$, we get

$$|X_T(f)|^2 = \frac{1}{4}[|G_T(f - f_c)|^2 + |G_T(f + f_c)|^2], \qquad f_c > B \quad (4.51)$$

Finally, substituting Eq. 4.51 in 4.47, and then using the definition of Eq. 4.45 for the power spectral density of the power signal $g(t)$, we get the desired result:

$$S_x(f) = \frac{1}{4}[S_g(f - f_c) + S_g(f + f_c)], \qquad f_c > B \quad (4.52)$$

PROPERTIES OF POWER SPECTRAL DENSITY

The role of power spectral density for power signals is similar to that of energy density for energy signals. Indeed, the power spectral density has properties that parallel those of the energy spectral density. In the sequel, we present the properties of power spectral density without proof; these properties may be verified by using arguments similar to those used in Section 4.1 for verifying the properties of energy spectral density.

PROPERTY 1

The power spectral density of a power signal $g(t)$ is a nonnegative real-valued function of frequency; that is,

$$S_g(f) \geq 0, \qquad \text{for all } f \quad (4.53)$$

PROPERTY 2

The power spectral density of a real-valued power signal $g(t)$ is an even function of frequency; that is,

$$S_g(-f) = S_g(f) \quad (4.54)$$

PROPERTY 3

The total area under the curve of the power spectral density of a power signal $g(t)$ equals the average signal power; that is,

$$P = \int_{-\infty}^{\infty} S_g(f)\, df \quad (4.55)$$

PROPERTY 4

When a power signal is transmitted through a linear time-invariant system, the power spectral density of the output equals the power spectral density of the input multiplied by the squared amplitude response of the system. That is, if $S_x(f)$ is the power spectral density of a power signal $x(t)$ applied to a linear time-invariant system of transfer function $H(f)$, the power spectral density $S_y(f)$ of the power signal $y(t)$ produced at the output of the system is defined by

$$S_y(f) = |H(f)|^2 S_x(f) \tag{4.56}$$

where $|H(f)|$ is the amplitude response of the system.

EXERCISE 7 Justify the validity of the input–output relation described in Eq. 4.56.

INTERPRETATION OF POWER SPECTRAL DENSITY

The input–output relation of Eq. 4.56 provides a basis for the physical interpretation of power spectral density, and therefore its measurement. Just as we did for the interpretation of energy spectral density, suppose a power signal $x(t)$ is applied to a band-pass filter followed by a power meter as in Fig. 4.6*a*. The filter has a narrow bandwidth Δf centered on some frequency f_c, as in Fig. 4.6*b*. Application of Eq. 4.56 yields the power spectral density of the resulting filter output $y(t)$ approximately as follows

$$S_y(f) \simeq \begin{cases} S_x(f_c) & f_c - \dfrac{\Delta f}{2} \leq |f| \leq f_c + \dfrac{\Delta f}{2} \\ 0, & \text{otherwise} \end{cases} \tag{4.57}$$

The average power of the filter output $y(t)$ is therefore approximately given by

$$P_y \simeq 2S_x(f_c)\,\Delta f \tag{4.58}$$

The evaluation of $S_x(f)$ is illustrated in Fig. 4.6*c*. Equivalently, we may write

$$S_x(f_c) \simeq \frac{P_y}{2\Delta f} \tag{4.59}$$

In other words, the power spectral density of the filter input $x(t)$ at some frequency f_c is equal to the average power of the filter output divided by $2\Delta f$, where Δf is the bandwidth of the filter centered on f_c. The factor of 2 accounts for the contributions of negative as well as positive frequency components.

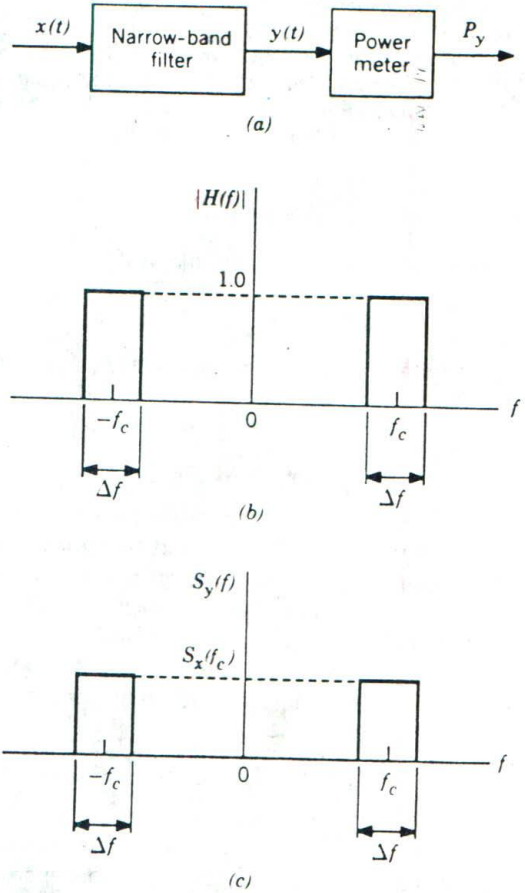


Figure 4.6
(a) *Circuit arrangement for measuring power spectral density.* (b) *Idealized amplitude response of narrow-band filter.* (c) *Power spectral density of the filter output.*

Equation 4.59 provides the basis for the measurement of power spectral density. Specifically, by varying the midband frequency f_c of the band-pass filter in Fig. 4.6*a*, and measuring the average power of the filter output for each setting of f_c, we may measure the power spectral density of a power signal (applied to the filter input) over a frequency band of interest.

4.4 CORRELATION OF POWER SIGNALS

We may develop a formula for the autocorrelation function of power signals by following a procedure similar to that described for the case of energy signals in Section 4.2. Specifically, we start with the defining equation (4.45) for the power spectral density $S_x(f)$ of a power signal $g(t)$, and rewrite it

in the form

$$S_g(f) = \lim_{T\to\infty} \frac{1}{2T} G_T(f)G_T^*(f) \tag{4.60}$$

where $G_T(f)$ is the Fourier transform of the truncated version $g_T(t)$ of the power signal $g(t)$. Next, we use the Fourier transform pair

$$g_T(\tau) \star g_T(-\tau) \rightleftharpoons G_T(f)\, G_T^*(f) \tag{4.61}$$

Multiplying both members of this pair by the factor $1/2T$ and then taking the limit as T approaches infinity, we have

$$\lim_{T\to\infty} \frac{1}{2T} g_T(\tau) \star g_T(-\tau) \rightleftharpoons \lim_{T\to\infty} \frac{1}{2T} G_T(f)G_T^*(f) \tag{4.62}$$

The function on the right side of this pair is recognized as the power spectral density of the power signal $g(t)$. Accordingly, we adopt the function on the left side of Eq. 4.62 as the autocorrelation function of the power signal $g(t)$, and thus write

$$R_g(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-\infty}^{\infty} g_T(t)g_T(t-\tau)\, dt \tag{4.63}$$

We may formulate the autocorrelation function $R_g(\tau)$ in terms of the power signal $g(t)$ itself by using the definition given in Eq. 4.40 for the truncated signal $g_T(t)$. By so doing, we define the autocorrelation function of a power signal $g(t)$ as follows

$$R_g(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} g(t)g(t-\tau)\, dt \tag{4.64}$$

PROPERTIES OF THE AUTOCORRELATION FUNCTION OF POWER SIGNALS

The autocorrelation function of a power signal has properties that are similar to those of the autocorrelation function of energy signals. Indeed, by following arguments similar to those presented in Section 4.2, we may readily establish the following properties of the autocorrelation function of power signals, which are presented without proof.

PROPERTY 1

The autocorrelation function of a real-valued power signal g(t) is a real-valued even function, as shown by

$$R_g(-\tau) = R_g(\tau) \tag{4.65}$$

PROPERTY 2

The value of the autocorrelation function of a power signal $g(t)$ at the origin is equal to the average power of the signal; that is,

$$R_g(0) = P \tag{4.66}$$

PROPERTY 3

The maximum value of the autocorrelation function of a power signal $g(t)$ occurs at the origin, as shown by

$$|R_g(\tau)| \leq R_g(0) \tag{4.67}$$

PROPERTY 4

For a power signal $g(t)$, the autocorrelation function and power spectral density form a Fourier transform pair; that is,

$$R_g(\tau) \rightleftharpoons S_g(f) \tag{4.68}$$

Equation 4.68 states that the power spectral density $S_g(f)$ of a power signal $g(t)$ is the Fourier transform of the autocorrelation function $R_g(\tau)$ of the signal, as shown in the expanded form:

$$S_g(f) = \int_{-\infty}^{\infty} R_g(\tau) \exp(-j2\pi f\tau)\, d\tau \tag{4.69}$$

Equation 4.68 also states that the autocorrelation function $R_g(\tau)$ is the inverse Fourier transform of the power spectral density $S_g(f)$, as shown in the expanded form:

$$R_g(\tau) = \int_{-\infty}^{\infty} S_g(f) \exp(j2\pi f\tau)\, df \tag{4.70}$$

Equations 4.69 and 4.70 are known as the *Einstein–Wiener–Khintchine relations*.[2] Given the autocorrelation function $R_g(\tau)$, we may use Eq. 4.69 to compute the power spectral density $S_g(f)$. Conversely, given the power

[2]Traditionally, Eqs. 4.69 and 4.70 have been referred to in the literature as the Wiener–Khintchine relations in recognition of pioneering work done by Wiener and Khintchine; for their original papers, see Wiener (1930) and Khintchine (1934). A recent discovery of a forgotten paper by Albert Einstein on time-series analysis (delivered at the Swiss Physical Society's February 1914 meeting in Basel) reveals that Einstein had discussed the autocorrelation function and its relationship to the spectral content of a time series many years before Wiener and Khintchine. For this very brief paper, see Einstein (1914). An English translation of Einstein's paper is reproduced in the *IEEE Acoustics, Speech, and Signal Processing Magazine*, vol. 4, October 1987. This particular issue also contains articles by W. A. Gardner and A. M. Yaglom, which elaborate on Einstein's original work.

spectral density $S_g(f)$, we may use Eq. 4.70 to compute the autocorrelation function $R_g(\tau)$.

EXERCISE 8 **Verify the properties of autocorrelation function of a power signal, described in Eqs. 4.65 through 4.68.**

EXAMPLE 5

Consider again the modulated wave $x(t)$, defined in Eq. 4.46, reproduced here for convenience:

$$x(t) = g(t) \cos(2\pi f_c t)$$

The signal $g(t)$ is a power signal band-limited to B hertz, where $B < f_c$. In this example, we evaluate the autocorrelation function of $x(t)$ in terms of that of $g(t)$.

We do the evaluation by using Property 4 of the autocorrelation function, namely, the fact that autocorrelation function and power spectral density form a Fourier transform. From Eq. 4.52 of Example 4, we have

$$S_x(f) = \frac{1}{4}[S_g(f - f_c) + S_g(f + f_c)]$$

Therefore, taking the inverse Fourier transform of both sides of the equation, we get

$$\begin{aligned} R_x(\tau) &= \frac{1}{4}[R_g(\tau) \exp(j2\pi f_c \tau) + R_g(\tau) \exp(-j2\pi f_c \tau)] \\ &= \frac{1}{2} R_g(\tau) \cos(2\pi f_c \tau) \end{aligned} \tag{4.71}$$

which is the desired result.

EXERCISE 9 **Using the relation of Eq. 4.71, show that the average power of the modulated signal $x(t)$ equals one-half the average power of the original signal $g(t)$.**

CROSS-CORRELATION OF POWER SIGNALS

We complete the discussion of correlation of power signals by considering their cross-correlation. Let $g_1(t)$ and $g_2(t)$ denote a pair of power signals.

We define the *cross-correlation* between $g_1(t)$ and $g_2(t)$ as

$$R_{12}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} g_1(t)g_2(t-\tau)\, dt \tag{4.72}$$

In a similar way we may define a second cross-correlation function $R_{21}(\tau)$.

The pair of power signals $g_1(t)$ and $g_2(t)$ are said to be *orthogonal* over the entire time interval if

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} g_1(t)g_2(t)\, dt = 0 \tag{4.73}$$

4.5 FLOWCHART SUMMARIES

In this section we summarize the significance of the time-frequency relations derived for energy and power signals.

Given an energy signal $g(t)$ of Fourier transform $G(f)$, we may summarize this relationship and its interplay with the formula of Eq. 4.4 for the energy spectral density $\Psi_g(f)$ and that of Eq. 4.27 for the autocorrelation function $R_g(\tau)$ as in Fig. 4.7. This chart clearly shows that whatever operation or sequence of operations is used to obtain the autocorrelation function $R_g(\tau)$ or the energy spectral density $\Psi_g(f)$, that operation or sequence of operations is *irreversible*. The implication of this is that when a signal $g(t)$ is converted to $R_g(\tau)$ or $\Psi_g(f)$, in general, information is lost about the original signal $g(t)$ or its Fourier transform $G(f)$. In going from $g(t)$ to $R_g(\tau)$, dependence on the physical time t is destroyed. In going from $G(f)$ to $\Psi_g(f)$, information on the phase spectrum of the signal is destroyed. This means that if two (or more) different signals have the same amplitude spectrum but different phase spectra, then they will have the same energy spectral density or, equivalently, the same autocorrelation function. In other words, for a given energy signal $g(t)$, there is a unique energy spectral density $\Psi_g(f)$ or, equivalently, a unique autocorrelation function $R_g(\tau)$. The converse of this statement, however, is not true.

The flowchart of Fig. 4.7 shows that given the energy signal $g(t)$, we may compute the energy spectral density $\Psi_g(f)$ in one of two equivalent ways:

1. We compute the Fourier transform $G(f)$, and then use the definition of Eq. 4.4.
2. We compute the autocorrelation function $R_g(\tau)$ using Eq. 4.27, and then compute the Fourier transform of $R_g(\tau)$.

The interrelations between time-domain and frequency-domain descriptions of power signals are analogous to those for energy signals. In particular, for a power signal $g(t)$ we may draw a chart similar to that of Fig.

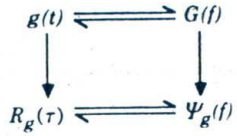


Figure 4.7
Flowchart summary of interrelations between time-domain and frequency-domain descriptions of energy signals.

4.7, except that the chart is now based on a truncated version $g_T(t)$ of the power signal. The point to note is that information is lost in the process of computing and retaining only the autocorrelation function $R_g(\tau)$ of the power signal or its power spectral density $\Psi_g(f)$. Moreover, given the signal $g(t)$, we may compute the power spectral density $S_g(f)$ using one of two equivalent procedures:

1. We compute the Fourier transform $G_T(f)$ of the power signal $g(t)$ for the interval $-T \leq t \leq T$ for large T, and then use Eq. 4.45 to compute the power spectral density $S_g(f)$.
2. We use Eq. 4.63 to compute the autocorrelation function $R_g(\tau)$, and then take the Fourier transform of $R_g(\tau)$.

EXERCISE 10 Given the energy signal $g(t)$, outline the two procedures that may be used to compute the autocorrelation function $R_g(\tau)$.

EXERCISE 11 Given the power signal $g(t)$, outline the two procedures that may be used to compute the autocorrelation function $R_g(\tau)$.

4.6 SPECTRAL CHARACTERISTICS OF PERIODIC SIGNALS

The definitions of power spectral density and autocorrelation function for power signals given in Eqs. 4.60 and 4.64 take on special forms for the case of periodic signals. These signals constitute an important class of power signals. Consider a periodic signal $g_p(t)$ of period T_0, represented in terms of its complex Fourier series as

$$g_p(t) = \sum_{n=-\infty}^{\infty} c_n \exp\left(\frac{j2\pi nt}{T_0}\right) \tag{4.74}$$

where the c_n are complex Fourier coefficients. For the situation at hand, the time average in the defining equation (4.39) for the average power of

the signal may be taken over one period, as shown by

$$P = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} g_p^2(t)\, dt \tag{4.75}$$

Correspondingly, the formula for the power spectral density given in Eq. 4.60 takes on a discrete form defined in terms of the complex Fourier coefficients as

$$S_{g_p}(f) = \sum_{n=-\infty}^{\infty} |c_n|^2\, \delta\left(f - \frac{n}{T_0}\right) \tag{4.76}$$

Naturally, the power spectral density $S_{g_p}(f)$ has all the properties listed in Section 4.3. Moreover, it is *a discrete function of frequency*, which is a consequence of the periodic nature of the signal $g_p(t)$. Since the total area under a curve of power spectral density equals the average power, we may define the total average power of the periodic signal $g_p(t)$ in terms of its frequency-domain description as

$$P = \sum_{n=-\infty}^{\infty} |c_n|^2 \tag{4.77}$$

This relation is known as *Parseval's power theorem*. It states that the average power of a periodic signal $g_p(t)$ is equal to the sum of the squared amplitudes of all the harmonic components of the signal $g_p(t)$. Note that the Parseval power theorem, as with the Rayleigh energy theorem, requires knowledge of the amplitude spectrum only.

The power spectral density of Eq. 4.76 has a delta function at zero frequency, which is weighted by $|c_0|^2$. The presence of this delta function implies that the periodic signal $g_p(t)$ has *dc power*, given by

$$P_{dc} = |c_0|^2 \tag{4.78}$$

The coefficient c_0 equals the *mean* or *average value* of the periodic signal $g_p(t)$; that is,

$$c_0 = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} g_p(t)\, dt \tag{4.79}$$

The *ac power* of the periodic signal $g_p(t)$ is defined by the sum of the weights associated with the remaining delta functions in $S_{g_p}(f)$, as shown by

$$P_{ac} = \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} |c_n|^2 \tag{4.80}$$

The square root of P_{ac} defines the *root mean square (rms) value* of the signal. Naturally, the sum of dc power P_{dc} and ac power P_{ac} equals the total average power P.

When the power signal of interest is periodic, the integrand in the defining equation (4.64) for the autocorrelation function $R_g(\tau)$ of the signal is likewise periodic. Hence, the time average in this formula may be taken over one period. Thus, we may express the autocorrelation function of a periodic signal $g_p(t)$ of period T_0 as

$$R_{g_p}(\tau) = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} g_p(t)g_p(t-\tau)\,dt \tag{4.81}$$

The autocorrelation function $R_{g_p}(\tau)$ exhibits all the properties listed in Section 4.4 for the autocorrelation function of power signals. In addition, *the autocorrelation function $R_{g_p}(\tau)$ is periodic with the same period as the periodic signal $g_p(t)$ itself;* that is

$$R_{g_p}(\tau) = R_{g_p}(\tau \pm nT_0), \qquad n = 1, 2, \ldots \tag{4.82}$$

EXAMPLE 6 SINUSOIDAL WAVE

Consider the *sinusoidal wave*

$$g_p(t) = A\cos(2\pi f_c t + \theta) \tag{4.83}$$

which is plotted in Fig. 4.8*a*; the period $T_0 = 1/f_c$. The requirement is to evaluate the power spectral density, average power, and autocorrelation function of this sinusoidal wave.

To express the given sinusoidal wave as a complex Fourier series, we use the formula for a cosine function in terms of a pair of complex exponentials. We thus write

$$g_p(t) = c_1 \exp(j2\pi f_c t) + c_{-1}\exp(-j2\pi f_c t)$$

where

$$c_1 = \frac{A}{2}\exp(j\theta)$$

and

$$c_{-1} = \frac{A}{2}\exp(-j\theta)$$

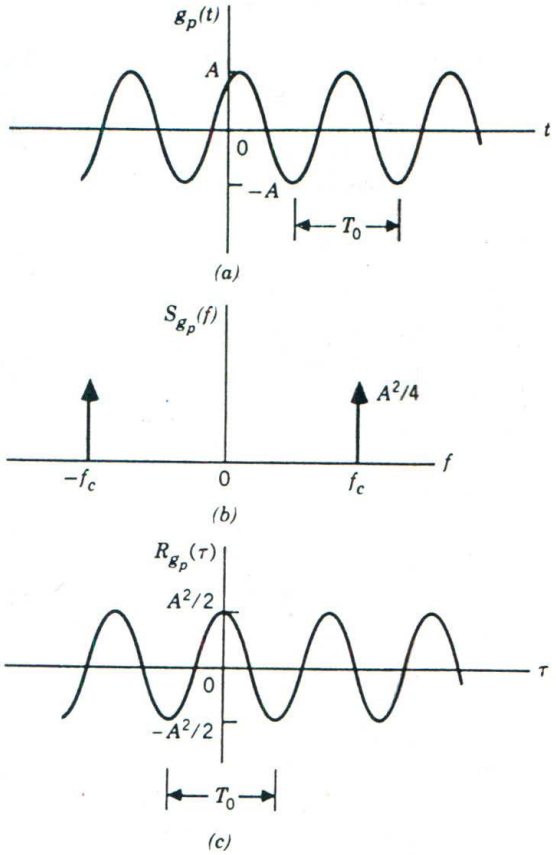


Figure 4.8
(a) *Sinusoidal wave.* (b) *Power spectral density.* (c) *Autocorrelation function.*

Hence, the use of Eq. 4.76 yields the power spectral duality:

$$S_{g_p}(f) = \frac{A^2}{4}\delta(f - f_c) + \frac{A^2}{4}\delta(f + f_c) \tag{4.84}$$

That is, the power spectral density of a sinusoidal wave consists of a pair of delta functions located at $f = \pm f_c$, both of which are weighted by the factor $A^2/4$, as depicted in Fig. 4.8*b*. Note that the power spectral density is independent of the phase θ of the sinusoidal wave.

By evaluating the total area under the power spectral density $S_{g_p}(f)$, we obtain the average power of the sinusoidal wave as

$$P = \frac{A^2}{2} \tag{4.85}$$

Finally, the use of Eq. 4.81 yields the autocorrelation function of the given sinusoidal wave as

$$\begin{aligned} R_{g_p}(\tau) &= A^2 f_c \int_{-1/2f_c}^{1/2f_c} \cos(2\pi f_c t + \theta) \cos(2\pi f_c t - 2\pi f_c \tau + \theta)\, dt \\ &= \frac{A^2 f_c}{2} \int_{-1/2f_c}^{1/2f_c} [\cos(2\pi f_c \tau) + \cos(4\pi f_c t - 2\pi f_c \tau + 2\theta)]\, dt \\ &= \frac{A^2}{2} \cos(2\pi f_c \tau) \end{aligned} \tag{4.86}$$

Equation 4.86 is plotted in Fig. 4.8*c*. It shows that the autocorrelation function of a sinusoidal wave is a sinusoidal function of τ, with the same period as the given sinusoidal wave. Moreover, putting $\tau = 0$ in Eq. 4.86, we find that $R_{g_p}(0) = P$, as expected.

EXERCISE 12 Show that the power spectral density and autocorrelation function of Eqs. 4.84 and 4.86 constitute a Fourier transform pair.

EXAMPLE 7 SQUARE WAVE

Consider next the square wave of Fig. 4.9*a*, one period of which is defined by

$$g_p(t) = \begin{cases} A, & -\dfrac{T_0}{4} \leq t \leq \dfrac{T_0}{4} \\ 0, & \text{for the remainder of the period} \end{cases} \tag{4.87}$$

The requirement is to determine the power spectral density and autocorrelation function of this square wave.

In Example 1, Chapter 2, we derived the formula for the complex Fourier coefficient c_n of a rectangular pulse train with arbitrary duty cycle; the result is given in Eq. 2.19. The square wave described here has a duty cycle of one-half. Hence, adapting Eq. 2.19 for this duty cycle, we find that the complex Fourier coefficient of the square wave of Fig. 4.9*a* is given by

$$c_n = \frac{A}{2} \operatorname{sinc}\left(\frac{n}{2}\right) \tag{4.88}$$

Substituting Eq. 4.88 in 4.76 yields the desired power spectral density of

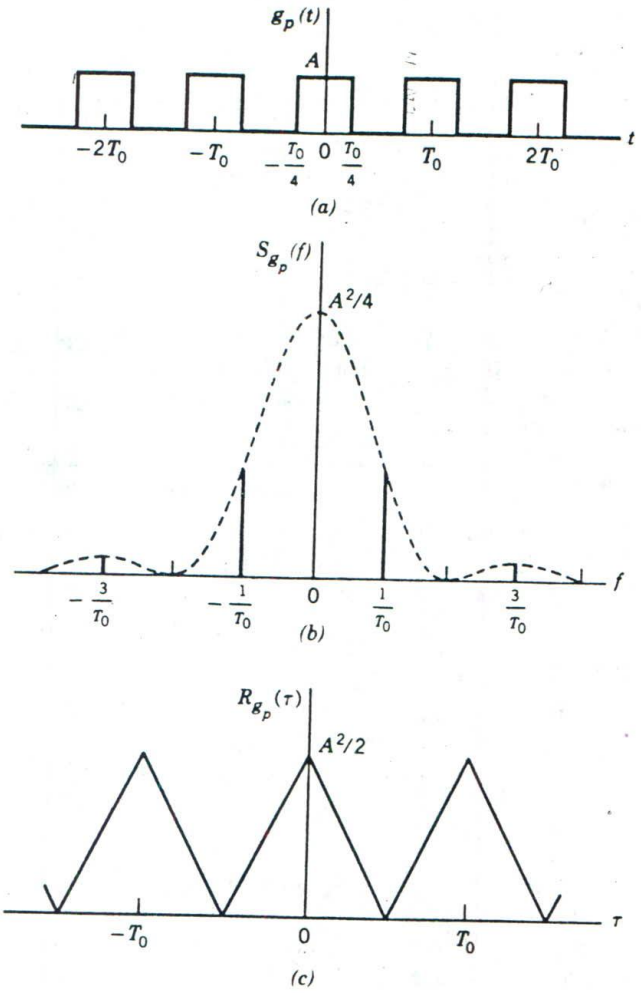


Figure 4.9
(a) *Square wave.* (b) *Power spectral density.* (c) *Autocorrelation function.*

the square wave as

$$S_{g_p}(f) = \frac{A^2}{4} \sum_{n=-\infty}^{\infty} \operatorname{sinc}^2\left(\frac{n}{2}\right) \delta\left(f - \frac{n}{T_0}\right) \tag{4.89}$$

which is plotted in Fig. 4.9*b*.

The most expedient approach for obtaining the autocorrelation function is to use the formula of Eq. 4.81. Figure 4.10 presents a graphical portrayal of the steps involved in the application of this formula for

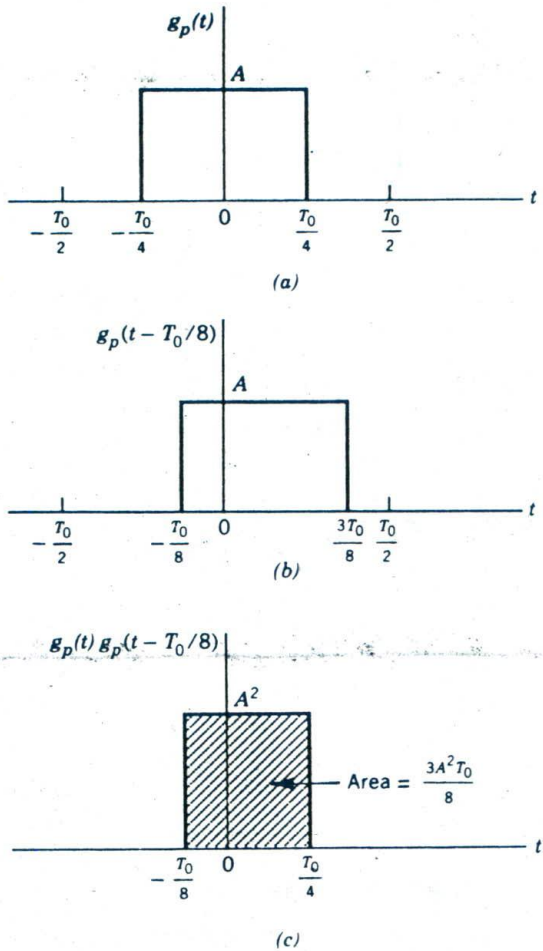


Figure 4.10
The computation of autocorrelation function $R_{g_p}(\tau)$ *for lag* $\tau = T_0/8$.

a delay $\tau = T_0/8$. Parts *a*, *b*, and *c* of the figure present plots of the square wave $g_p(t)$, its delayed version $g_p(t - T_0/8)$, and the product $g_p(t)g_p(t - T_0/8)$, respectively, for the period $-(T_0/2) \leq t \leq (T_0/2)$. The area under the product $g_p(t)g_p(t - T_0/8)$ for this period is shown shaded in Fig. 4.10*c*. Evaluating this area and scaling it by the factor $1/T_0$, we get

$$R_{g_p}\left(\frac{T_0}{8}\right) = \frac{3A^2}{8}$$

Proceeding in a similar manner for other values of delay τ, we obtain

$$R_{g_p}(\tau) = \begin{cases} \dfrac{A^2}{2}\left(1 + \dfrac{2\tau}{T_0}\right), & -\dfrac{T_0}{2} \leq \tau \leq 0 \\ \dfrac{A^2}{2}\left(1 - \dfrac{2\tau}{T_0}\right), & 0 \leq \tau \leq \dfrac{T_0}{2} \end{cases} \tag{4.90}$$

Recognizing that the autocorrelation function of a periodic wave with period T_0 is also periodic with the same period, we find that the use of Eq. 4.90 yields the plot shown in Fig. 4.9*c* for the autocorrelation function of the given square wave.

EXERCISE 13 Use Eq. 4.89 to illustrate the properties of the power spectral density of a periodic signal.

EXERCISE 14 Use Eq. 4.90 to illustrate the properties of the autocorrelation function of a periodic signal.

EXERCISE 15 Determine the power spectral density and autocorrelation function of a rectangular wave, one period of which is defined by

$$g_p(t) = \begin{cases} A, & -\dfrac{T_0}{8} \leq t \leq \dfrac{T_0}{8} \\ 0, & \text{for the remainder of the period} \end{cases}$$

4.7 SPECTRAL CHARACTERISTICS OF RANDOM SIGNALS AND NOISE

Random signals constitute another important class of power signals. We say a signal is random if there is *uncertainty* about the signal before it actually occurs. Such a signal may be viewed as belonging to an *ensemble* of signals, the generation of which is governed by a mechanism that is *probabilistic* in nature. Hence, no two signals in the ensemble exhibit the same variation with time. Each waveform (signal) in the ensemble is referred to as a *sample function*, and the ensemble of all possible sample functions is referred to as a *random process*.

Let $x(t)$ denote a sample function of a random process $X(t)$. Figure 4.11*a* shows a plot of the waveform of $x(t)$. Since $x(t)$ is a power signal, its Fourier transform does not exist. This necessitates dealing with a trun-

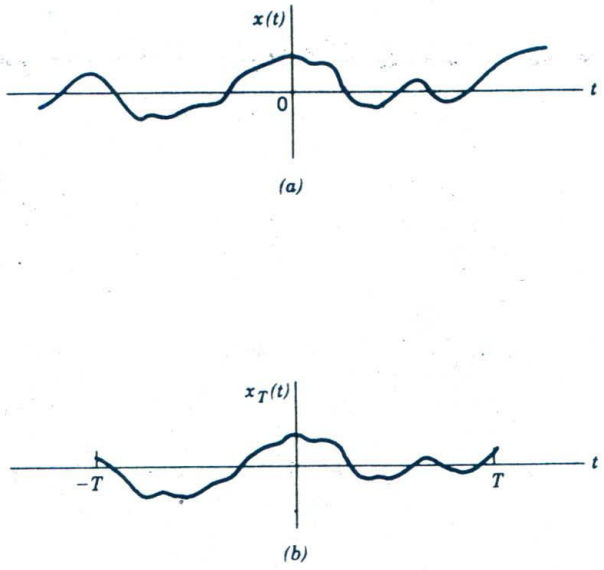


Figure 4.11
(a) *Sample function of a random process.* (b) *Truncated version of the sample function.*

cated version of the sample function, namely,

$$x_T(t) = \begin{cases} x(t), & -T \leq t \leq T \\ 0, & \text{otherwise} \end{cases} \tag{4.91}$$

Figure 4.11*b* depicts the truncated signal $x_T(t)$. From the discussion presented in Section 4.3, we note that the time average power spectral density of the sample function $x(t)$ over the interval $-T \leq t \leq T$ is $|X_T(f)|^2/2T$, where $X_T(f)$ is the Fourier transform of $x_T(t)$. This time-averaged power spectral density depends on the particular sample function $x(t)$ drawn from the random process $X(t)$. Accordingly, we must perform an *ensemble averaging* operation, and then take the limit as T approaches infinity. The value of frequency f is held fixed while averaging over the ensemble. The ensemble averaging operation requires using the probability distribution of the ensemble.[3] For the purpose of our present discussion, it is sufficient to acknowledge the ensemble averaging operation by using the operator E, commonly referred to as the *expectation operator*. We thus write the *ensemble-averaged* or *mean value* of $|X_T(f)|^2$ simply as $E[|X_T(f)|^2]$ and the

[3]The issue of ensemble averaging is considered in Chapter 8.

corresponding power spectral density of the random process $X(t)$ as

$$S_X(f) = \lim_{T \to \infty} \frac{1}{2T} E[|X_T(f)|^2] \tag{4.92}$$

It is important to note that in Eq. 4.92 the ensemble averaging must be performed before the limit is taken. Also, we have used an uppercase letter as the subscript for the power spectral density in Eq. 4.92 to distinguish this definition of power spectral density for a random process from that for a power signal of deterministic form.

Our involvement with random processes in this book will be in the context of noise analysis of communication systems. The term *noise* is used customarily to designate unwanted waveforms that tend to disturb the transmission and processing of signals in communication systems, and over which we have incomplete control. In practice, we find that there are many potential sources of noise in a communication system. The sources of noise may be external to the system (e.g., atmospheric noise, galactic noise, man-made noise), or internal to the system. The second category includes an important type of noise that arises owing to *spontaneous fluctuations* of current or voltage in electrical circuits. This type of noise, in one way or another, is present in every communication system and represents a basic limitation on the reliable transmission of information. It originates at the front end of the receiver part of the system; hence, it is commonly referred to as *receiver noise*.[4] It is also referred to as *channel noise*.

WHITE NOISE

The noise analysis of communication systems is customarily based on an idealized form of a noise process called *white noise*, the power spectral density of which is independent of frequency. The adjective *white* is used in the sense that white light contains equal amounts of all frequencies within the visible band of electromagnetic radiation. We denote the power spectral density of a white-noise process $W(t)$ as

$$S_W(f) = \frac{N_0}{2} \tag{4.93}$$

where the factor 1/2 has been included to indicate that half the power is associated with positive frequencies and half with negative frequencies, as illustrated in Fig. 4.12*a*. The dimensions of N_0 are in watts per hertz. The parameter N_0 is usually measured at the input stage of the receiver of a communication system.

[4]For a discussion of various types of noise encountered in communication systems, see Appendix C.

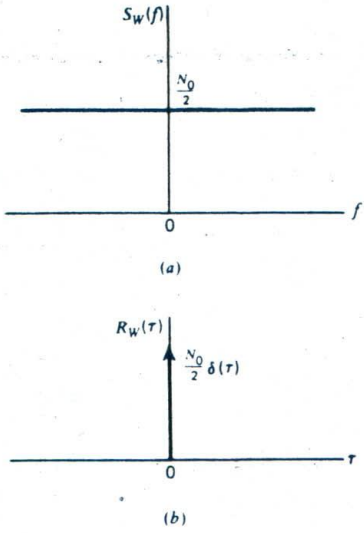


Figure 4.12
Characteristics of white noise. (a) *Power spectral density.* (b) *Autocorrelation function.*

The absence of a delta function in the power spectral density of Fig. 4.12*a* at the origin means that the white noise so described has no dc power. That is, its mean or average value is zero.

Since the autocorrelation function is the inverse Fourier transform of the power spectral density, it follows that for white noise

$$R_W(\tau) = \frac{N_0}{2}\delta(\tau) \tag{4.94}$$

That is, the autocorrelation function of white noise consists of a delta function weighted by the factor $N_0/2$ and occurring at $\tau = 0$, as in Fig. 4.12*b*. We note that $R_W(\tau)$ is zero for $\tau \neq 0$. Accordingly, any two different samples of white noise, no matter how close together in time they are taken, are uncorrelated.

Strictly speaking, white noise has infinite average power and, as such, it is not physically realizable. Nevertheless, white noise has convenient mathematical properties and therefore is useful in system analysis.

The utility of a white-noise process is parallel to that of an impulse function or delta function in the analysis of linear systems. The effect of an impulse is observed only after it has been passed through a system with finite bandwidth. Likewise, the effect of white noise is observed only after passing through a system with finite bandwidth. We may state, therefore, that as long as the bandwidth of a noise process at the input of a system

is appreciably larger than that of the system itself, we may model the noise process as white noise.

EXAMPLE 8 IDEAL LOW-PASS FILTERED WHITE NOISE

Suppose that a white-noise process $W(t)$ of zero mean and power spectral density $N_0/2$ is applied to an ideal low-pass filter of bandwidth B and a passband amplitude response of 1. The power spectral density of the noise process $N(t)$ appearing at the filter output is therefore (see Fig. 4.13*a*)

$$S_N(f) = \begin{cases} \dfrac{N_0}{2}, & -B < f < B \\ 0, & |f| > B \end{cases} \tag{4.95}$$

The autocorrelation function of $N(t)$ is the inverse Fourier transform of

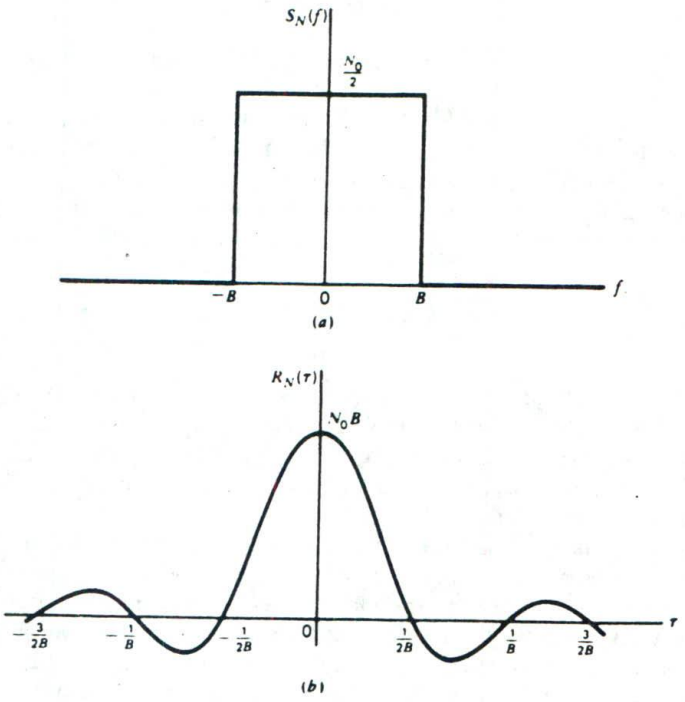


Figure 4.13
Characteristics of low-pass filtered white noise. (a) *Power spectral density.* (b) *Autocorrelation function.*

the power spectral density shown in Fig. 4.13*a*:

$$R_N(\tau) = \int_{-B}^{B} \frac{N_0}{2} \exp(j2\pi f\tau)\, df$$
$$= N_0 B \operatorname{sinc}(2B\tau) \tag{4.96}$$

This autocorrelation function is plotted in Fig. 4.13*b*. We see that $R_N(\tau)$ has its maximum value of N_0B at the origin, and it passes through zero at $\tau = \pm n/2B$, where $n = 1, 2, 3, \ldots$.

EXAMPLE 9 RC LOW-PASS FILTERED WHITE NOISE

Consider next a white-noise process $W(t)$ of zero mean and power spectral density $N_0/2$ applied to a low-pass filter, as in Fig. 4.14*a*. The transfer function of the filter is

$$H(f) = \frac{1}{1 + j2\pi fRC} \tag{4.97}$$

The power spectral density of the noise $N(t)$ appearing at the low-pass RC filter output is therefore (see Fig. 4.14*b*)

$$S_N(f) = \frac{N_0/2}{1 + (2\pi fRC)^2} \tag{4.98}$$

From Example 4 of Chapter 2, we have

$$\exp(-|\tau|) \rightleftharpoons \frac{2}{1 + (2\pi f)^2} \tag{4.99}$$

Therefore, using the time-scaling property of the Fourier transform, we

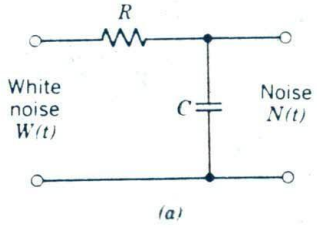
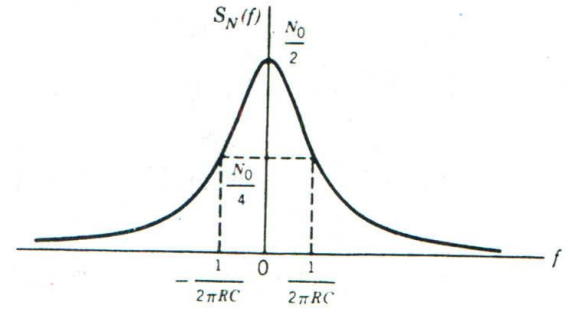
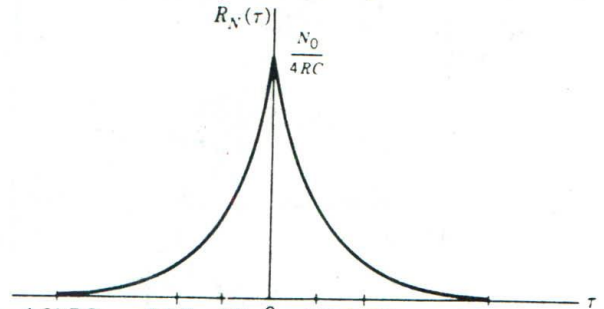


Figure 4.14
Characteristics of RC-filtered white noise. (a) *Low-pass RC filter.* (b) *Power spectral density of filter output* N(t). (c) *Autocorrelation function of* N(t).



(b)



(c)

Figure 4.14 (*continued*)

find that the autocorrelation function of the filtered noise process $N(t)$ is

$$R_N(\tau) = \frac{N_0}{4RC} \exp\left(-\frac{|\tau|}{RC}\right) \tag{4.100}$$

which is plotted in Fig. 4.14c.

EXERCISE 16 Using the autocorrelation function of Eq. 4.100, find the average power of the RC filter output.

EXERCISE 17 Using the power spectral density of Eq. 4.98, find the average power of the RC filter output. Check this result against that of Exercise 16.

EXAMPLE 10 AUTOCORRELATION OF A SINUSOIDAL WAVE PLUS WHITE NOISE

Consider a random process $X(t)$ consisting of a sinusoidal wave component and a white-noise process of zero mean and power spectral density $N_0/2$. A sample function (i.e., single realization) of $X(t)$ is denoted by

$$x(t) = A\cos(2\pi f_c t + \theta) + w(t) \tag{4.101}$$

The phase θ of the sinusoidal component may lie anywhere inside the interval $-\pi \leq \theta \leq \pi$ with equal likelihood. The problem is to determine the autocorrelation function of the random process $X(t)$ represented by the sample function $x(t)$.

The two components of $x(t)$ originate from independent sources. Therefore, the autocorrelation function of $X(t)$ is the sum of the individual autocorrelation functions of the sinusoidal wave and white-noise components. In Example 6, we showed that the autocorrelation function of the sinusoidal component is equal to $(A^2/2)\cos(2\pi f_c \tau)$. The autocorrelation function of the white-noise component is equal to $(N_0/2)\delta(\tau)$. We may therefore write

$$R_X(\tau) = \frac{A^2}{2}\cos(2\pi f_c \tau) + \frac{N_0}{2}\delta(\tau) \tag{4.102}$$

which is plotted in Fig. 4.15. We thus see that for $|\tau| > 0$, the autocorrelation function of the random process $X(t)$ is the same as that of the sinusoidal wave component. This shows that by determining the autocorrelation function of $X(t)$ we can detect the presence of a periodic signal component that is corrupted by additive white noise.

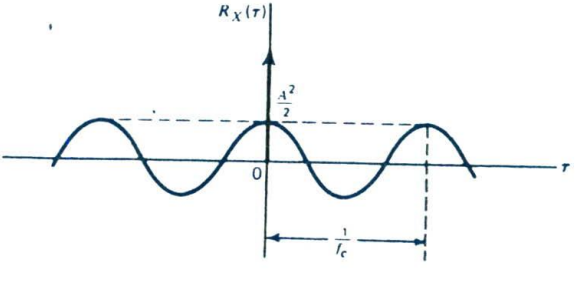


Figure 4.15
Autocorrelation function of sinusoidal wave plus white noise.

4.8 NOISE-EQUIVALENT BANDWIDTH

In Example 8 we observed that when a source of white noise of zero mean and power spectral density $N_0/2$ is connected across the input of an ideal low-pass filter of bandwidth B and passband amplitude response of one, the average output noise power [or equivalently $R_N(0)$] is equal to N_0B. In Example 10 we observed that when such a similar noise source is connected to the input of the simple RC low-pass filter of Fig. 4.14*a*, the corresponding value of the average output noise power is equal to $N_0/(4RC)$. For this filter, the half-power or 3-dB bandwidth is equal to $1/(2\pi RC)$. We may therefore make two important observations. First, filtered white noise has *finite* average power. Second, the average power is proportional to bandwidth.

We may generalize these observations to include all kinds of low-pass filters by defining a noise-equivalent bandwidth as follows. Suppose that we have a source of white noise of zero mean and power spectral density $N_0/2$ connected to the input of an arbitrary low-pass filter of transfer function $H(f)$. The resulting average output noise power is therefore

$$\begin{aligned} P_N &= \frac{N_0}{2}\int_{-\infty}^{\infty} |H(f)|^2\, df \\ &= N_0\int_0^{\infty} |H(f)|^2\, df \end{aligned} \tag{4.103}$$

where, in the last line, we have made use of the fact that the amplitude response $|H(f)|$ is an even function of frequency. Consider next the same source of white noise connected to the input of an ideal low-pass filter of zero-frequency response $H(0)$ and bandwidth B_N. In this case, the average output noise power is

$$P_N = N_0 B_N H^2(0) \tag{4.104}$$

Equation 4.104 shows that the filtered noise power P_N is finite and proportional to bandwidth B_N. The bandwidth B_N is called the *noise-equivalent bandwidth* for a low-pass filter; its definition follows directly from Eqs. 4.103 and 4.104 as

$$B_N = \frac{\displaystyle\int_0^{\infty} |H(f)|^2\, df}{H^2(0)} \tag{4.105}$$

Thus the procedure for calculating the noise-equivalent bandwidth consists of replacing the arbitrary low-pass filter of transfer function $H(f)$ by an equivalent ideal low-pass filter of zero-frequency response $H(0)$ and bandwidth B_N, as illustrated in Fig. 4.16.

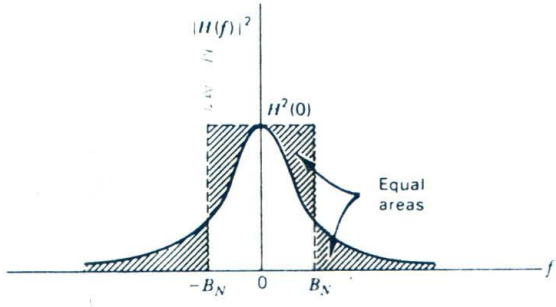


Figure 4.16
The definition of noise-equivalent bandwidth for low-pass filter.

In a similar way, we may define a noise-equivalent bandwidth for a band-pass filter, as illustrated in Fig. 4.17; this figure depicts the squared amplitude response of the filter for positive frequencies only. Thus, the noise-equivalent bandwidth for a band-pass filter may be defined as

$$B_N = \frac{\int_0^\infty |H(f)|^2 \, df}{|H(f_c)|^2} \tag{4.106}$$

where $|H(f_c)|$ is the center-frequency amplitude response of the filter.

We may combine the definitions of Eqs. 4.105 and 4.106 for the noise-

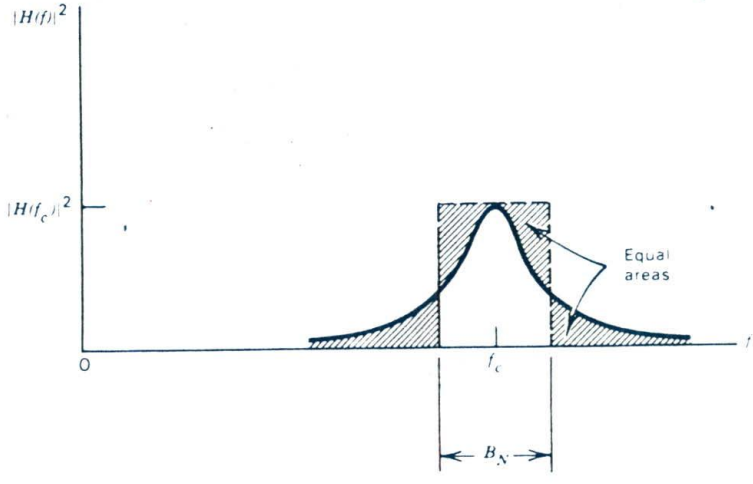


Figure 4.17
The definition of noise equivalent bandwidth for band-pass filters; only the response for positive frequencies is shown.

equivalent bandwidth of low-pass and band-pass filters into a single formula:

$$B_N = \frac{1}{g_a}\int_0^\infty |H(f)|^2\, df \tag{4.107}$$

where $|H(f)|$ is the amplitude response of the filter. The parameter g_a in Eq. 4.107 is the *maximum available power gain* of the filter, defined by

$$\begin{aligned} g_a &= \text{maximum value of } |H(f)|^2 \\ &= \begin{cases} |H(0)|^2, & \text{low-pass filter} \\ |H(f_c)|^2, & \text{band-pass filter} \end{cases} \end{aligned} \tag{4.108}$$

Correspondingly, we may express the output noise power of a filter (for both positive and negative frequencies) as

$$P_N = N_0 g_a B_N \tag{4.109}$$

where $N_0/2$ is the noise power spectral density at the filter input. According to Eq. 4.109, the effect of passing white noise through a filter may be separated into two parts:

1. The maximum available power gain of the filter, g_a.
2. The noise-equivalent bandwidth B_N, representing *relative frequency selectivity* of the filter.

Eq. 4.109 also shows that, whether the filter of interest is low-pass or band-pass, the filtered noise power P_N is proportional to the noise-equivalent bandwidth B_N. Hence, as a general rule, we may state that the effect of noise in a system (e.g., communication receiver) is reduced by narrowing the system bandwidth.

EXERCISE 18 What is the noise-equivalent bandwidth of the *RC* low-pass filter of Fig. 4.14*a*? Express your answer in terms of the 3-dB bandwidth of the filter.

PROBLEMS

P4.1 Energy Spectral Density

Problem 1 Consider the decaying exponential pulse

$$g(t) = \begin{cases} \exp(-at), & t > 0 \\ \dfrac{1}{2}, & t = 0 \\ 0, & t < 0 \end{cases}$$

Find the percentage of the total energy of $g(t)$ contained inside the frequency band $-W \leq f \leq W$, where $W = a/2\pi$.

Problem 2 Show that the two different pulses defined in parts *a* and *b* of Fig. P4.1 have the same energy spectral density:

$$\psi_g(f) = \frac{4A^2T^2\cos^2(\pi Tf)}{\tau^2(4T^2f^2 - 1)^2}$$

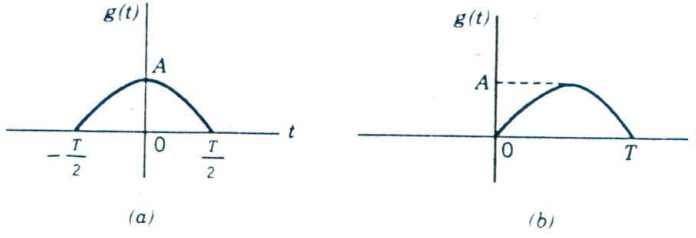


Figure P4.1

P4.2 Correlation of Energy Signals

Problem 3 Determine and sketch the autocorrelation functions of the following exponential pulses:

(a) $g(t) = \exp(-at)u(t)$
(b) $g(t) = \exp(-a|t|)$
(c) $g(t) = \exp(-at)u(t) - \exp(at)u(-t)$

where $u(t)$ is the unit step function, and $u(-t)$ is its time-reversed version.

Problem 4 Determine and sketch the autocorrelation function of a Gaussian pulse defined by

$$g(t) = \frac{1}{t_0}\exp\left(-\frac{\pi t^2}{t_0^2}\right)$$

Problem 5 The Fourier transform of a signal is defined by $|\text{sinc}(f)|$. Show that the autocorrelation function of this signal is triangular in form.

Problem 6 Specify two distinctly different pulse signals that have exactly the same autocorrelation function.

Problem 7 Consider a signal $g(t)$ defined by

$$g(t) = A_0 + A_1\cos(2\pi f_1 t + \theta_1) + A_2\cos(2\pi f_2 t + \theta_2)$$

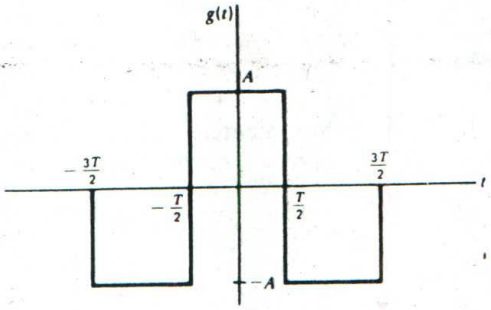


Figure P4.2

(a) Determine the autocorrelation function $R_g(\tau)$ of this signal.
(b) What is the value of $R_g(0)$?
(c) Has any information about $g(t)$ been lost in obtaining the autocorrelation function?

Problem 8 Determine the autocorrelation function of the triplet pulse shown in Fig. P4.2

Problem 9 Let $G(f)$ denote the Fourier transform of a real-valued energy signal $g(t)$, and $R_g(\tau)$ its autocorrelation function. Show that

$$\int_{-\infty}^{\infty} \left[\frac{dR_g(\tau)}{d\tau}\right]^2 d\tau = 4\pi^2 \int_{-\infty}^{\infty} f^2 |G(f)|^4 \, df$$

Problem 10 Determine the cross-correlation function $R_{12}(\tau)$ of the pair of rectangular pulses shown in Fig. P4.3, and sketch it. What is $R_{21}(\tau)$?

Problem 11 Determine the cross-correlation function $R_{12}(\tau)$ of the rectangular pulse $g_1(t)$ and triplet pulse $g_2(t)$ shown in Fig. P4.4, and sketch it. What is $R_{21}(\tau)$? Are these signals orthogonal? Why?

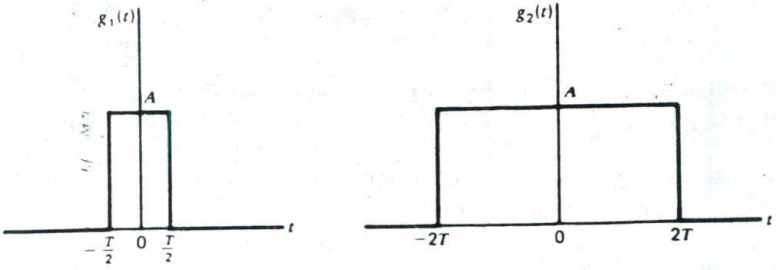


Figure P4.3

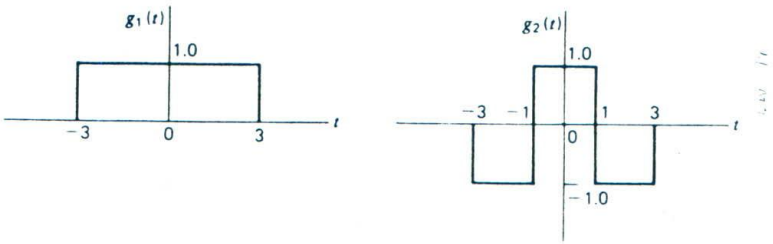


Figure P4.4

Problem 12 Consider two signals $g_1(t)$ and $g_2(t)$. These two signals are delayed by amounts equal to t_1 and t_2 seconds, respectively. Show that the time delays are additive in convolving the pair of delayed signals, whereas they are subtractive in cross-correlating them.

P4.3 Power Spectral Density

Problem 13 Consider the truncated version of a complex exponential, defined by

$$g_T(t) = A \exp(j2\pi f_c t) \operatorname{rect}\left(\frac{t}{2T}\right)$$

where $\operatorname{rect}(t/2T)$ is a rectangular function of unit amplitude and duration $2T$. Find the power spectral density of $g_T(t)$ for finite T. What is the limiting value of this power spectral density as T approaches infinity?

Problem 14 Figure P4.5 shows the power spectral density of a power signal $g(t)$. Find the average power of the signal.

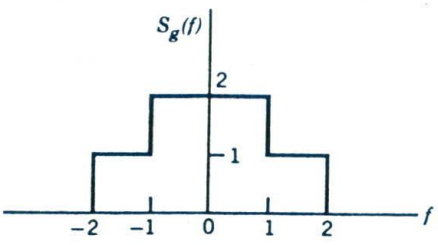


Figure P4.5

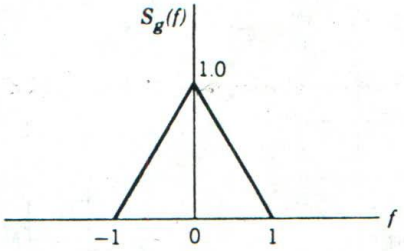


Figure P4.6

P4.4 Correlation of Power Signals

Problem 15 Find the autocorrelation function of the truncated version of a complex exponential, defined in Problem 13. What is the limiting value of this autocorrelation as T approaches infinity?

Problem 16 Find the autocorrelation function of a power signal $g(t)$ whose power spectral density is depicted in Fig. P4.6. What is the value of this autocorrelation function at the origin?

P4.6 Spectral Characteristics of Periodic Signals

Problem 17 Consider the square wave shown in Fig. P4.7. Find the power spectral density, average power, and autocorrelation function of this square wave. Does the wave have dc power? Explain your answer.

Problem 18 Consider two periodic signals $g_{p1}(t)$ and $g_{p2}(t)$ that have the following complex Fourier series representations:

$$g_{p1}(t) = \sum_{n=-\infty}^{\infty} c_{1,n} \exp\left(-\frac{j2\pi nt}{T_0}\right)$$

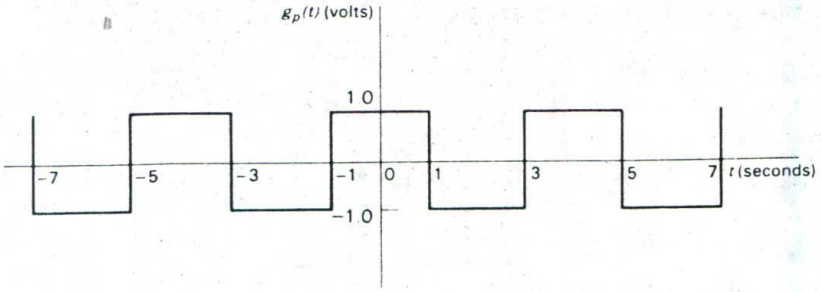


Figure P4.7

and

$$g_{p2}(t) = \sum_{n=-\infty}^{\infty} c_{2,n} \exp\left(-\frac{j2\pi nt}{T_0}\right)$$

The two signals have a common period equal to T_0.

Using the following definition of cross-correlation for a pair of periodic signals,

$$R_{12}(\tau) = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} g_{p1}(t) g_{p2}(t - \tau)\, dt$$

show that the prescribed pair of periodic signals satisfies the Fourier transform pair

$$R_{12}(\tau) \rightleftharpoons \sum_{n=-\infty}^{\infty} c_{1,n} c_{2,n}^{*}\, \delta\left(f - \frac{n}{T_0}\right)$$

P4.7 Spectral Characteristics of Random Signals and Noise

Problem 19 The power spectral density of a random process $X(t)$ is shown in Fig. P4.8.

(a) What is the dc power contained in this random process?
(b) What is the ac power contained in it?

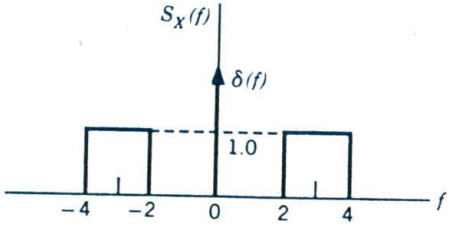


Figure P4.8

Problem 20 A white noise process of zero mean and power spectral density $N_0/2$ is applied to the low-pass RL filter shown in Fig. 4.9. Deter-

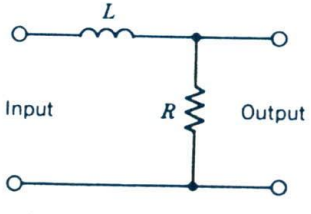


Figure P4.9

mine the power spectral density and autocorrelation function of the output.

Problem 21 Consider a white-noise process of zero mean and power spectral density $N_0/2$ applied to the input of the system shown in Fig. P4.10.

(a) Find the power spectral density of the random process at the output of the system.
(b) What is the average power of this output?

Hint: You may use Eq. 4.52, interpreted for a random process, to evaluate the power spectral density of the low-pass filter input.

P4.8 Noise-Equivalent Bandwidth

Problem 22 Find the noise-equivalent bandwidth for the low-pass *RL* filter shown in Fig. P4.9.

Problem 23 A white-noise process of power spectral density $N_0/2$ is applied to a *Butterworth* low-pass filter of order n with its amplitude response

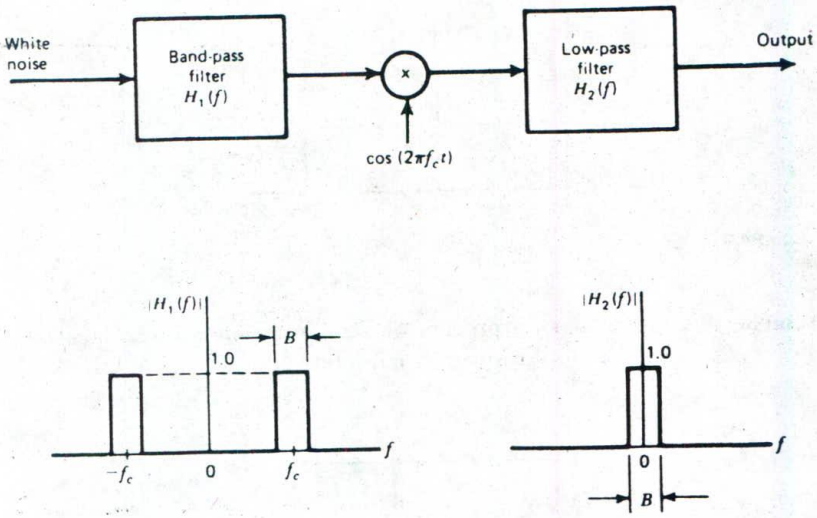


Figure P4.10

defined by

$$|H(f)| = \frac{1}{[1 + (f/f_0)^{2n}]^{1/2}}$$

(a) Determine the noise-equivalent bandwidth for this low-pass filter.
(b) What is the limiting value of the noise-equivalent bandwidth as n approaches infinity?

DIGITAL CODING OF ANALOG WAVEFORMS

To transport an information-bearing signal from one point to another over a communication channel, we may use digital or analog techniques. As mentioned in Chapter 1, the use of digital communications offers several important advantages as compared to analog communications. In particular, a digital communication system offers the following highly attractive features:

1. *Ruggedness* to channel noise and external interference, unmatched by any analog communication system.
2. *Flexible* operation of the system.
3. *Integration* of diverse sources of information into a common format.
4. *Security* of information in the course of its transmission from source to destination.

For these reasons, digital communications have become the dominant form of communication technology in our society.

To handle the transmission of analog message signals (e.g., voice and video signals) by digital means, the signal has to undergo an *analog-to-digital conversion*. In the next section, we present an overview of three important methods of analog-to-digital conversion, which are known as *pulse-code modulation, differential pulse-code modulation,* and *delta modulation*. Their detailed descriptions are presented in subsequent sections of the chapter.

5.1 DIGITAL PULSE MODULATION

The process of analog-to-digital conversion is sometimes referred to as *digital pulse modulation*. The use of the terminology "pulse modulation" is justified by virtue of the fact that the first operation performed in the conversion of an analog signal into digital form involves the representation of the signal by a sequence of uniformly spaced *pulses,* the amplitude of which is *modulated* by the signal. Naturally, the pulse-repetition frequency must be chosen in accordance with the *sampling theorem*. In both pulse-code modulation and differential pulse-code modulation, the pulse-repetition frequency or the sampling rate is chosen to be *slightly greater* than the Nyquist rate (i.e., greater than twice the highest frequency component) of the analog signal. In delta modulation, on the other hand, the sampling rate is purposely chosen to be *much greater* than the Nyquist rate. The reason for such a choice in the latter case is to increase correlation between adjacent samples derived from the information-bearing analog signal and thereby to simplify the physical implementation of the delta modulation process. The distinguishing feature between pulse-code modulation and differential pulse-code modulation is that in the latter case, additional circuitry (designed to perform linear prediction) is used to exploit the correlation between adjacent samples of the analog signal so as to reduce the transmitted bit rate.

Figure 5.1 summarizes the comparison between delta modulation, pulse-code modulation, and differential pulse-code modulation in the context of two important system features: *circuit complexity* and *transmitted bit rate*. The bit rate refers to the rate at which *bits* (*bi*nary digi*ts*) constituting the digital version of an analog information-bearing signal are transmitted over the communication channel.

Pulse-code modulation is usually viewed as a *benchmark* against which other methods of digital pulse modulation are measured in performance and circuit complexity. It is therefore appropriate that we begin our study of digital pulse modulation by considering the operations involved in pulse-code modulation, which we do in the next section.

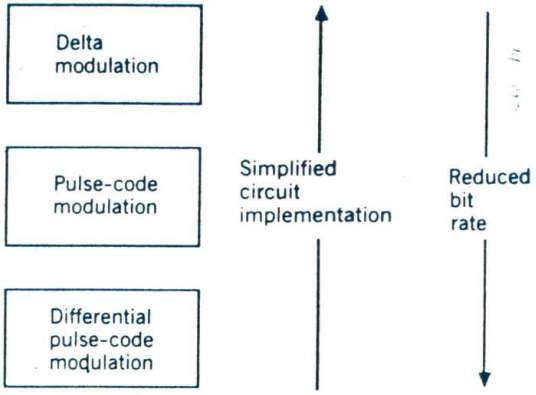


Figure 5.1
Diagrammatic comparison of the three basic forms of digital pulse modulation.

5.2 PULSE-CODE MODULATION

Pulse-code modulation[1] (PCM) is complex in the sense that the message signal is subjected to a great number of operations. The essential operations in the transmitter of a PCM system are *sampling*, *quantizing*, and *encoding*, as shown in Fig. 5.2. The quantizing and encoding operations are usually performed in the same circuit, which is called an *analog-to-digital converter*. The essential operations in the receiver are *regeneration* of impaired signals, *decoding*, and *demodulation* of the train of quantized samples. These operations are usually performed in the same circuit, which is called a *digital-to-analog converter*. At intermediate points along the transmission route from the transmitter to the receiver, *regenerative repeaters* are used to reconstruct (regenerate) the transmitted sequence of coded pulses in order to combat the accumulated effects of signal distortion and noise.

Quantizing refers to the use of a *finite set of amplitude levels* and the selection of a level nearest to a particular sample value of the message signal as the *representation* for it. This operation, combined with sampling, permits the use of *coded pulses* for representing the message signal. Indeed, it is the combined use of quantizing and coding that distinguishes pulse-code modulation from analog modulation techniques.

In the next three sections, we discuss the operations of sampling, quantizing, and coding, in that order.

[1]Pulse-code modulation is the oldest method for analog-to-digital conversion. It was invented by Reeves in 1937. For a historical account of this invention, see Reeves (1975).

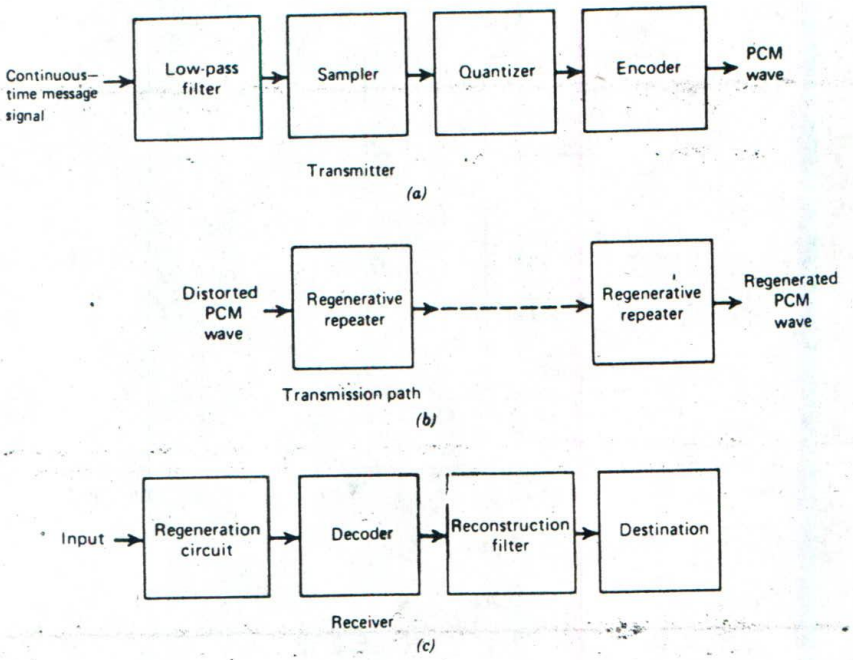


Figure 5.2
The basic elements of a PCM system. (a) *Transmitter.* (b) *Transmission path.* (c) *Receiver.*

5.3 SAMPLING

The sampling operation is performed in accordance with the *sampling theorem*. Specifically, we may state the sampling theorem for band-limited signals of finite energy in two equivalent parts (see Section 2.7):

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

Part 1 of the sampling theorem is exploited in the transmitter; part 2 of the theorem is exploited in the receiver. The sampling rate $2W$ is called the *Nyquist rate*, and its reciprocal $1/2W$ is called the *Nyquist interval*.

The derivation of the sampling theorem, presented in Section 2.7, was based on the assumption that the message signal of interest is strictly band-limited. In practice, however, the amplitude spectrum of the signal approaches zero asymptotically as the frequency approaches infinity, as il-

lustrated in Fig. 5.3*a*. This factor gives rise to an effect called *aliasing* or *fold-over,* which refers to a high-frequency component in the spectrum of the message signal apparently taking on the identity of a lower frequency in the spectrum of a sampled version of the signal.

The aliasing effect is illustrated in Fig. 5.3*b*. This figure shows the message spectrum and two frequency-shifted replicas of it; one replica is shifted to the right by the sampling rate $f_s = 2W$, and the other replica is shifted to the left by f_s. These replicas are manifestations of the periodic spectrum that results from sampling the message signal at the rate f_s; see Section 2.7. Inspection of the spectrum of the sampled signal, which is the sum of the message spectrum and its frequency-shifted replicas, shows that we are no longer able to recover the original message spectrum without distortion, owing to the presence of aliasing.

The presence of aliasing results in signal *distortion*. To combat the effects of aliasing in practice, we use two corrective measures:

1. Prior to sampling, a low-pass *pre-alias filter* is used to attenuate those high-frequency components of the signal that lie outside the band of interest.
2. The filtered signal is sampled at a rate higher than the Nyquist rate.

Figure 5.4 is the block diagram of a system for performing the sampling process. The low-pass pre-alias filter is included at the input of the sampling system, in accordance with point 1. The sampling rate is determined in accordance with point 2 by setting the pulse repetition frequency f_s of the

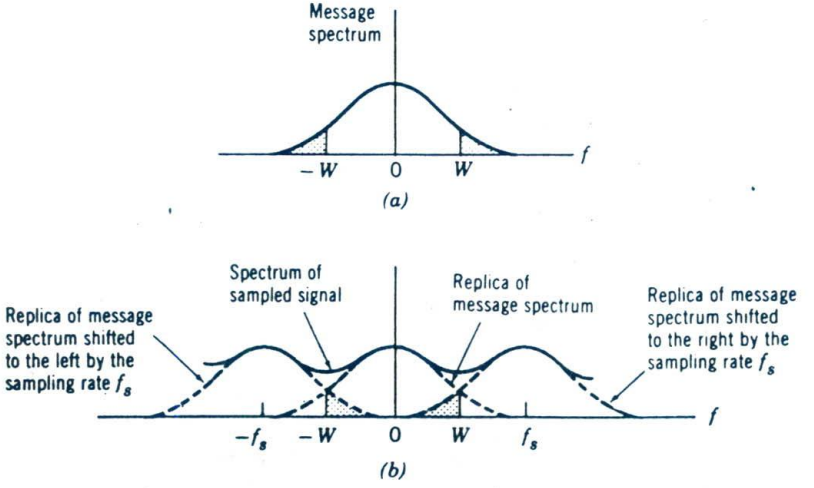


Figure 5.3
The aliasing effect. (a) ***Message spectrum.*** (b) ***Spectrum of sampled signal.***

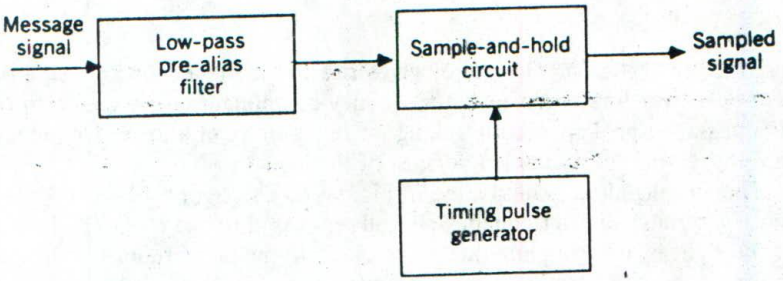


Figure 5.4
Practical sampling circuit arrangement.

timing pulse generator at a value greater than the Nyquist rate $2W$, where W is the pre-alias filter bandwidth.

SAMPLE-AND-HOLD CIRCUIT

The generation of samples is actually performed by a functional block termed the *sample-and-hold circuit* in Fig. 5.4. This circuit produces *flat-top samples* rather than the idealized instantaneous samples as postulated by the sampling theorem. Basically, the sample-and-hold circuit consists of two field-effect transistor (FET) switches and a capacitor connected together as in Fig. 5.5*a*. The "sampling switch" is closed briefly by a short pulse applied to gate G_1 of one transistor. The capacitor is thereby quickly charged up to a voltage equal to the instantaneous sample value of the incoming signal. It holds the sampled voltage until discharged by a pulse applied to gate G_2 of the other transistor. The output of the sample-and-hold circuit thus consists of a sequence of flat-top samples, as depicted in Fig. 5.5*b*.

PULSE-AMPLITUDE MODULATION

The sequence of flat-top samples depicted as $s(t)$ in Fig. 5.5*b* represents a pulse-amplitude modulated wave. In *pulse-amplitude modulation* (*PAM*), *the amplitudes of regularly spaced rectangular pulses vary with the instantaneous sample values of an analog message signal in a one-to-one fashion.*[2]

[2]Pulse-amplitude modulation is one basic type of *analog pulse modulation*. There are two other basic types of analog pulse modulation: pulse-duration modulation and pulse-position modulation. In *pulse-duration modulation* (*PDM*), *the samples of the message signal are used to vary the duration of the individual rectangular pulses*. This form of modulation is also referred to as pulse-width modulation or pulse-length modulation. In *pulse-position modulation* (*PPM*) *the position of a pulse relative to its unmodulated time of occurrence is varied in accordance with the message signal*. For a detailed discussion of analog pulse modulation techniques, see Carlson (1986, Chapter 10) or Black (1953).

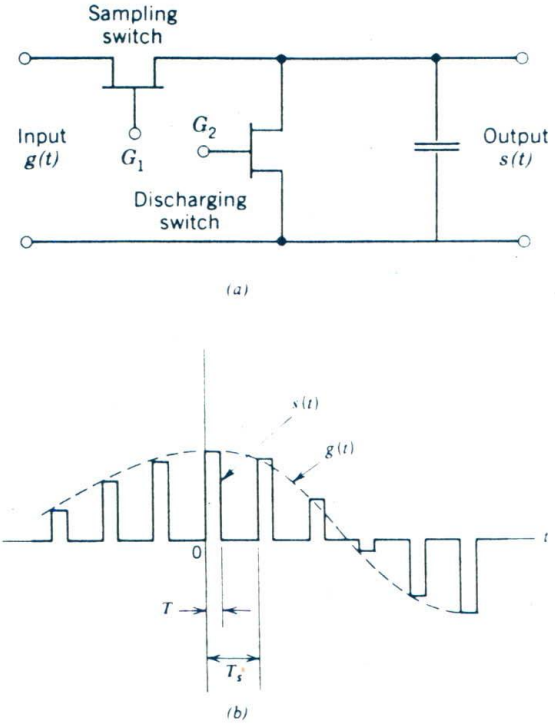


Figure 5.5
(a) *Sample-and-hold circuit.* (b) *Flat-top samples.*

The waveform denoted by $s(t)$ in Fig. 5.5*b* befits this definition exactly. Note that in PAM the carrier wave consists of a periodic train of rectangular pulses, and the carrier frequency (i.e., the pulse repetition frequency) is the same as the sampling rate.

For a mathematical representation of the PAM wave $s(t)$, we may write

$$s(t) = \sum_{n=-\infty}^{\infty} g(nT_s)h(t - nT_s) \tag{5.1}$$

The term $h(t)$ is a rectangular pulse of unit amplitude and duration T, defined as follows (see Fig. 5.6*a*)

$$h(t) = \begin{cases} 1, & 0 < t < T \\ \frac{1}{2}, & t = 0, t = T \\ 0, & \text{otherwise} \end{cases} \tag{5.2}$$

The term $g(nT_s)$ is the value of the input signal $g(t)$ at time $t = nT_s$. The *instantaneously sampled* version of the signal $g(t)$ is given by

$$g_\delta(t) = \sum_{n=-\infty}^{\infty} g(nT_s)\delta(t - nT_s) \tag{5.3}$$

Convolving $g_\delta(t)$ with the pulse $h(t)$, we get

$$\begin{aligned} g_\delta(t) \star h(t) &= \int_{-\infty}^{\infty} g_\delta(\tau)h(t - \tau)\, d\tau \\ &= \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} g(nT_s)\delta(\tau - nT_s)h(t - \tau)\, d\tau \\ &= \sum_{n=-\infty}^{\infty} g(nT_s) \int_{-\infty}^{\infty} \delta(\tau - nT_s)h(t - \tau)\, d\tau \end{aligned}$$

Using the sifting property of the delta function, we thus obtain

$$g_\delta(t) \star h(t) = \sum_{n=-\infty}^{\infty} g(nT_s)h(t - nT_s) \tag{5.4}$$

Therefore, from Eqs. 5.1 and 5.4 it follows that $s(t)$ is mathematically equivalent to the convolution of $g_\delta(t)$, the instantaneously sampled version of $g(t)$, and the pulse $h(t)$, as shown by

$$s(t) = g_\delta(t) \star h(t) \tag{5.5}$$

Taking the Fourier transform of both sides of Eq. 5.5 and recognizing that the convolution of two time functions is transformed into the multiplication of their respective Fourier transforms, we get

$$S(f) = G_\delta(f)H(f) \tag{5.6}$$

where $S(f) = \mathrm{F}[s(t)]$, $G_\delta(f) = \mathrm{F}[g_\delta(t)]$, and $H(f) = \mathrm{F}[h(t)]$. In Section 2.7 we showed that instantaneous sampling of the time function $g(t)$ introduces periodicity into the spectrum, as described in Eq. 2.131. This equation is reproduced here in the form

$$G_\delta(f) = f_s \sum_{m=-\infty}^{\infty} G(f - mf_s) \tag{5.7}$$

where $f_s = 1/T_s$ is the sampling rate. Therefore, substitution of Eq. 5.7 into 5.6 yields

$$S(f) = f_s \sum_{m=-\infty}^{\infty} G(f - mf_s)H(f) \tag{5.8}$$

where $G(f) = \mathrm{F}[g(t)]$.

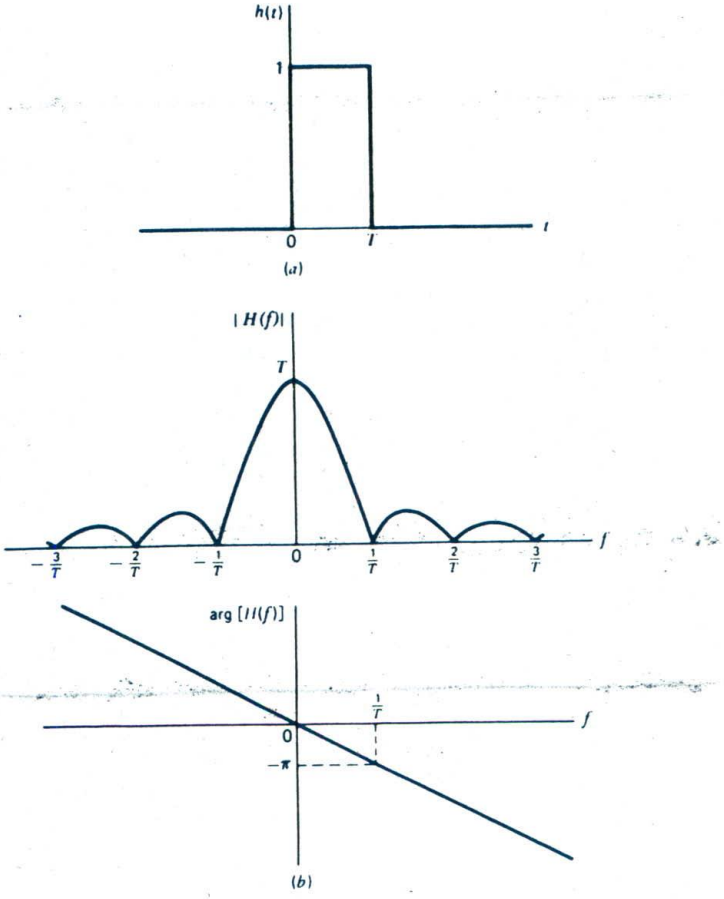


Figure 5.6
(a) *Rectangular pulse* h(t). (b) *Spectrum* H(f).

Finally, suppose that $g(t)$ is strictly band-limited and that the sampling rate f_s is greater than the Nyquist rate. Then, passing $s(t)$ through a low-pass reconstruction filter, we find that the spectrum of the resulting filter output is equal to $G(f)H(f)$. This is equivalent to passing the original analog signal $g(t)$ through a low-pass filter of transfer function $H(f)$.

From Eq. 5.2 we find that

$$H(f) = T\,\mathrm{sinc}(fT)\exp(-j\pi fT) \tag{5.9}$$

which is plotted in Fig. 5.6*b*. Hence, we see that by using pulse-amplitude modulation to represent an analog message signal we introduce *amplitude distortion* as well as a *delay* of $T/2$. This effect is similar to that caused by the finite size of the scanning aperture in television and facsimile. Ac-

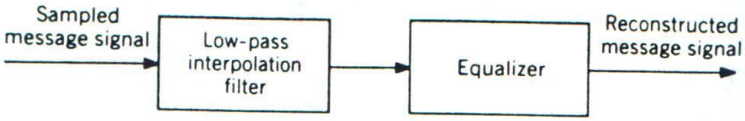


Figure 5.7
Block diagram of reconstruction circuit.

cordingly, the distortion caused by the use of flat-top samples in the generation of a PAM wave, as in Fig. 5.5*b*, is referred to as the *aperture effect.*

EXERCISE 1 What happens to the transfer function $H(f)/T$ of Eq. 5.9 as the pulse duration T approaches zero?

RECONSTRUCTION

Since sampling of the incoming message signal is the first basic operation performed in a PCM transmitter, *reconstruction* of the message signal is the final operation performed in the PCM receiver. Figure 5.7 is a block diagram of the circuitry used to perform this reconstruction. It consists of two components connected in cascade. The first component is a low-pass *interpolation filter* with a bandwidth that equals the message bandwidth W. The second component is an *equalizer* that corrects for the aperture effect due to flat-top sampling in the sample-and-hold circuit. The equalizer has the effect of decreasing the in-band loss of the interpolation filter as the frequency increases in such a manner as to compensate for the aperture effect. Ideally, the amplitude response of the equalizer is given by

$$\frac{1}{|H(f)|} = \frac{1}{T\,\mathrm{sinc}(fT)} = \frac{1}{T}\,\frac{\pi fT}{\sin(\pi fT)}$$

where $H(f)$ is the transfer function defined in Eq. 5.9. The amount of equalization needed in practice is usually small.

EXAMPLE 1

At $f = f_s/2$, which corresponds to the highest frequency component of the message signal for a sampling rate equal to the Nyquist rate, we find from Eq. 5.9 that the amplitude response of the equalizer normalized to that at zero frequency is equal to

$$\frac{1}{\mathrm{sinc}(0.5T/T_s)} = \frac{\cdot(\pi/2)(T/T_s)}{\sin[(\pi/2)T/T_s)]}$$

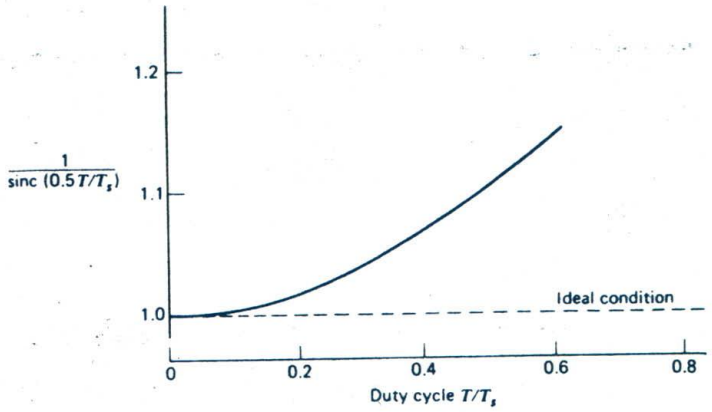


Figure 5.8
Normalized equalization (to compensate for aperture effect) plotted versus T/T_s.

where the ratio T/T_s is equal to the duty cycle of the sampling pulses. In Fig. 5.8 this result is plotted as a function of T/T_s. Ideally, it should be equal to 1 for all values of T/T_s. For a duty cycle of 10%, it is equal to 1.0041. It follows therefore that for duty cycles of less than 10% the aperture effect becomes negligible.

5.4 *QUANTIZING*

A continuous signal, such as 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. It is not necessary in fact to transmit the exact amplitudes of the samples. Any human sense (the ear or the eye), as ultimate receiver, can only detect finite intensity differences. This means that the original continuous 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 PCM. Clearly, if we assign the discrete amplitude levels with sufficiently close spacing, we may make the approximated signal practically indistinguishable from the original continuous signal.

The conversion of an analog (continuous) sample of the signal into a digital (discrete) form is called the *quantizing* process. Graphically, the quantizing process means that a straight line representing the relation between the input and output of a linear continuous system is replaced by a *staircase* characteristic, as in Fig. 5.9*a*. The difference between two adjacent discrete values is called a *quantum* or *step size*. Signals applied to a *quantizer*, with the input–output characteristic of Fig. 5.9*a*, are sorted into

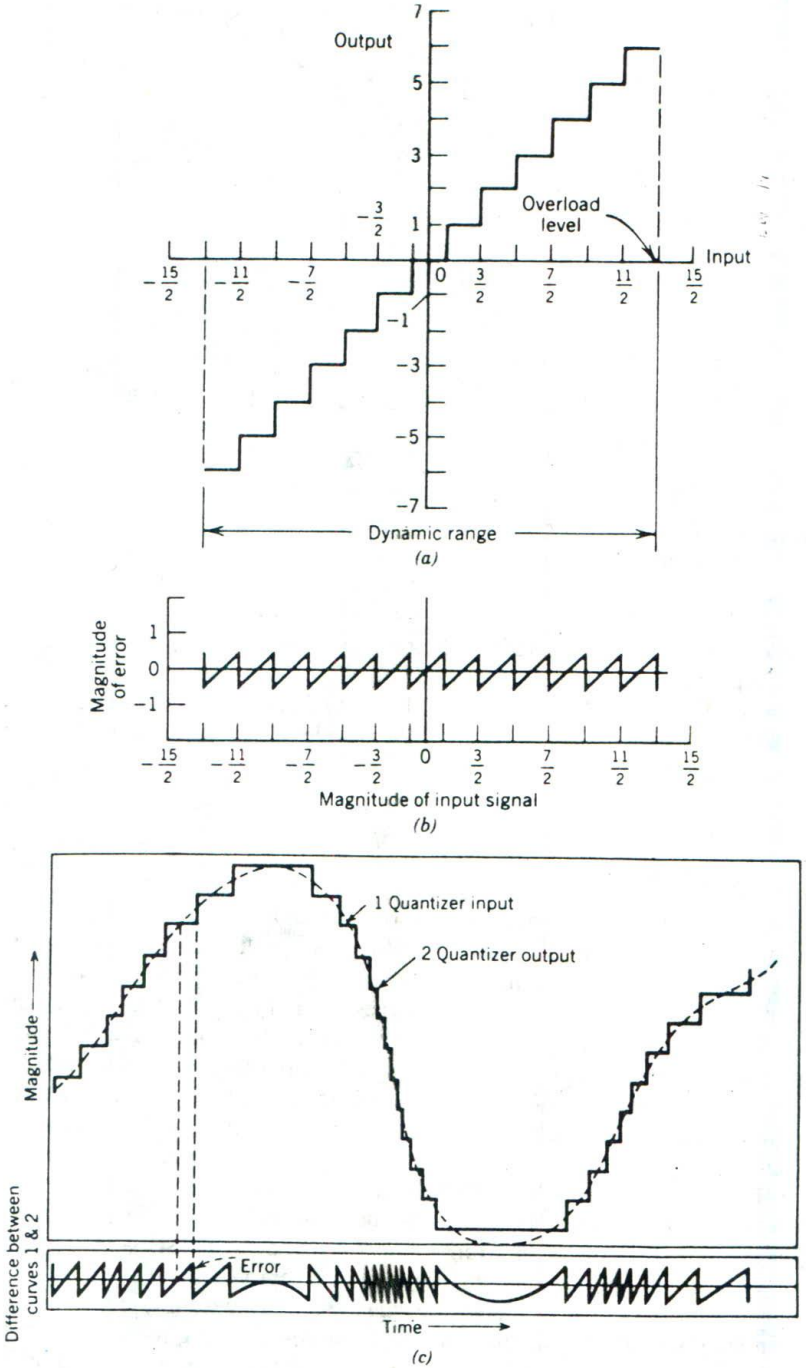


Figure 5.9
The quantizing principle. (a) *Quantizing characteristic.* (b) *Characteristic of errors in quantizing.* (c) *A quantized signal wave and the corresponding error curve. This figure is adapted from Bennett (1948) by permission of AT and T.*

amplitude slices (the treads of the staircase), and all input signals within plus or minus half a quantum step of the midvalue of a slice are replaced in the output by the midvalue in question.

The *quantizing error* consists of the difference between the input and output signals of the quantizer. It is apparent that the maximum instantaneous value of this error is half of one quantum step, and the total range of variation is from minus half a step to plus half a step. In part *b* of Fig. 5.9 the error is shown plotted as a function of the input signal. In part *c* of the figure typical variations of the quantizer input, the quantizer output, and the difference between them (i.e., the quantizing error) as functions of time are indicated.

A quantizer having the input-output amplitude characteristic of Fig. 5.9*a* is said to be of the *midtread* type, because the origin lies in the middle of a tread of the staircase-like graph. According to this characteristic, the quantizer output may be expressed as $i\Delta$, where $i = 0, \pm 1, \pm 2, \ldots, \pm K$. These discrete amplitude values of the quantizer output are called *representation levels*. A quantizer of the midtread type has an *odd* number of representation levels, as shown by

$$L = 2K + 1 \tag{5.10}$$

The *dynamic range* or *peak-to-peak excursion* of the quantizer input is $L\Delta$. One half of this excursion defines the absolute value of the *overload level* of the quantizer. Clearly, the amplitude of the quantizer input must not exceed the overload level; otherwise, *overload distortion* results.

In the quantizer example illustrated in Fig. 5.9*a*, the step size Δ equals 1; the integer K is 6; and the number of representation levels L is 13. The corresponding absolute value of the overload level is 13/2.

QUANTIZING NOISE

Quantizing noise or *quantizing error* is produced in the transmitting end of a PCM system by *rounding off the sampled values of a continuous message signal to the nearest representation level.* We assume a quantizing process with a uniform step size denoted by Δ volts, so that the representation levels are at $0, \pm\Delta, \pm 2\Delta, \pm 3\Delta, \ldots$. Consider a particular sample at the quantizer input, with an amplitude that lies in the range $i\Delta - (\Delta/2)$ to $i\Delta + (\Delta/2)$, where i is an integer (positive or negative, including zero) and $i\Delta$ defines the corresponding quantizer output. We thus have a region of *uncertainty* of width Δ, centered about $i\Delta$, as illustrated in Fig. 5.10. Let q_e denote the value of the error produced by the quantizing process. Then the amplitude of the sample at the quantizer input is $i\Delta + q_e$. It is apparent that with a random input signal, the quantizing error q_e varies randomly within the interval $-\Delta/2 \leq q_e \leq \Delta/2$.

When the quantization is fine enough (say, the number of representation levels is greater than 64), the distortion produced by quantizing noise affects

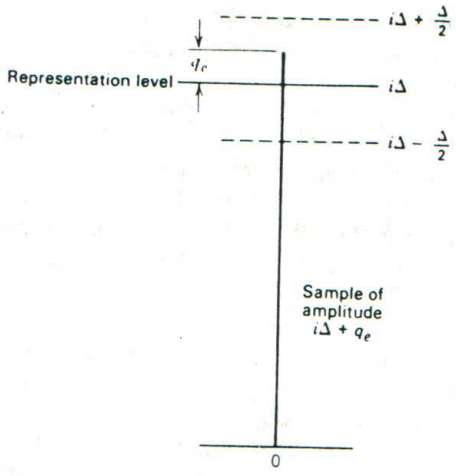


Figure 5.10
Illustrating the quantizing error q_e.

the performance of a PCM system as though it were an additive independent source of noise with zero mean and mean-square value determined by the quantizer step size Δ. The reason for this is that the power spectral density of the quantizing noise in the quantizer output is practically independent of that of the message signal over a wide range of input signal amplitudes. Furthermore, for a message signal of a root mean-square value that is large compared to a quantum step, it is found that the power spectral density of the quantizing noise has a large bandwidth compared with the signal bandwidth. Thus, with the quantizing noise uniformly distributed throughout the signal band, its interfering effect on a signal is similar to that of thermal noise. (Thermal noise is discussed in Appendix C.)

We say that the quantizing noise is *uniformly distributed* when the error may take on a sample value q_e anywhere inside the interval $(-\Delta/2, \Delta/2)$ with equal likelihood. Under this assumption, we may determine the *average power of the quantizing noise* by averaging q_e^2, the squared quantizing error, over all possible values of q_e. We may thus express the *average power of quantizing noise*, P_q, as follows

$$\begin{aligned} P_q &= \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} q_e^2\, dq_e \\ &= \frac{\Delta^2}{12} \end{aligned} \tag{5.11}$$

Thus, the average power of quantizing noise grows as the square of the step size Δ. This is perhaps the most often used result in quantization.

The step size Δ is under the designer's control. Hence, the signal distortion due to quantizing noise can be controlled by choosing the step size Δ small enough, as illustrated in the following example.

EXAMPLE 2 SIGNAL-TO-QUANTIZING NOISE RATIO FOR SINUSOIDAL MODULATION

Consider the special case of a *full-load* sinusoidal modulating wave of amplitude A_m, which uses all the representation levels provided. The average signal power is $A_m^2/2$. The peak-to-peak excursion of the quantizer input is $2A_m$, because the modulating wave swings between $-A_m$ and A_m. Assuming that the number of representation levels equals L, the quantizer step size is

$$\Delta = \frac{2A_m}{L} \tag{5.12}$$

Therefore Eq. 5.11 gives the average quantizing noise power as

$$P_q = \frac{A_m^2}{3L^2} \tag{5.13}$$

Thus the *output signal-to-quantizing noise ratio* of the PCM system, for a full-load test tone, is

$$(\text{SNR})_0 = \frac{A_m^2/2}{A_m^2/3L^2} = \frac{3L^2}{2} \tag{5.14}$$

Expressing the signal-to-quantizing noise ratio in decibels, we get

$$10 \log_{10}(\text{SNR})_0 = 1.8 + 20 \log_{10} L \tag{5.15}$$

Hence, the output signal-to-noise ratio of a PCM system in decibels, due to quantizing noise, increases logarithmically with the number of representation levels, L.

TABLE 5.1

Number of Representation Levels, L	Code Word Length n	Signal-to-Quantizing Noise Ratio, dB
32	5	31.8
64	6	37.8
128	7	43.8
256	8	49.8

For various values of L, the corresponding values of signal-to-quantizing noise ratio are as given in Table 5.1. The second column of this table corresponds to the binary code word length, an issue that is considered in Section 5.5.

EXERCISE 2 **A sinusoidal signal is transmitted using PCM. The output signal-to-quantizing noise ratio is required to be 55.8 dB. Find the minimum number of representation levels L required to achieve this performance.**

COMPANDING

The quantizing process based on Fig. 5.9*a* uses a *uniform separation* between the representation levels. In certain applications, however, it is preferable to use a variable separation between the representation levels. For example, the ratio of voltage levels 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. The excursions of the voice signal into the large amplitude ranges, which occur in practice relatively infrequently, can be taken care of by using a *nonuniform quantizer*. Such a quantizer is designed so that the step size increases as the separation from the origin of the input-output amplitude characteristic is increased. We thus find that the weak passages, which need more protection, are favored at the expense of the loud passages. In this way, a nearly uniform percentage precision is achieved throughout the amplitude range of the input signal, with the result that fewer steps are needed than would be the case if a uniform quantizer were used.

The use of a nonuniform quantizer is equivalent to passing the baseband signal through a *compressor* and then applying the compressed signal to a uniform quantizer. A particular form of compression law that is used in practice is the so-called μ-*law* defined by

$$|v_2| = \frac{\log(1 + \mu|v_1|)}{\log(1 + \mu)} \tag{5.16}$$

where v_1 and v_2 are normalized input and output voltages, and μ is a positive constant. Figure 5.11*a* plots the μ-law for varying μ. The case of uniform quantization corresponds to $\mu = 0$. For a given value of μ, the reciprocal slope of the compression curve, which defines the quantum steps, is

$$\frac{d|v_1|}{d|v_2|} = \frac{\log(1 + \mu)}{\mu}(1 + \mu|v_1|) \tag{5.17}$$

We see therefore that the μ-law is neither strictly linear nor strictly logarithmic, but it is approximately linear at low input levels corresponding to

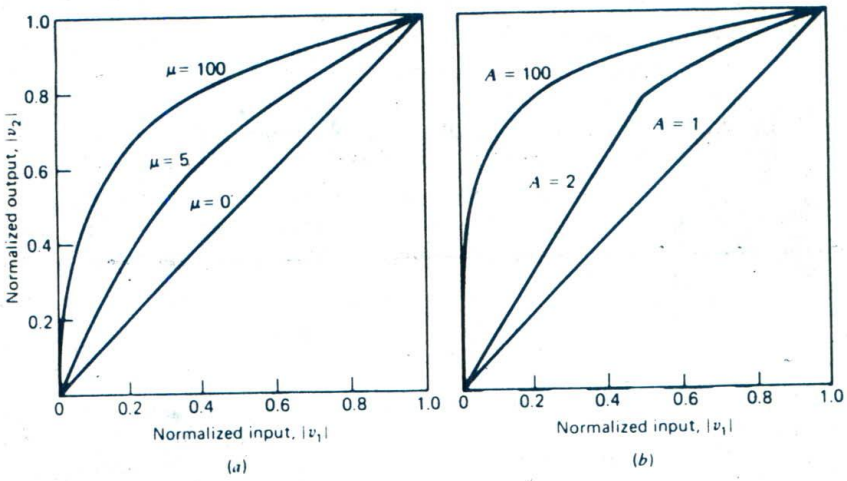


Figure 5.11
Compression laws. (a) μ-*law.* (b) A-*law.*

$\mu|v_1| \ll 1$, and approximately logarithmic at high input levels corresponding to $\mu|v_1| \gg 1$.

Another compression law that is used in practice is the so-called *A-law* defined by

$$|v_2| = \begin{cases} \dfrac{A|v_1|}{1 + \log A} & 0 \leq |v_1| \leq \dfrac{1}{A} \\ \dfrac{1 + \log(A|v_1|)}{1 + \log A} & \dfrac{1}{A} \leq |v_1| \leq 1 \end{cases} \tag{5.18}$$

which is shown plotted in Fig. 5.11*b*. Practical values of A (as of μ in the μ-law) tend to be in the vicinity of 100. The case of uniform quantization corresponds to $A = 1$. The reciprocal slope of this compression curve is

$$\frac{d|v_1|}{d|v_2|} = \begin{cases} \dfrac{1 + \log A}{A} & 0 \leq |v_1| \leq \dfrac{1}{A} \\ (1 + \log A)|v_1| & \dfrac{1}{A} \leq |v_1| \leq 1 \end{cases} \tag{5.19}$$

Thus the quantum steps over the central linear segment, which have a dominant effect on small signals, are diminished by the factor $A/(1 + \log A)$. This is typically about 25 dB in practice, as compared with uniform quantization.

To restore the signal samples to their correct relative levels, 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 inverse so that, except for the effect of quantization, the expander output is equal to the compressor input. The combination of a *compressor* and an *expander* is called a *compander*.

In actual PCM systems, the companding circuitry does not produce an exact replica of the nonlinear compression curves shown in Fig. 5.11. 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. This form of approximation is illustrated in Section 5.11.

5.5 CODING

In combining the processes of sampling and quantizing, the specification of a continuous message signal becomes limited to a discrete sequence of values, but not in the form best suited to transmission over a line or radio path. To fully exploit the advantages of sampling and quantizing, we require the use of an *encoding process* to translate the discrete sequence of sample values to a more appropriate form of signal. Any plan for representing each element of this discrete set of values as a particular arrangement of discrete events is called a *code*. One of the discrete events in a code is called a *code element* or *symbol*. For example, the presence or absence of

TABLE 5.2

Ordinal Number of Representation Level	Level Number Expressed as Sum of Powers of 2	Binary Number
0		0 0 0 0
1	2^0	0 0 0 1
2	2^1	0 0 1 0
3	$2^1 + 2^0$	0 0 1 1
4	2^2	0 1 0 0
5	$2^2 + 2^0$	0 1 0 1
6	$2^2 + 2^1$	0 1 1 0
7	$2^2 + 2^1 + 2^0$	0 1 1 1
8	2^3	1 0 0 0
9	$2^3 + 2^0$	1 0 0 1
10	$2^3 + 2^1$	1 0 1 0
11	$2^3 + 2^1 + 2^0$	1 0 1 1
12	$2^3 + 2^2$	1 1 0 0
13	$2^3 + 2^2 + 2^0$	1 1 0 1
14	$2^3 + 2^2 + 2^1$	1 1 1 0
15	$2^3 + 2^2 + 2^1 + 2^0$	1 1 1 1

a pulse is a symbol. A particular arrangement of symbols used in a code to represent a single value of the discrete set is called a *code word* or *character*.

In a *binary code*, each symbol may be either of two distinct values or kinds, such as the presence or absence of a pulse. The two symbols of a binary code are customarily denoted as 0 and 1. In a *ternary code*, each symbol may be one of three distinct values or kinds, and so on for other codes. However, *the maximum advantage over the effects of noise in a transmission medium is obtained by using a binary code, because a binary symbol withstands a relatively high level of noise and is easy to regenerate.*

Suppose that, in a binary code, each code word consists of n *bits*; the bit is an acronym for *bi*nary digi*t*. Then, using such a code, we may represent a total of 2^n distinct numbers. For example, a sample quantized into one of 128 levels may be represented by a 7-bit code word. There are several ways of establishing a one-to-one correspondence between representation levels and code words. A convenient one is to express the ordinal number of the representation level as a binary number. In the binary number system, each digit has a place-value that is a power of 2, as illustrated in Table 5.2 for the case of $n = 4$.

EXAMPLE 3 SIGNAL-TO-QUANTIZING NOISE RATIO FOR SINUSOIDAL MODULATION (CONTINUED)

In this example, we reformulate the output signal-to-quantizing noise ratio of Eq. 5.15 for a PCM system operating with a full-load test tone. This equation is reproduced here for convenience

$$10 \log_{10}(\text{SNR})_o = 1.8 + 20 \log_{10} L, \text{ dB}$$

where L is the number of representation levels used in the system. Assuming the use of an n-bit binary code word, we may define L for a quantizer of the midtread type as

$$\begin{aligned} L &= 2^n - 1 \\ &\simeq 2^n, \qquad n \text{ large} \end{aligned} \tag{5.20}$$

Accordingly, we may redefine the output signal-to-quantizing noise ratio in terms of the code word length n as

$$10 \log_{10}(\text{SNR})_o = (1.8 + 6n), \text{ dB} \tag{5.21}$$

For various values of n, the corresponding values of signal-to-quantizing noise ratio are as given in Table 5.1. The formula of Eq. 5.21 states that each bit in the code word of a PCM system contributes 6 dB to the output signal-to-quantizing noise ratio.

EXAMPLE 4 TRADEOFF BETWEEN CHANNEL BANDWIDTH AND SIGNAL-TO-QUANTIZING NOISE RATIO

We may develop further insight into the performance of a PCM system by examining the relationship between the signal-to-quantizing noise ratio and transmission bandwidth requirement of a binary PCM system. For the purpose of this evaluation, we will again consider the use of a sinusoidal modulating wave.

From our discussion of the sampling process, we have seen that a message signal of bandwidth W requires a minimum sampling rate of $2W$. With each signal sample represented by an n-bit code word, the bit duration T_b has a maximum value of $1/2nW$. In Section 6.4, it is shown that the channel bandwidth B required to transmit a pulse of this duration is given by

$$B = \kappa n W \tag{5.22}$$

where κ is a constant with a value between 1 and 2.

Expressing the output signal-to-quantizing noise ratio simply as a ratio, we have from Eqs. 5.14 and 5.20:

$$(\text{SNR})_0 = \frac{3}{2}(4^n) \tag{5.23}$$

Accordingly, using Eqs. 5.22 and 5.23, we get

$$(\text{SNR})_0 = \frac{3}{2}(4^{B/\kappa W}) \tag{5.24}$$

This relation shows that a PCM system is capable of improving the output signal-to-noise ratio *exponentially* with the bandwidth expansion ratio B/W.

EXERCISE 3 A *television (TV) signal* with a bandwidth of 4.2 MHz is transmitted using binary PCM. The number of representation levels is 512. Calculate the following parameters:

(a) The code word length.
(b) The final bit rate.
(c) The transmission bandwidth, assuming that $\kappa = 2$ in Eq. 5.22.

EXERCISE 4 The frequency content of a *studio-quality audio signal* that we like to hear extends from 20 Hz to 20 kHz. For professional use, the

signal is sampled at the rate of 48 × 10^3 samples per second. The standard code word used for conversion into a PCM format is 16 bits per sample. What is the final bit rate for digital storage of the signal?

DIGITAL FORMATS

To send the encoded digital data over a channel, we require the use of a *format* or *waveform* for representing the data.[3] In this context, we have a number of formats available to us. Figure 5.12 illustrates some commonly used ones for the example of binary sequence 0110100011. Specifically, we have illustrated the following formats:

(a) Symbol 1 is represented by transmitting a pulse of constant amplitude for the duration of the symbol, and symbol 0 is represented by switching off the pulse, as in Fig. 5.12*a*. This type of format is referred to as *on–off* or *unipolar signaling*.

(b) Symbols 1 and 0 are represented by pulses of equal positive and negative amplitudes, as in Fig. 5.12*b*. This type of format is referred to as *polar signaling*.

(c) A rectangular pulse (half-symbol wide) is used for a 1 and no pulse for a 0, as in Fig. 5.12*c*. This type of format is called *return-to-zero* (RZ) *signaling*.

(d) Positive and negative pulses (of equal amplitude) are used alternately for symbol 1, and no pulse for symbol 0, as in Fig. 5.12*d*. This type of format is called *bipolar signaling*. A useful property of this method of signaling is that the power spectrum of the transmitted signal has no dc component and relatively insignificant low-frequency components.

(e) Symbol 1 is represented by a positive pulse followed by a negative pulse, with both pulses being of equal amplitude and half-symbol wide; for symbol 0, the polarities of these pulses are reversed, as in Fig. 5.12*e*. This type of format is called a *split-phase* or *Manchester code*. It also suppresses the dc component and has relatively insignificant low-frequency components.

Note that the polar signal waveform of Fig. 5.12*b* and the Manchester code of Fig. 5.12*e* are examples of *nonreturn-to-zero* (*NRZ*) *signaling*.

The binary code is a special case of *M-ary code*. In practice, we usually find that M, the number of symbols in the code, is an integer power of 2. Then, each code word in the M-ary code carries the equivalent information of $\log_2 M$ bits. Consider, for example, a *four-level* (*quaternary*) *code* (i.e., $M = 4$). In such a code, we may identify four distinct

[3]Digital formats (waveforms) are also referred to in the literature as *line* or *transmission codes*.

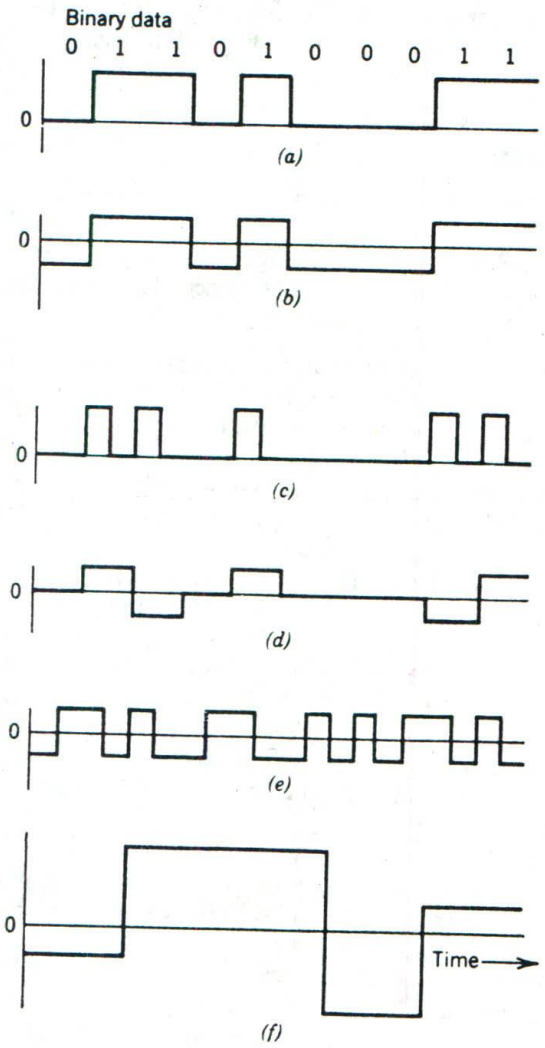


Figure 5.12
Electrical representations of binary data. (a) *On-off signaling.* (b) *Polar signaling.* (c) *Return-to-zero signaling.* (d) *Bipolar signaling.* (e) *Split-phase or Manchester code.* (f) *four-level Gray coding.*

dibits (pairs of bits). In Table 5.3*a*, we show two arrangements for the four possible dibits together with their individual electrical representations. In particular, the dibits are shown in both their *natural code* and *Gray code*. Using the notations of the Gray code in Table 5.3*a*, the binary sequence 0110100011 is thus represented by the waveform shown in Fig. 5.12*f*. To obtain this waveform, the given sequence is viewed as a new sequence of dibits, namely, 01,10,10,00,11, and each dibit is represented in accordance with the Gray code of Table 5.3*a*.

TABLE 5.3 Examples of Natural and Gray Codes

(a) Four-level code

Code Word Number	Natural Code	Gray Code	Electrical Representation
0	00	00	$-\frac{3}{2}$
1	01	01	$-\frac{1}{2}$
2	10	11	$+\frac{1}{2}$
3	11	10	$+\frac{3}{2}$

(b) Eight-level Code

Code Word Number	Natural Code	Gray Code	Electrical Representation
0	000	000	$-\frac{7}{2}$
1	001	001	$-\frac{5}{2}$
2	010	011	$-\frac{3}{2}$
3	011	010	$-\frac{1}{2}$
4	100	110	$+\frac{1}{2}$
5	101	111	$+\frac{3}{2}$
6	110	101	$+\frac{5}{2}$
7	111	100	$+\frac{7}{2}$

The distinguishing feature of a Gray code[4] is that there is a *one-bit change* as we move from one code word to another. This is well illustrated in the two Gray codes shown in Table 5.3 for $M = 4,8$. Note that in Table 5.3*b* for $M = 8$, for example, the rule of a one-bit change per transition applies not only to all the transitions for code word 0 to

[4]The origin of Gray codes goes back to the development of the rotary form of mechanical encoders known as *shaft encoders*. The use of Gray encoding makes it possible for electromechanical arrangements to change from one code word to another by changing the state of a single digit. With natural encoding, on the other hand, two or more digits may be required to change state *simultaneously*, which is difficult for electromechanical devices.

code word 1, from code word 1 to code word 2, and so on up to the transition from code word 6 to code word 7, but also to the "wrap-around" transition from code word 7 to code word 0. This wrap-around feature makes a Gray encoder *cyclic* in nature.

The choice of a particular digital waveform is influenced by the application of interest. Nevertheless, it is highly desirable for a digital waveform to have the following properties:

1. *Timing content.* The transmitted digital waveform should have adequate timing content to permit the extraction of *clock* information required for the purpose of synchronizing the receiver to the transmitter.
2. *Ruggedness.* The waveform should possess immunity to channel noise and interference for prescribed channel bandwidth and transmitted power.
3. *Error detection capability.* The waveform should permit the detection of errors that may occur in the course of transmission due to the presence of channel noise.
4. *Matched power spectrum.* The power spectral density of the transmitted digital waveform should match the frequency response of the channel as closely as possible so as to minimize signal distortion.
5. *Transparency.* The correct transmission of digital data over a channel should be transparent to the pattern of 1's and 0's contained in the data.

It is for these reasons that we find, for example, the bipolar format has become the standard for transmitting binary encoded PCM data over telephone channels.

EXERCISE 5 Rank the six digital waveforms depicted in Fig. 5.12 in increasing order of transmission bandwidth requirement.

DECODING

The first operation in the receiver is to regenerate (i.e., reshape and clean up) the received pulses. These clean pulses are then regrouped into code words and decoded (i.e., mapped back) into a quantized PAM signal. The *decoding* process involves generating a pulse the amplitude of which is the linear sum of all the pulses in the code word, with each pulse weighted by its place-value (2^0, 2^1, 2^2, 2^3, . . .) in the code.

It is noteworthy that every operation performed in the transmitter of a PCM system, except for the quantizing operation, is reversible. Specifically, the operations of sampling and encoding performed in the transmitter are reversed by performing decoding and interpolation (in that order) in the

receiver. On the other hand, quantizing is an irreversible process that manifests itself by destroying information; once quantizing noise is introduced in the transmitter, there is nothing we can do in the receiver to make up for the loss of information thereby incurred.

5.6 REGENERATION

The most important feature of PCM systems lies in the ability to control the effects of distortion and noise produced by transmitting a PCM wave through a channel. This capability is accomplished by reconstructing the PCM wave by means of a chain of *regenerative repeaters* sufficiently close along the transmission route. As illustrated in Fig. 5.13, three basic functions are performed by 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 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 signal-to-noise ratio is a maximum. The decision device is enabled at the sampling times determined by the timing circuitry. It makes its decision based on whether or not the amplitude of the quantized pulse plus noise exceeds a predetermined voltage level. Thus, for example, in a PCM system with on–off signaling, the repeater makes a decision in each bit interval as to whether or not a pulse is present. If the decision is "yes" a clean new pulse is transmitted to the next repeater. If, on the other hand, the decision is "no," a clean base line is transmitted. In this way, the accumulation of distortion and noise in a repeater span is 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 presence of transmission noise and interference causes the repeater to make wrong decisions occasionally, thereby introducing *bit errors* into the regenerated signal; we will have more to say on this issue in Chapter 10.

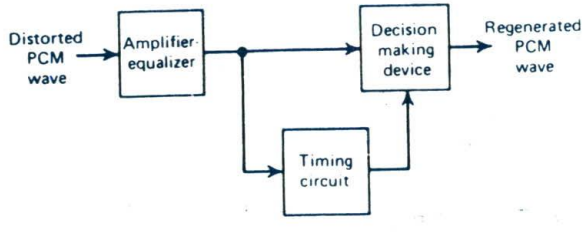


Figure 5.13
Block diagram of a regenerative repeater.

2. If the spacing between received pulses deviates from its assigned value, a *jitter* is introduced into the regenerated pulse position, thereby causing distortion.

5.7 DIFFERENTIAL PULSE-CODE MODULATION

When a voice or video signal is sampled at a rate slightly higher than the Nyquist rate, the resulting sampled signal is found to exhibit a high 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. When these highly correlated samples are encoded, as in a standard PCM system, the resulting encoded signal contains *redundant information*. This means that symbols that are not absolutely essential to the transmission of information are generated as a result of the encoding process. By removing this redundancy before encoding, we obtain a more efficient coded signal.

Now, if we know a sufficient part of a redundant signal, we may infer the rest, or at least make the most probable estimate. In particular, if we know the past behavior of a signal up to a certain point in time, it is possible to make some inference about its future values; such a process is commonly called *prediction*. Suppose then a message signal $m(t)$ is sampled at the rate $1/T_s$ to produce a sequence of correlated samples T_s seconds apart; which is denoted by $\{m(nT_s)\}$. The fact that it is possible to predict future values of the signal $m(t)$ provides motivation for the *differential quantization* scheme shown in Fig. 5.14*a*. In this scheme, the input to the quantizer is

$$e(nT_s) = m(nT_s) - \hat{m}(nT_s) \tag{5.25}$$

which is the difference between the unquantized input sample $m(nT_s)$ and a prediction of it, denoted by $\hat{m}(nT_s)$. This predicted value is produced by using a *prediction filter* with an input, as we will see, that consists of a quantized version of the message sample $m(nT_s)$. The difference signal $e(nT_s)$ is called a *prediction error*, since it is the amount by which the prediction filter fails to predict the input exactly.

By encoding the quantizer output, as in Fig. 5.14*a*, we obtain a variation of PCM, known as *differential pulse-code modulation* (DPCM). It is this encoded signal that is used for transmission.

The quantizer output may be represented as

$$e_q(nT_s) = e(nT_s) + q_e(nT_s) \tag{5.26}$$

where $q_e(nT_s)$ is the quantizing error. According to Fig. 5.14*a*, the quantizer output $e_q(nT_s)$ is added to the predicted value $\hat{m}(nT_s)$ to produce the prediction-filter input

$$m_q(nT_s) = \hat{m}(nT_s) + e_q(nT_s) \tag{5.27}$$

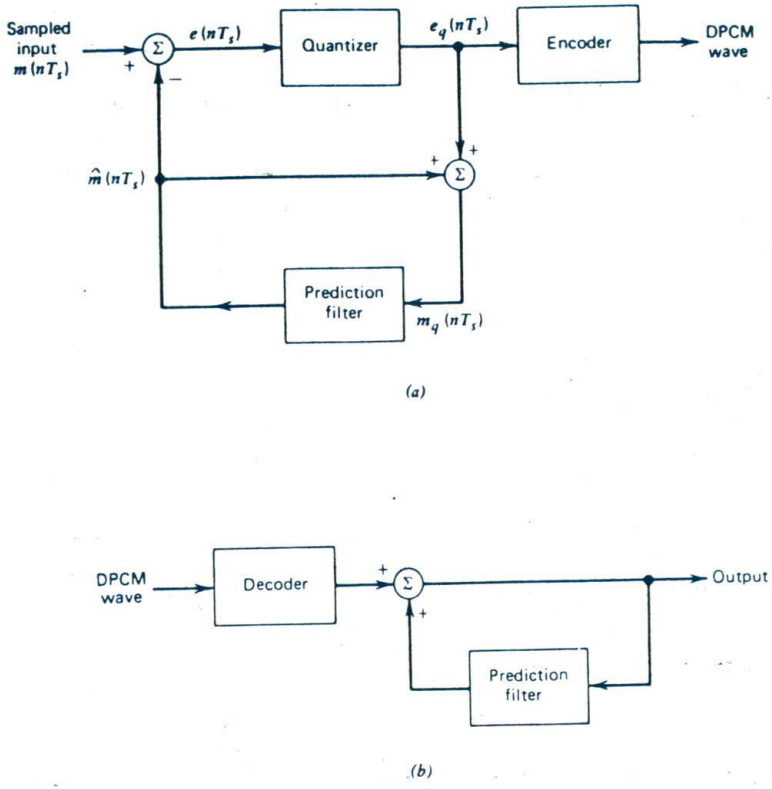


Figure 5.14
DPCM system. (a) *Transmitter.* (b) *Receiver.*

Substituting Eq. 5.26 in 5.27, we get

$$m_q(nT_s) = \hat{m}(nT_s) + e(nT_s) + q_e(nT_s) \tag{5.28}$$

However, from Eq. 5.25 we observe that $\hat{m}(nT_s)$ plus $e(nT_s)$ is equal to the incoming message sample $m(nT_s)$. Therefore, we may rewrite Eq. 5.28 as follows

$$m_q(nT_s) = m(nT_s) + q_e(nT_s) \tag{5.29}$$

which represents a quantized version of $m(nT_s)$. That is, irrespective of the properties of the prediction filter, the quantized sample, $m_q(nT_s)$, at the prediction filter input, differs from the sample $m(nT_s)$ of the original message signal $m(t)$ by the quantizing error $q_e(nT_s)$. Accordingly, if the prediction is good, the average power of the prediction error sequence $\{e(nT_s)\}$ will be smaller than that of the message sequence $\{m(nT_s)\}$. Hence,

a quantizer with a given number of levels can be adjusted to produce a quantizing error sequence with a smaller average power than would be possible if the incoming message sequence were quantized directly as in a standard PCM system.

The receiver for reconstructing the quantized version of the input is shown in Fig. 5.14*b*. It consists of a decoder to reconstruct the quantized error sequence. The quantized version of the original input is reconstructed from the decoder output using the same prediction filter as used in the transmitter of Fig. 5.14*a*. In the absence of transmission noise, we find that the encoded signal at the receiver input is identical to the encoded signal at the transmitter ouput. The corresponding receiver output differs from the original message signal only by the quantizing error incurred as a result of quantizing the prediction error.

From the foregoing analysis we observe that, in a noise-free environment, the prediction filters in the transmitter and receiver operate on the same sequence of samples, $\{m_q(nT_s)\}$. It is with this purpose in mind that a feedback path is added to the quantizer in the transmitter, as shown in Fig. 5.14*a*.

The average power of the message sequence $\{m(nT_s)\}$ is given by

$$P_m = \frac{1}{N}\sum_{n=0}^{N-1} m^2(nT_s) \tag{5.30}$$

where N is the *length* of the message sequence. The average power of the quantizing error sequence $\{q_e(nT_s)\}$, also assumed to be of length N, is given by

$$P_q = \frac{1}{N}\sum_{n=0}^{N-1} q_e^2(nT_s) \tag{5.31}$$

We may thus define the *output signal-to-quantizing noise ratio* of a DPCM system as

$$(\text{SNR})_0 = \frac{P_m}{P_q} \tag{5.32}$$

It is clear that we may rewrite Eq. 5.32 as

$$(\text{SNR})_0 = \frac{P_m}{P_e}\frac{P_e}{P_q} = G_p(\text{SNR})_Q$$

where $(\text{SNR})_Q$ is the *signal-to-quantizing noise ratio* defined by

$$(\text{SNR})_Q = \frac{P_e}{P_q} \tag{5.33}$$

and G_p is the *prediction gain* produced by the differential quantization scheme, defined by

$$G_p = \frac{P_m}{P_e} \tag{5.34}$$

The quantity G_p, when greater than unity, represents the gain in signal-to-noise ratio that is due to the differential quantization scheme of Fig. 5.14. Now for a given message signal, the average power P_m is fixed, so that G_p is maximized by minimizing the average prediction error power P_e. Accordingly, our objective should be to design the prediction filter so as to minimize P_e, while the signal-to-quantizing noise ratio is maintained constant.

THE PREDICTION FILTER

One approach to specify the nature of the prediction filters in the transmitter and the receiver of the DPCM system shown in Fig. 5.14 is to use a *tapped-delay-line filter* as the basis of the design. An advantage of this approach is that it leads to tractable mathematics, and it is simple to implement. Thus the predicted value $\hat{m}(nT_s)$ is modeled as a linear combination of past values of the quantized input as shown by (see Fig. 5.15)

$$\hat{m}(nT_s) = \sum_{k=1}^{p} w_k m_q(nT_s - kT_s) \tag{5.35}$$

where the tap weights $w_1, w_2, \ldots, w_p$ define the prediction filter coefficients, and p is the *order* of the prediction filter. Substitution of Eq. 5.35 in 5.25 yields the prediction error

$$e(nT_s) = m(nT_s) - \sum_{k=1}^{p} w_k m_q(nT_s - kT_s) \tag{5.36}$$

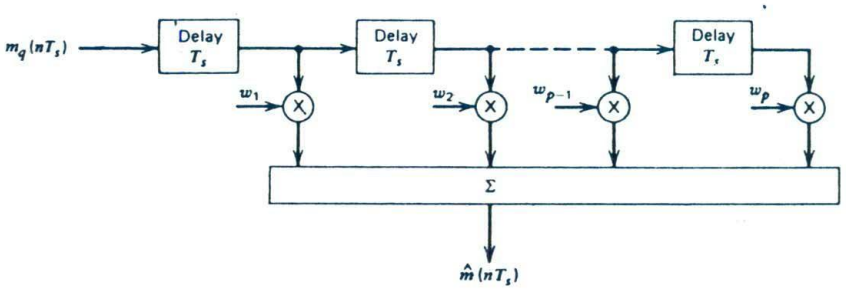


Figure 5.15
Tapped-delay-line filter used as a prediction filter.

The mathematical basis for the design of the prediction filter is that of *minimizing the average prediction-error power with respect to the tap weights of the filter.*[5]

5.8 DELTA MODULATION

The exploitation of signal correlations in DPCM suggests the further possibility of oversampling a message signal (i.e., at a rate much higher than the Nyquist rate) to purposely increase the correlation between adjacent samples of the signal. This would permit the use of a simple quantizing strategy for constructing the encoded signal. *Delta modulation* (DM), which is the one-bit (or two-level) version of DPCM, is precisely such a scheme.

In its simple form, DM provides a staircase approximation to the oversampled version of an incoming message signal, as illustrated in Fig. 5.16*a*. The difference between the input and the approximation is quantized into only two representation levels, namely, $\pm\delta$, corresponding to positive and negative differences. Thus, if the approximation falls below the signal at any sampling epoch, it is increased by δ. If, on the other hand, the approximation lies above the signal, it is diminished by δ. Provided that the signal does not change too rapidly from sample to sample, we find that the staircase approximation remains within $\pm\delta$ of the input signal.

Denoting the input signal as $m(t)$ and the staircase approximation as $m_q(t)$, the basic principle of delta modulation may be formalized in the following set of discrete-time relations:

$$e(nT_s) = m(nT_s) - m_q(nT_s - T_s) \tag{5.37}$$

$$e_q(nT_s) = \delta \operatorname{sgn}[e(nT_s)] \tag{5.38}$$

and

$$m_q(nT_s) = m_q(nT_s - T_s) + e_q(nT_s) \tag{5.39}$$

[5]The result of this minimization is a set of simultaneous equations, expressed in matrix form as follows

$$\begin{bmatrix} 1 & \rho(T_s) & \cdots & \rho(pT_s - T_s) \\ \rho(T_s) & 1 & \cdots & \rho(pT_s - 2T_s) \\ \vdots & \vdots & & \vdots \\ \rho(pT_s - T_s) & \rho(pT_s - 2T_s) & \cdots & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix} = \begin{bmatrix} \rho(T_s) \\ \rho(2T_s) \\ \vdots \\ \rho(pT_s) \end{bmatrix}$$

Here it is assumed that the ouput signal-to-noise ratio, $(\mathrm{SNR})_o$, is large compared to unity. The parameter $\rho(kT_s)$ is the *normalized autocorrelation function* of the prediction filter's input signal for a lag of kT_s, as shown by

$$\rho(kT_s) = \frac{R_M(kT_s)}{R_M(0)}, \qquad k = 0, 1, \ldots, p$$

where the subscript M refers to the input. Hence, given the set of autocorrelation functions $\{R_M(kT_s),\ k = 0, 1, \ldots, p\}$, we may compute the prediction filter's tap weights. For a detailed treatment of prediction filters, see the following books: Jayant and Noll (1984) and Haykin (1986).

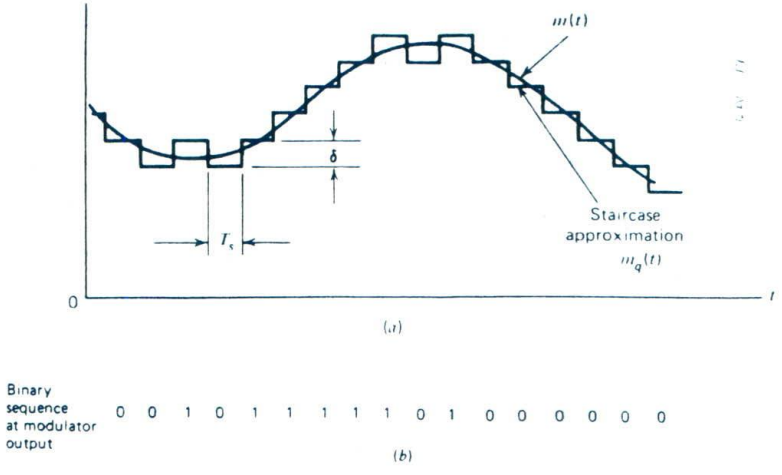


Figure 5.16
Delta modulation

where T_s is the sampling period; $e(nT_s)$ is an error signal representing the difference between the present sample value $m(nT_s)$ of the input signal and the latest approximation to it, namely, $\hat{m}(nT_s) = m_q(nT_s - T_s)$; and $e_q(nT_s)$ is the quantized version of $e(nT_s)$. The quantizer output $e_q(nT_s)$ is the desired DM wave for varying n.

Part *a* of Fig. 5.16 illustrates the way in which the staircase approximation $m_q(t)$ follows variations in the input signal $m(t)$ in accordance with Eqs. 5.37 through 5.39, and part *b* of the figure displays the corresponding binary sequence at the delta modulator output. It is apparent that in a delta modulation system the rate of information transmission is simply equal to the sampling rate $1/T_s$.

The principal virtue of delta modulation is its simplicity. It may be generated by applying the sampled version of the incoming message signal to a modulator that involves a summer, quantizer, and accumulator interconnected as shown in Fig. 5.17*a*. Details of the modulator follow directly from Eqs. 5.37 and 5.39. In particular, the quantizer consists of a *hard limiter* with an input–output relation defined by Eq. 5.38, which is depicted in Fig. 5.18. The quantizer output is applied to an *accumulator*, producing the result

$$\begin{aligned} m_q(nT_s) &= \delta \sum_{i=1}^{n} \operatorname{sgn}[e(iT_s)] \\ &= \sum_{i=1}^{n} e_q(iT_s) \end{aligned} \tag{5.40}$$

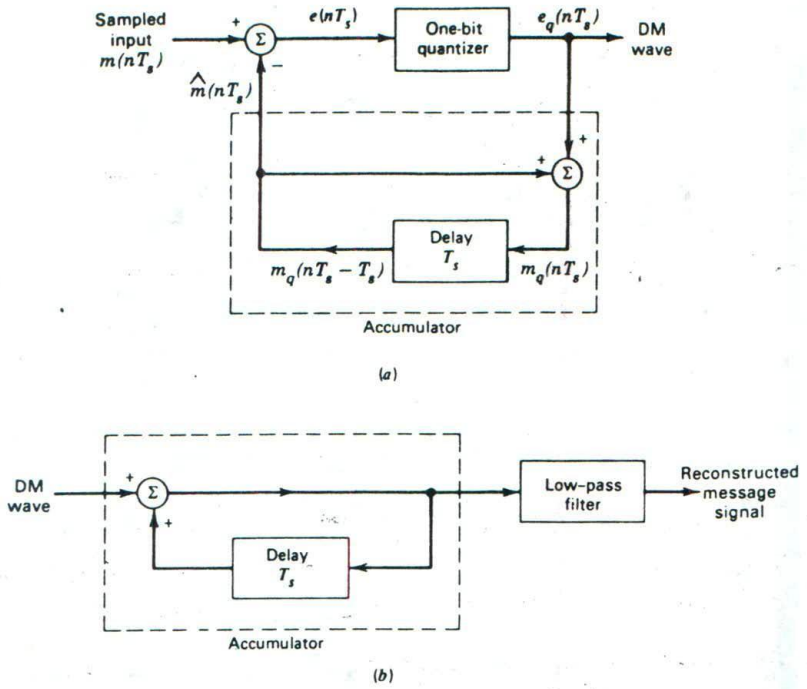


Figure 5.17
DM system. (a) *Transmitter.* (b) *Receiver.*

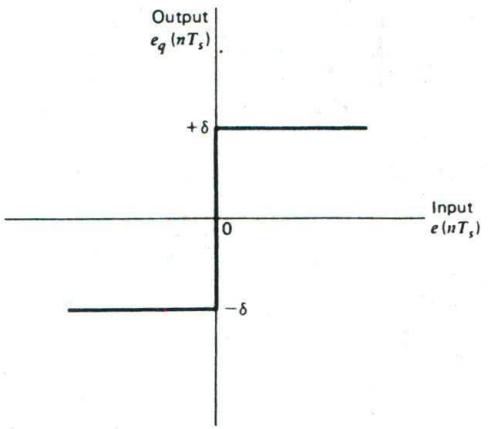


Figure 5.18
Input–output characteristic of quantizer for DM system.

which is obtained by solving Eqs. 5.38 and 5.39 for $m_q(nT_s)$. Thus, at the sampling instant nT_s, the accumulator increments the approximation by an amount equal to δ in a positive or negative direction, depending on the algebraic sign of the error signal $e(nT_s)$. If the input signal $m(nT_s)$ is greater than the most recent approximation $\hat{m}(nT_s)$, a positive increment $+\delta$ is applied to the approximation. If, on the other hand, the input signal is smaller, a negative increment $-\delta$ is applied to the approximation. In this way, the accumulator does the best it can to track the input samples by one step at a time. In the receiver, shown in Fig. 5.17*b*, the staircase approximation $m_q(t)$ is reconstructed by passing the sequence of positive and negative pulses, produced at the decoder output, through an accumulator in a manner similar to that used in the transmitter. The out-of-band quantizing noise in the high-frequency staircase waveform $m_q(t)$ is rejected by passing it through a low-pass filter with a bandwidth equal to the original message bandwidth.

In comparing the DPCM and DM networks of Fig. 5.14 and 5.17, we note that they are similar, except for two important differences, namely, the use of a one-bit (two-level) quantizer in the delta-modulator and the replacement of the prediction filter by a single delay element.

QUANTIZING NOISE

Delta modulation systems are subject to two types of quantizing error: (1) slope overload distortion, and (2) granular noise. We first discuss the cause of slope overload distortion, and then granular noise.

We observe that Eq. 5.40 is the digital equivalent of integration in the sense that it represents the accumulation of positive and negative increments of magnitude δ. Also, denoting the quantizing error by $q_e(nT_s)$, as shown by,

$$m_q(nT_s) = m(nT_s) + q_e(nT_s) \tag{5.41}$$

we observe from Eq. 5.37 that the input to the quantizer is

$$e(nT_s) = m(nT_s) - m(nT_s - T_s) - q_e(nT_s - T_s) \tag{5.42}$$

Thus, except for the quantizing error $q_e(nT_s - T_s)$, the quantizer input is a *first backward difference* of the input signal, which may be viewed as a digital approximation to the derivative of the input signal or, equivalently, as the inverse of the digital integration process. If we consider the maximum slope of the original input waveform $m(t)$, it is clear that in order for the quantized sequence $\{m_q(nT_s)\}$ to increase as fast as the message sequence $\{m(nT_s)\}$ in a region of maximum slope of $m(t)$, we require that the condition

$$\frac{\delta}{T_s} \geq \max \left| \frac{d\,m(t)}{dt} \right| \tag{5.43}$$

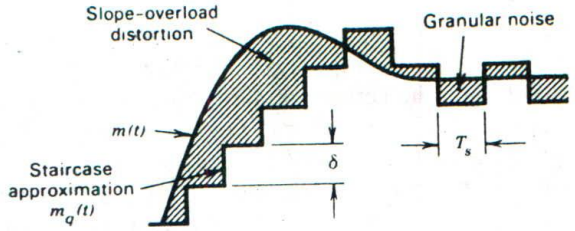


Figure 5.19
Quantizing error in delta modulation.

be satisfied. Otherwise, we find that the absolute value of the representation level δ is too small for the staircase approximation $m_q(t)$ to follow a steep segment of the input waveform $m(t)$, with the result that $m_q(t)$ falls behind $m(t)$, as illustrated in Fig. 5.19. This condition is called *slope overload*, and the resulting quantizing error is called *slope-overload distortion* (noise). Note that since the maximum slope of the staircase approximation $m_q(t)$ is fixed by the value of δ, increases and decreases in $m_q(t)$ tend to occur along straight lines. For this reason, a delta modulator using a fixed δ is often referred to as a *linear delta modulator*.

In contrast to slope-overload distortion, *granular noise* occurs when δ is too large relative to the local slope characteristics of the input waveform $m(t)$, thereby causing the staircase approximation $m_q(t)$ to hunt around a relatively flat segment of the input waveform; this phenomenon is also illustrated in Fig. 5.19. The granular noise is analogous to quantizing noise in a PCM system.

We thus see that there is a need to have a large δ so as to accommodate a wide dynamic range, whereas a small δ is required for the accurate representation of relatively low-level signals. It is therefore clear that the choice of the optimum δ that minimizes the mean-square value of the quantizing error in linear delta modulation will be the result of a compromise between slope overload distortion and granular noise.

EXERCISE 6 From Fig. 5.18 we see that the *step size* of a linear delta modulator is

$$\Delta = 2\delta$$

What is the average power of the granular noise expressed in terms of δ?

5.9 DISCUSSION

In this section we discuss the advantages and disadvantages of DPCM and DM, compared with standard PCM, for the encoding of voice and television signals.

The "relative" behavior of standard PCM and DPCM systems is much the same with either uniform or logarithmic quantizing, because the repertoire of signals consists of waveforms similar in character, but differing in mean level. In the case of voice signals, it is found that the signal-to-quantizing noise advantage of DPCM over standard PCM is in the neighborhood of 4–11 dB. The greatest improvement occurs in going from no prediction to first-order prediction, with some additional gain resulting from increasing the order of the prediction filter up to 4 or 5, after which little additional gain is obtained. Since 6 dB of quantizing noise is equivalent to 1 bit per sample, by virtue of Eq. 5.21, the advantage of DPCM may also be expressed in terms of bit rate. For a constant signal-to-quantizing noise ratio, and assuming a sampling rate of 8 kHz, the use of DPCM may provide a saving of about 8–16 kilobits per second (1 to 2 bits per sample) over standard PCM.

In the case of television signals, DPCM provides more of an advantage for high-resolution television systems than for low-resolution systems. For monochrome entertainment television, DPCM provides a signal-to-quantizing noise ratio of approximately 12 dB higher than standard PCM. For a constant signal-to-quantizing noise ratio, and assuming a sampling rate of 9 MHz, this represents a saving of about 18 megabits per second (2 bits per sample) by DPCM over PCM.

Turning next to delta modulation, subjective voice tests and noise measurements have shown that a DM system operating at 40 kilobits per second is equivalent to a standard PCM system operating with a sampling rate of 8 kHz and 5 bits per sample. At lower bit rates, DM is better than the standard PCM (the latter still using 8-kHz sampling and a reduced number of bits per sample), but at higher bit rates PCM is superior to DM. The quality of 5-bit PCM is low for most purposes in telephony. For telephone quality voice signals, it has become conventional to use 8-bit PCM. Equivalent voice quality with DM can be obtained only by using bit rates much higher than 64 kilobits per second.

Also, in a delta modulation system, operating on voice signals under optimum conditions, the SNR is increased by 9 dB by doubling the bit rate. By comparison, in the case of standard PCM, we achieve a 6 dB increase in SNR for each *added* bit. For example, by doubling the bit rate from 40 to 80 kilobits per second, the SNR is increased by 9 dB using DM. On the other hand, if PCM is employed and the bit rate is similarly doubled by increasing the number of bits per sample from 5 to 10 (keeping the sampling rate fixed at 8 kHz), the SNR is improved by 30 dB. Thus the increase of SNR with bit rate is much more dramatic for PCM than for DM.

The use of delta modulation is therefore recommended only in certain special circumstances: (1) if it is necessary to reduce the bit rate below 40 kilobits per second and limited voice quality is tolerable; or (2) if extreme circuit simplicity is of overriding importance and the accompanying use of a high-bit rate is acceptable. Note that since delta modulation uses a high

sampling rate, there is no need for employing a pre-alias filter prior to sampling in the transmitter.

ADAPTIVE DIGITAL CODING OF WAVEFORMS

From the discussion presented on PCM using a uniform quantizer with a fixed step size, we see that we have a dilemma in quantizing speech signals. On the one hand, we wish to choose the quantization step size large enough to accommodate the maximum peak-to-peak range of the input signal with the lowest possible number of representation levels. On the other hand, we would like to make the quantization step size small enough to minimize the average power of the quantizing noise. This issue is further compounded by the fact that the amplitude of the speech signal can vary over a wide range, depending on the speaker, the communication environment, and within a given utterance, from *voiced* to *unvoiced sounds*.[6] One approach to accommodating these conflicting requirements is to use a nonuniform quantizer; this approach is commonly used in PCM systems for telephony as described in Section 5.4. An alternative approach is to use an *adaptive quantizer,* wherein the step size is varied automatically so as to match the average power of the input speech signal; this second approach is commonly used in *adaptive DPCM (ADPCM) systems.*

In ADPCM systems used in telephony, the prediction filter is also adaptive. An *adaptive prediction filter* is responsive to changing level and spectrum of the input speech signal. The variation of performance with speakers and speech material, together with variations in signal level inherent in the speech communication process, make the combined use of adaptive quantization and adaptive prediction necessary to achieve best performance over a wide range of speakers and speaking situations.

It is of interest to note that improvements in circuit design and technology have made it possible for ADPCM to provide toll quality speech coding at 32 kb/s[7]; this corresponds to a sampling rate of 8 kHz and 4 bits per sample. By "toll quality" we mean the quality of commercial telephone service. This performance is comparable to that of 64 kb/s PCM incorporating the use of μ-law (logarithmic) companding with $\mu = 255$. However, unlike log-PCM, the performance of the ADPCM system is very signal dependent.

Finally, we should mention that a delta modulator may also be made adaptive, wherein the variable step size increases during a steep segment

[6] *Voiced sounds* are produced by forcing air through the glottis with the tension of the vocal cords adjusted so that they vibrate in a relaxation oscillation, thereby producing quasiperiodic pulses of air that excite the vocal tract. *Fricative* or *unvoiced sounds* are generated by forming a constriction at some point in the vocal tract (usually toward the mouth end) and forcing air through the constriction at a high enough velocity to produce turbulence. This creates a broad-spectrum noise source to excite the vocal tract.

[7] Jayant and Noll, 1984, pp. 309–311.

of the input signal and decreases when the modulator is quantizing an input signal with a slowly varying segment. In this way the step size is adapted to the level of the input signal. The resulting system is called an *adaptive delta modulator* (ADM).

5.10 TIME-DIVISION MULTIPLEXING

The sampling theorem enables us to transmit the complete information contained in a band-limited message signal by using samples of the message signal taken uniformly at a rate that is usually slightly higher than the Nyquist rate. An important feature of the sampling process is a conservation of time. That is, the transmission of the message samples engages the transmission channel for only a fraction of the sampling interval on a periodic basis, and in this way some of the time interval between adjacent samples is cleared for use by other independent message sources on a time-shared basis. We thereby obtain a *time-division multiplex system* (TDM), which enables the joint use of a common transmission channel by a plurality of independent message sources without mutual interference.

The concept of TDM is illustrated by the block diagram shown in Fig. 5.20. Each input message signal is first restricted in bandwidth by a low-pass filter to remove the frequencies that are nonessential to an adequate signal representation. The low-pass filter outputs are then applied to a *commutator* that is usually implemented using electronic switching circuitry. The function of the commutator is two-fold: (1) to take a narrow sample of each of the N input messages at a rate $1/T_s$ that is slightly higher than $2W$, where W is the cutoff frequency of the low-pass input filter, and (2) to sequentially interleave these N samples inside a sampling interval T_s. Indeed, this latter function is the essence of the time-division multiplexing operation. Following the commutation process, the multiplexed signal is applied to a *pulse modulator,* (e.g., pulse-amplitude modulator), the purpose of which is to transform the multiplexed signal into a form suitable for transmission over the common channel. It is clear that the use of time-division multiplexing introduces a *bandwidth expansion factor N,* because the scheme must squeeze N samples derived from N independent message sources into a time slot equal to one sampling interval. At the receiving end of the system, the received signal is applied to a *pulse demodulator,* which performs the inverse operation of the pulse modulator. The narrow samples produced at the pulse demodulator output are distributed to the appropriate low-pass reconstruction filters by means of a decommutator, which operates in *synchronism* with the commutator in the transmitter. This synchronization is essential for the satisfactory operation of the system.

The TDM system is highly sensitive to dispersion in the common transmission channel, that is, to variations of amplitude with frequency or nonlinear phase response. Accordingly, accurate equalization of both the amplitude and phase responses of the channel is necessary to ensure a satisfactory

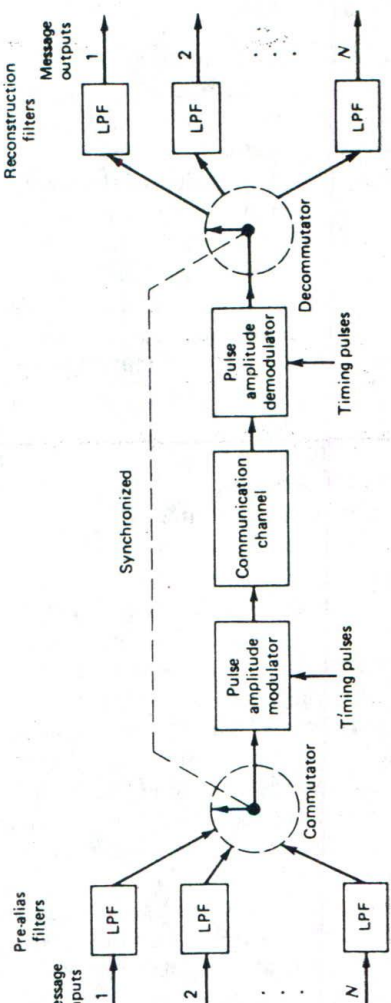


Figure 5.20
Block diagram of TDM system.

operation of the system. This issue is discussed in Chapter 6. To a first approximation, however, TDM is immune to amplitude nonlinearities in the channel as a source of crosstalk, because the different message signals are not simultaneously impressed on the channel.

5.11 *APPLICATION: DIGITAL MULTIPLEXERS FOR TELEPHONY*

In the previous section we introduced the idea of time-division multiplexing whereby a group of analog signals (e.g., voice signals) are sampled sequentially in time at a *common* sampling rate and then multiplexed for transmission over a common line. In this section we consider the *multiplexing of digital signals*[8] at different bit rates. This enables us to combine several digital signals, such as computer outputs, digitized voice signals, and digitized facsimile and television signals, into a single data stream (at a considerably higher bit rate than any of the inputs). Figure 5.21 shows a conceptual diagram of the digital multiplexing–demultiplexing operation.

The multiplexing of digital signals may be accomplished by using a *bit-by-bit interleaving procedure* with a selector switch that sequentially takes a bit from each incoming line and then applies it to the high-speed common line. At the receiving end of the system the output of this common line is separated out into its individual low-speed components and then delivered to their respective destinations.

Two major groups of digital multiplexers are used in practice:

1. One group of multiplexers is designed to combine relatively low-speed digital signals, up to a maximum rate of 4800 bits per second, into a higher speed multiplexed signal with a rate of up to 9600 bits per second. These multiplexers are used primarily to transmit data over voice-grade channels of a telephone network. Their implementation requires the use of *modems* in order to convert the digital format into an analog format suitable for transmission over telephone channels. The operation of a modem (modulator–demodulator) is covered in Section 7.14.

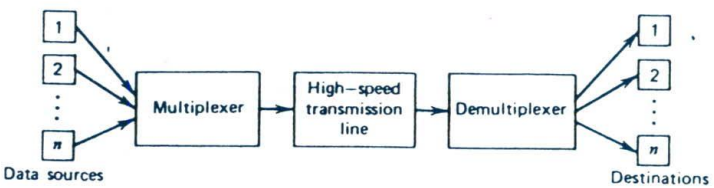


Figure 5.21
Conceptual diagram of multiplexing–demultiplexing.

[8]For more detailed information on the multiplexing of digital signals, see Bell Telephone Laboratories (1970).

2. The second group of multiplexers, designed to operate at much higher bit rates, forms part of the data transmission service generally provided by communication carriers. For example, Fig. 5.22 is a block diagram of the digital hierarchy based on the T1 carrier, which has been developed by the Bell System. The T1 carrier, described later on, is designed to operate at 1.544 megabits per second, the T2 at 6.312 megabits per second, the T3 at 44.736 megabits per second, and the T4 at 274.176 megabits per second. The system is thus made up of various combinations of lower order T-carrier subsystems. These subsystems are designed to accommodate the transmission of voice signals, Picturephone® service, and television signals using PCM, as well as (direct) digital signals from data terminal equipment.

There are some basic problems involved in the design of a digital multiplexer, irrespective of its grouping:

1. Digital signals cannot be directly interleaved into a format that allows for their eventual separation unless their bit rates are locked to a common clock. Accordingly, provision has to be made for *synchronization* of the incoming digital signals, so that they can be properly interleaved.
2. The multiplexed signal must include some form of *framing*, so that its individual components can be identified at the receiver.
3. The multiplexer has to handle small variations in the bit rates of the incoming digital signals. For example, a 1000-kilometer coaxial cable carrying 3×10^8 pulses per second will have about 1 million pulses in transit, with each pulse occupying about 1 meter of the cable. A 0.01%

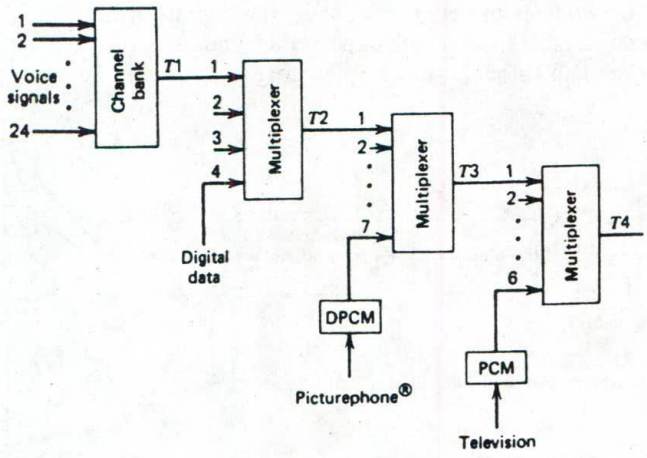


Figure 5.22
Digital hierarchy, Bell system.

variation in the propagation delay, produced by a 1°F decrease in temperature, will result in 100 fewer pulses in the cable. Clearly, these pulses must be absorbed by the multiplexer.

To cater to the requirements of synchronization and rate adjustment to accommodate small variations in the input data rates, we may use a technique known as *bit stuffing*. The idea here is to have the outgoing bit rate of the multiplexer slightly higher than the sum of the maximum expected bit rates of the input channels. This is achieved by stuffing in additional non-information-carrying pulses. All incoming digital signals are stuffed with a number of bits sufficient to raise their respective bit rates to that of a locally generated clock. To accomplish bit stuffing, each incoming digital signal or bit stream is fed into an *elastic store* at the multiplexer. The elastic store is a device that stores a bit stream in such a manner that the stream may be read out at a rate different from the rate at which it is read in. At the demultiplexer, the stuffed bits must obviously be removed from the multiplexed signal. This requires some method of identifying the stuffed bits. To illustrate one such method, and also to show one method of providing frame synchronization, we describe the signal format of the Bell System *M12 multiplexer*, which is designed to combine four T1 bit streams into one T2 bit stream. We begin the description by considering the T1 system.

T1 SYSTEM

The *T1 carrier system* is designed to accommodate 24 voice channels, primarily for short-distance, heavy usage in metropolitan areas. The T1 system was pioneered by the Bell System in the United States in the early 1960s, and with its introduction the shift to digital communication facilities started. The T1 system has been adopted for use throughout the United States, Canada, and Japan. It forms the basis for a complete hierarchy of higher-order multiplexed systems that are used for either long-distance transmission or transmission in heavily populated urban centers.

A voice signal (male or female) is essentially limited to a band from 300 to 3400 Hz in the sense that frequencies outside this band do not contribute much to articulation efficiency. Indeed, telephone circuits that respond to this range of frequencies give quite satisfactory service. Accordingly, it is customary to pass the voice signal through a low-pass filter with a cutoff frequency of about 3.4 kHz prior to sampling. Hence, the nominal value of the Nyquist rate is 6.8 kHz. The filtered voice signal is usually sampled at a slightly higher rate, namely, 8 kHz, which is the *standard* sampling rate in digital telephony.

For companding, the T1 system uses a *piecewise-linear* characteristic (consisting of 15 linear segments) to approximate the logarithmic μ-law of Eq. 5.16 with the constant $\mu = 255$. This approximation is constructed in such a way that the segment end points lie on the compression curve

computed from Eq. 5.16, and their projections onto the vertical axis are spaced uniformly. Table 5.4 gives the projections of the segment end points onto the horizontal axis and the step sizes of the individual segments. The table is normalized to 8159, so that all values are represented as integer numbers. Segment 0 of the approximation is a linear segment, passing through the origin; it contains a total of 31 uniform representation levels. Linear segments $1a$, $2a$, . . . , $7a$ lie above the horizontal axis, whereas linear segments $1b$, $2b$, . . . , $7b$ lie below the horizontal axis; each of these 14 segments contains 16 uniform representation levels. For colinear segment 0 the representation levels at the compressor input are ± 1, ± 3, . . . , ± 31, and the corresponding compressor output levels are 0, ± 1, . . . , ± 15. For linear segments $1a$ and $1b$ the representation levels at the compressor input are ± 33, ± 35, . . . , ± 95, and the corresponding compressor output levels are ± 16, ± 17, . . . , ± 31, and so on for the other linear segments.

There are a total of $31 + 14 \times 16 = 255$ output levels associated with the 15-segment companding characteristic described herein. To accommodate this number of output levels, each of the 24 voice channels uses a binary code with an 8-bit word. The first bit indicates whether the input voice sample is positive or negative. The next three bits of the code word identify the particular segment inside which the amplitude of the input voice sample lies, and the last four bits identify the actual quantizing step inside that segment.

With a sampling rate of 8 kHz, each frame of the multiplexed signal occupies a period of 125 μs. In particular, it consists of twenty-four 8-bit words, plus a single bit that is added at the end of the frame for the purpose of synchronization. Hence, each frame consists of a total of $24 \times 8 + 1 = 193$ bits. Correspondingly, the duration of each bit equals 0.647 μs, and the resultant transmission rate is 1.544 megabits per second.

In addition to the voice signal, a telephone system must also pass special

TABLE 5.4 The 15-Segment μ-law Companding Characteristic ($\mu = 255$)

Linear Segment Number	Step Size	Projections of Segment End Point onto the Horizontal Axis
0	2	± 31
$1a$, $1b$	4	± 95
$2a$, $2b$	8	± 223
$3a$, $3b$	16	± 479
$4a$, $4b$	32	± 991
$5a$, $5b$	64	± 2015
$6a$, $6b$	128	± 4063
$7a$, $7b$	256	± 8159

supervisory signals to the far end. This *signaling information* is needed to transmit dial pulses, as well as telephone off-hook/on-hook signals. In the T1 system this requirement is accomplished as follows. Every sixth frame, the least significant (i.e., the eighth) bit of each voice channel is deleted and a *signaling bit* is inserted in its place, thereby yielding an average $7\frac{5}{8}$-bit operation for each voice input. The sequence of signaling bits is thus transmitted at a rate equal to the sampling rate divided by six, that is, 1.333 kilobits per second.

M12 MULTIPLEXER

Figure 5.23 illustrates the signal format of the M12 multiplexer. Each frame is subdivided into four subframes. The first subframe (first line in Fig. 5.23) is transmitted, then the second, the third, and the fourth, in that order.

Bit-by-bit interleaving of the incoming four T1 bit streams is used to accumulate a total of 48 bits, 12 from each input. A *control bit* is then inserted by the multiplexer. Each frame contains a total of 24 control bits, separated by sequences of 48 data bits. Three types of control bits are used in the M12 multiplexer to provide synchronization and frame indication, and to identify which of the four input signals has been stuffed. These control bits are labeled as F, M, and C in Fig. 5.23. Their functions are:

1. The F-control bits, two per subframe, constitute the *main* framing pulses. The subscripts on the F-control bits denote the actual bit (0 or 1) transmitted. Thus the main framing sequence is $F_0F_1F_0F_1F_0F_1F_0F_1$ or 01010101.
2. The M-control bits, one per subframe, form *secondary* framing pulses to identify the four subframes. Here again the subscripts on the M-control bits denote the actual bit (0 or 1) transmitted. Thus the secondary framing sequence is $M_0M_1M_1M_1$ or 0111.
3. The C-control bits, three per subframe are *stuffing indicators*. In particular, C_{I} refers to input channel I, C_{II} refers to input channel II, and so forth. For example, the three C-control bits following M_0 in the first subframe are stuffing indicators for the first T1 signal. The insertion of a stuffed bit in this T1 signal is indicated by setting all three C-control bits to 1. To indicate no stuffing, all three are set to 0. If the three C-control bits indicate stuffing, the stuffed bit is located in the position of

$$
\begin{array}{llllllllllll}
M_0 & [48] & C_{\text{I}} & [48] & F_0 & [48] & C_{\text{I}} & [48] & C_{\text{I}} & [48] & F_1 & [48] \\
M_1 & [48] & C_{\text{II}} & [48] & F_0 & [48] & C_{\text{II}} & [48] & C_{\text{II}} & [48] & F_1 & [48] \\
M_1 & [48] & C_{\text{III}} & [48] & F_0 & [48] & C_{\text{III}} & [48] & C_{\text{III}} & [48] & F_1 & [48] \\
M_1 & [48] & C_{\text{IV}} & [48] & F_0 & [48] & C_{\text{IV}} & [48] & C_{\text{IV}} & [48] & F_1 & [48]
\end{array}
$$

Figure 5.23
Signal format of Bell system M12 multiplexer.

the first information bit associated with the first T1 signal that follows the F_1-control bit in the same subframe. In a similar way, the second, third, and fourth T1 signals may be stuffed, as required. By using *majority logic decoding* in the receiver, a single error in any of the three C-control bits can be detected. This form of decoding means simply that the majority of the C-control bits determine whether an all-one or all-zero sequence was transmitted. Thus three 1's or combinations of two 1's and a 0 indicate that a stuffed bit is present in the information sequence, following the control bit F_1 in the pertinent subframe. On the other hand, three 0's or combinations of two 0's and a 1 indicate that no stuffing is used.

The demultiplexer at the receiving M12 unit first searches for the main framing sequence $F_0F_1F_0F_1F_0F_1F_0F_1$. This establishes identity for the four input T1 signals and also for the M- and C-control bits. From the $M_0M_1M_1M_1$ sequence, the correct framing of the C-control bits is verified. Finally, the four T1 signals are properly demultiplexed and destuffed.

This signal format has two safeguards:

1. It is possible, although unlikely, that with just the $F_0F_1F_0F_1F_0F_1F_0F_1$ sequence, one of the incoming T1 signals may contain a similar sequence. This could then cause the receiver to lock onto the wrong sequence. The presence of the $M_0M_1M_1M_1$ sequence provides verification of the genuine $F_0F_1F_0F_1F_0F_1F_0F_1$ sequence, thereby ensuring that the four T1 signals are properly demultiplexed.
2. The single-error correction capability built into the C-control bits ensures that the four T1 signals are properly destuffed.

EXAMPLE 5: CAPACITY OF M12 MULTIPLEXER

The capacity of the M12 multiplexer to accommodate small variations in the input data rates can be calculated from the format of Fig. 5.23. In each M frame, defined as the interval containing one cycle of $M_0M_1M_1M_1$ bits, one bit can be stuffed into each of four input T1 signals. Each such signal has $12 \times 6 \times 4 = 288$ positions in each M frame. Also the T1 signal has a bit rate equal to 1.544 megabits per second. Hence, each input can be incremented by

$$1.544 \times 10^6 \times \frac{1}{288} = 5.4 \text{ kilobits/s}$$

This result is much larger than the expected change in the bit rate of the incoming T1 signal. It follows therefore that the use of only one stuffed bit per input channel in each frame is sufficient to accommodate expected variations in the input signal rate.

The local clock that determines the outgoing bit rate also determines the nominal *stuffing rate S*, defined as the average number of bits stuffed per channel in any frame. The M12 multiplexer is designed for $S = 1/3$. Accordingly, the nominal bit rate of the T2 line is

$$1.544 \times 4 \times \frac{49}{48} \times \frac{288}{288\text{-}S} = 6.312 \text{ megabits/s}$$

This also ensures that the nominal T2 clock frequency is a multiple of 8 kHz (the nominal sampling rate of a voice signal), which is a desirable feature.

EXERCISE 7 Given that the data rate for one Picturephone® service is 6:312 megabits per second, and that for one television service is 44.736 megabits per second, determine the capacity of each Bell Telephone system level measured in terms of the number of (a) voice, (b) picturephone, or (c) television channels that it can accommodate.

PROBLEMS

P5.3 Sampling

Problem 1 Figure P5.1 depicts the spectrum of a message signal $m(t)$. The signal is *undersampled* at a rate of 1.5 Hz.

(a) Sketch the spectrum of the sampled version of this signal.
(b) The sampled signal is passed through an idealized low-pass interpolation filter of bandwidth 1 Hz. Sketch the spectrum of the resulting filter output.

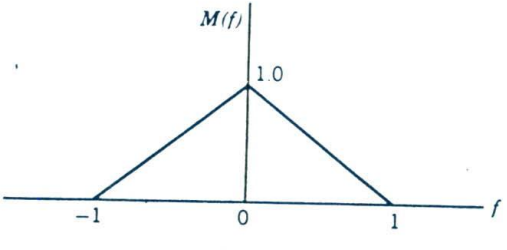


Figure P5.1

Problem 2 Consider the operation of a sample-and-hold circuit with the following parameters:

Message bandwidth, $W = 1$ Hz
Sampling period, $T_s = 0.4$ s
Pulse duration, $T = 0.2$ s

(a) Calculate the amplitude distortion produced by the aperture effect (arising from the use of flat-top samples) at the highest frequency component of the message signal.
(b) Find the amplitude response of the equalizer required to compensate for the aperture effect.

P5.4 Quantizing

Problem 3

(a) A sinusoidal signal, with an amplitude of 3.25 V is applied to a uniform quantizer of the *midtread* type with output values of 0, ±1, ±2, ±3 V, as in Fig. P5.2*a*. 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 *midriser* type with output values ±0.5, ±1.5, ±2.5, ±3.5 V, as in Fig. P5.2*b*.

Problem 4 The signal

$$m(t) = 6 \sin(2\pi t) \text{ volts}$$

is transmitted using a 4-bit binary PCM system. The quantizer is of the midriser type, with a step size of 1 V. Sketch the resulting sequence of quantized samples for one complete cycle of the input. Assume a sampling

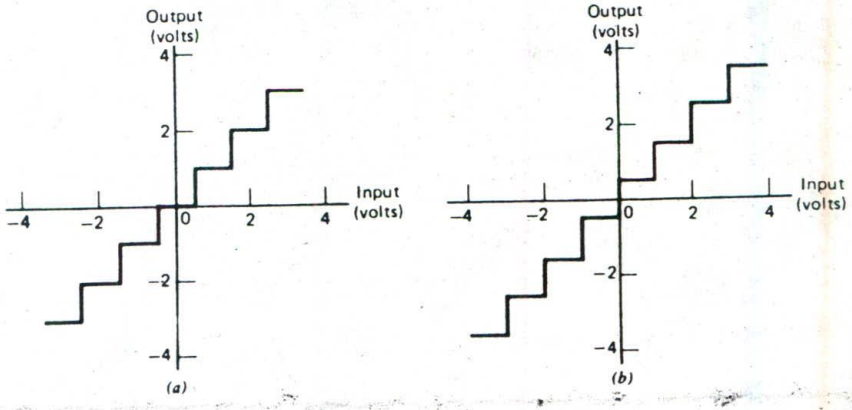


Figure P5.2

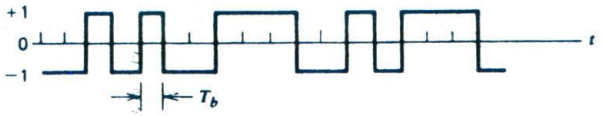


Figure P5.3

rate of four samples per second, with samples taken at $t = \pm 1/8, \pm 3/8, \pm 5/8, \ldots,$ seconds.

P5.5 Coding

Problem 5 Consider the following binary sequences:

(a) An alternating sequence of 1's and 0's.
(b) A long sequence of 1's followed by a long sequence of 0's.
(c) A long sequence of 1's followed by a single 0 and than a long sequence of 1's.

Sketch the waveform for each of these sequences using the following methods of representing symbols 1 and 0:

(a) On–off signaling.
(b) Polar signaling.
(c) Return-to-zero signaling.
(d) Bipolar signaling.
(e) Manchester code.

Problem 6 Figure P5.3 shows a PCM wave in which the amplitude levels of +1 V and −1 V are used to represent binary symbols 1 and 0, respectively. The code word used consists of three bits. Find the sampled version of an analog signal from which this PCM wave is derived.

Problem 7 The bipolar waveform of Fig. 5.12*d*, representing the binary sequence 0110100011, is transmitted over a noisy channel. The received waveform is shown in Fig. P5.4, which contains a single error. Locate the position of this error, giving reasons for your answer.

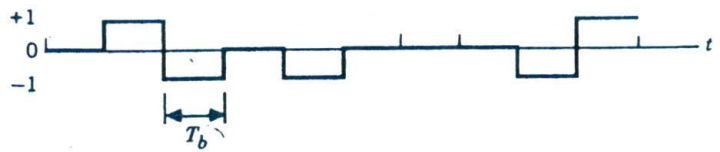


Figure P5.4

Problem 8 A PCM system uses a uniform quantizer followed by a 7-bit binary encoder. The bit rate of the system is equal to 50 megabits per second.

(a) What is the maximum message bandwidth for which the system operates satisfactorily?

(b) Determine the output signal-to-quantizing noise ratio when a full-load sinusoidal modulating wave of frequency 1 MHz is applied to the input.

P5.7 Differential Pulse-Code Modulation

Problem 9 In the DPCM system depicted in Fig. P5.5, show that in the absence of noise in the channel, the transmitting and receiving prediction filters operate on slightly different input signals.

P5.8 Delta Modulation

Problem 10 Consider a sine wave of frequency f_m and amplitude A_m, applied to a delta modulator with representation levels $\pm\delta$. 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?

Problem 11 The ramp signal $m(t) = at$ is applied to a delta modulator that operates with a sampling period T_s and representation levels $\pm\delta$.

(a) Show that slope-overload distortion occurs if $\delta < aT_s$.

(b) Sketch the modulator output for the following three values of δ:

(i) $\delta = 0.75\, aT_s$

(ii) $\delta = aT_s$

(iii) $\delta = 1.25\, aT_s$

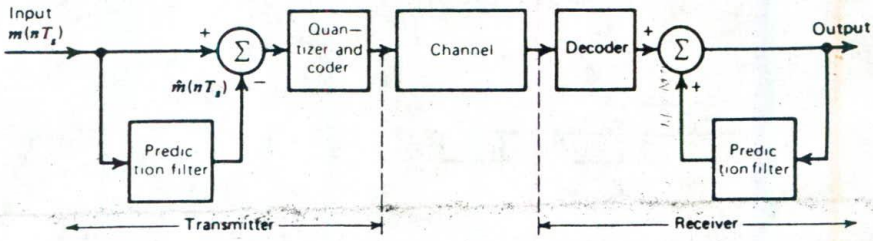


Figure P5.5

Problem 12 Consider a speech signal with maximum frequency of 3.4 kHz and maximum amplitude of 1 V. This speech signal is applied to a delta modulator with its bit rate set at 20 kilobits per second. Discuss the choice of an appropriate step size for the modulator.

P5.10 Time-Division Multiplexing

Problem 13 Six independent message sources of bandwidths W, W, $2W$, $2W$, $3W$, and $3W$ hertz are to be transmitted on a time-division multiplexed basis using a common communication channel.

(a) Set up a scheme for accomplishing this multiplexing requirement, with each message signal sampled at its Nyquist rate.
(b) Determine the minimum transmission bandwidth of the channel.

Problem 14 Twenty-four voice signals are sampled uniformly and then time-division multiplexed. 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.

INTERSYMBOL INTERFERENCE AND ITS CURES

When digital data (of whatever origin) is transmitted over a *band-limited channel*, dispersion in the channel gives rise to a troublesome form of interference called intersymbol interference. As the name implies, *intersymbol interference* refers to interference caused by the time response of the channel *spilling over* from one symbol into another. This has the effect of introducing deviations (errors) between the data sequence reconstructed at the receiver output and the original data sequence applied to the transmitter input. Hence, unless corrective measures are taken, intersymbol interference may impose a limit on the attainable rate of data transmission that is far below the physical capability of the channel.

In this chapter, we study the *intersymbol interference problem* and the use of *baseband pulse shaping* as the solution to the problem. The term "baseband" is used to designate the band of frequencies representing the original signal as delivered by a source of information.

6.1 BASEBAND TRANSMISSION OF BINARY DATA

For the *baseband transmission of digital data,* the use of *discrete pulse-amplitude modulation* (PAM) provides the most efficient form of discrete pulse modulation in terms of power and bandwidth use. In discrete PAM, *the amplitude of the transmitted pulses is varied in a discrete manner in accordance with the given digital data.*

The basic elements of a *baseband binary* PAM *system* are shown in Fig. 6.1. The signal applied to the input of the system consists of a binary data sequence $\{b_k\}$ with a *bit duration* of T_b seconds; b_k is in the form of 1 or 0. This signal is applied to a pulse generator, producing the pulse waveform

$$x(t) = \sum_{k=-\infty}^{\infty} A_k g(t - kT_b) \tag{6.1}$$

where $g(t)$ denotes a *shaping pulse* with its value at time $t = 0$ defined by

$$g(0) = 1$$

The amplitude A_k depends on the identity of the input bit b_k; specifically, we assume that

$$A_k = \begin{cases} +a, & \text{if the input bit } b_k \text{ is symbol 1} \\ -a, & \text{if the input bit } b_k \text{ is symbol 0} \end{cases} \tag{6.2}$$

The PAM signal $x(t)$ passes through a *transmitting filter* of transfer function $H_T(f)$. The resulting filter output defines the transmitted signal, which is modified in a deterministic fashion as a result of transmission through the channel of transfer function $H_C(f)$. The signal at the receiver input is passed through a *receiving filter* of transfer function $H_R(f)$. This filter output is sampled *synchronously* with the transmitter, with the sampling instants being determined by a *clock* or *timing signal* that is usually extracted from the receiving filter output. Finally, the sequence of samples thus obtained is used to reconstruct the original data sequence by means of a *decision device.* The amplitude of each sample is compared to a *threshold.* If the threshold is exceeded, a decision is made in favor of symbol 1 (say). If the threshold is not exceeded, a decision is made in favor of symbol 0. If the sample amplitude equals the threshold exactly, the symbol may be chosen as 0 or 1 without affecting overall performance. In such an event, we will choose symbol 0.

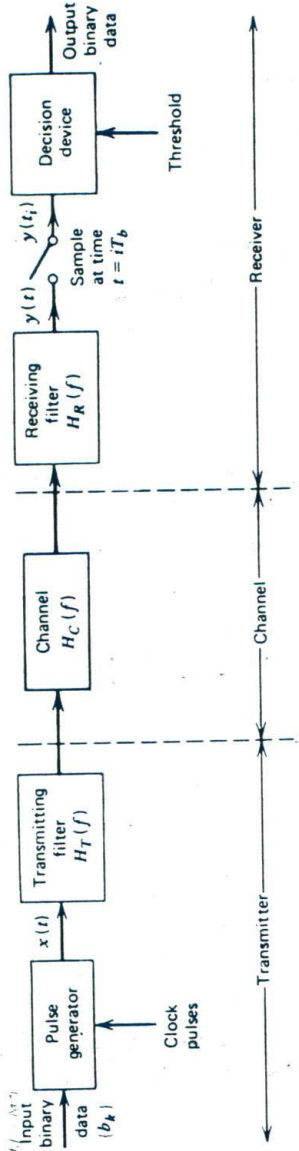


Figure 6.1
Baseband binary data transmission system.

The model shown in Fig. 6.1 represents not only a data transmission system inherently baseband in nature (e.g., data transmission over a coaxial cable) but also the baseband *equivalent* of a linear modulation system used to transmit data over a band-pass channel (e.g., telephone channel). In the latter case, the baseband equivalent model of the data transmission system is developed by using the ideas presented in Section 3.5. Linear modulation techniques for transmitting digital data over band-pass channels are considered in Chapter 7.

6.2 THE INTERSYMBOL INTERFERENCE PROBLEM

For the present discussion, we assume that the channel is *noiseless*. We do so in order to focus attention on the effects of imperfections in the frequency response of the channel (i.e., dispersion of the pulse shape by the channel) on data transmission through the channel. The effect of channel noise on the receiver output is considered in Chapter 10.

The receiving filter output in Fig. 6.1 may be written as[1]

$$y(t) = \mu \sum_{k=-\infty}^{\infty} A_k p(t - kT_b) \tag{6.3}$$

where μ is a scaling factor. The pulse $p(t)$ has a shape different from that of $g(t)$, but it is normalized such that

$$p(0) = 1$$

The pulse $\mu A_k p(t)$ is the response of the cascade connection of the transmitting filter, the channel, and the receiving filter, which is produced by the pulse $A_k g(t)$ applied to the input of this cascade connection. Therefore, we may relate $p(t)$ to $g(t)$ in the frequency domain as follows (after cancelling the common factor A_k)

$$\mu P(f) = G(f)H_T(f)H_C(f)H_R(f) \tag{6.4}$$

where $P(f)$ and $G(f)$ are the Fourier transforms of $p(t)$ and $g(t)$, respectively.

The receiving filter output $y(t)$ is sampled at time $t_i = iT_b$ (with i taking

[1]To be precise, an arbitrary time delay t_0 should be included in the argument of the pulse $p(t - kT_b)$ in Eq. 6.3 to represent the effect of transmission delay through the system. For convenience, we have put this delay equal to zero in Eq. 6.3.

on integer values), yielding

$$y(t_i) = \mu \sum_{k=-\infty}^{\infty} A_k p[(i-k)T_b]$$

$$= \mu A_i + \mu \sum_{\substack{k=-\infty \\ k \neq i}}^{\infty} A_k p[(i-k)T_b] \qquad i = 0, \pm 1, \pm 2, \ldots \quad (6.5)$$

In Eq. 6.5 the first term μA_i represents the contribution of the ith transmitted bit. The second term represents the residual effect of all other transmitted bits on the decoding of the ith received bit; this residual effect is called *intersymbol interference* (ISI).

In the absence of ISI (and, of course, channel noise), we observe from Eq. 6.5 that

$$y(t_i) = \mu A_i \quad (6.6)$$

which shows that, under these conditions, the ith transmitted bit can be decoded correctly. The unavoidable presence of ISI in the system, however, introduces errors in the decision device at the receiver output. Therefore, in the design of the transmitting and receiving filters, the objective is to minimize the effects of ISI, and thereby deliver the digital data to its destination with the smallest error rate possible.

Typically, the channel transfer function $H_C(f)$ and the pulse spectrum $G(f)$ are specified, and the problem is to determine the transfer functions of the transmitting and receiving filters, $H_T(f)$ and $H_R(f)$, so as to enable the receiver to correctly decode the received sequence of sample values $\{y(t_i)\}$ in accordance with Eq. 6.6. Deviation from this ideal condition is caused by the presence of intersymbol interference that arises owing to dispersion of the pulse shape by the channel. To solve the problem, we have to exercise control over intersymbol interference, an issue that is discussed next.

6.3 IDEAL SOLUTION

Control of intersymbol interference in the system is achieved in the time domain by controlling the function $p(t)$, or in the frequency domain by controlling $P(f)$. One signal waveform that produces *zero* intersymbol interference is defined by the *sinc function:*

$$p(t) = \frac{\sin(2\pi B_0 t)}{2\pi B_0 t} = \text{sinc}(2B_0 t) \quad (6.7)$$

where

$$B_0 = \frac{1}{2T_b} \tag{6.8}$$

The parameter B_0 is called the *Nyquist bandwidth;* it defines the minimum transmission bandwidth for zero intersymbol interference. According to Eq. 6.8, the Nyquist bandwidth B_0 is equal to one half of the *bit rate* $1/T_b$. Note the analogy between this relation and the sampling theorem for strictly band-limited signals. (The sampling theorem was discussed in Sections 2.7 and 5.3).

The frequency response $P(f)$, representing the Fourier transform of the pulse $p(t)$ of Eq. 6.7, is defined by

$$P(f) = \begin{cases} \dfrac{1}{2B_0}, & 0 \leq |f| < B_0 \\ 0 & B_0 < |f| \end{cases} \tag{6.9}$$

This means that no frequencies of absolute value exceeding half the bit rate are needed. The function $p(t)$ can be regarded as the impulse response of an ideal low-pass filter with an amplitude response of $1/(2B_0)$ in the passband and a bandwidth B_0. The function $p(t)$ has its peak value at the origin and goes through zero at integer multiples of the bit duration T_b. It is apparent that if the received waveform $y(t)$ is sampled at the instants of time $t = 0, \pm T_b, \pm 2T_b, \ldots$, then the pulses defined by $A_i p(t - iT_b)$ with arbitrary amplitude A_i and $i = 0, \pm 1, \pm 2, \ldots$, will not interfere with each other.

Although this ideal choice of pulse shape for $p(t)$ achieves economy in bandwidth in that it solves the problem of zero intersymbol interference with the minimum bandwidth possible, there are two difficulties that make its use for system design impractical:

1. It requires that frequency response $P(f)$ be flat from $-B_0$ to B_0, and zero elsewhere. This is physically unrealizable, and very difficult to approximate in practice because of the abrupt transitions at $\pm B_0$.
2. The time function $p(t)$ decreases as $1/|t|$ for large $|t|$, resulting in a slow rate of decay. This is caused by the discontinuity of $P(f)$ at $\pm B_0$. Accordingly, there is practically no margin of error in sampling times in the receiver.

To evaluate the effect of this *timing error,* consider the sample of $y(t)$ at $t = \Delta t$, where Δt is the timing error. To simplify the analysis, we have put

the correct sampling time t_i equal to zero. We thus obtain, in the absence of noise:

$$\begin{aligned} y(\Delta t) &= \mu \sum_k A_k p(\Delta t - kT_b) \\ &= \mu \sum_k A_k \operatorname{sinc}[2B_0(\Delta t - kT_b)] \end{aligned}$$

Since $2B_0T_b = 1$, we may rewrite this relation as

$$\begin{aligned} y(\Delta t) &= \mu \sum_k A_k \operatorname{sinc}(2B_0 \Delta t - k) \\ &= \mu A_0 \operatorname{sinc}(2B_0 \Delta t) + \mu \frac{\sin(2\pi B_0 \Delta t)}{\pi} \sum_{\substack{k \\ k \neq 0}} \frac{(-1)^k A_k}{2B_0 \Delta t - k} \end{aligned} \tag{6.10}$$

The first term on the right side of Eq. 6.10 defines the desired symbol, whereas the remaining series represents the intersymbol interference caused by the timing error Δt in sampling the signal $y(t)$. In certain cases, it is possible for this series to diverge, thereby causing erroneous decisions in the receiver.

We therefore have to look to other pulse shapes not only to combat the intersymbol interference problem but also to do so in a feasible way. In the next section, we present one such solution that is a natural extension of the minimum-bandwidth (ideal) solution just described.

6.4 RAISED COSINE SPECTRUM

The solution we have in mind differs from the ideal solution in one important respect: the overall frequency response $P(f)$ decreases toward zero gradually rather than abruptly. In particular, $P(f)$ consists of a *flat* portion and a *rolloff* portion that has the form of a raised-cosine function, as follows[2]

$$P(f) = \begin{cases} \dfrac{1}{2B_0}, & 0 \leq |f| < f_1 \\ \dfrac{1}{4B_0}\left\{1 + \cos\left[\dfrac{\pi(|f| - f_1)}{2B_0 - 2f_1}\right]\right\}, & f_1 < |f| < 2B_0 - f_1 \\ 0, & 2B_0 - f_1 < |f| \end{cases} \tag{6.11}$$

[2]The solution described in Eq. 6.11 was first proposed by Nyquist (1928) in his studies of telegraph transmission theory.

The frequency f_1 and the Nyquist bandwidth B_0 are related by

$$\alpha = 1 - \frac{f_1}{B_0} \tag{6.12}$$

which is called the *rolloff factor*. For $\alpha = 0$, i.e., $f_1 = B_0$, we get the minimum bandwidth solution described in Section 6.3.

The frequency response $P(f)$, normalized by multiplying it by $2B_0$, is plotted in Fig. 6.2*a* for three values of α, namely, 0, 0.5, and 1. We see that for $\alpha = 0.5$ or 1, the function $P(f)$ cuts off gradually as compared with an ideal low-pass filter (corresponding to $\alpha = 0$), and it is therefore easier to realize in practice. Also the function $P(f)$ exhibits odd symmetry about the cutoff frequency B_0 of the ideal low-pass filter.

The time response $p(t)$, that is, the inverse Fourier transform of $P(f)$, is defined by

$$p(t) = \text{sinc}(2B_0t)\,\frac{\cos(2\pi\alpha B_0t)}{1 - 16\alpha^2B_0^2t^2} \tag{6.13}$$

This function consists of the product of two factors: the factor $\text{sinc}(2B_0t)$ associated with the ideal solution, and a second factor that decreases as $1/|t|^2$ for large $|t|$. The first factor ensures zero crossings of $p(t)$ at the desired sampling instants of time $t = iT$ with i an integer (positive and negative). The second factor reduces the tails of the pulse considerably below that obtained from the ideal low-pass filter, so that the transmission of binary waves using such pulses is relatively insensitive to sampling time errors. In fact, the amount of intersymbol interference resulting from this timing error decreases as the rolloff factor α is increased from zero to unity.

The time response $p(t)$ is plotted in Fig. 6.2*b* for $\alpha = 0$, 0.5 and 1. For the special case of $\alpha = 1$, the function $p(t)$ simplifies as

$$p(t) = \frac{\text{sinc}(4B_0t)}{1 - 16B_0^2t^2} \tag{6.14}$$

This time response exhibits two interesting properties:

1. At $t = \pm T_b/2 = \pm 1/4B_0$, we have $p(t) = 0.5$; that is, the pulse width measured at half amplitude is exactly equal to the bit duration T_b.
2. There are zero crossings at $t = \pm 3T_b/2, \pm 5T_b/2, \ldots$ in addition to the usual zero crossings at the sampling times $t = \pm T_b, \pm 2T_b, \ldots$.

These two properties are particularly useful in generating a timing signal from the received signal for the purpose of synchronization.

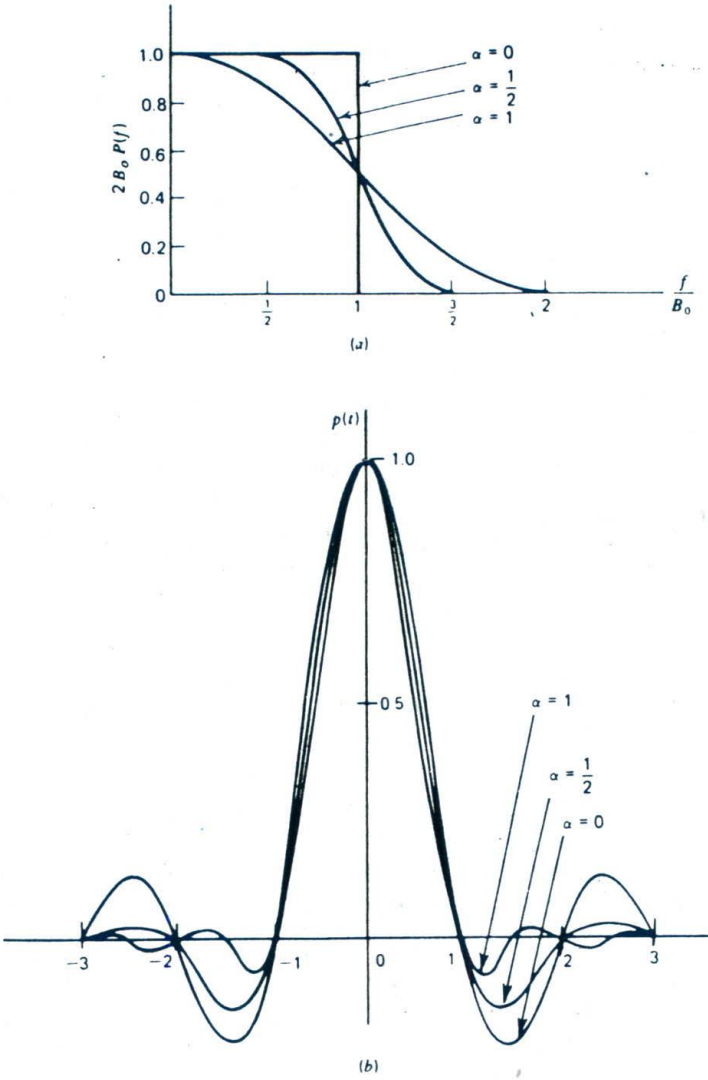


Figure 6.2
Responses for different rolloff factors. (a) Frequency response. (b) Time response.

EXERCISE 1 Given the frequency response $P(f)$ defined in Eq. 6.11, show that the inverse Fourier transform $p(t)$ is as given in Eq. 6.13.

TRANSMISSION BANDWIDTH REQUIREMENT

From Eq. 6.11 we see that the nonzero portion of the frequency response $P(f)$, resulting from use of the raised cosine spectrum, is limited to the interval $(0, 2B_0 - f_1)$ for positive frequencies. Accordingly, the transmission bandwidth required by using the raised cosine spectrum is given by

$$B = 2B_0 - f_1 \tag{6.15}$$

Eliminating the frequency f_1 between Eqs. 6.12 and 6.15, we get

$$B = B_0(1 + \alpha) \tag{6.16}$$

where B_0 is the Nyquist bandwidth and α is the rolloff factor. Thus, the transmission bandwidth requirement of the raised cosine solution exceeds that of the ideal solution by an amount equal to αB_0. Note that the ratio of the *excess bandwidth* (resulting from the raised cosine solution) to the Nyquist bandwidth (required by the ideal solution) equals the rolloff factor α.

The following two cases, one ideal and the other practical, are of particular interest:

1. When the rolloff factor α is zero, the excess bandwidth αB_0 is reduced to zero, thereby permitting the transmission bandwidth B to assume its minimum value B_0.
2. When the rolloff factor α is unity, the excess bandwidth is increased to B_0. Correspondingly, the transmission bandwidth B is doubled, compared to the (ideal) case 1.

EXAMPLE 1 BANDWIDTH REQUIREMENTS OF THE T1 SYSTEM

In Chapter 5 we described the signal format for the T1 carrier system that is used to multiplex 24 independent voice inputs, based on an 8-bit PCM word. It was shown that the bit duration of the resulting time-division multiplexed signal (including a framing bit) is

$$T_b = 0.647\ \mu\text{s}$$

The bit rate of the T1 system is

$$R_b = \frac{1}{T_b} = 1.544\ \text{Mb/s}$$

Assuming an ideal low-pass characteristic for the channel, it follows that the Nyquist bandwidth of the T1 system is

$$B_0 = \frac{1}{2T_b} = 772\ \text{kHz}$$

This is the minimum transmission bandwidth of the T1 system for zero intersymbol interference. However, a more realistic value for the transmission bandwidth B is obtained by using a raised cosine spectrum with $\alpha = 1$. In this case, we find that

$$B = \frac{1}{T_b} = 1.544 \text{ MHz}$$

EXERCISE 2 **Calculate the transmission bandwidth requirement of the M12 multiplexer described in Section 5.11. Assume the use of a raised cosine spectrum with rolloff factor $\alpha = 1$ for the baseband pulse shaping.**

6.5 CORRELATIVE CODING

Thus far we have treated intersymbol interference as an undesirable phenomenon that produces a degradation in system performance. Indeed, its very name connotes a nuisance effect. Nevertheless, by adding intersymbol interference to the transmitted signal in a controlled manner, it is possible to achieve a signaling rate of $2B_0$ symbols per second in a channel of bandwidth B_0 hertz. Such schemes are called *correlative coding* or *partial-response signaling* schemes.[3] The design of these schemes is based on the premise that since the intersymbol interference that is introduced into the transmitted signal is known, its effect can be accounted for at the receiver. Thus correlative coding may be regarded as a practical means of achieving the theoretical maximum signaling rate of $2B_0$ symbols per second in a bandwidth of B_0 hertz, using realizable and perturbation-tolerant filters.

In this section, we illustrate the basic idea of correlative coding by considering two specific examples: *duobinary signaling* and *modified duobinary signaling*. Duobinary signaling employs a correlation span of one binary digit, whereas modified duobinary signaling employs a correlation span of two binary digits; the use of "duo" is intended to imply doubling of the transmission capacity of a straight binary system.

DUOBINARY SIGNALING

Consider a binary input sequence $\{b_k\}$ consisting of uncorrelated binary digits each having duration T_b seconds, with symbol 1 represented by a

[3]Correlative coding and partial response signaling are synonomous; both terms are used in the literature. The idea of correlative coding was originated by Lender (1963). For an overview on correlative coding, see Pasupathy (1977).

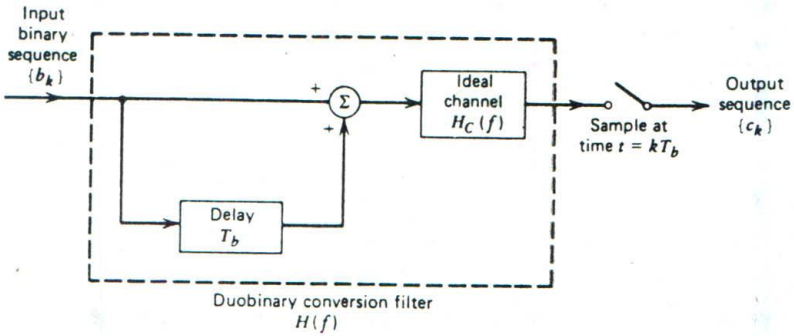


Figure 6.3
Duobinary signaling scheme.

pulse of amplitude +1 V, and symbol 0 by a pulse amplitude −1 V. When this sequence is applied to a *duobinary encoder*, it is converted into a *three-level output*, namely, −2, 0, and +2 V. To produce this transformation, we may express the digit c_k at the duobinary coder output as the sum of the present binary digit b_k and its previous value b_{k-1}, as shown by

$$c_k = b_k + b_{k-1} \tag{6.17}$$

One of the effects of the transformation described by Eq. 6.17 is to change the input sequence $\{b_k\}$ of uncorrelated binary digits into a sequence $\{c_k\}$ of correlated digits. This correlation between the adjacent transmitted levels may be viewed as introducing intersymbol interference into the transmitted signal in an artificial manner. However, this intersymbol interference is under the designer's control; this is the basis of correlative coding.

Figure 6.3 depicts the block diagram of a *duobinary encoder*, including a band-limited channel assumed to be ideal. The binary sequence $\{b_k\}$ is first passed through a simple filter consisting of the parallel combination of a direct path and an ideal element producing a *delay* of T_b seconds, where T_b is the bit duration. For every unit impulse applied to the input of this filter, we get two unit impulses spaced T_b seconds apart at the filter output. The output of this filter in response to the incoming binary sequence $\{b_k\}$ is then passed through the channel of transfer function $H_C(f)$. A continuous waveform is therefore produced at the channel output. The resulting waveform is sampled uniformly every T_b seconds, thereby producing the duobinary encoded sequence $\{c_k\}$. Note that the effect of the channel is included in this encoding operation.

The cascade connection of the delay-line filter and the channel is called a *duobinary conversion filter*. In Fig. 6.3, we have enclosed this filter inside a dashed rectangle. The response of the filter may be characterized in terms of an overall transfer function $H(f)$, which is evaluated next.

An ideal delay element, producing a delay of T_b seconds, has the transfer function $\exp(-j2\pi f T_b)$, so that the transfer function of the delay-line filter shown in Fig. 6.3 is $1 + \exp(-j2\pi f T_b)$. Hence, the overall transfer function of this filter connected in cascade with the ideal channel $H_C(f)$ is

$$\begin{aligned} H(f) &= H_C(f)[1 + \exp(-j2\pi f T_b)] \\ &= H_C(f)[\exp(j\pi f T_b) + \exp(-j\pi f T_b)] \exp(-j\pi f T_b) \\ &= 2H_C(f) \cos(\pi f T_b) \exp(-j\pi f T_b) \end{aligned} \tag{6.18}$$

For an ideal channel of bandwidth $B_0 = 1/2T_b$, we have

$$H_C(f) = \begin{cases} 1, & |f| \leq 1/2T_b \\ 0, & \text{otherwise} \end{cases} \tag{6.19}$$

Thus the overall frequency response has the form of a half-cycle cosine function, as shown by

$$H(f) = \begin{cases} 2 \cos(\pi f T_b) \exp(-j\pi f T_b), & |f| \leq 1/2T_b \\ 0, & \text{otherwise} \end{cases} \tag{6.20}$$

for which the amplitude response and phase response are as shown in parts *a* and *b* of Fig. 6.4, respectively. An advantage of this frequency response is that it can be easily approximated in practice.

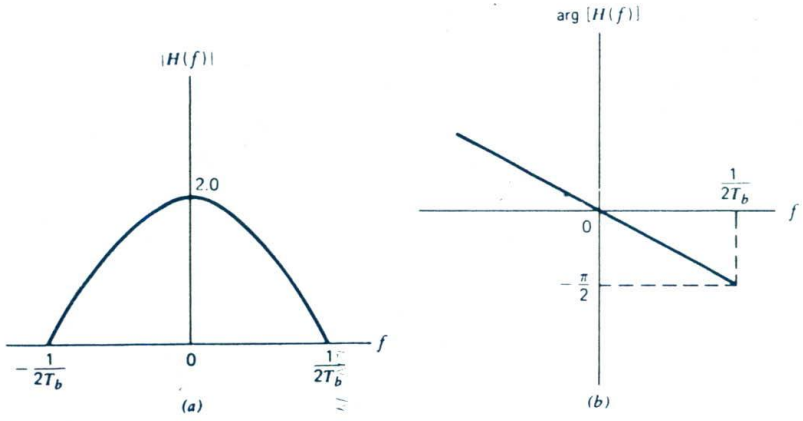


Figure 6.4
Frequency response of the duobinary conversion filter. (a) Amplitude response. (b) Phase response.

The corresponding value of the impulse response consists of two sinc pulses, time-displaced by T_b seconds, as shown by (except for a scaling factor)

$$\begin{aligned} h(t) &= \frac{\sin(\pi t/T_b)}{\pi t/T_b} + \frac{\sin[\pi(t - T_b)/T_b]}{\pi(t - T_b)/T_b} \\ &= \frac{\sin(\pi t/T_b)}{\pi t/T_b} - \frac{\sin(\pi t/T_b)}{\pi(t - T_b)/T_b} \\ &= \frac{T_b^2 \sin(\pi t/T_b)}{\pi t(T_b - t)} \end{aligned} \tag{6.21}$$

which is shown plotted in Fig. 6.5. We see that the overall impulse response $h(t)$ has only *two* distinguishable values at the sampling instants.

The original data $\{b_k\}$ may be detected from the duobinary-coded sequence $\{c_k\}$ by subtracting the previous decoded binary digit from the currently received digit c_k in accordance with Eq. 6.17. Specifically, letting $\hat{b}_k$ represent the *estimate* of the original binary digit b_k as conceived by the receiver at time $t = kT_b$, we have

$$\hat{b}_k = c_k - \hat{b}_{k-1} \tag{6.22}$$

It is apparent that if c_k is received without error and if the previous estimate $\hat{b}_{k-1}$ at time $t = (k - 1)T_b$ also corresponds to a correct decision, then the current estimate $\hat{b}_k$ will be correct too. The technique of using a stored estimate of the previous symbol in the estimation of the current symbol is called *decision feedback*.

We observe that the detection procedure as described here is essentially an inverse of the operation of the simple filter at the transmitter. However, a major drawback of this detection process is that once errors are made, they tend to propagate. This is because a decision on the current binary

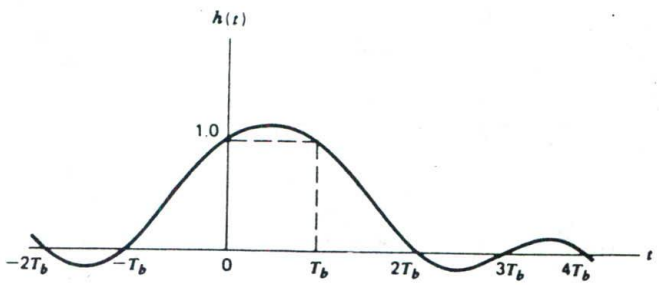


Figure 6.5
Impulse response of duobinary conversion filter.

digit b_k depends on the correctness of the decision made on the previous binary digit b_{k-1}.

A practical means of avoiding this error propagation is to use *precoding* before the duobinary coding, as shown in Fig. 6.6. The precoding operation performed on the input binary sequence $\{b_k\}$ converts it into another binary sequence $\{a_k\}$ defined by

$$a_k = b_k \oplus a_{k-1} \tag{6.23}$$

where the symbol $\oplus$ denotes *modulo-two addition* of the binary digits b_k and a_{k-1}. This addition is equivalent to the EXCLUSIVE OR operation. An EXCLUSIVE OR gate operates as follows. The output of a two-input EXCLUSIVE OR gate is a 1 if exactly one input is a 1; otherwise, the output remains a 0. The resulting precoder output $\{a_k\}$ is next applied to the duobinary coder, thereby producing the sequence $\{c_k\}$ that is related to $\{a_k\}$ as follows

$$c_k = a_k + a_{k-1} \tag{6.24}$$

Note that unlike the linear operation of duobinary coding, precoding is a nonlinear operation.

We assume that symbol 1 at the precoder output in Fig. 6.6 is represented by $+1$ V and symbol 0 by -1 V. Therefore, from Eqs. 6.23 and 6.24, we find that

$$c_k = \begin{cases} \pm 2 \text{ V}, & \text{if } b_k \text{ is represented by symbol 0} \\ 0 \text{ V} & \text{if } b_k \text{ is represented by symbol 1} \end{cases} \tag{6.25}$$

which is illustrated in Example 2. From Eq. 6.25 we deduce the following decision rule for constructing the decoded binary sequence $\{\hat{b}_k\}$ at the

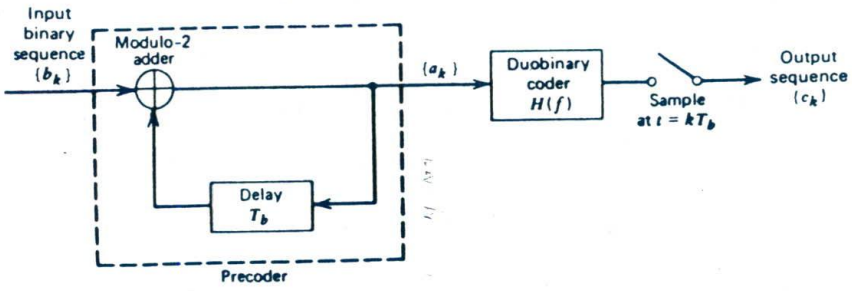


Figure 6.6
A precoded duobinary scheme. Details of the duobinary coder are given in Fig. 6.3.

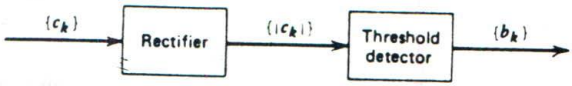


Figure 6.7
Detector for recovering original binary sequence from the precoded duobinary coder output.

receiver output:

$$\hat{b}_k = \begin{cases} \text{symbol 0}, & \text{if } |c_k| > 1 \text{ V} \\ \text{symbol 1}, & \text{if } |c_k| \leq 1 \text{ V} \end{cases} \tag{6.26}$$

According to Eq. 6.26, the detector (decoder) consists of a rectifier, the output of which is compared to a threshold of 1 V, and the original binary sequence $\{b_k\}$ is thereby detected. A block diagram of the detector is shown in Fig. 6.7. A useful feature of this detector is that no knowledge of any input sample other than the present one is required. Hence, error propagation cannot occur in the detector of Fig. 6.7.

Moreover, we may note the following two points:

1. In the absence of channel noise, the decoded sequence $\{\hat{b}_k\}$ derived from Eq. 6.26 is exactly the same as the original binary sequence $\{b_k\}$ at the transmitter input.
2. The use of Eq. 6.23 requires the addition of an extra bit to the precoded sequence $\{a_k\}$. The decoded sequence $\{\hat{b}_k\}$ is invariant to the use of a 1 or a 0 for this extra bit.

EXAMPLE 2

Consider the input binary sequence 0010110. To proceed with the precoding of this sequence, which involves feeding the precoder output back to the input, we add an extra bit to the precoder output. This extra bit is chosen arbitrarily as a bit 1. Hence, using Eq. 6.23, we find that the sequence $\{a_k\}$ at the precoder output is as shown in row 2 of Table 6.1. We assume that symbol 1 is represented by $+1$ V and symbol 0 by -1 V. Accordingly, the precoder output has the amplitudes shown in row 3. Finally, using Eq. 6.24, we find that the duobinary coder output has the amplitudes given in row 4 of Table 6.1.

To detect the original binary sequence, we apply the decision rule of Eq. 6.26, and so obtain the sequence given in row 5 of Table 6.1. This shows that, in the absence of noise, the original binary sequence is detected correctly.

TABLE 6.1

Input binary sequence $\{b_k\}$		0	0	1	0	1	1	0
Precoded binary sequence $\{a_k\}$	1	1	1	0	0	1	0	0
Polar representation of sequence $\{a_k\}$	+1	+1	+1	−1	−1	+1	−1	−1
Duobinary coder output, $\{c_k\}$		2	2	0	−2	0	0	−2
Decoded binary sequence $\{\hat{b}_k\}$		0	0	1	0	1	1	0

EXERCISE 3 Repeat the calculations of Table 6.1, assuming that the extra bit at the beginning of the precoded sequence $\{a_k\}$ is a 0. Hence, show that the decoded sequence $\{\hat{b}_k\}$ is unaffected by this change (compared to the initial bit used in Example 2).

EXERCISE 4 The duobinary, ternary, and bipolar signaling techniques have one common feature: They all employ three amplitude levels. In what way does the duobinary technique differ from the other two?

MODIFIED DUOBINARY SIGNALING

In the duobinary signaling technique just described, the transfer function $H(f)$, and consequently the power spectral density of the transmitted pulse, is nonzero at the origin. In some applications, this is an undesirable feature. We may correct for this drawback by using the *modified duobinary signaling* technique, which involves a correlation span of two binary digits. This is achieved by subtracting input binary digits spaced $2T_b$ seconds apart. Specifically, the output of the modified duobinary conversion filter is related to the sequence $\{a_k\}$ at its input as follows

$$c_k = a_k - a_{k-2} \tag{6.27}$$

Here, again, we find that a three-level signal is generated. If $a_k = \pm 1$ V, as assumed previously, c_k takes on one of three values: 2, 0, and −2 V.

Figure 6.8 depicts the complete block diagram of a *modified duobinary encoder*, incorporating an appropriate precoder and a band-limited channel assumed to be ideal. Here again the channel is included as an integral part of the encoding operation. The incoming binary sequence $\{b_k\}$ produces a continuous waveform at the channel output. This waveform is therefore

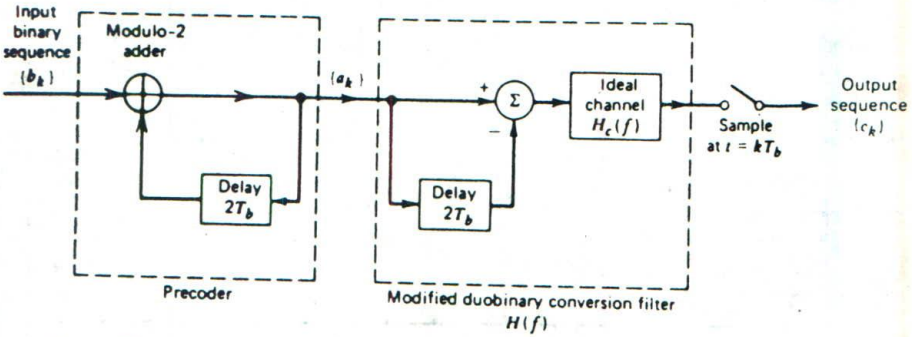


Figure 6.8
Modified duobinary signaling scheme.

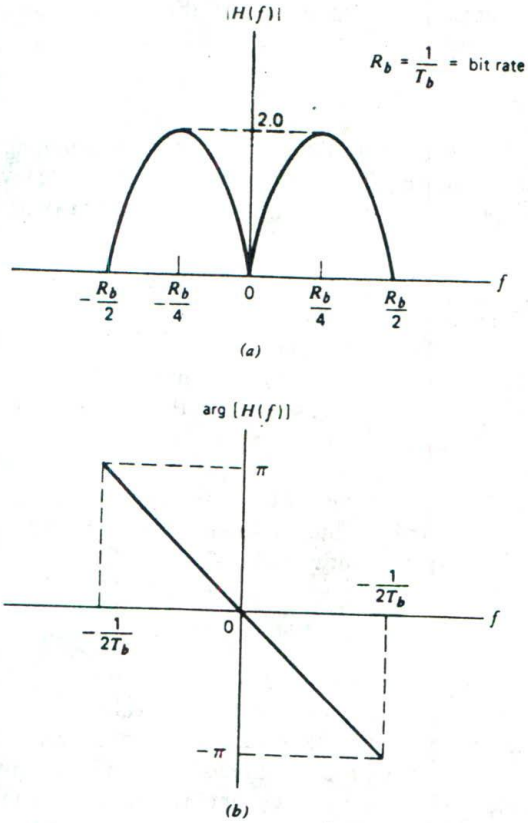


Figure 6.9
Frequency response of modified duobinary conversion filter. (a) Amplitude response. (b) Phase response.

sampled uniformly every T_b seconds to produce the modified duobinary encoded sequence $\{c_k\}$.

Let $H(f)$ denote the overall transfer function of the *modified duobinary conversion filter* that consists of the cascade connection of the delay-line-filter and the channel; this filter is enclosed inside the second dashed rectangle in Fig. 6.8. Hence, we may write

$$\begin{aligned} H(f) &= H_C(f)[1 - \exp(-j4\pi f T_b)] \\ &= 2jH_C(f) \sin(2\pi f T_b) \exp(-j2\pi f T_b) \end{aligned} \tag{6.28}$$

where $H_C(f)$ is defined in Eq. 6.19. We, therefore, have an overall frequency response in the form of a half-cycle sine function, as shown by

$$H(f) = \begin{cases} 2j \sin(2\pi f T_b) \exp(-j2\pi f T_b), & |f| \leq 1/2T_b \\ 0, & \text{elsewhere} \end{cases} \tag{6.29}$$

The corresponding amplitude response and phase response of the modified duobinary-coder are as shown in parts *a* and *b* of Fig. 6.9, respectively. Note that the phase response depicted in Fig. 6.9*b* does not include the constant 90°-phase shift due to the multiplying factor j in Eq. 6.29. A useful feature of the modified duobinary coder is the fact that its output has no dc component. This property is important since, in practice, many communication channels cannot transmit a dc component.

The impulse response of the modified duobinary coder consists of two sinc pulses that are time-displaced by $2T_b$ seconds, as shown by (except for a scaling factor)

$$\begin{aligned} h(t) &= \frac{\sin(\pi t/T_b)}{\pi t/T_b} - \frac{\sin[\pi(t - 2T_b)/T_b]}{\pi(t - 2T_b)/T_b} \\ &= \frac{\sin(\pi t/T_b)}{\pi t/T_b} - \frac{\sin(\pi t/T_b)}{\pi(t - 2T_b)/T_b} \\ &= \frac{2T_b^2 \sin(\pi t/T_b)}{\pi t(2T_b - t)} \end{aligned} \tag{6.30}$$

This impulse response is plotted in Fig. 6.10, which shows that it has *three* distinguishable levels at the sampling instants.

To eliminate the possibility of error propagation in the modified duobinary system, we use a precoding procedure similar to that used for the duobinary case. Specifically, prior to the generation of the modified duobinary signal, a modulo-two logical addition is used on signals $2T_b$ seconds apart, as shown by (see Fig. 6.8)

$$a_k = b_k \oplus a_{k-2} \tag{6.31}$$

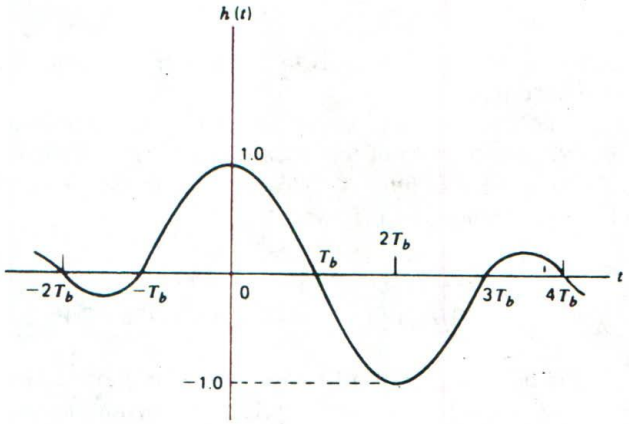


Figure 6.10
Impulse response of the modified duobinary conversion filter.

where $\{b_k\}$ is the input binary sequence and $\{a_k\}$ is the sequence at the precoder output. The sequence $\{a_k\}$ thus produced is then applied to the modified duobinary conversion filter.

In Fig. 6.8, the output digit c_k equals 0, +2, or −2V, assuming the use of a polar representation for the precoded sequence $\{a_k\}$. Also we find that the decoded (detected) digit $\hat{b}_k$ at the receiver output may be extracted from c_k by disregarding the polarity of c_k. Specifically, we may write

$$\hat{b}_k = \begin{cases} \text{symbol 1 if } |c_k| > 1 \text{ V} \\ \text{symbol 0 if } |c_k| \leq 1 \text{ V} \end{cases} \tag{6.32}$$

As with the duobinary signaling, we may note the following:

1. In the absence of channel noise, the decoded binary sequence $\{\hat{b}_k\}$ is exactly the same as the original binary sequence $\{b_k\}$ at the transmitter input.
2. The use of Eq. 6.31 requires the addition of two extra bits to the precoded sequence $\{a_k\}$. The composition of the decoded sequence $\{\hat{b}_k\}$ using Eq. 6.32 is invariant to the selection made for these two bits.

EXERCISE 5 Consider again the binary sequence 0010110 used to illustrate the operation of the duobinary signaling scheme in Example 2. Using this sequence as the input $\{b_k\}$, calculate the following sequences for the modified duobinary signaling scheme of Fig. 6.8:

(a) The sequence $\{a_k\}$ at the precoder output.

(b) The polar representation of $\{a_k\}$.

(c) The sequence $\{c_k\}$ at the modified duobinary conversion filter output, assuming the addition of bits 11 at the beginning of the precoded sequence $\{a_k\}$.

(d) The decoded sequence $\{\hat{b}_k\}$ at the receiver output. Compare this sequence with the original binary sequence $\{b_k\}$.

EXERCISE 6 Repeat the calculations of Exercise 5, assuming that the bits added at the beginning of the precoded sequence $\{a_k\}$ are 00. Hence, show that the decoded sequence $\{\hat{b}_k\}$ is unaffected by this choice of initial bits for the sequence $\{a_k\}$.

6.6 BASEBAND TRANSMISSION OF M-ARY DATA

In the baseband binary PAM system of Fig. 6.1, the output of the pulse generator consists of binary pulses, that is, pulses with one of two possible amplitude levels. On the other hand, in a *baseband M-ary* version of the system, the output of the pulse generator takes on one of M possible amplitude levels with $M > 2$; the digital waveform of a *quaternary system* (that is, $M = 4$) is illustrated in Fig. 5.12*f*. In an M-ary system, the information source emits a sequence of symbols from an alphabet that consists of M symbols. Each amplitude level at the pulse generator output corresponds to a distinct symbol, so that there are M distinct amplitude levels to be transmitted.

Consider then an M-ary PAM system with a signal alphabet that contains M symbols, with the *symbol duration* denoted by T seconds. We refer to $1/T$ as the *signaling rate* of the system, which is expressed in *symbols per second* or *bauds*. It is informative to relate the signaling rate of this system to that of an equivalent binary PAM system for which the value of M is 2 and the bit duration is T_b seconds. The binary PAM system transmits data at the rate of $1/T_b$ *bits per second*. We also observe that in the case of a *quaternary* PAM system, for example, the four possible symbols may be identified with the dibits 00, 10, 11, and 01. We thus see that each symbol represents 2 bits of data and 1 baud is equal to 2 bits per second. We may generalize this result by stating that in an M-ary PAM system, 1 baud is equal to $\log_2 M$ bits per second, and the symbol duration T of the M-ary PAM system is related to the bit duration T_b of the equivalent binary PAM system as follows:

$$T = T_b \log_2 M \tag{6.33}$$

Therefore, in a given channel bandwidth, we find that by using an M-ary PAM system we are able to transmit data at a rate that is $\log_2 M$ faster than the corresponding binary PAM system.

However, this improvement in bandwidth use is attained at a price.

Specifically, the transmitted power must be increased by a factor equal to $M^2/\log_2 M$, compared to a binary PAM system, if we are to realize the same performance in the presence of channel noise.[4] Also, system complexity is increased.

EXERCISE 7 An M-ary PAM system uses a raised cosine spectrum with rolloff factor α. Show that the signaling rate of the system is given by

$$\frac{1}{T} = \frac{2 \log_2 M}{1 + \alpha} B$$

where B is the channel bandwidth.

6.7 EYE PATTERN

One way to study intersymbol interference in a PCM or data transmission system experimentally is to apply the received wave to the vertical deflection plates of an oscilloscope and to apply a sawtooth wave at the transmitted symbol rate $1/T$ to the horizontal deflection plates. The waveforms in successive symbol intervals are thereby translated into one interval on the oscilloscope display, as illustrated in Fig. 6.11 for the case of a binary wave for which $T = T_b$. The resulting display is called an *eye pattern* because of its resemblance to the human eye for binary waves. The interior region of the eye pattern is called the *eye opening*.

An eye pattern provides a great deal of information about the performance of the pertinent system, as described here (see Fig. 6.12):

1. The width of the eye opening defines the time interval over which the received wave can be sampled without error from intersymbol interference. It is apparent that the preferred time for sampling is the instant of time at which the eye is open widest.
2. The sensitivity of the system to timing error is determined by the rate of closure of the eye as the sampling time is varied.
3. The height of the eye opening, at a specified sampling time, defines the margin over channel noise.

[4]The performance of a data transmission system in the presence of channel noise is usually measured in terms of the average probability of symbol error. When M is much larger than 2 and the average probability of symbol error is small compared to unity, an M-ary PAM system requires a transmitted power larger than in a binary PAM by a factor of $M^2/\log_2 M$. For a proof of this result, see Haykin (1988), pp. 78–80.

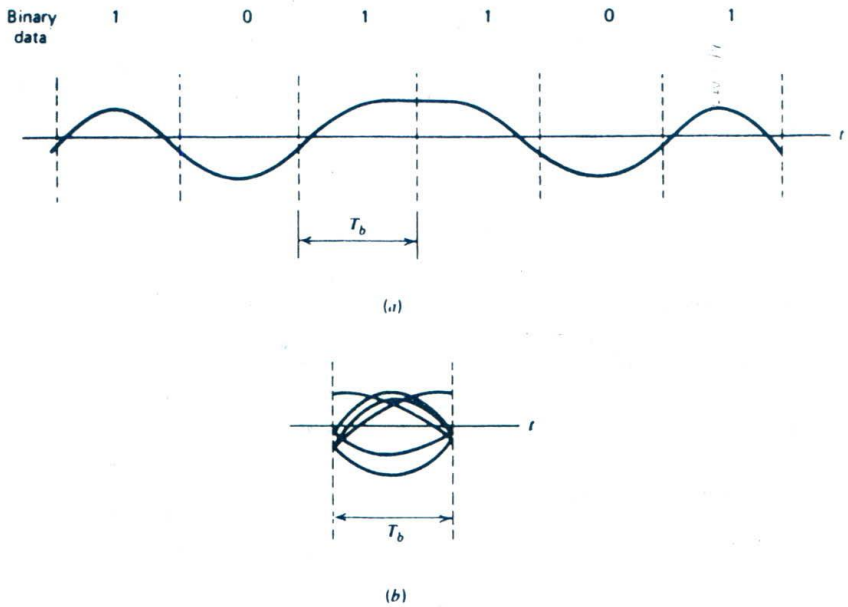


Figure 6.11
(a) Distorted binary wave. (b) Eye pattern.

When the effect of intersymbol interference is severe, traces from the upper portion of the eye pattern cross traces from the lower portion, with the result that the eye is completely closed. In such a situation, it is impossible to avoid errors due to the combined presence of intersymbol interference and channel noise in the system.

In the case of an M-ary system, the eye pattern contains $(M - 1)$ eye

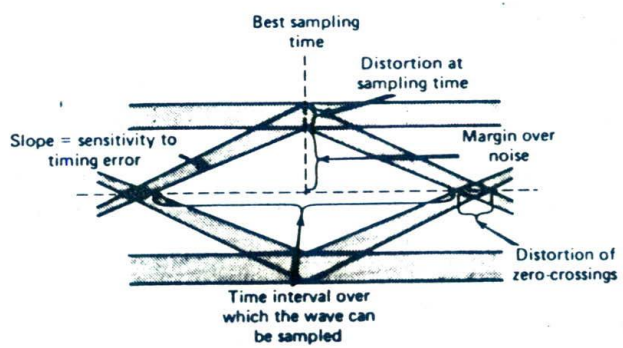


Figure 6.12
Interpretation of the eye pattern.

openings stacked up vertically one on the other, where M is the number of discrete amplitude levels used to construct the transmitted signal. In a strictly linear system with truly random data, all these eye openings would be identical. In practice, however, it is often possible to discern asymmetries in the eye pattern, which are caused by nonlinearities in the transmission channel.

6.8 ADAPTIVE EQUALIZATION

A study of baseband data transmission would be incomplete without some discussion of the *equalization problem*. By equalization we mean the process of correcting channel-induced signal distortion. Equalization is of paramount importance in the *high-speed transmission* of digital data over a band-limited channel. In this final section of the chapter, we briefly discuss the need for equalization in the context of data transmission over a voice-grade telephone channel, which is essentially linear and is also characterized by a limited bandwidth and a high signal-to-noise ratio.

An efficient approach to high-speed data transmission over such a channel involves the combined use of two basic forms of modulation:

1. *Discrete pulse-amplitude modulation* (PAM): In this operation, the amplitudes of successive pulses in a periodic train (acting as a carrier) are varied in a discrete fashion in accordance with the incoming data stream.
2. *Linear modulation:* In this second operation, the amplitude or phase of a sinusoidal carrier is varied in accordance with the discrete PAM signal resulting from the first stage of modulation. The selection of a specific type of linear modulation is made with the aim of conserving channel bandwidth. Linear modulation schemes for data transmission are considered in Sections 7.15 and 10.7.

At the receiving end of the system, the received wave is demodulated, and then synchronously sampled and quantized. As a result of dispersion of the pulse shape by the channel, however, we find that the number of detectable amplitude levels is often limited by intersymbol interference rather than by additive noise. In principle, if the channel is known precisely, it is virtually always possible to make the intersymbol interference (at the sampling instants) arbitrarily small by using a suitable pair of transmitting and receiving filters, so as to control the overall pulse shape in the manner described in Section 6.4. The transmitting filter is placed directly before the modulator, whereas the receiving filter is placed directly after the demodulator. Thus, insofar as intersymbol interference is concerned, we may consider the data transmission as being essentially baseband.

However, in a switched telephone network, we find that two factors contribute to the distribution of pulse distortion on different link connections: (1) differences in the transmission characteristics of the individual

links that may be switched together, and (2) differences in the number of links in a connection. The result is that the telephone channel is random in the sense of being one of an ensemble of possible channels. Consequently, the use of a fixed pair of transmitting and receiving filters designed on the basis of average channel characteristics may not adequately reduce intersymbol interference. To realize the full transmission capability of a telephone channel, there is need for *adaptive equalization*. An *equalizer* is a filter that compensates for the dispersive effects of a channel. The process of equalization is said to be *adaptive* when the equalizer is capable of adjusting its coefficients continuously during the transmission of data; it does so by operating on the received signal (channel output) in accordance with some *algorithm*.

Among the philosophies for adaptive equalization of data transmission systems, we have *prechannel equalization* at the transmitter and *postchannel equalization* at the receiver. Because the first approach requires a feedback channel, we consider only adaptive equalization at the receiving end of the system. This equalization can be achieved, prior to data transmission, by training the filter with the guidance of a suitable *training sequence* transmitted through the channel so as to adjust the filter parameters to optimum values. The typical telephone channel changes little during an average data call, so that precall equalization with a training sequence is sufficient in most cases encountered in practice. The equalizer is positioned after the receiving filter in the receiver.

Figure 6.13 shows a popular structure used to design adaptive equalizers. The structure is a tapped-delay-line filter that consists of a set of delay elements, a set of multipliers connected to the delay-line taps, a corresponding set of adjustable tap weights, and a summer for adding the multiplier outputs. Let the sequence $\{x(nT)\}$, appearing at the output of the receiving filter, be applied to the input of this tapped-delay-line filter, producing the output (see Fig. 6.13).

$$y(nT) = \sum_{i=0}^{M-1} w_i x(nT - iT) \tag{6.34}$$

where w_i is the weight at the ith tap, and M is the total number of taps. These M tap weights constitute the adaptive filter coefficients. The tap spacing is chosen equal to the symbol duration T of the transmitted signal or the reciprocal of the signaling rate.

The adaptation of the filter may be achieved by proceeding as follows:

1. A *known sequence* $\{d(nT)\}$ is transmitted, and in the receiver the resulting *response sequence* $\{y(nT)\}$ is obtained by measuring the filter output at the sampling instants.
2. Viewing the known transmitted sequence $\{d(nT)\}$ as the *desired response*, the differences between it and the response sequence $\{y(nT)\}$

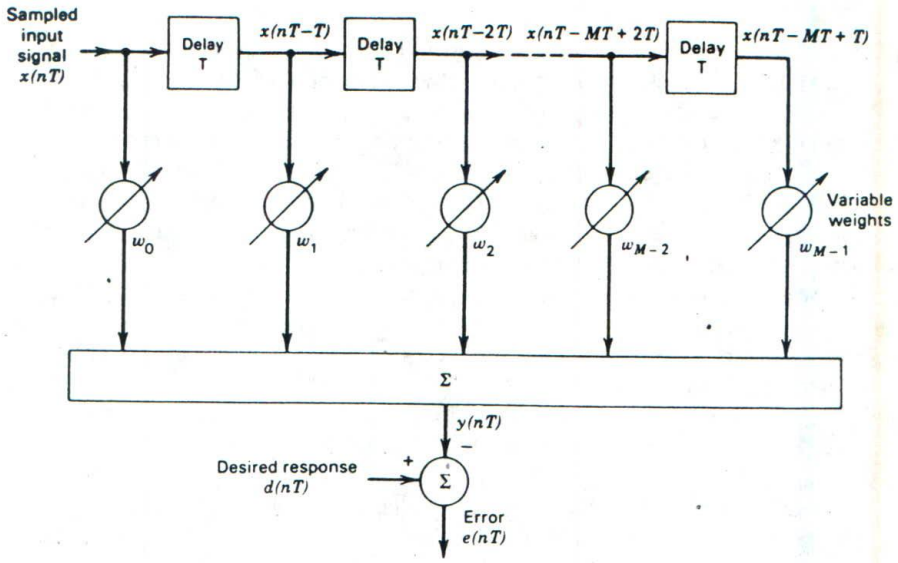


Figure 6.13
Elements of an adaptive filter.

is computed. This difference is called the *error sequence*, denoted by $\{e(nT)\}$; thus,

$$e(nT) = d(nT) - y(nT), \qquad n = 0, 1, \ldots N - 1 \quad (6.35)$$

where N is the total *length* of the sequence.

3. The error sequence $\{e(nT)\}$ is used to estimate the direction in which the weights $\{w_i\}$ of the filter are changed so as to make them approach their optimum settings $\{w_{oi}\}$.

We assume that all sequences (signals) of interest are real valued. A criterion appropriate for optimization is the *total error energy* defined by

$$\mathcal{E} = \sum_{n=0}^{N-1} e^2(nT) \quad (6.36)$$

The *optimum values* of the tap weights, namely, $w_{o0}, w_{o1}, \ldots w_{o,M-1}$ result when the total error energy $\mathcal{E}$ is minimized.

The solution to this optimization problem may be developed in the form of an *algorithm* that adjusts the tap weights of the filter in a *recursive* manner, which means that the tapped-delay-line filter assumes a *time-varying* form. In particular, the present estimate of each tap weight is updated by incrementing it by a *correction term* proportional to the error

signal at that time. Thus, starting from some arbitrary *initial condition*, the algorithm *learns* (about the operating channel conditions) from the incoming data, sample by sample, and thereby automatically adjusts the tap weights toward the optimum solution.

A simple and yet effective solution to this adaptation procedure is provided by the *least-mean-square (LMS) algorithm*.[5] According to the LMS algorithm, the tap weights are adapted as follows:

$$\hat{w}_i(nT + T) = \hat{w}_i(nT) + \mu e(nT)x(nT - iT) \tag{6.37}$$

where $i = 0, 1, \ldots, M - 1$, and $\hat{w}_i(nT)$ is the *present estimate* of the optimum weight w_{oi} for tap i at time nT, and $\hat{w}_i(nT + T)$ is the *updated estimate*. The parameter μ in Eq. 6.37 is called the *adaptation constant*. In particular, it controls the amount of *correction* applied to the old estimate $\hat{w}_i(nT)$ to produce the updated estimate $\hat{w}_i(nT + T)$. In addition to the parameter μ, the correction depends on the filter input $x(nT - iT)$ and the error signal $e(nT)$, both measured at time nT. Thus, by a proper choice of the adaptation constant μ, the use of the recursive equation (6.37) helps the adjustment of the tap weights move toward their optimum settings in a step-by-step fashion. Typically, for the starting condition, all the tap weights of the equalizer are set equal to zero.

The LMS algorithm requires knowledge of the desired response $d(nT)$ and the filter response $y(nT)$ to form the error signal $e(nT)$ in accordance with Eq. 6.35. For $y(nT)$, we may use Eq. 6.34 with $\hat{w}_i(nT)$ substituted for w_i. However, by the very nature of data communications, the desired response (providing a frame of reference for the adaptation process) originates at the channel input, which is separated physically from the receiver where the adaptive equalization is preformed. There are two methods in which a *replica* of the desired response $d(nT)$ may be obtained, as illustrated in Fig. 6.14. These two methods and their applicability are described in the following paragraphs.

In the first method, a replica of the desired response is *stored* in the receiver. Naturally, the generator of this stored reference has to be *synchronized* with the known transmitted sequence. The use of a stored reference is well suited for the *initial training* of the equalizer. This operation of the equalizer corresponds to position 1 of the switch in Fig. 6.14. (In Section 8.9 we describe a pseudo-random sequence known as a *linear maximal sequence* that may be used for this purpose.)

In the second method, the output from a decision device in the receiver is used. Under normal operating conditions, the decisions made by the receiver are correct with high probability. This means that the *error esti-*

[5]For a detailed discussion of the LMS and other adaptive filtering algorithms, see the following references: Haykin (1986), and Widrow and Stearns (1985).

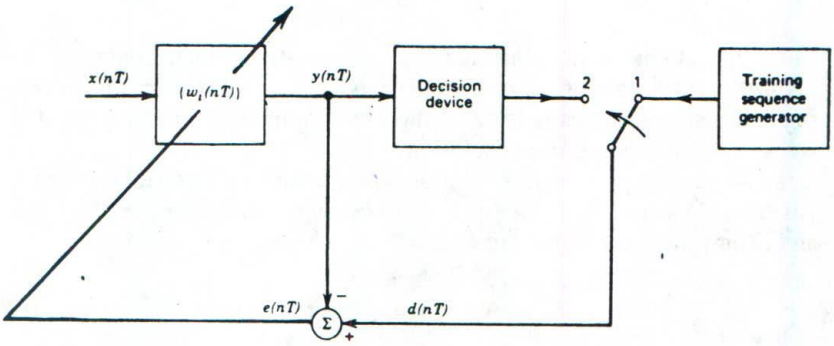


Figure 6.14
Illustrating the two modes of operation of an adaptive equalizer: Position 1 of the switch corresponds to the training mode. Position 2 corresponds to the decision-directed mode.

mates thus obtained are correct most of the time, thereby permitting the adaptive equalizer to operate satisfactorily. This second method of operation is referred to as the *decision-directed mode* of the adaptive equalizer; it corresponds to position 2 of the switch in Fig. 6.14. It is well suited for *tracking* relatively slow variations in channel characteristics during the course of transmission.

The adaptive equalizer depicted in Fig. 6.14 represents a *closed-loop feedback system,* irrespective of its mode of operation. As such, there is a tendency for the adaptive equalizer to become unstable. To ensure *stability,* care has to be exercised in the value assigned to the adaptation constant μ in the time update of Eq. 6.37. On the one hand, μ must be large enough to ensure a reasonably fast rate of convergence of the LMS algorithm. On the other hand, it must be small enough to make it possible for the LMS algorithm to track slow statistical variations in the channel.

PROBLEMS

P6.3 Ideal Solution

Problem 1 The pulse shape $p(t)$ of a baseband binary PAM system is defined by

$$p(t) = \text{sinc}\left(\frac{t}{T_b}\right)$$

where T_b is the bit duration of the input binary data. The amplitude levels at the pulse generator output are $+1$ V or -1 V, depending on whether

the binary symbol at the input is 1 or 0, respectively. Sketch the waveform at the output of the receiving filter in response to the input data 001101001.

P6.4 Raised Cosine Spectrum

Problem 2 An analog signal is sampled, quantized, and encoded into a binary PCM wave. The specifications of the PCM system include the following:

Sampling rate = 8 kHz
Number of representation levels = 64

The PCM wave is transmitted over a baseband channel using discrete pulse-amplitude modulation. Determine the minimum bandwidth required for transmitting the PCM wave if each pulse is allowed to take on the following number of amplitude levels:

(a) 2
(b) 4
(c) 8

Problem 3 The raised cosine pulse spectrum for a rolloff factor of unity is given by

$$P(f) = \begin{cases} \dfrac{1}{2B_0}\cos^2\left(\dfrac{\pi f}{4B_0}\right), & 0 \leq |f| < 2B_0 \\ 0, & 2B_0 \leq |f| \end{cases}$$

Show that the time response $p(t)$, the inverse Fourier transform of $P(f)$, is

$$p(t) = \frac{\operatorname{sinc}(4B_0 t)}{1 - 16B_0^2 t^2}$$

Problem 4 A computer puts out binary data at the rate of 56 kilobits per second. The computer output is transmitted using a baseband binary PAM system that is designed to have a raised cosine pulse spectrum. Determine the transmission bandwidth required for each of the following rolloff factors:

(a) $\alpha = 0.25$
(b) $\alpha = 0.5$
(c) $\alpha = 0.75$
(d) $\alpha = 1.0$

Problem 5 A binary PAM wave is to be transmitted over a low-pass channel with an absolute maximum bandwidth of 75 kHz. The bit duration is 10 μs. Find a raised cosine spectrum that satisfies these requirements.

P6.5 Correlative Coding

Problem 6 The binary data 001101001 is applied to the input of a duobinary system.

(a) Construct the duobinary coder output and corresponding receiver output, without a precoder.
(b) Suppose that owing to error during transmission, the level at the receiver input produced by the second input digit is reduced to zero. Construct the new receiver output.

Problem 7 Repeat Problem 6, assuming the use of a precoder in the transmitter.

Problem 8 The binary data 011100101 is applied to the input of a modified duobinary system.

(a) Construct the modified duobinary coder output and corresponding receiver output, without a precoder.
(b) Suppose that owing to error during transmission, the level produced by the third input digit is zero. Construct the new receiver output.

Problem 9 Repeat Problem 8, assuming the use of a precoder in the transmitter.

Problem 10 Using conventional analog filter design methods, it is difficult to approximate the frequency response of the modified duobinary system defined by Eq. 6.29. To get around this problem, we may use the arrangement shown in Fig. P6.1. Justify the validity of this scheme.

P6.6 Baseband Transmission of *M*-ary Data

Problem 11 Repeat Problem 4, given that each set of three successive binary digits in the computer output are coded into one of eight possible

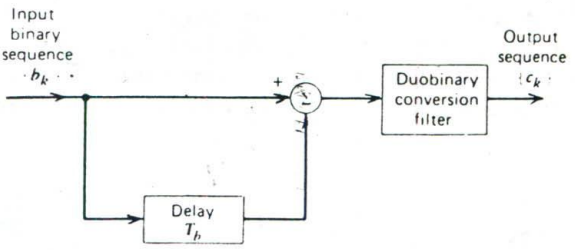


Figure P6.1

amplitude levels, and the resulting signal is transmitted by using an 8-level PAM system designed to have a raised cosine pulse spectrum.

Problem 12 An analog signal is sampled, quantized, and encoded into a binary PCM wave. The number of representation levels used is 128. A synchronizing pulse is added at the end of each code word representing a sample of the analog signal. The resulting PCM wave is transmitted over a channel of bandwidth 12 kHz using a *quaternary* PAM system with a raised cosine pulse spectrum. The rolloff factor is unity.

(a) Find the rate (in bits per second) at which information is transmitted through the channel.
(b) Find the rate at which the analog signal is sampled. What is the maximum possible value for the highest frequency component of the analog signal?

P6.7 Eye Pattern

Problem 13 A binary wave using polar signaling is generated by representing symbol 1 by a pulse of amplitude +1 V and symbol 0 by a pulse of amplitude −1 V; in both cases the pulse duration equals the bit duration. This signal is applied to a low-pass *RC* filter with transfer function:

$$H(f) = \frac{1}{1 + jf/f_0}$$

Construct the eye pattern for the filter output for the following sequences:

(a) Alternating 1's and 0's.
(b) A long sequence of 1's followed by a long sequence of 0's.
(c) A long sequence of 1's followed by a single 0 and then a long sequence of 1's.

Assume a bit rate of $2f_0$ bits per second.

Problem 14 The binary sequence 011010 is transmitted through a channel having a raised cosine characteristic with a rolloff factor of unity. Assume the use of polar signaling, with symbols 1 and 0 represented by +1 and −1 V, respectively.

(a) Construct, to scale, the received wave, and indicate the best sampling times for regeneration.
(b) Construct the eye pattern for this received wave and show that it is completely open.
(c) Determine the zero crossings of the received wave.