4

CHARACTERIZATION OF COMMUNICATION SIGNALS AND SYSTEMS

Signals can be categorized in a number of different ways, such as random versus deterministic, discrete time versus continuous time, discrete amplitude versus continuous amplitude, lowpass versus bandpass, finite energy versus infinite energy, finite average power versus infinite average power, etc. In this chapter, we treat the characterization of signals and systems that are usually encountered in the transmission of digital information over a communication channel. In particular, we introduce the representation of various forms of digitally modulated signals and describe their spectral characteristics.

We begin with the characterization of bandpass signals and systems, including the mathematical representation of bandpass stationary stochastic processes. Then, we present a vector space representation of signals. We conclude with the representation of digitally modulated signals and their spectral characteristics.

4-1 REPRESENTATION OF BANDPASS SIGNALS AND SYSTEMS

Many digital information-bearing signals are transmitted by some type of carrier modulation. The channel over which the signal is transmitted is limited in bandwidth to an interval of frequencies centered about the carrier, as in double-sideband modulation, or adjacent to the carrier, as in single-sideband modulation. Signals and channels (systems) that satisfy the condition that their bandwidth is much smaller than the carrier frequency are termed *narrowband bandpass signals and channels (systems)*. The modulation performed at the

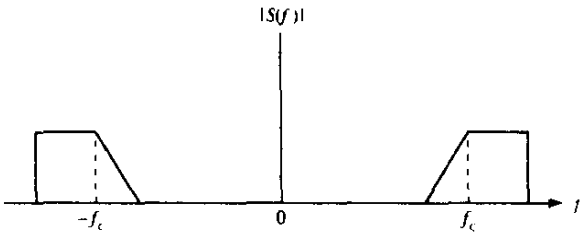


FIGURE 4-1-1 Spectrum of a bandpass signal.

transmitting end of the communication system to generate the bandpass signal and the demodulation performed at the receiving end to recover the digital information involve frequency translations. With no loss of generality and for mathematical convenience, it is desirable to reduce all bandpass signals and channels to equivalent lowpass signals and channels. As a consequence, the results of the performance of the various modulation and demodulation techniques presented in the subsequent chapters are independent of carrier frequencies and channel frequency bands. The representation of bandpass signals and systems in terms of equivalent lowpass waveforms and the characterization of bandpass stationary stochastic processes are the main topics of this section.

4-1-1* Representation of Bandpass Signals

Suppose that a real-valued signal $s(t)$ has a frequency content concentrated in a narrow band of frequencies in the vicinity of a frequency f_c, as shown in Fig. 4-1-1. Our objective is to develop a mathematical representation of such signals. First, we construct a signal that contains only the positive frequencies in $s(t)$. Such a signal may be expressed as

$$S_+(f) = 2u(f)S(f) \tag{4-1-1}$$

where $S(f)$ is the Fourier transform of $s(t)$ and $u(f)$ is the unit step function. The equivalent time-domain expression for (4-1-1) is

$$\begin{aligned} s_+(t) &= \int_{-\infty}^{\infty} S_+(f)e^{j2\pi ft}\,df \\ &= F^{-1}[2u(f)] \bigstar F^{-1}[S(f)] \end{aligned} \tag{4-1-2}$$

The signal $s_+(t)$ is called the *analytic signal* or the *pre-envelope* of $s(t)$. We note that $F^{-1}[S(f)] = s(t)$ and

$$F^{-1}[2u(f)] = \delta(t) + \frac{j}{\pi t} \tag{4-1-3}$$

Hence,

$$s_+(t) = \left[\delta(t) + \frac{j}{\pi t}\right] \bigstar s(t)$$

$$= s(t) + j\frac{1}{\pi t} \bigstar s(t) \tag{4-1-4}$$

We define $\hat{s}(t)$ as

$$\hat{s}(t) = \frac{1}{\pi t} \bigstar s(t)$$

$$= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{s(\tau)}{t-\tau} d\tau \tag{4-1-5}$$

The signal $\hat{s}(t)$ may be viewed as the output of the filter with impulse response

$$h(t) = \frac{1}{\pi t}, \quad -\infty < t < \infty \tag{4-1-6}$$

when excited by the input signal $s(t)$. Such a filter is called a *Hilbert transformer*. The frequency response of this filter is simply

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

$$= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1}{t} e^{-j2\pi ft}\, dt$$

$$= \begin{cases} -j & (f > 0) \\ 0 & (f = 0) \\ j & (f < 0) \end{cases} \tag{4-1-7}$$

We observe that $|H(f)| = 1$ and that the phase response $\Theta(f) = -\frac{1}{2}\pi$ for $f > 0$ and $\Theta(f) = \frac{1}{2}\pi$ for $f < 0$. Therefore, this filter is basically a 90° phase shifter for all frequencies in the input signal.

The analytic signal $s_+(t)$ is a bandpass signal. We may obtain an equivalent lowpass representation by performing a frequency translation of $S_+(f)$. Thus, we define $S_l(f)$ as

$$S_l(f) = S_+(f + f_c) \tag{4-1-8}$$

The equivalent time-domain relation is

$$s_l(t) = s_+(t)e^{-j2\pi f_c t}$$

$$= [s(t) + j\hat{s}(t)]e^{-j2\pi f_c t} \tag{4-1-9}$$

or, equivalently,

$$s(t) + j\hat{s}(t) = s_l(t)e^{j2\pi f_c t} \tag{4-1-10}$$

In general, the signal $s_l(t)$ is complex-valued (see Problem 4-5), and may be expressed as

$$s_l(t) = x(t) + jy(t) \tag{4-1-11}$$

If we substitute for $s_l(t)$ in (4-1-11) and equate real and imaginary parts on each side, we obtain the relations

$$s(t) = x(t)\cos 2\pi f_c t - y(t)\sin 2\pi f_c t \tag{4-1-12}$$

$$\hat{s}(t) = x(t)\sin 2\pi f_c t + y(t)\cos 2\pi f_c t \tag{4-1-13}$$

The expression (4-1-12) is the desired form for the representation of a bandpass signal. The low-frequency signal components $x(t)$ and $y(t)$ may be viewed as amplitude modulations impressed on the carrier components $\cos 2\pi f_c t$ and $\sin 2\pi f_c t$, respectively. Since these carrier components are in phase quadrature, $x(t)$ and $y(t)$ are called the *quadrature components* of the bandpass signal $s(t)$.

Another representation of the signal in (4-1-12) is

$$\begin{aligned} s(t) &= \text{Re}\{[x(t) + jy(t)]e^{j2\pi f_c t}\} \\ &= \text{Re}[s_l(t)e^{j2\pi f_c t}] \end{aligned} \tag{4-1-14}$$

where Re denotes the real part of the complex-valued quantity in the brackets following. The lowpass signal $s_l(t)$ is usually called the *complex envelope* of the real signal $s(t)$, and is basically the *equivalent lowpass signal.*

Finally, a third possible representation of a bandpass signal is obtained by expressing $s_l(t)$ as

$$s_l(t) = a(t)e^{j\theta(t)} \tag{4-1-15}$$

where

$$a(t) = \sqrt{x^2(t) + y^2(t)} \tag{4-1-16}$$

$$\theta(t) = \tan^{-1}\frac{y(t)}{x(t)} \tag{4-1-17}$$

Then

$$\begin{aligned} s(t) &= \text{Re}[s_l(t)e^{j2\pi f_c t}] \\ &= \text{Re}[a(t)e^{j[2\pi f_c t + \theta(t)]}] \\ &= a(t)\cos[2\pi f_c t + \theta(t)] \end{aligned} \tag{4-1-18}$$

The signal $a(t)$ is called the *envelope* of $s(t)$, and $\theta(t)$ is called the *phase* of $s(t)$. Therefore, (4-1-12), (4-1-14), and (4-1-18) are equivalent representations of bandpass signals.

The Fourier transform of $s(t)$ is

$$\begin{aligned} S(f) &= \int_{-\infty}^{\infty} s(t)e^{-j2\pi ft}\,dt \\ &= \int_{-\infty}^{\infty} \{\text{Re}[s_l(t)e^{j2\pi f_c t}]\}e^{-j2\pi ft}\,dt \end{aligned} \tag{4-1-19}$$

Use of the identity

$$\text{Re}(\xi) = \tfrac{1}{2}(\xi + \xi^*) \tag{4-1-20}$$

in (4-1-19) yields the result

$$S(f) = \frac{1}{2}\int_{-\infty}^{\infty} [s_l(t)e^{j2\pi f_c t} + s_l^*(t)e^{-j2\pi f_c t}]e^{-j2\pi ft}\,dt$$
$$= \tfrac{1}{2}[S_l(f - f_c) + S_l^*(-f - f_c)] \qquad (4\text{-}1\text{-}21)$$

where $S_l(f)$ is the Fourier transform of $s_l(t)$. This is the basic relationship between the spectrum of the real bandpass signal $S(f)$ and the spectrum of the equivalent lowpass signal $S_l(f)$.

The energy in the signal $s(t)$ is defined as

$$\mathcal{E} = \int_{-\infty}^{\infty} s^2(t)\,dt$$
$$= \int_{-\infty}^{\infty} \{\text{Re}\,[s_l(t)e^{j2\pi f_c t}]\}^2\,dt \qquad (4\text{-}1\text{-}22)$$

When the identity in (4-1-20) is used in (4-1-22), we obtain the following result:

$$\mathcal{E} = \frac{1}{2}\int_{-\infty}^{\infty} |s_l(t)|^2\,dt$$
$$+ \frac{1}{2}\int_{-\infty}^{\infty} |s_l(t)|^2 \cos[4\pi f_c t + 2\theta(t)]\,dt \qquad (4\text{-}1\text{-}23)$$

Consider the second integral in (4-1-23). Since the signal $s(t)$ is narrowband, the real envelope $a(t) \equiv |s_l(t)|$ or, equivalently, $a^2(t)$ varies slowly relative to the rapid variations exhibited by the cosine function. A graphical illustration of the integrand in the second integral of (4-1-23) is shown in Fig. 4-1-2. The value of the integral is just the net area under the cosine function modulated by $a^2(t)$. Since the modulating waveform $a^2(t)$ varies slowly relative to the cosine function, the net area contributed by the second integral is very small relative to the value of the first integral in (4-1-23) and, hence, it can be

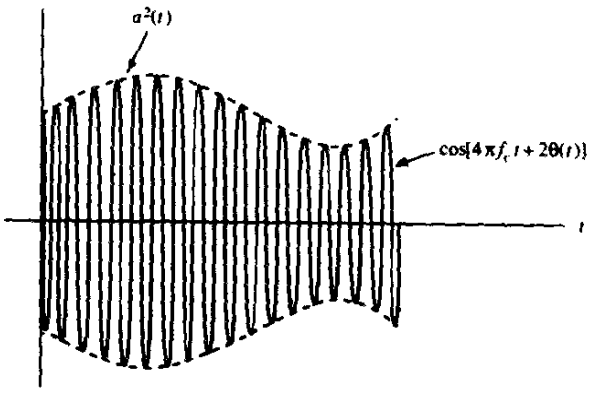


FIGURE 4-1-2 The signal $a^2(t)\cos[4\pi f_c t + 2\theta(t)]$.

neglected. Thus, for all practical pruposes, the energy in the bandpass signal $s(t)$, expressed in terms of the equivalent lowpass signal $s_l(t)$, is

$$\mathcal{E} = \frac{1}{2}\int_{-\infty}^{\infty} |s_l(t)|^2\, dt \tag{4-1-24}$$

where $|s_l(t)|$ is just the envelope $a(t)$ of $s(t)$.

4-1-2 Representation of Linear Bandpass Systems

A linear filter or system may be described either by its impulse response $h(t)$ or by its frequency response $H(f)$, which is the Fourier transform of $h(t)$. Since $h(t)$ is real,

$$H^*(-f) = H(f) \tag{4-1-25}$$

Let us define $H_l(f - f_c)$ as

$$H_l(f - f_c) = \begin{cases} H(f) & (f > 0) \\ 0 & (f < 0) \end{cases} \tag{4-1-26}$$

Then

$$H_l^*(-f - f_c) = \begin{cases} 0 & (f > 0) \\ H^*(-f) & (f < 0) \end{cases} \tag{4-1-27}$$

Using (4-1-25), we have

$$H(f) = H_l(f - f_c) + H_l^*(-f - f_c) \tag{4-1-28}$$

which resembles (4-1-21) except for the factor $\frac{1}{2}$. The inverse transform of $H(f)$ in (4-1-28) yields $h(t)$ in the form

$$\begin{aligned} h(t) &= h_l(t)e^{j2\pi f_c t} + h_l^*(t)e^{-j2\pi f_c t} \\ &= 2\,\mathrm{Re}\,[h_l(t)e^{j2\pi f_c t}] \end{aligned} \tag{4-1-29}$$

where $h_l(t)$ is the inverse Fourier transform of $H_l(f)$. In general, the impulse response $h_l(t)$ of the equivalent lowpass system is complex-valued.

4-1-3 Response of a Bandpass System to a Bandpass Signal

In Sections 4-1-1 and 4-1-2, we have shown that narrowband bandpass signals and systems can be represented by equivalent lowpass signals and systems. In this section, we demonstrate that the output of a bandpass system to a

bandpass input signal is simply obtained from the equivalent lowpass input signal and the equivalent lowpass impulse response of the system.

Suppose that $s(t)$ is a narrowband bandpass signal and $s_l(t)$ is the equivalent lowpass signal. This signal excites a narrowband bandpass system characterized by its bandpass impulse response $h(t)$ or by its equivalent lowpass impulse response $h_l(t)$. The output of the bandpass system is also a bandpass signal, and, therefore, it can be expressed in the form

$$r(t) = \text{Re}\,[r_l(t)e^{j2\pi f_c t}] \tag{4-1-30}$$

where $r(t)$ is related to the input signal $s(t)$ and the impulse response $h(t)$ by the convolution integral

$$r(t) = \int_{-\infty}^{\infty} s(\tau)h(t-\tau)\,d\tau \tag{4-1-31}$$

Equivalently, the output of the system, expressed in the frequency domain, is

$$R(f) = S(f)H(f) \tag{4-1-32}$$

Substituting from (4-1-21) for $S(f)$ and from (4-1-28) for $H(f)$, we obtain the result

$$R(f) = \tfrac{1}{2}[S_l(f-f_c) + S_l^*(-f-f_c)][H_l(f-f_c) + H_l^*(-f-f_c)] \tag{4-1-33}$$

When $s(t)$ is a narrowband signal and $h(t)$ is the impulse response of a narrowband system, $S_l(f-f_c) \approx 0$ and $H_l(f-f_c) = 0$ for $f < 0$. It follows from this narrowband condition that

$$S_l(f-f_c)H_l^*(-f-f_c) = 0, \qquad S_l^*(-f-f_c)H_l(f-f_c) = 0$$

Therefore, (4-1-33) simplifies to

$$\begin{aligned} R(f) &= \tfrac{1}{2}[S_l(f-f_c)H_l(f-f_c) + S_l^*(-f-f_c)H_l^*(-f-f_c)] \\ &= \tfrac{1}{2}[R_l(f-f_c) + R_l^*(-f-f_c)] \end{aligned} \tag{4-1-34}$$

where

$$R_l(f) = S_l(f)H_l(f) \tag{4-1-35}$$

is the output spectrum of the equivalent lowpass system excited by the equivalent lowpass signal. It is clear that the time domain relation for the output $r_l(t)$ is given by the convolution of $s_l(t)$ with $h_l(t)$. That is,

$$r_l(t) = \int_{-\infty}^{\infty} s_l(\tau)h_l(t-\tau)\,d\tau \tag{4-1-36}$$

The combination of (4-1-36) with (4-1-30) gives the relationship between the bandpass output signal $r(t)$ and the equivalent lowpass time functions $s_l(t)$ and $h_l(t)$. This simple relationship allows us to ignore any linear frequency translations encountered in the modulation of a signal for purposes of matching its spectral content to the frequency allocation of a particular channel. Thus, for mathematical convenience, we shall deal only with the transmission of equivalant lowpass signals through equivalent lowpass channels.

4-1-4 Representation of Bandpass Stationary Stochastic Processes

The representation of bandpass signals presented in Section 4-1-1 applied to deterministic signals. In this section, we extend the representation to sample functions of a bandpass stationary stochastic process. In particular, we derive the important relations between the correlation functions and power spectra of the bandpass signal and the correlation functions and power spectra of the equivalent lowpass signal.

Suppose that $n(t)$ is a sample function of a wide-sense stationary stochastic process with zero mean and power spectral density $\Phi_{nn}(f)$. The power spectral density is assumed to be zero outside of an interval of frequencies centered around $\pm f_c$, where f_c is termed the *carrier frequency*. The stochastic process $n(t)$ is said to be a *narrowband bandpass process* if the width of the spectral density is much smaller than f_c. Under this condition, a sample function of the process $n(t)$ can be represented by any of the three equivalent forms given in Section 4-1-1, namely,

$$n(t) = a(t)\cos[2\pi f_c t + \theta(t)] \tag{4-1-37}$$

$$= x(t)\cos 2\pi f_c t - y(t)\sin 2\pi f_c t \tag{4-1-38}$$

$$= \text{Re}[z(t)e^{j2\pi f_c t}] \tag{4-1-39}$$

where $a(t)$ is the envelope and $\theta(t)$ is the phase of the real-valued signal, $x(t)$ and $y(t)$ are the quadrature components of $n(t)$, and $z(t)$ is called the *complex envelope of* $n(t)$.

Let us consider the form given by (4-1-38) in more detail. First, we observe that if $n(t)$ is zero mean, then $x(t)$ and $y(t)$ must also have zero mean values. In addition, the stationarity of $n(t)$ implies that the autocorrelation and cross-correlation functions of $x(t)$ and $y(t)$ satisfy the following properties:

$$\phi_{xx}(\tau) = \phi_{yy}(\tau) \tag{4-1-40}$$

$$\phi_{xy}(\tau) = -\phi_{yx}(\tau) \tag{4-1-41}$$

That these two properties follow from the stationarity of $n(t)$ is now demonstrated. The autocorrelation function $\phi_{nn}(\tau)$ of $n(t)$ is

$$\begin{aligned} E[n(t)n(t+\tau)] &= E\{[x(t)\cos 2\pi f_c t - y(t)\sin 2\pi f_c t] \\ &\quad \times [x(t+\tau)\cos 2\pi f_c(t+\tau) \\ &\quad\quad - y(t+\tau)\sin 2\pi f_c(t+\tau)]\} \\ &= \phi_{xx}(\tau)\cos 2\pi f_c t \cos 2\pi f_c(t+\tau) \\ &\quad + \phi_{yy}(\tau)\sin 2\pi f_c t \sin 2\pi f_c(t+\tau) \\ &\quad - \phi_{xy}(\tau)\sin 2\pi f_c t \cos 2\pi f_c(t+\tau) \\ &\quad - \phi_{yx}(\tau)\cos 2\pi f_c t \sin 2\pi f_c(t+\tau) \end{aligned} \tag{4-1-42}$$

Use of the trigonometric identities

$$\begin{aligned} \cos A \cos B &= \tfrac{1}{2}[\cos(A-B) + \cos(A+B)] \\ \sin A \sin B &= \tfrac{1}{2}[\cos(A-B) - \cos(A+B)] \\ \sin A \cos B &= \tfrac{1}{2}[\sin(A-B) + \sin(A+B)] \end{aligned} \tag{4-1-43}$$

in (4-1-42) yields the result

$$\begin{aligned} E[n(t)n(t+\tau)] &= \tfrac{1}{2}[\phi_{xx}(\tau) + \phi_{yy}(\tau)]\cos 2\pi f_c \tau \\ &\quad + \tfrac{1}{2}[\phi_{xx}(\tau) - \phi_{yy}(\tau)]\cos 2\pi f_c(2t+\tau) \\ &\quad - \tfrac{1}{2}[\phi_{yx}(\tau) - \phi_{xy}(\tau)]\sin 2\pi f_c \tau \\ &\quad - \tfrac{1}{2}[\phi_{yx}(\tau) + \phi_{xy}(\tau)]\sin 2\pi f_c(2t+\tau) \end{aligned} \tag{4-1-44}$$

Since $n(t)$ is stationary, the right-hand side of (4-1-44) must be independent of t. But this condition can only be satisfied if (4-1-40) and (4-1-41) hold. As a consequence, (4-1-44) reduces to

$$\phi_{nn}(\tau) = \phi_{xx}(\tau)\cos 2\pi f_c \tau - \phi_{yx}(\tau)\sin 2\pi f_c \tau \tag{4-1-45}$$

We note that the relation between the autocorrelation function $\phi_{nn}(\tau)$ of the bandpass process and the autocorrelation and cross-correlation functions $\phi_{xx}(\tau)$ and $\phi_{yx}(\tau)$ of the quadrature components is identical in form to (4-1-38), which expresses the bandpass process in terms of the quadrature components.

The autocorrelation function of the equivalent lowpass process

$$z(t) = x(t) + jy(t) \tag{4-1-46}$$

is defined as

$$\phi_{zz}(\tau) = \tfrac{1}{2}E[z^*(t)z(t+\tau)] \tag{4-1-47}$$

Substituting (4-1-46) into (4-1-47) and performing the expectation operation, we obtain

$$\phi_{zz}(\tau) = \tfrac{1}{2}[\phi_{xx}(\tau) + \phi_{yy}(\tau) - j\phi_{xy}(\tau) + j\phi_{yx}(\tau)] \tag{4-1-48}$$

Now if the symmetry properties given in (4-1-40) and (4-1-41) are used in (4-1-48), we obtain

$$\phi_{zz}(\tau) = \phi_{xx}(\tau) + j\phi_{yx}(\tau) \tag{4-1-49}$$

which relates the autocorrelation function of the complex envelope to the autocorrelation and cross-correlation functions of the quadrature components. Finally, we incorporate the result given by (4-1-49) into (4-1-45), and we have

$$\phi_{nn}(\tau) = \text{Re}\,[\phi_{zz}(\tau)e^{j2\pi f_c\tau}] \tag{4-1-50}$$

Thus, the autocorrelation function $\phi_{nn}(\tau)$ of the bandpass stochastic process is uniquely determined from the autocorrelation function $\phi_{zz}(\tau)$ of the equivalent lowpass process $z(t)$ and the carrier frequency f_c.

The power density spectrum $\Phi_{nn}(f)$ of the stochastic process $n(t)$ is the Fourier transform of $\phi_{nn}(\tau)$. Hence,

$$\begin{aligned}\Phi_{nn}(f) &= \int_{-\infty}^{\infty} \{\text{Re}\,[\phi_{zz}(\tau)e^{j2\pi f_c\tau}]\}e^{-j2\pi f\tau}\,d\tau \\ &= \tfrac{1}{2}[\Phi_{zz}(f - f_c) + \Phi_{zz}(-f - f_c)]\end{aligned} \tag{4-1-51}$$

where $\Phi_{zz}(f)$ is the power density spectrum of the equivalent lowpass process $z(t)$. Since the autocorrelation function of $z(t)$ satisfies the property $\phi_{zz}(\tau) = \phi^*_{zz}(-\tau)$, it follows that $\Phi_{zz}(f)$ is a real-valued function of frequency.

Properties of the Quadrature Components It was just demonstrated above that the cross-correlation function of the quadrature components $x(t)$ and $y(t)$ of the bandpass stationary stochastic process $n(t)$ satisfies the symmetry condition in (4-1-41). Furthermore, any cross-correlation function satisfies the condition

$$\phi_{yx}(\tau) = \phi_{xy}(-\tau) \tag{4-1-52}$$

From these two conditions, we conclude that

$$\phi_{xy}(\tau) = -\phi_{xy}(-\tau) \tag{4-1-53}$$

That is, $\phi_{xy}(\tau)$ is an odd function of τ. Consequently, $\phi_{xy}(0) = 0$, and, hence, $x(t)$ and $y(t)$ are uncorrelated (for $\tau = 0$, only). Of course, this does not mean that the processes $x(t)$ and $y(t + \tau)$ are uncorrelated for all τ, since that would imply that $\phi_{xy}(\tau) = 0$ for all τ. If, indeed, $\phi_{xy}(\tau) = 0$ for all τ, then $\phi_{zz}(\tau)$ is real and the power spectral density $\Phi_{zz}(f)$ satisfies the condition

$$\Phi_{zz}(f) = \Phi_{zz}(-f) \tag{4-1-54}$$

and vice versa. That is, $\Phi_{zz}(f)$ is symmetric about $f = 0$.

In the special case in which the stationary stochastic process $n(t)$ is gaussian, the quadrature components $x(t)$ and $y(t+\tau)$ are jointly gaussian. Moreover, for $\tau = 0$, they are statistically independent, and, hence, their joint probability density function is

$$p(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \tag{4-1-55}$$

where the variance σ^2 is defined as $\sigma^2 = \phi_{xx}(0) = \phi_{yy}(0) = \phi_{nn}(0)$.

Representation of White Noise White noise is a stochastic process that is defined to have a flat (constant) power spectral density over the entire frequency range. This type of noise cannot be expressed in terms of quadrature components, as a result of its wideband character.

In problems concerned with the demodulation of narrowband signals in noise, it is mathematically convenient to model the additive noise process as white and to represent the noise in terms of quadrature components. This can be accomplished by postulating that the signals and noise at the receiving terminal have passed through an ideal bandpass filter, having a passband that includes the spectrum of the signals but is much wider. Such a filter will introduce negligible, if any, distortion on the signal but it does eliminate the noise frequency components outside of the passband.

The noise resulting from passing the white noise process through a spectrally flat (ideal) bandpass filter is termed *bandpass white noise* and has the power spectral density depicted in Fig. 4-1-3. Bandpass white noise can be represented by any of the forms given in (4-1-37), (4-1-38), and (4-1-39). The equivalent lowpass noise $z(t)$ has a power spectral density

$$\Phi_{zz}(f) = \begin{cases} N_0 & (|f| \leq \frac{1}{2}B) \\ 0 & (|f| > \frac{1}{2}B) \end{cases} \tag{4-1-56}$$

and its autocorrelation function is

$$\phi_{zz}(\tau) = N_0 \frac{\sin \pi B\tau}{\pi\tau} \tag{4-1-57}$$

The limiting form of $\phi_{zz}(\tau)$ as B approaches infinity is

$$\phi_{zz}(\tau) = N_0\delta(\tau) \tag{4-1-58}$$

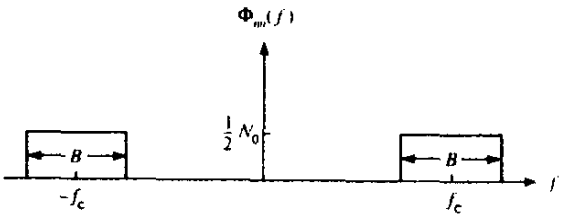


FIGURE 4-1-3 Bandpass noise with a flat spectrum.

The power spectral density for white noise and bandpass white noise is symmetric about $f = 0$, so $\phi_{vv}(\tau) = 0$ for all τ. Therefore,

$$\phi_{zz}(\tau) = \phi_{xx}(\tau) = \phi_{yy}(\tau) \tag{4-1-59}$$

That is, the quadrature components $x(t)$ and $y(t)$ are uncorrelated for all time shifts τ and the autocorrelation functions of $z(t)$, $x(t)$, and $y(t)$ are all equal.

4-2 SIGNAL SPACE REPRESENTATIONS

In this section, we demonstrate that signals have characteristics that are similar to vectors and develop a vector representation for signal waveforms. We begin with some basic definitions and concepts involving vectors.

4-2-1 Vector Space Concepts

A vector $\mathbf{v}$ in an n-dimensional space is characterized by its n components $[v_1\ v_2\ \ldots\ v_n]$. It may also be represented as a linear combination of *unit vectors* or *basis vectors* $\mathbf{e}_i$, $1 \leq i \leq n$, i.e.,

$$\mathbf{v} = \sum_{i=1}^{n} v_i \mathbf{e}_i \tag{4-2-1}$$

where, by definition, a unit vector has length unity and v_i is the projection of the vector $\mathbf{v}$ onto the unit vector $\mathbf{e}_i$.

The *inner product* of two n-dimensional vectors $\mathbf{v}_1 = [v_{11}\ v_{12}\ \ldots\ v_{1n}]$ and $\mathbf{v}_2 = [v_{21}\ v_{22}\ \ldots\ v_{2n}]$ is defined as

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = \sum_{i=1}^{n} v_{1i} v_{2i} \tag{4-2-2}$$

Two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal if $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. More generally, a set of m vectors $\mathbf{v}_k$, $1 \leq k \leq m$, are orthogonal if

$$\mathbf{v}_i \cdot \mathbf{v}_j = 0 \tag{4-2-3}$$

for all $1 \leq i, j \leq m$ and $i \neq j$.

The *norm* of a vector $\mathbf{v}$ is denoted by $\|\mathbf{v}\|$ and is defined as

$$\|\mathbf{v}\| = (\mathbf{v} \cdot \mathbf{v})^{1/2} = \sqrt{\sum_{i=1}^{n} v_i^2} \tag{4-2-4}$$

which is simply its length. A set of m vectors is said to be *orthonormal* if the vectors are orthogonal and each vector has a unit norm. A set of m vectors is said to be *linearly independent* if no one vector can be represented as a linear combination of the remaining vectors.

Two n-dimensional vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ satisfy the *triangle inequality*

$$\|\mathbf{v}_1 + \mathbf{v}_2\| \leq \|\mathbf{v}_1\| + \|\mathbf{v}_2\| \tag{4-2-5}$$

with equality if $\mathbf{v}_1$ and $\mathbf{v}_2$ are in the same direction, i.e., $\mathbf{v}_1 = a\mathbf{v}_2$ where a is a

positive real scalar. From the triangle inequality there follows the *Cauchy–Schwartz inequality*

$$|\mathbf{v}_1 \cdot \mathbf{v}_2| \leq \|\mathbf{v}_1\| \, \|\mathbf{v}_2\| \tag{4-2-6}$$

with equality if $\mathbf{v}_1 = a\mathbf{v}_2$. The norm square of the sum of two vectors may be expressed as

$$\|\mathbf{v}_1 + \mathbf{v}_2\|^2 = \|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2 + 2\mathbf{v}_1 \cdot \mathbf{v}_2 \tag{4-2-7}$$

If $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal then $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$ and, hence,

$$\|\mathbf{v}_1 + \mathbf{v}_2\|^2 = \|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2 \tag{4-2-8}$$

This is the Pythagorean relation for two orthogonal n-dimensional vectors.

From matrix algebra, we recall that a linear transformation in an n-dimensional vector space is a matrix transformation of the form

$$\mathbf{v}' = \mathbf{A}\mathbf{v} \tag{4-2-9}$$

where the matrix $\mathbf{A}$ transforms the vector $\mathbf{v}$ into some vector $\mathbf{v}'$. In the special case where $\mathbf{v}' = \lambda\mathbf{v}$, i.e.,

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v} \tag{4-2-10}$$

where λ is some (positive or negative) scalar, the vector $\mathbf{v}$ is called an *eigenvector* of the transformation and λ is the corresponding *eigenvalue.*

Finally, let us review the Gram–Schmidt procedure for constructing a set of orthonormal vectors from a set of n-dimensional vectors $\mathbf{v}_i$, $1 \leq i \leq m$. We begin by arbitrarily selecting a vector from the set, say $\mathbf{v}_1$. By normalizing its length, we obtain the first vector, say

$$\mathbf{u}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} \tag{4-2-11}$$

Next, we may select $\mathbf{v}_2$ and, first, subtract the projection of $\mathbf{v}_2$ onto $\mathbf{u}_1$. Thus, we obtain

$$\mathbf{u}_2' = \mathbf{v}_2 - (\mathbf{v}_2 \cdot \mathbf{u}_1)\mathbf{u}_1 \tag{4-2-12}$$

Then, we normalize the vector $\mathbf{u}_2'$ to unit length. This yields

$$\mathbf{u}_2 = \frac{\mathbf{u}_2'}{\|\mathbf{u}_2'\|} \tag{4-2-13}$$

The procedure continues by selecting $\mathbf{v}_3$ and subtracting the projections of $\mathbf{v}_3$ into $\mathbf{u}_1$ and $\mathbf{u}_2$. Thus, we have

$$\mathbf{u}_3' = \mathbf{v}_3 - (\mathbf{v}_3 \cdot \mathbf{u}_1)\mathbf{u}_1 - (\mathbf{v}_3 \cdot \mathbf{u}_2)\mathbf{u}_2 \tag{4-2-14}$$

Then, the orthonormal vector $\mathbf{u}_3$ is

$$\mathbf{u}_3 = \frac{\mathbf{u}_3'}{\|\mathbf{u}_3'\|} \tag{4-2-15}$$

By continuing this procedure, we shall construct a set of n_1, orthonormal vectors, where $n_1 \leq n$, in general. If $m < n$ then $n_1 \leq m$, and if $m \geq n$ then $n_1 \leq n$.

4-2-2 Signal Space Concepts

As in the case of vectors, we may develop a parallel treatment for a set of signals defined on some interval $[a, b]$. The *inner product* of two generally complex-valued signals $x_1(t)$ and $x_2(t)$ is denoted by $\langle x_1(t), x_2(t)\rangle$ and defined as

$$\langle x_1(t), x_2(t)\rangle = \int_a^b x_1(t)x_2^*(t)\,dt \tag{4-2-16}$$

The signals are orthogonal if their inner product is zero.

The *norm* of a signal is defined as

$$\|x(t)\| = \left(\int_a^b |x(t)|^2\,dt\right)^{1/2} \tag{4-2-17}$$

A set of m signals are *orthonormal* if they are orthogonal and their norms are all unity. A set of m signals is *linearly independent*, if no signal can be represented as a linear combination of the remaining signals.

The *triangle inequality* for two signals is simply

$$\|x_1(t) + x_2(t)\| \leq \|x_1(t)\| + \|x_2(t)\| \tag{4-2-18}$$

and the *Cauchy–Schwartz inequality* is

$$\left|\int_a^b x_1(t)x_2^*(t)\,dt\right| \leq \left|\int_a^b |x_1(t)|^2\,dt\right|^{1/2} \left|\int_a^b |x_2(t)|^2\,dt\right|^{1/2} \tag{4-2-19}$$

with equality when $x_2(t) = ax_1(t)$, where a is any complex number.

4-2-3 Orthogonal Expansions of Signals

In this section, we develop a vector representation for signal waveforms, and, thus, we demonstrate an equivalence between a signal waveform and its vector representation.

Suppose that $s(t)$ is a deterministic, real-valued signal with finite energy

$$\mathscr{E}_s = \int_{-\infty}^{\infty} [s(t)]^2\,dt \tag{4-2-20}$$

Furthermore, suppose that there exists a set of functions $\{f_n(t), n = 1, 2, \ldots, N\}$ that are orthonormal in the sense that

$$\int_{-\infty}^{\infty} f_n(t)f_m(t)\,dt = \begin{cases} 0 & (m \neq n) \\ 1 & (m = n) \end{cases} \tag{4-2-21}$$

We may approximate the signal $s(t)$ by a weighted linear combination of these functions, i.e.,

$$\hat{s}(t) = \sum_{k=1}^{K} s_k f_k(t) \tag{4-2-22}$$

where $\{s_k,\ 1 \leq k \leq K\}$ are the coefficients in the approximation of $s(t)$. The approximation error incurred is

$$e(t) = s(t) - \hat{s}(t) \tag{4-2-23}$$

Let us select the coefficients $\{s_k\}$ so as to minimize the energy $\mathscr{E}_e$ of the approximation error. Thus,

$$\begin{aligned} \mathscr{E}_e &= \int_{-\infty}^{\infty} [s(t) - \hat{s}(t)]^2\, dt \\ &= \int_{-\infty}^{\infty} \left[s(t) - \sum_{k=1}^{K} s_k f_k(t) \right]^2 dt \end{aligned} \tag{4-2-24}$$

The optimum coefficients in the series expansion of $s(t)$ may be found by differentiating (4-2-24) with respect to each of the coefficients $\{s_k\}$ and setting the first derivatives to zero. Alternatively, we may use a well-known result from estimation theory based on the mean-square-error criterion, which, simply stated, is that the minimum of $\mathscr{E}_e$ with respect to the $\{s_k\}$ is obtained when the error is orthogonal to each of the functions in the series expansion. Thus,

$$\int_{-\infty}^{\infty} \left[s(t) - \sum_{k=1}^{K} s_k f_k(t) \right] f_n(t)\, dt = 0, \qquad n = 1, 2, \ldots, K \tag{4-2-25}$$

Since the functions $\{f_n(t)\}$ are orthonormal, (4-2-25) reduces to

$$s_n = \int_{-\infty}^{\infty} s(t) f_n(t)\, dt, \qquad n = 1, 2, \ldots, K \tag{4-2-26}$$

Thus, the coefficients are obtained by projecting the signal $s(t)$ onto each of the functions $\{f_n(t)\}$. Consequently, $\hat{s}(t)$ is the projection of $s(t)$ onto the K-dimensional signal space spanned by the functions $\{f_n(t)\}$. The minimum mean square approximation error is

$$\begin{aligned} \mathscr{E}_{\min} &= \int_{-\infty}^{\infty} e(t) s(t)\, dt \\ &= \int_{-\infty}^{\infty} [s(t)]^2\, dt - \int_{-\infty}^{\infty} \sum_{k=1}^{K} s_k f_k(t) s(t)\, dt \\ &= \mathscr{E}_s - \sum_{k=1}^{k} s_k^2 \end{aligned} \tag{4-2-27}$$

which is nonnegative, by definition.

When the minimum mean square approximation error $\mathscr{E}_{\min} = 0$,

$$\mathscr{E}_s = \sum_{k=1}^{K} s_k^2 = \int_{-\infty}^{\infty} [s(t)]^2 \, dt \tag{4-2-28}$$

Under the condition that $\mathscr{E}_{\min} = 0$, we may express $s(t)$ as

$$s(t) = \sum_{k=1}^{K} s_k f_k(t) \tag{4-2-29}$$

where it is understood that equality of $s(t)$ to its series expansion holds in the sense that the approximation error has zero energy.

When every finite energy signal can be represented by a series expansion of the form in (4-2-29) for which $\mathscr{E}_{\min} = 0$, the set of orthonormal functions $\{f_n(t)\}$ is said to be *complete*.

Example 4-2-1: Trigonometric Fourier Series

A finite energy signal $s(t)$ that is zero everywhere except in the range $0 \leq t \leq T$ and has a finite number of discontinuities in this interval, can be represented in a Fourier series as

$$s(t) = \sum_{k=0}^{\infty} \left(a_k \cos \frac{2\pi k t}{T} + b_k \sin \frac{2\pi k t}{T} \right) \tag{4-2-30}$$

where the coefficients $\{a_k, b_k\}$ that minimize the mean square error are given by

$$\begin{aligned} a_k &= \frac{1}{\sqrt{T}} \int_0^T s(t) \cos \frac{2\pi k t}{T} \, dt \\ b_k &= \frac{1}{\sqrt{T}} \int_0^T s(t) \sin \frac{2\pi k t}{T} \, dt \end{aligned} \tag{4-2-31}$$

The set of trigonometric functions $\{\sqrt{2/T} \cos 2\pi k t/T, \sqrt{2/T} \sin 2\pi k t/T\}$ is complete, and, hence, the series expansion results in zero mean square error. These properties are easily established from the development given above.

Gram–Schmidt Procedure Now suppose that we have a set of finite energy signal waveforms $\{s_i(t), i = 1, 2, \ldots, M\}$ and we wish to construct a set of orthonormal waveforms. The Gram–Schmidt orthogonalization procedure allows us to construct such a set. We begin with the first waveform $s_1(t)$, which is assumed to have energy $\mathscr{E}_1$. The first waveform is simply constructed as

$$f_1(t) = \frac{s_1(t)}{\sqrt{\mathscr{E}_1}} \tag{4-2-32}$$

Thus, $f_1(t)$ is simply $s_1(t)$ normalized to unit energy.

The second waveform is constructed from $s_2(t)$ by first computing the projection of $f_1(t)$ onto $s_2(t)$, which is

$$c_{12} = \int_{-\infty}^{\infty} s_2(t) f_1(t)\, dt \tag{4-2-33}$$

Then, $c_{12} f_1(t)$ is subtracted from $s_2(t)$ to yield

$$f_2'(t) = s_2(t) - c_{12} f_1(t) \tag{4-2-34}$$

This waveform is orthogonal to $f_1(t)$ but it does not have unit energy. If $\mathscr{E}_2$ denotes the energy of $f_2'(t)$, the normalized waveform that is orthogonal to $f_1(t)$ is

$$f_2(t) = \frac{f_2'(t)}{\sqrt{\mathscr{E}_2}} \tag{4-2-35}$$

In general, the orthogonalization of the kth function leads to

$$f_k(t) = \frac{f_k'(t)}{\sqrt{\mathscr{E}_k}} \tag{4-2-36}$$

where

$$f_k'(t) = s_k(t) - \sum_{i=1}^{k-1} c_{ik} f_i(t) \tag{4-2-37}$$

and

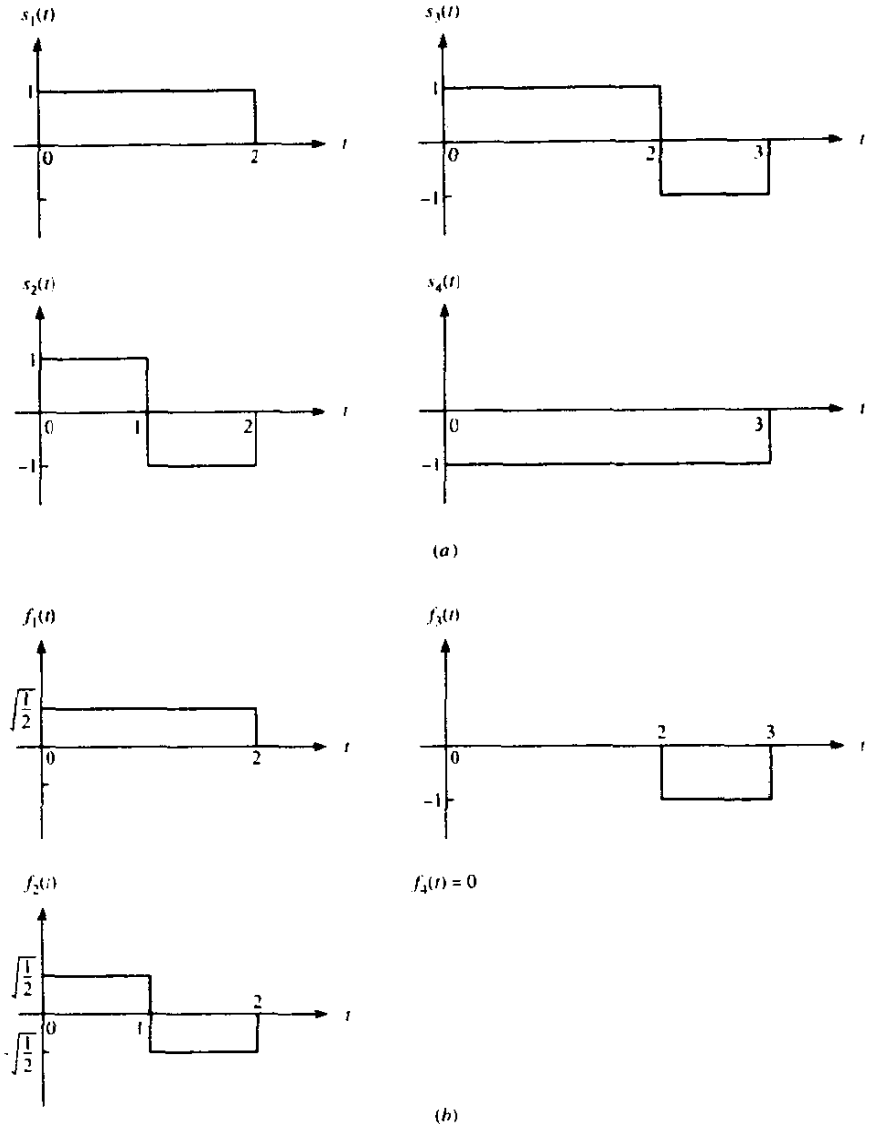
$$c_{ik} = \int_{-\infty}^{\infty} s_k(t) f_i(t)\, dt, \qquad i = 1, 2, \ldots, k-1 \tag{4-2-38}$$

Thus, the orthogonalization process is continued until all the M signal waveforms $\{s_i(t)\}$ have been exhausted and $N \leq M$ orthonormal waveforms have been constructed. The dimensionality N of the signal space will be equal to M if all the signal waveforms are linearly independent, i.e., none of the signals waveforms is a linear combination of the other signal waveforms.

Example 4-2-2

Let us apply the Gram–Schmidt procedure to the set of four waveforms illustrated in Fig. 4-2-1(*a*). The waveform $s_1(t)$ has energy $\mathscr{E}_1 = 2$, so that $f_1(t) = \sqrt{\frac{1}{2}}\, s_1(t)$. Next, we observe that $c_{12} = 0$; hence, $s_2(t)$ and $f_1(t)$ are orthogonal. Therefore, $f_2(t) = s_2(t)/\sqrt{\mathscr{E}_2} = \sqrt{\frac{1}{2}}\, s_2(t)$. To obtain $f_3(t)$, we compute c_{13} and c_{23}, which are $c_{13} = \sqrt{2}$ and $c_{23} = 0$. Thus,

$$f_3'(t) = s_3(t) - \sqrt{2} f_1(t) = \begin{cases} -1 & (2 \leq t \leq 3) \\ 0 & \text{(otherwise)} \end{cases}$$



Gram–Schmidt orthogonalization of the signals $\{s_i(t), i = 1, 2, 3, 4\}$ and the corresponding orthogonal signals.

Since $f_3'(t)$ has unit energy, it follows that $f_3(t) = f_3'(t)$. In determining $f_4(t)$, we find that $c_{14} = -\sqrt{2}$, $c_{24} = 0$, and $c_{34} = 1$. Hence,

$$f_4'(t) = s_4(t) + \sqrt{2}\, f_1(t) - f_3(t) = 0$$

Consequently, $s_4(t)$ is a linear combination of $f_1(t)$ and $f_3(t)$ and, hence, $f_4(t) = 0$. The three orthonormal functions are illustrated in Fig. 4-2-1(b).

Once we have constructed the set of orthonormal waveforms $\{f_n(t)\}$, we can express the M signals $\{s_n(t)\}$ as linear combinations of the $\{f_n(t)\}$. Thus, we may write

$$s_k(t) = \sum_{n=1}^{N} s_{kn} f_n(t), \qquad k = 1, 2, \ldots, M \tag{4-2-39}$$

and

$$\mathcal{E}_k = \int_0^T [s_k(t)]^2 \, dt = \sum_{n=1}^{N} s_{kn}^2 = \|\mathbf{s}_k\|^2 \tag{4-2-40}$$

Based on the expression in (4-2-39), each signal may be represented by the vector

$$\mathbf{s}_k = [s_{k1} \; s_{k2} \; \cdots \; s_{kN}] \tag{4-2-41}$$

or, equivalently, as a point in the N-dimensional signal space with coordinates $\{s_{ki}, i = 1, 2, \ldots, N\}$. The energy in the kth signal is simply the square of the length of the vector or, equivalently, the square of the Euclidean distance from the origin to the point in the N-dimensional space. Thus, any signal can be represented geometrically as a point in the signal space spanned by the orthonormal functions $\{f_n(t)\}$.

Example 4-2-3

Let us obtain the vector representation of the four signals shown in Fig. 4-2-1(a) by using the orthonormal set of functions in Fig. 4-2-1(b). Since the dimensionality of the signal space is $N = 3$, each signal is described by three components. The signal $s_1(t)$ is characterized by the vector $\mathbf{s}_1 = (\sqrt{2}, 0, 0)$. Similarly, the signals $s_2(t)$, $s_3(t)$, and $s_4(t)$ are characterized by the vectors $\mathbf{s}_2 = (0, \sqrt{2}, 0)$, $\mathbf{s}_3 = (\sqrt{2}, 0, 1)$, and $\mathbf{s}_4 = (-\sqrt{2}, 0, 1)$, respectively. These vectors are shown in Fig. 4-2-2. Their lengths are $|\mathbf{s}_1| = \sqrt{2}$, $|\mathbf{s}_2| = \sqrt{2}$,

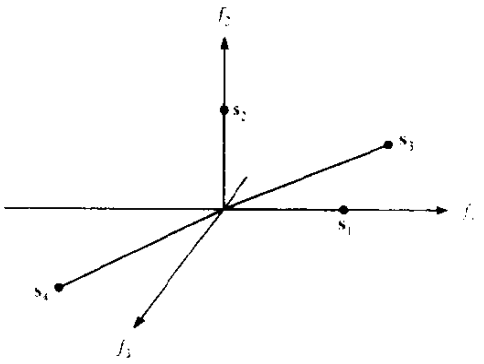


FIGURE 4-2-2 The four signal vectors represented as points in three dimensional function space.

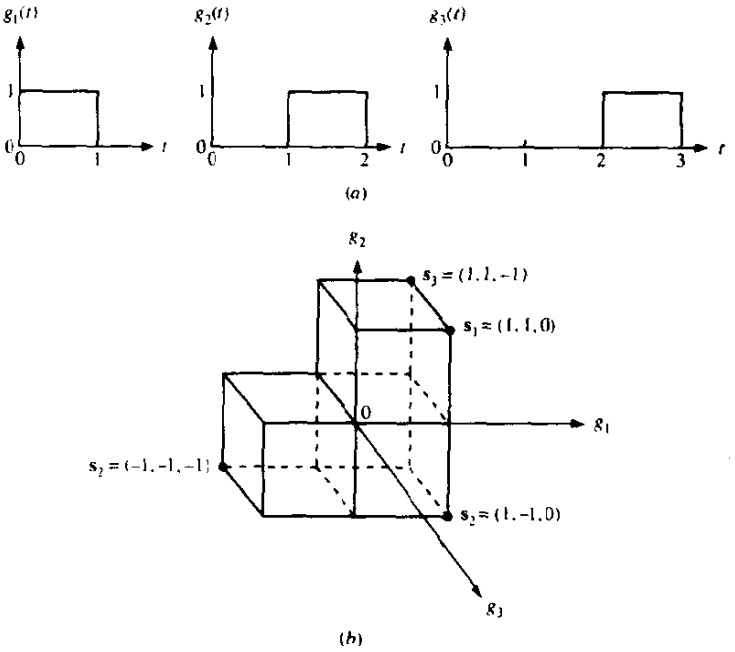
$|\mathbf{s}_3| = \sqrt{3}$, and $|\mathbf{s}_4| = \sqrt{3}$, and the corresponding signal energies are $\mathscr{E}_k = |\mathbf{s}_k|^2$, $k = 1, 2, 3, 4$.

We have demonstrated that a set of M finite energy waveforms $\{s_n(t)\}$ can be represented by a weighted linear combination of orthonormal functions $\{f_n(t)\}$ of dimensionality $N \leq M$. The functions $\{f_n(t)\}$ are obtained by applying the Gram–Schmidt orthogonalization procedure on $\{s_n(t)\}$. It should be emphasized, however, that the functions $\{f_n(t)\}$ obtained from the Gram–Schmidt procedure are not unique. If we alter the order in which the orthogonalization of the signals $\{s_n(t)\}$ is performed, the orthonormal waveforms will be different and the corresponding vector representation of the signals $\{s_n(t)\}$ will depend on the choice of the orthonormal functions $\{f_n(t)\}$. Nevertheless, the vectors $\{\mathbf{s}_n\}$ will retain their geometrical configuration and their lengths will be invariant to the choice of orthonormal functions $\{f_n(t)\}$.

Example 4-2-4

An alternative set of orthonormal functions for the four signals in Fig. 4-2-1 is illustrated in Fig. 4-2-3(a). By using these functions to expand $\{s_n(t)\}$, we

FIGURE 4-2-3 An alternative set of orthonormal functions for the four signals in Fig. 4-2-1(a) and the corresponding signal points.



obtain the corresponding vectors $\mathbf{s}_1 = (1, 1, 0)$, $\mathbf{s}_2 = (1, -1, 0)$, $\mathbf{s}_3 = (1, 1, -1)$, and $\mathbf{s}_4 = (-1, -1, -1)$, which are shown in Fig. 4-2-3(*b*). Note that the vector lengths are identical to those obtained from the orthonormal functions $\{f_n(t)\}$.

The orthogonal expansions described above were developed for real-valued signal waveforms. The extension to complex-valued signal waveforms is left as an exercise for the reader (see Problems 4-6 and 4-7).

Finally, let us consider the case in which the signal waveforms are bandpass and represented as

$$s_m(t) = \text{Re}\,[s_{lm}(t)e^{j2\pi f_c t}], \qquad m = 1, 2, \ldots, M \tag{4-2-42}$$

where $\{s_{lm}(t)\}$ denote the equivalent lowpass signals. Recall that the signal energies may be expressed either in terms of $s_m(t)$ or $s_{lm}(t)$, as

$$\begin{aligned} \mathcal{E}_m &= \int_{-\infty}^{\infty} s_m^2(t)\,dt \\ &= \frac{1}{2}\int_{-\infty}^{\infty} |s_{lm}(t)|^2\,dt \end{aligned} \tag{4-2-43}$$

The similarity between any pair of signal waveforms, say $s_m(t)$ and $s_k(t)$, is measured by the normalized cross-correlation

$$\frac{1}{\sqrt{\mathcal{E}_m \mathcal{E}_k}}\int_{-\infty}^{\infty} s_m(t)s_k(t)\,dt = \text{Re}\left\{\frac{1}{2\sqrt{\mathcal{E}_m \mathcal{E}_k}}\int_{-\infty}^{\infty} s_{lm}(t)s_{lk}^*(t)\,dt\right\} \tag{4-2-44}$$

We define the complex-valued cross-correlation coefficient ρ_{km} as

$$\rho_{km} = \frac{1}{2\sqrt{\mathcal{E}_m \mathcal{E}_k}}\int_{-\infty}^{\infty} s_{lm}^*(t)s_{lk}(t)\,dt \tag{4-2-45}$$

Then,

$$\text{Re}\,(\rho_{km}) = \frac{1}{\sqrt{\mathcal{E}_m \mathcal{E}_k}}\int_{-\infty}^{\infty} s_m(t)s_k(t)\,dt \tag{4-2-46}$$

or, equivalently,

$$\text{Re}\,(\rho_{km}) = \frac{\mathbf{s}_m \cdot \mathbf{s}_k}{\|\mathbf{s}_m\|\,\|\mathbf{s}_k\|} = \frac{\mathbf{s}_m \cdot \mathbf{s}_k}{\sqrt{\mathcal{E}_m \mathcal{E}_k}} \tag{4-2-47}$$

The cross-correlation coefficients between pairs of signal waveforms or signal vectors comprise one set of parameters that characterize the similarity

of a set of signals. Another related parameter is the Euclidean distance $d_{km}^{(e)}$ between a pair of signals, defined as

$$\begin{aligned} d_{km}^{(e)} &= \|\mathbf{s}_m - \mathbf{s}_k\| \\ &= \left\{\int_{-\infty}^{\infty} [s_m(t) - s_k(t)]^2\, dt\right\}^{1/2} \\ &= \{\mathscr{E}_m + \mathscr{E}_k - 2\sqrt{\mathscr{E}_m \mathscr{E}_k}\,\mathrm{Re}\,(\rho_{km})\}^{1/2} \end{aligned} \tag{4-2-48}$$

When $\mathscr{E}_m = \mathscr{E}_k = \mathscr{E}$ for all m and k, this expression simplifies to

$$d_{km}^{(e)} = \{2\mathscr{E}[1 - \mathrm{Re}\,(\rho_{km})]\}^{1/2} \tag{4-2-49}$$

Thus, the Euclidean distance is an alternative measure of the similarity (or dissimilarity) of the set of signal waveforms or the corresponding signal vectors.

In the following section, we describe digitally modulated signals and make use of the signal space representation for such signals. We shall observe that digitally modulated signals, which are classified as linear, are conveniently expanded in terms of two orthonormal basis functions of the form

$$\begin{aligned} f_1(t) &= \sqrt{\frac{2}{T}}\cos 2\pi f_c t \\ f_2(t) &= -\sqrt{\frac{2}{T}}\sin 2\pi f_c t \end{aligned} \tag{4-2-50}$$

Hence, if $s_{lm}(t)$ is expressed as $s_{lm}(t) = x_l(t) + jy_l(t)$, it follows that $s_m(t)$ in (4-2-42) may be expressed as

$$s_m(t) = x_l(t) f_1(t) + y_l(t) f_2(t) \tag{4-2-51}$$

where $x_l(t)$ and $y_l(t)$ represent the signal modulations.

4-3 REPRESENTATION OF DIGITALLY MODULATED SIGNALS

In the transmission of digital information over a communications channel, the modulator is the interface device that maps the digital information into analog waveforms that match the characteristics of the channel. The mapping is generally performed by taking blocks of $k = \log_2 M$ binary digits at a time from the information sequence $\{a_n\}$ and selecting one of $M = 2^k$ deterministic, finite energy waveforms $\{s_m(t),\ m = 1, 2, \ldots, M\}$ for transmission over the channel.

When the mapping from the digital sequence $\{a_n\}$ to waveforms is performed under the constraint that a waveform transmitted in any time interval depends on one or more previously transmitted waveforms, the modulator is said to have *memory*. On the other hand, when the mapping

from the sequence $\{a_n\}$ to the waveforms $\{s_m(t)\}$ is performed without any constraint on previously transmitted waveforms, the modulator is called *memoryless.*

In addition to classifying the modulator as either memoryless or having memory, we may classify it as either *linear* or *nonlinear.* Linearity of a modulation method requires that the principle of superposition applies in the mapping of the digital sequence into successive waveforms. In nonlinear modulation, the superposition principle does not apply to signals transmitted in successive time intervals. We shall begin by describing memoryless modulation methods.

4-3-1 Memoryless Modulation Methods

As indicated above, the modulator in a digital communication system maps a sequence of binary digits into a set of corresponding signal waveforms. These waveforms may differ in either amplitude or in phase or in frequency, or some combination of two or more signal parameters. We consider each of these signal types separately, beginning with digital pulse amplitude modulation (PAM). In all cases, we assume that the sequence of binary digits at the input to the modulator occurs at a rate of R bits/s.

Pulse Amplitude Modulated (PAM) Signals In digital PAM, the signal waveforms may be represented as

$$\begin{aligned} s_m(t) &= \text{Re}\,[A_m g(t) e^{j2\pi f_c t}] \\ &= A_m g(t) \cos 2\pi f_c t, \quad m = 1, 2, \ldots, M, \quad 0 \leq t \leq T \end{aligned} \tag{4-3-1}$$

where $\{A_m, 1 \leq m \leq M\}$ denote the set of M possible amplitudes corresponding to $M = 2^k$ possible k-bit blocks or *symbols.* The signal amplitudes A_m take the discrete values (levels)

$$A_m = (2m - 1 - M)d, \qquad m = 1, 2, \ldots, M \tag{4-3-2}$$

where $2d$ is the distance between adjacent signal amplitudes. The waveform $g(t)$ is a real-valued signal pulse whose shape influences the spectrum of the transmitted signal, as we shall observe later. The symbol rate for the PAM signal is R/k. This is the rate at which changes occur in the amplitude of the carrier to reflect the transmission of new information. The time interval $T_b = 1/R$ is called the *bit interval* and the time interval $T = k/R = kT_b$ is called the *symbol interval.*

The M PAM signals have energies

$$\begin{aligned} \mathcal{E}_m &= \int_0^T s_m^2(t)\, dt \\ &= \tfrac{1}{2} A_m^2 \int_0^T g^2(t)\, dt \\ &= \tfrac{1}{2} A_m^2 \mathcal{E}_g \end{aligned} \tag{4-3-3}$$

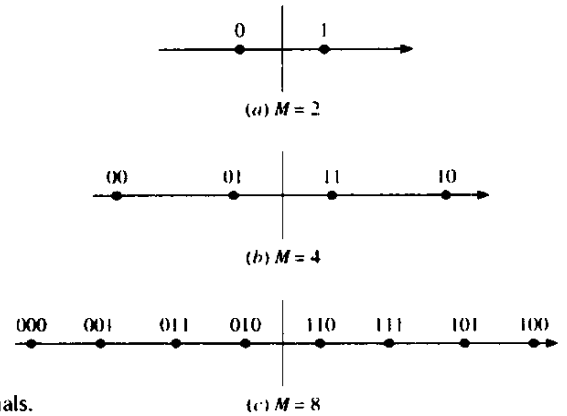


FIGURE 4-3-1 Signal space diagram for digital PAM signals.

where $\mathscr{E}_g$ denotes the energy in the pulse $g(t)$. Clearly, these signals are one-dimensional ($N = 1$), and, hence, are represented by the general form

$$s_m(t) = s_m f(t) \tag{4-3-4}$$

where $f(t)$ is defined as the unit-energy signal waveform given as

$$f(t) = \sqrt{\frac{2}{\mathscr{E}_g}}\, g(t) \cos 2\pi f_c t \tag{4-3-5}$$

and

$$s_m = A_m \sqrt{\tfrac{1}{2}\mathscr{E}_g}, \qquad m = 1, 2, \ldots, M \tag{4-3-6}$$

The corresponding signal space diagrams for $M = 2$, $M = 4$ and $M = 8$ are shown in Fig. 4-3-1. Digital PAM is also called *amplitude-shift keying* (ASK).

The mapping or assignment of k information bits to the $M = 2^k$ possible signal amplitudes may be done in a number of ways. The preferred assignment is one in which the adjacent signals amplitudes differ by one binary digit as illustrated in Fig. 4-3-1. This mapping is called *Gray encoding*. It is important in the demodulation of the signal because the most likely errors caused by noise involve the erroneous selection of an adjacent amplitude to the transmitted signal amplitude. In such a case, only a single bit error occurs in the k-bit sequence.

We note that the Euclidean distance between any pair of signal points is

$$\begin{aligned} d_{mn}^{(e)} &= \sqrt{(s_m - s_n)^2} \\ &= \sqrt{\tfrac{1}{2}\mathscr{E}_g}\,|A_m - A_n| \\ &= d\sqrt{2\mathscr{E}_g}\,|m - n| \end{aligned} \tag{4-3-7}$$

Hence, the distance between a pair of adjacent signal points, i.e., the minimum Euclidean distance, is

$$d_{\min}^{(e)} = d\sqrt{2\mathscr{E}_g} \tag{4-3-8}$$

The carrier-modulated PAM signal represented by (4-3-1) is a double-sideband (DSB) signal and requires twice the channel bandwidth of the equivalent lowpass signal for transmission. Alternatively, we may use single-sideband (SSB) PAM, which has the representation (lower or upper sideband).

$$s_m(t) = \text{Re}\{A_m[g(t) \pm j\hat{g}(t)]e^{j2\pi f_c t}\}, \quad m = 1, 2, \ldots, M \tag{4-3-9}$$

where $\hat{g}(t)$ is the Hilbert transform of $g(t)$. Thus, the bandwidth of the SSB signal is half that of the DSB signal.

The digital PAM signal is also appropriate for transmission over a channel that does not require carrier modulation. In this case, the signal waveform may be simply represented as

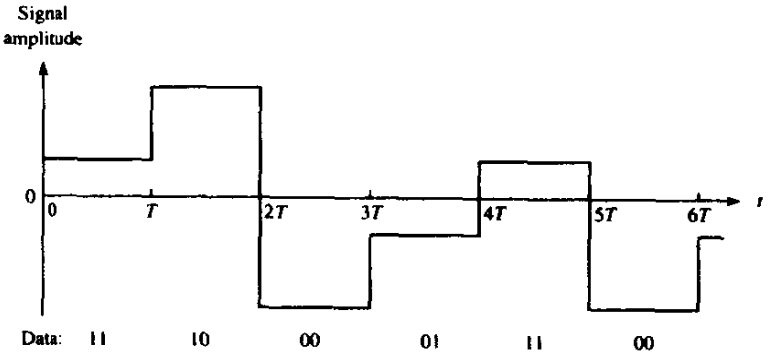
$$s_m(t) = A_m g(t), \quad m = 1, 2, \ldots, M \tag{4-3-10}$$

This is now called a *baseband* signal. For example a four-amplitude level baseband PAM signal is illustrated in Fig. 4-3-2(*a*). The carrier-modulated version of the signal is shown in Fig. 4-3-2(*b*).

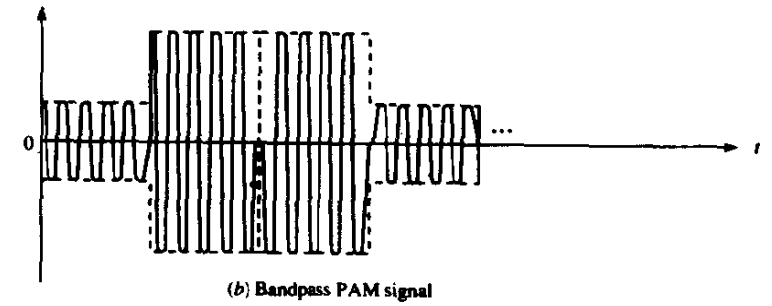
In the special case of $M = 2$ signals, the binary PAM waveforms have the special property that

$$s_1(t) = -s_2(t)$$

FIGURE 4-3-2 Baseband and bandpass PAM signals.



(*a*) Baseband PAM signal



(*b*) Bandpass PAM signal

Hence, these two signals have the same energy and a cross-correlation coefficient of -1. Such signals are called *antipodal.*

Phase-Modulated Signals In digital phase modulation, the M signal waveforms are represented as

$$\begin{aligned} s_m(t) &= \text{Re}\left[g(t)e^{j2\pi(m-1)/M}e^{j2\pi f_c t}\right], \quad m = 1, 2, \ldots, M, \quad 0 \le t \le T \\ &= g(t)\cos\left[2\pi f_c t + \frac{2\pi}{M}(m-1)\right] \\ &= g(t)\cos\frac{2\pi}{M}(m-1)\cos 2\pi f_c t - g(t)\sin\frac{2\pi}{M}(m-1)\sin 2\pi f_c t \end{aligned} \tag{4-3-11}$$

where $g(t)$ is the signal pulse shape and $\theta_m = 2\pi(m-1)/M$, $m = 1, 2, \ldots, M$, are the M possible phases of the carrier that convey the transmitted information. Digital phase modulation is usually called *phase-shift keying* (PSK).

We note that these signal waveforms have equal energy, i.e.,

$$\begin{aligned} \mathscr{E} &= \int_0^T s_m^2(t)\,dt \\ &= \frac{1}{2}\int_0^T g^2(t)\,dt = \tfrac{1}{2}\mathscr{E}_g \end{aligned} \tag{4-3-12}$$

Furthermore, the signal waveforms may be represented as a linear combination of two-orthonormal signal waveforms, $f_1(t)$ and $f_2(t)$, i.e.,

$$s_m(t) = s_{m1}f_1(t) + s_{m2}f_2(t) \tag{4-3-13}$$

where

$$f_1(t) = \sqrt{\frac{2}{\mathscr{E}_g}}\,g(t)\cos 2\pi f_c t \tag{4-3-14}$$

$$f_2(t) = -\sqrt{\frac{2}{\mathscr{E}_g}}\,g(t)\sin 2\pi f_c t \tag{4-3-15}$$

and the two-dimensional vectors $\mathbf{s}_m = [s_{m1}\ \ s_{m2}]$ are given by

$$\mathbf{s}_m = \left[\sqrt{\frac{\mathscr{E}_g}{2}}\cos\frac{2\pi}{M}(m-1) \quad \sqrt{\frac{\mathscr{E}_g}{2}}\sin\frac{2\pi}{M}(m-1)\right], \quad m = 1, 2, \ldots, M \tag{4-3-16}$$

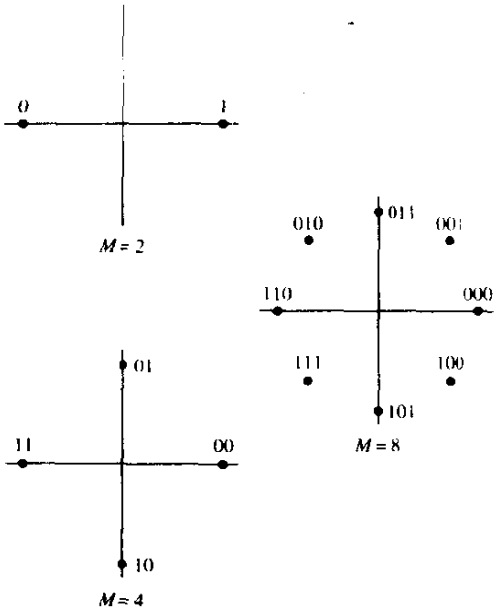


FIGURE 4-3-3 Signal space diagrams for PSK signals.

Signal space diagrams for $M = 2$, 4, and 8 are shown in Fig. 4-3-3. We note that $M = 2$ corresponds to one-dimensional signals, which are identical to binary PAM signals.

As is the case of PAM, the mapping or assignment of k information bits to the $M = 2^k$ possible phases may be done in a number of ways. The preferred assignment is Gray encoding, so that the most likely errors caused by noise will result in a single bit error in the k-bit symbol.

The Euclidean distance between signal points is

$$d_{mn}^{(c)} = |\mathbf{s}_m - \mathbf{s}_n| = \left\{ \mathscr{E}_g \left[1 - \cos \frac{2\pi}{M} (m - n) \right] \right\}^{1/2} \tag{4-3-17}$$

The minimum Euclidean distance corresponds to the case in which $|m - n| = 1$, i.e., adjacent signal phases. In this case,

$$d_{\min}^{(c)} = \sqrt{\mathscr{E}_g \left(1 - \cos \frac{2\pi}{M} \right)} \tag{4-3-18}$$

Quadrature Amplitude Modulation The bandwidth efficiency of PAM/SSB can also be obtained by simultaneously impressing two separate k-bit symbols from the information sequence $\{a_n\}$ on two quadrature carriers

$\cos 2\pi f_c t$ and $\sin 2\pi f_c t$. The resulting modulation technique is called quadrature PAM or QAM, and the corresponding signal waveforms may be expressed as

$$\begin{aligned} s_m(t) &= \text{Re}\,[(A_{mc} + jA_{ms})g(t)e^{j2\pi f_c t}], \quad m = 1, 2, \ldots, M, \quad 0 \leq t \leq T \\ &= A_{mc}g(t)\cos 2\pi f_c t - A_{ms}g(t)\sin 2\pi f_c t \end{aligned} \tag{4-3-19}$$

where A_{mc} and A_{ms} are the information-bearing signal amplitudes of the quadrature carriers and $g(t)$ is the signal pulse.

Alternatively, the QAM signal waveforms may be expressed as

$$\begin{aligned} s_m(t) &= \text{Re}\,[V_m e^{j\theta_m} g(t) e^{j2\pi f_c t}] \\ &= V_m g(t)\cos(2\pi f_c t + \theta_m) \end{aligned} \tag{4-3-20}$$

where $V_m = \sqrt{A_{mc}^2 + A_{ms}^2}$ and $\theta_m = \tan^{-1}(A_{ms}/A_{mc})$. From this expression, it is apparent that the QAM signal waveforms may be viewed as combined amplitude and phase modulation.

In fact, we may select any combination of M_1-level PAM and M_2-phase PSK to construct an $M = M_1 M_2$ combined PAM–PSK signal constellation. If $M_1 = 2^n$ and $M_2 = 2^m$, the combined PAM–PSK signal constellation results in the simultaneous transmission of $m + n = \log M_1 M_2$ binary digits occurring at a symbol rate $R/(m + n)$. Examples of signal space diagrams for combined PAM–PSK are shown in Fig. 4-3-4, for $M = 8$ and $M = 16$.

As in the case of PSK signals, the QAM signal waveforms may be represented as a linear combination of two orthonormal signal waveforms, $f_1(t)$ and $f_2(t)$, i.e.,

$$s_m(t) = s_{m1} f_1(t) + s_{m2} f_2(t) \tag{4-3-21}$$

where

$$\begin{aligned} f_1(t) &= \sqrt{\frac{2}{\mathscr{E}_g}}\, g(t)\cos 2\pi f_c t \\ f_2(t) &= -\sqrt{\frac{2}{\mathscr{E}_g}}\, g(t)\sin 2\pi f_c t \end{aligned} \tag{4-3-22}$$

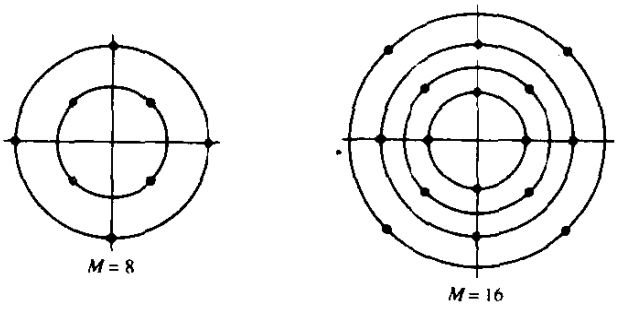


FIGURE 4-3-4 Examples of combined PAM–PSK signal space diagrams.

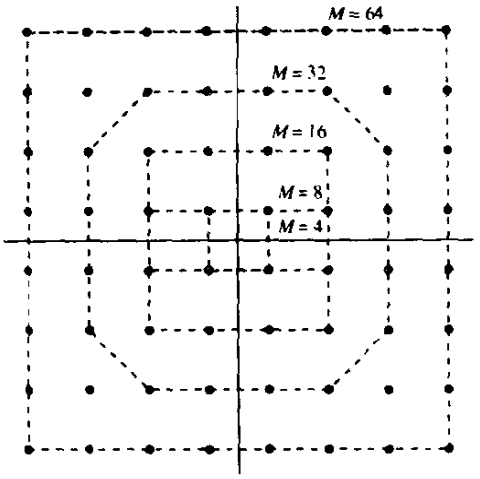


FIGURE 4-3-5 Several signal space diagrams for rectangular QAM.

and

$$\begin{aligned}\mathbf{s}_m &= [s_{m1} \quad s_{m2}] \\ &= [A_{mc}\sqrt{\tfrac{1}{2}\mathscr{E}_g} \quad A_{ms}\sqrt{\tfrac{1}{2}\mathscr{E}_g}]\end{aligned} \tag{4-3-23}$$

$\mathscr{E}_g$ is the energy of the signal pulse $g(t)$.

The Euclidean distance between any pair of signal vectors is

$$\begin{aligned}d_{mn}^{(e)} &= |\mathbf{s}_m - \mathbf{s}_n| \\ &= \sqrt{\tfrac{1}{2}\mathscr{E}_g[(A_{mc} - A_{nc})^2 + (A_{ms} - A_{ns})^2]}\end{aligned} \tag{4-3-24}$$

In the special case where the signal amplitudes takes the set of discrete values $\{(2m - 1 - M)d,\ m = 1, 2, \ldots, M\}$, the signal space diagram is rectangular, as shown in Fig. 4-3-5. In this case, the Euclidean distance between adjacent points, i.e., the minimum distance, is

$$d_{\min}^{(e)} = d\sqrt{2\mathscr{E}_g} \tag{4-3-25}$$

which is the same result as for PAM.

Multidimensional Signals It is apparent from the discussion above that the digital modulation of the carrier amplitude and phase allows us to construct signal waveforms that correspond to two-dimensional vectors and signal space diagrams. If we wish to construct signal waveforms corresponding to higher-dimensional vectors, we may use either the time domain or the frequency domain or both in order to increase the number of dimensions.

Suppose we have N-dimensional signal vectors. For any N, we may subdivide a time interval of length $T_1 = NT$ into N subintervals of length $T = T_1/N$. In each subinterval of length T, we may use binary PAM (a one-dimensional signal) to transmit an element of the N-dimensional signal

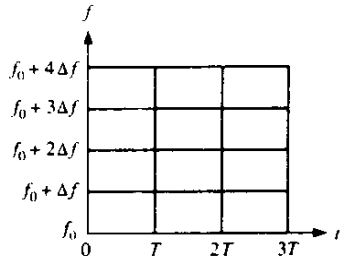


FIGURE 4-3-6 Subdivision of time and frequency axes into distinct slots.

vector. Thus, the N time slots are used to transmit the N-dimensional signal vector. If N is even, a time slot of length T may be used to simultaneously transmit two components of the N-dimensional vector by modulating the amplitude of quadrature carriers independently by the corresponding components. In this manner, the N-dimensional signal vector is transmitted in $\frac{1}{2}NT$ seconds ($\frac{1}{2}N$ time slots).

Alternatively, a frequency band of width $N\,\Delta f$ may be subdivided into N frequency slots each of width Δf. An N-dimensional signal vector can be transmitted over the channel by simultaneously modulating the amplitude of N carriers, one in each of the N frequency slots. Care must be taken to provide sufficient frequency separation Δf between successive carriers so that there is no cross talk interference among the signals on the N carriers. If quadrature carriers are used in each frequency slot, the N-dimensional vector (even N) may be transmitted in $\frac{1}{2}N$ frequency slots, thus reducing the channel bandwidth utilization by a factor of 2.

More generally, we may use both the time and frequency domains jointly to transmit an N-dimensional signal vector. For example, Fig. 4-3-6 illustrates a subdivision of the time and frequency axes into 12 slots. Thus, an $N = 12$-dimensional signal vector may be transmitted by PAM or an $N = 24$-dimensional signal vector may be transmitted by use of two quadrature carriers (QAM) in each slot.

Orthogonal Multidimensional Signals As a special case of the construction of multidimensional signals, let us consider the construction of M equal-energy orthogonal signal waveforms that differ in frequency, and are represented as

$$\begin{aligned} s_m(t) &= \text{Re}\,[s_{lm}(t)e^{j2\pi f_c t}], \quad m = 1, 2, \ldots, M, \quad 0 \leq t \leq T \\ &= \sqrt{\frac{2\mathscr{E}}{T}} \cos\,[2\pi f_c t + 2\pi m\,\Delta f\, t] \end{aligned} \tag{4-3-26}$$

where the equivalent lowpass signal waveforms are defined as

$$s_{lm}(t) = \sqrt{\frac{2\mathscr{E}}{T}} e^{j2\pi m\,\Delta f t}, \quad m = 1, 2, \ldots, M, \quad 0 \leq t \leq T \tag{4-3-27}$$

This type of frequency modulation is called *frequency-shift keying* (FSK).

These waveforms are characterized as having equal energy and cross-correlation coefficients

$$\rho_{km} = \frac{2\mathscr{E}/T}{2\mathscr{E}} \int_0^T e^{j2\pi(m-k)\Delta f t}\, dt$$

$$= \frac{\sin \pi T(m-k)\,\Delta f}{\pi T(m-k)\,\Delta f} e^{j\pi T(m-k)\,\Delta f} \tag{4-3-28}$$

The real part of ρ_{km} is

$$\rho_r \equiv \text{Re}\,(\rho_{km}) = \frac{\sin\,[\pi T(m-k)\,\Delta f]}{\pi T(m-k)\,\Delta f} \cos\,[\pi T(m-k)\,\Delta f]$$

$$= \frac{\sin\,[2\pi T(m-k)\,\Delta f]}{2\pi T(m-k)\,\Delta f} \tag{4-3-29}$$

First, we observe that $\text{Re}\,(\rho_{km}) = 0$ when $\Delta f = 1/2T$ and $m \neq k$. Since $|m-k| = 1$ corresponds to adjacent frequency slots, $\Delta f = 1/2T$ represents the minimum frequency separation between adjacent signals for orthogonality of the M signals. Plots of $\text{Re}\,(\rho_{km})$ versus Δf and $|\rho_{km}|$ versus Δf are shown in Fig. 4-3-7. Note that $|\rho_{km}| = 0$ for multiples of $1/T$ whereas $\text{Re}\,(\rho_{km}) = 0$ for multiples of $1/2T$.

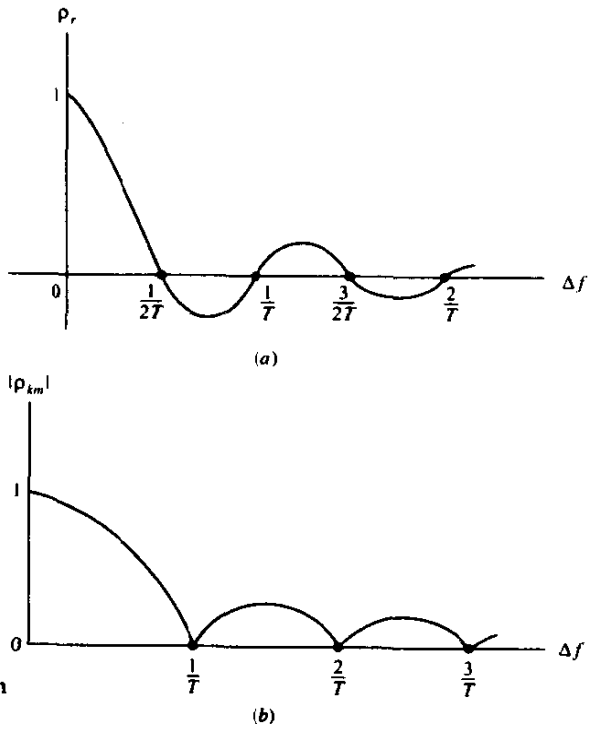


FIGURE 4-3-7 Cross-correlation coefficient as a function of frequency separation for FSK signals.

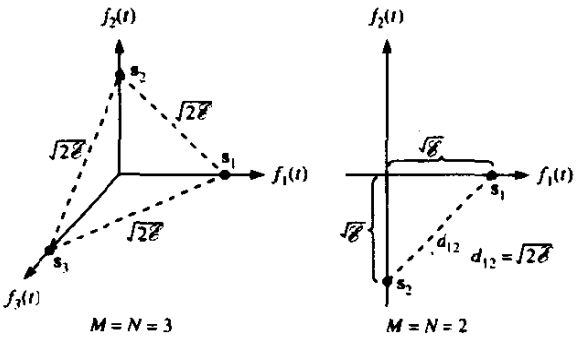


FIGURE 4-3-8 Orthogonal signals for $M = N = 3$ and $M = N = 2$.

For the case in which $\Delta f = 1/2T$, the M FSK signals are equivalent to the N-dimensional vectors

$$\begin{aligned} \mathbf{s}_1 &= [\sqrt{\mathscr{E}} \quad 0 \quad 0 \quad \ldots \quad 0 \quad 0] \\ \mathbf{s}_2 &= [0 \quad \sqrt{\mathscr{E}} \quad 0 \quad \ldots \quad 0 \quad 0] \\ &\vdots \quad \vdots \\ \mathbf{s}_N &= [0 \quad 0 \quad 0 \quad \ldots \quad 0, \quad \sqrt{\mathscr{E}}] \end{aligned} \tag{4-3-30}$$

where $N = M$. The distance between pairs of signals is

$$d_{km}^{(e)} = \sqrt{2\mathscr{E}} \quad \text{for all } m, k \tag{4-3-31}$$

which is also the minimum distance. Figure 4-3-8 illustrates the signal space diagram for $M = N = 2$ and $M = N = 3$.

Biorthogonal Signals A set of M biorthogonal signals can be constructed from $\frac{1}{2}M$ orthogonal signals by simply including the negatives of the orthogonal signals. Thus, we require $N = \frac{1}{2}M$ dimensions for the construction of a set of M biorthogonal signals. Figure 4-3-9 illustrates the biorthogonal signals for $M = 4$ and 6.

We note that the correlation between any pair of waveforms is either $\rho_r = -1$ or 0. The corresponding distances are $d = 2\sqrt{\mathscr{E}}$ or $\sqrt{2\mathscr{E}}$, with the latter being the minimum distance.

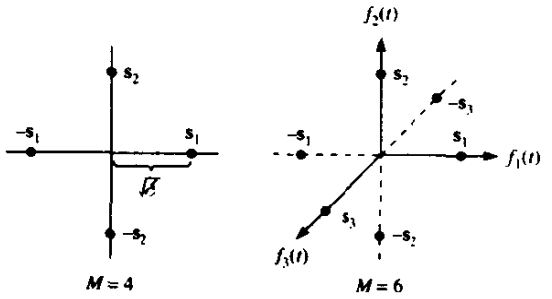


FIGURE 4-3-9 Signal space diagrams for $M = 4$ and $M = 6$ biorthogonal signals.

Simplex Signals Suppose we have a set of M orthogonal waveforms $\{s_m(t)\}$ or, equivalently, their vector representation $\{\mathbf{s}_m\}$. Their mean is

$$\bar{\mathbf{s}} = \frac{1}{M}\sum_{m=1}^{M} \mathbf{s}_m \tag{4-3-32}$$

Now, let us construct another set of M signals by subtracting the mean from each of the M orthogonal signals. Thus,

$$\mathbf{s}'_m = \mathbf{s}_m - \bar{\mathbf{s}}, \quad m = 1, 2, \ldots, M \tag{4-3-33}$$

The effect of the subtraction is to translate the origin of the m orthogonal signals to the point $\bar{\mathbf{s}}$.

The resulting signal waveforms are called *simplex signals* and have the following properties. First, the energy per waveform is

$$\begin{aligned} |\mathbf{s}'_m|^2 &= |\mathbf{s}_m - \bar{\mathbf{s}}|^2 \\ &= \mathscr{E} - \frac{2}{M}\mathscr{E} + \frac{1}{M}\mathscr{E} \\ &= \mathscr{E}\left(1 - \frac{1}{M}\right) \end{aligned} \tag{4-3-34}$$

Second, the cross-correlation of any pair of signals is

$$\begin{aligned} \mathrm{Re}\,(\rho_{mn}) &= \frac{\mathbf{s}'_m \cdot \mathbf{s}'_n}{|\mathbf{s}'_m|\,|\mathbf{s}'_n|} \\ &= \frac{-1/M}{1 - 1/M} = -\frac{1}{M-1} \end{aligned} \tag{4-3-35}$$

for all m, n. Hence, the set of simplex waveforms is *equally correlated* and requires less energy, by the factor $1 - 1/M$, than the set of orthogonal waveforms. Since only the origin was translated, the distance between any pair of signal points is maintained at $d = \sqrt{2\mathscr{E}}$, which is the same as the distance between any pair of orthogonal signals.

Figure 4-3-10 illustrates the simplex signals for $M = 2, 3$, and 4. Note that the signal dimensionality is $N = M - 1$.

Signal Waveforms from Binary Codes A set of M signaling waveforms can be generated from a set of M binary code words of the form

$$\mathbf{C}_m = [c_{m1} \quad c_{m2} \quad \ldots \quad c_{mN}], \quad m = 1, 2, \ldots, M \tag{4-3-36}$$

where $c_{mj} = 0$ or 1 for all m and j. Each component of a code word is mapped into an elementary binary PSK waveform as follows:

$$\begin{aligned} c_{mj} = 1 &\Rightarrow s_{mj}(t) = \sqrt{\frac{2\mathscr{E}_c}{T_c}}\cos 2\pi f_c t \quad (0 \le t \le T_c) \\ c_{mj} = 0 &\Rightarrow s_{mj}(t) = -\sqrt{\frac{2\mathscr{E}_c}{T_c}}\cos 2\pi f_c t \quad (0 \le t \le T_c) \end{aligned} \tag{4-3-37}$$

where $T_c = T/N$ and $\mathscr{E}_c = \mathscr{E}/N$. Thus, the M code words $\{C_m\}$ are mapped into a set of M waveforms $\{s_m(t)\}$.

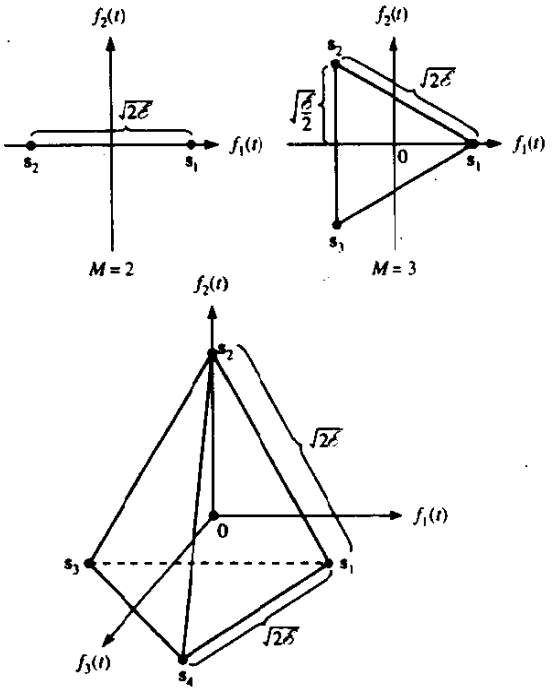


FIGURE 4-3-10 Signal space diagrams for M-ary simplex signals.

The waveforms can be represented in vector form as

$$\mathbf{s}_m = [s_{m1} \quad s_{m2} \quad \cdots \quad s_{mN}], \qquad m = 1, 2, \ldots, M \tag{4-3-38}$$

where $s_{mj} = \pm\sqrt{\mathscr{E}/N}$ for all m and j. N is called the block length of the code, and it is also the dimension of the M waveforms.

We note that there are 2^N possible waveforms that can be constructed from the 2^N possible binary code words. We may select a subset of $M < 2^N$ signal waveforms for transmission of the information. We also observe that the 2^N possible signal points correspond to the vertices of an N-dimensional hypercube with its center at the origin. Figure 4-3-11 illustrates the signal points in $N = 2$ and 3 dimensions.

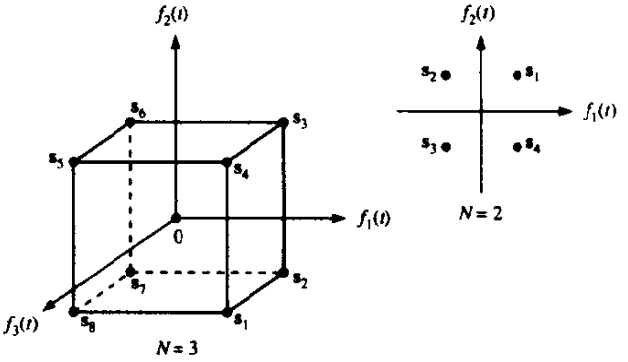


FIGURE 4-3-11 Signal space diagrams for signals generated from binary codes.

Each of the M waveforms has energy $\mathscr{E}$. The cross-correlation between any pair of waveforms depends on how we select the M waveforms from the 2^N possible waveforms. This topic is treated in Chapter 7. Clearly, any adjacent signal points have a cross-correlation coefficient

$$\rho_r = \frac{\mathscr{E}(1 - 2/N)}{\mathscr{E}} = \frac{N-2}{N} \tag{4-3-39}$$

and a corresponding distance of

$$\begin{aligned} d^{(e)} &= \sqrt{2\mathscr{E}(1-\rho_r)} \\ &= \sqrt{4\mathscr{E}/N} \end{aligned} \tag{4-3-40}$$

This concludes our discussion of memoryless modulation signals.

4-3-2 Linear Modulation with Memory

The modulation signals introduced in the previous section were classified as memoryless, because there was no dependence between signals transmitted in non-overlapping symbol intervals. In this section, we present some modulation signals in which there is dependence between the signals transmitted in successive symbol intervals. This signal dependence is usually introduced for the purpose of shaping the spectrum of the transmitted signal so that it matches the spectral characteristics of the channel. Signal dependence between signals transmitted in different signal intervals is generally accomplished by encoding the data sequence at the input to the modulator by means of a *modulation code,* as described in Chapter 9.

In this section, we shall present examples of modulation signals with memory and characterize their memory in terms of Markov chains. We shall confine our treatment to baseband signals. The generalization to bandpass signals is relatively straightforward.

Figure 4-3-12 illustrates three different baseband signals and the corresponding data sequence. The first signal, called NRZ, is the simplest. The binary information digit 1 is represented by a rectangular pulse of polarity A and the binary digit zero is represented by a rectangular pulse of polarity $-A$.

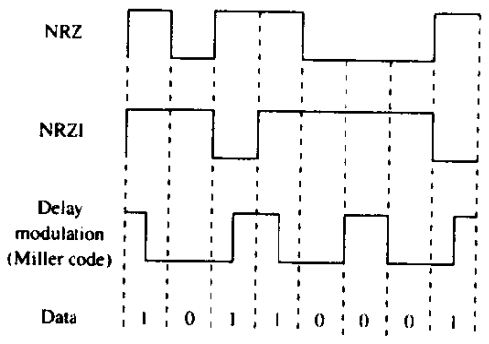


FIGURE 4-3-12 Baseband signals.

Hence, the NRZ modulation is memoryless and is equivalent to a binary PAM or a binary PSK signal in a carrier-modulated system.

The NRZI signal is different from the NRZ signal in that transitions from one amplitude level to another occur only when a 1 is transmitted. The amplitude level remains unchanged when a zero is transmitted. This type of signal encoding is called *differential encoding*. The encoding operation is described mathematically by the relation

$$b_k = a_k \oplus b_{k-1} \tag{4-3-41}$$

where $\{a_k\}$ is the binary information sequence into the encoder, $\{b_k\}$ is the output sequence of the encoder, and $\oplus$ denotes addition modulo 2. When $b_k = 1$, the transmitted waveform is a rectangular pulse of amplitude A, and when $b_k = 0$, the transmitted waveform is a rectangular pulse of amplitude $-A$. Hence, the output of the encoder is mapped into one of two waveforms in exactly the same manner as for the NRZ signal.

The differential encoding operation introduces memory in the signal. The combination of the encoder and the modulator operations may be represented by a state diagram (a Markov chain) as shown in Fig. 4-3-13. The state diagram may be described by two transition matrices corresponding to the two possible input bits {0, 1}. We note that when $a_k = 0$, the encoder stays in the same state. Hence, the state transition matrix for a zero is simply

$$\mathbf{T}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{4-3-42}$$

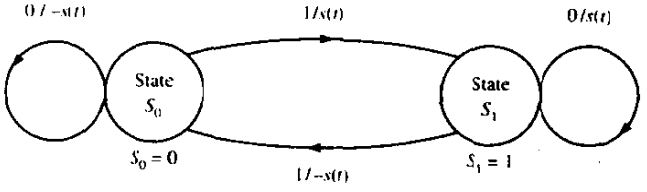
where $t_{ij} = 1$ if a_k results in a transition from state i to state j, $i = 1, 2$, and $j = 1, 2$; otherwise, $t_{ij} = 0$. Similarly, the state transition matrix for $a_k = 1$ is

$$\mathbf{T}_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{4-3-43}$$

Thus, these two state transition matrices characterize the NRZI signal.

Another way to display the memory introduced by the precoding operation is by means of a trellis diagram. The trellis diagram for the NRZI signal is

FIGURE 4-3-13 *State diagram for the NRZI signal.*



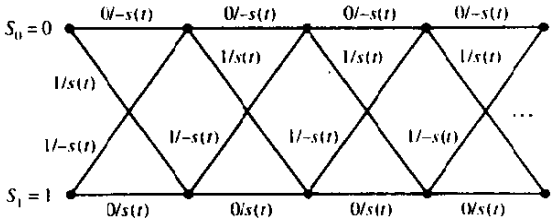


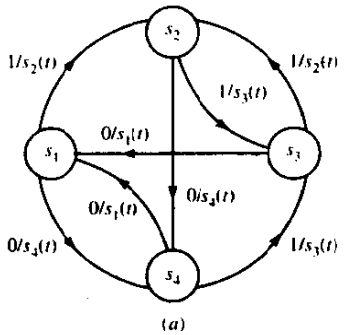
FIGURE 4-3-14 The trellis diagram for the NRZI signal.

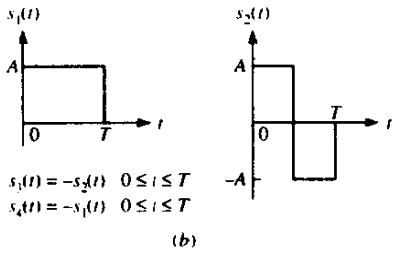
illustrated in Fig. 4-3-14. The trellis provides exactly the same information concerning the signal dependence as the state diagram, but also depicts a time evolution of the state transitions.

The signal generated by delay modulation also has memory. As shown in Chapter 9, delay modulation is equivalent to encoding the data sequence by a run-length-limited code called a *Miller code* and using NRZI to transmit the encoded data. This type of digital modulation has been used extensively for digital magnetic recording and in carrier modulation systems employing binary PSK. The signal may be described by a state diagram that has four states as shown in Fig. 4-3-15(*a*). There are two elementary waveforms $s_1(t)$ and $s_2(t)$ and their negatives $-s_1(t)$ and $-s_2(t)$, which are used for transmitting the binary information. These waveforms are illustrated in Fig. 4-3-15(*b*). The mapping from bits to corresponding waveforms is illustrated in the state diagram. The state transition matrices that characterize the memory of this encoding and modulation method are easily obtained from the state diagram in Fig. 4-3-15. When $a_k = 0$, we have

$$\mathbf{T}_1 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \tag{4-3-44}$$

FIGURE 4-3-15 State diagram (*a*) and basic waveforms (*b*) for delay modulated (Miller-encoded) signal.





and when $a_k = 1$, the transition matrix is

$$\mathbf{T}_2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{4-3-45}$$

Thus, these two 4×4 state transition matrices characterize the state diagram for the Miller-encoded signal.

Modulation techniques with memory such as NRZI and Miller coding are generally characterized by a K-state Markov chain with *stationary state probabilities* $\{p_i, i = 1, 2, \ldots, K\}$ and *transition probabilities* $\{p_{ij}, i, j = 1, 2, \ldots, K\}$. Associated with each transition is a signal waveform $s_j(t)$, $j = 1, 2, \ldots, K$. Thus, the transition probability p_{ij} denotes the probability that signal waveform $s_j(t)$ is transmitted in a given signaling interval after the transmission of the signal waveform $s_i(t)$ in the previous signaling interval. The transition probabilities may be arranged in matrix form as

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1K} \\ p_{21} & p_{22} & \cdots & p_{2K} \\ \vdots & \vdots & & \vdots \\ p_{K1} & p_{K2} & \cdots & p_{KK} \end{bmatrix} \tag{4-3-46}$$

where $\mathbf{P}$ is called the *transition probability matrix.*

The transition probability matrix is easily obtained from the transition matrices $\{\mathbf{T}_i\}$ and the corresponding probabilities of occurrence of the input bits (or, equivalently, the stationary state transition probabilities $\{p_i\}$). The general relationship may be expressed as

$$\mathbf{P} = \sum_{i=1}^{2} q_i \mathbf{T}_i \tag{4-3-47}$$

where $q_1 = P(a_k = 0)$ and $q_2 = P(a_k = 1)$.

For the NRZI signal with equal state probabilities $p_1 = p_2 = \frac{1}{2}$ and transition matrices given by (4-3-42) and (4-3-43), the transition probability matrix is

$$\mathbf{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \tag{4-3-48}$$

Similarly, the transition probability matrix for the Miller-coded signal with equally likely symbols ($q_1 = q_2 = \frac{1}{2}$ or, equivalently, $p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$) is

$$\mathbf{P} = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix} \tag{4-3-49}$$

The transition probability matrix is useful in the determination of the spectral

characteristics of digital modulation techniques with memory, as we shall observe in Section 4-4.

4-3-3 Nonlinear Modulation Methods with Memory

In this section, we consider a class of digital modulation methods in which the phase of the signal is constrained to be continuous. This constraint results in a phase or frequency modulator that has memory. The modulation method is also nonlinear.

Continuous-Phase FSK (CPFSK) A conventional FSK signal is generated by shifting the carrier by an amount $f_n = \frac{1}{2}\Delta f I_n$, $I_n = \pm 1, \pm 3, \ldots, \pm(M-1)$, to reflect the digital information that is being transmitted. This type of FSK signal was described in Section 4-3-1, and it is memoryless. The switching from one frequency to another may be accomplished by having $M = 2^k$ separate oscillators tuned to the desired frequencies and selecting one of the M frequencies according to the particular k-bit symbol that is to be transmitted in a signal interval of duration $T = k/R$ seconds. However, such abrupt switching from one oscillator output to another in successive signaling intervals results in relatively large spectral side lobes outside of the main spectral band of the signal and, consequently, this method requires a large frequency band for transmission of the signal.

To avoid the use of signals having large spectral side lobes, the information-bearing signal frequency modulates a single carrier whose frequency is changed continuously. The resulting frequency-modulated signal is phase-continuous and, hence, it is called *continuous-phase* FSK (CPFSK). This type of FSK signal has memory because the phase of the carrier is constrained to be continuous.

In order to represent a CPFSK signal, we begin with a PAM signal

$$d(t) = \sum_n I_n g(t - nT) \tag{4-3-50}$$

where $\{I_n\}$ denotes the sequence of amplitudes obtained by mapping k-bit blocks of binary digits from the information sequence $\{a_n\}$ into the amplitude levels $\pm 1, \pm 3, \ldots, \pm(M-1)$ and $g(t)$ is a rectangular pulse of amplitude $1/2T$ and duration T seconds. The signal $d(t)$ is used to frequency-modulate the carrier. Consequently, the equivalent lowpass waveform $v(t)$ is expressed as

$$v(t) = \sqrt{\frac{2\mathcal{E}}{T}} \exp\left\{j\left[4\pi T f_d \int_{-\infty}^{t} d(\tau)\, d\tau + \phi_0\right]\right\} \tag{4-3-51}$$

where f_d is the *peak frequency deviation* and ϕ_0 is the initial phase of the carrier.

The carrier-modulated signal corresponding to (4-3-51) may be expressed as

$$s(t) = \sqrt{\frac{2\mathcal{E}}{T}} \cos\left[2\pi f_c t + \phi(t; \mathbf{I}) + \phi_0\right] \tag{4-3-52}$$

where $\phi(t; \mathbf{I})$ represents the time-varying phase of the carrier, which is defined as

$$\phi(t; \mathbf{I}) = 4\pi T f_d \int_{-\infty}^{t} d(\tau)\, d\tau$$
$$= 4\pi T f_d \int_{-\infty}^{t} \left[\sum_n I_n g(\tau - nT)\right] d\tau \tag{4-3-53}$$

Note that, although $d(t)$ contains discontinuities, the integral of $d(t)$ is continuous. Hence, we have a continuous-phase signal. The phase of the carrier in the interval $nT \leq t \leq (n+1)T$ is determined by integrating (4-3-53). Thus,

$$\phi(t; \mathbf{I}) = 2\pi f_d T \sum_{k=-\infty}^{n-1} I_k + 2\pi f_d (t - nT) I_n$$
$$= \theta_n + 2\pi h I_n q(t - nT) \tag{4-3-54}$$

where h, θ_n, and $q(t)$ are defined as

$$h = 2f_d T \tag{4-3-55}$$

$$\theta_n = \pi h \sum_{k=-\infty}^{n-1} I_k \tag{4-3-56}$$

$$q(t) = \begin{cases} 0 & (t < 0) \\ t/2T & (0 \leq t \leq T) \\ \frac{1}{2} & (t > T) \end{cases} \tag{4-3-57}$$

We observe that θ_n represents the accumulation (memory) of all symbols up to time $(n-1)T$. The parameter h is called the *modulation index*.

Continuous-Phase Modulation (CPM) When expressed in the form of (4-3-54), CPFSK becomes a special case of a general class of continuous-phase modulated (CPM) signals in which the carrier phase is

$$\phi(t; \mathbf{I}) = 2\pi \sum_{k=-\infty}^{n} I_k h_k q(t - kT), \quad nT \leq t \leq (n+1)T \tag{4-3-58}$$

where $\{I_k\}$ is the sequence of M-ary information symbols selected from the alphabet $\pm 1, \pm 3, \ldots, \pm(M-1)$, $\{h_k\}$ is a sequence of modulation indices, and $q(t)$ is some normalized waveform shape.

When $h_k = h$ for all k, the modulation index is fixed for all symbols. When the modulation index varies from one symbol to another, the CPM signal is called *multi-h*. In such a case, the $\{h_k\}$ are made to vary in a cyclic manner through a set of indices.

The waveform $q(t)$ may be represented in general as the integral of some pulse $g(t)$, i.e.,

$$q(t) = \int_0^t g(\tau)\, d\tau \tag{4-3-59}$$

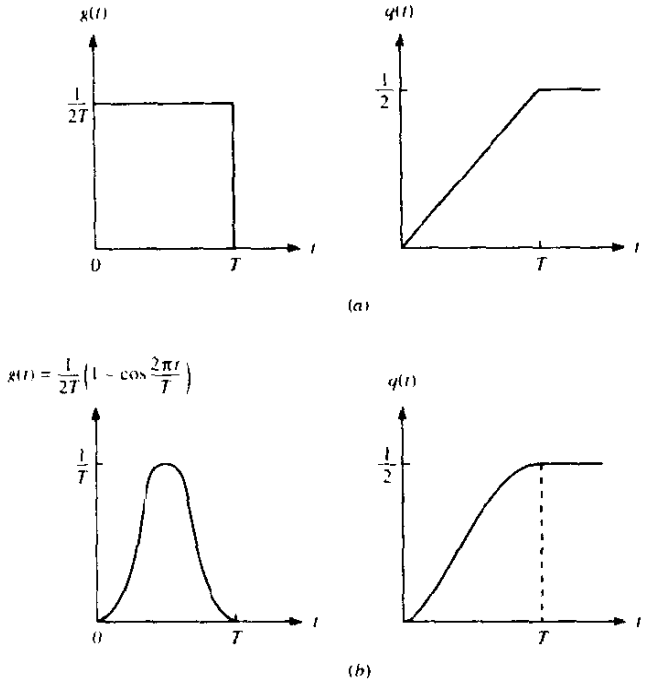


FIGURE 4-3-16 Pulse shapes for full response CPM (*a*, *b*) and partial response CPM (*c*, *d*).

If $g(t) = 0$ for $t > T$, the CPM signal is called *full response CPM*. If $g(t) \neq 0$ for $t > T$, the modulated signal is called *partial response CPM*. Figure 4-3-16 illustrates several pulse shapes for $g(t)$, and the corresponding $q(t)$. It is apparent that an infinite variety of CPM signals can be generated by choosing different pulse shapes $g(t)$ and by varying the modulation index h and the alphabet size M.

It is instructive to sketch the set of phase trajectories $\phi(t; \mathbf{I})$ generated by all possible values of the information sequence $\{I_n\}$. For example, in the case of CPFSK with binary symbols $I_n = \pm 1$, the set of phase trajectories beginning at time $t = 0$ is shown in Fig. 4-3-17. For comparison, the phase trajectories for quaternary CPFSK are illustrated in Fig. 4-3-18. These phase diagrams are called *phase trees*. We observe that the phase trees for CPFSK are piecewise linear as a consequence of the fact that the pulse $g(t)$ is rectangular. Smoother phase trajectories and phase trees are obtained by using pulses that do not contain discontinuities, such as the class of raised cosine pulses. For example, a phase trajectory generated by the sequence $(1, -1, -1, -1, 1, 1, -1, 1)$ for a partial response, raised cosine pulse of length $3T$ is illustrated in Fig. 4-3-19. For comparison, the corresponding phase trajectory generated by CPFSK is also shown.

The phase trees shown in these figures grow with time. However, the phase

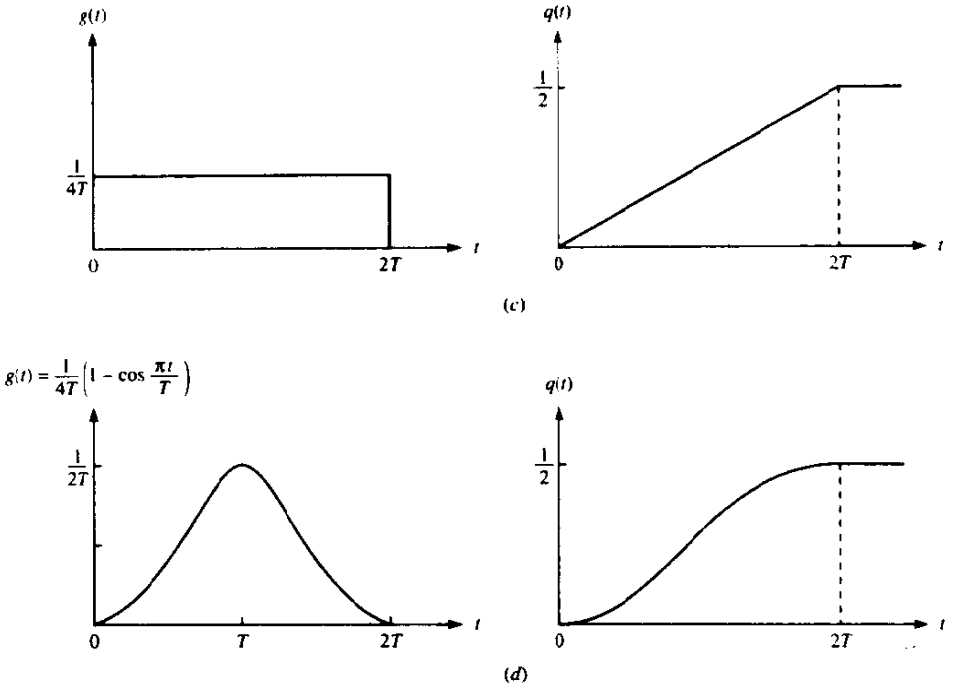


FIGURE 4-3-16 (*Continued*).

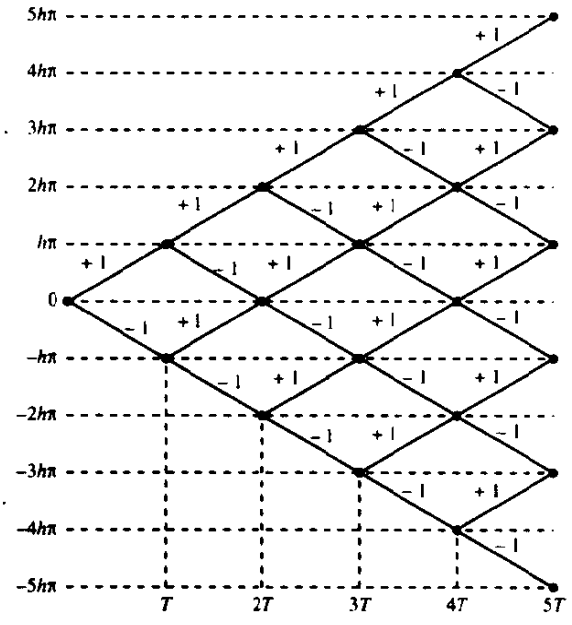


FIGURE 4-3-17 Phase trajectory for binary CPFSK.

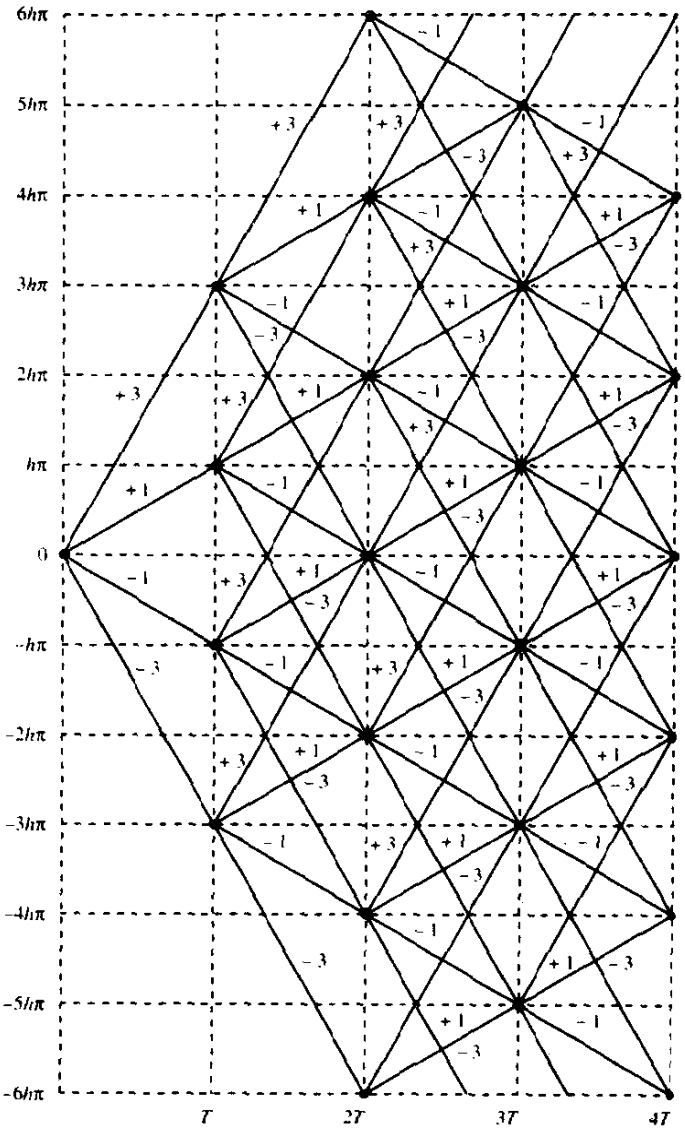


FIGURE 4-3-18 Phase trajectory for quaternary CPFSK.

of the carrier is unique only in the range from $\phi = 0$ to $\phi = 2\pi$ or, equivalently, from $\phi = -\pi$ to $\phi = \pi$. When the phase trajectories are plotted modulo 2π, say in the range $(-\pi, \pi)$, the phase tree collapses into a structure called a *phase trellis*. To properly view the phase trellis diagram, we may plot the two quadrature components $x_c(t; \mathbf{I}) = \cos \phi(t; \mathbf{I})$ and $x_s(t; \mathbf{I}) = \sin \phi(t; \mathbf{I})$ as functions of time. Thus, we generate a three-dimensional plot in which the quadrature components x_c and x_s appear on the surface of a cylinder of unit radius. For example, Fig. 4-3-20 illustrates the phase trellis or phase cylinder

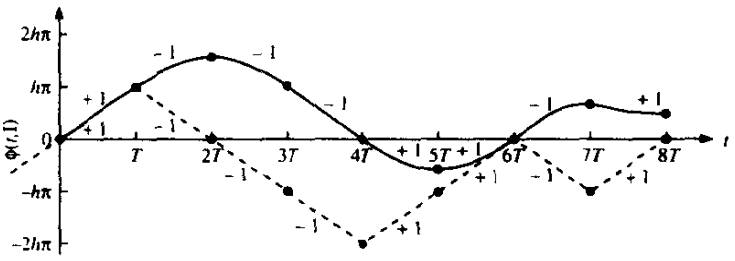


FIGURE 4-3-19 Phase trajectories for binary CPFSK (dashed) and binary, partial response CPM based on raised cosine pulse of length $3T$ (solid). [*From Sundberg (1986), © 1986 IEEE.*]

obtained with binary modulation, a modulation index $h = \frac{1}{2}$, and a raised cosine pulse of length $3T$.

Simpler representations for the phase trajectories can be obtained by displaying only the terminal values of the signal phase at the time instants $t = nT$. In this case, we restrict the modulation index of the CPM signal to be rational. In particular, let us assume that $h = m/p$, where m and p are relatively prime integers. Then, a full response CPM signal at the time instants $t = nT$ will have the terminal phase states

$$\Theta_s = \left\{0, \frac{\pi m}{p}, \frac{2\pi m}{p}, \ldots, \frac{(p-1)\pi m}{p}\right\} \tag{4-3-60}$$

when m is even and

$$\Theta_s = \left\{0, \frac{\pi m}{p}, \frac{2\pi m}{p}, \ldots, \frac{(2p-1)\pi m}{p}\right\} \tag{4-3-61}$$

when m is odd. Hence, there are p terminal phase states when m is even and $2p$ states when m is odd. On the other hand, when the pulse shape extends

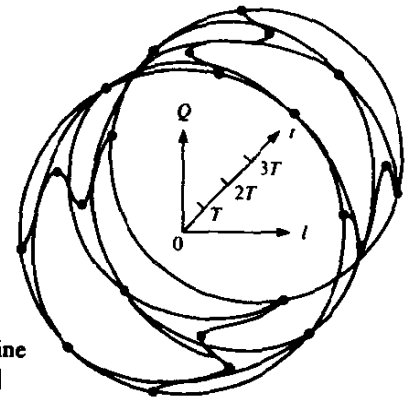


FIGURE 4-3-20 Phase cylinder for binary CPM with $h = \frac{1}{2}$ and a raised cosine pulse of length $3T$. [*From Sundberg (1986), © 1986 IEEE.*]

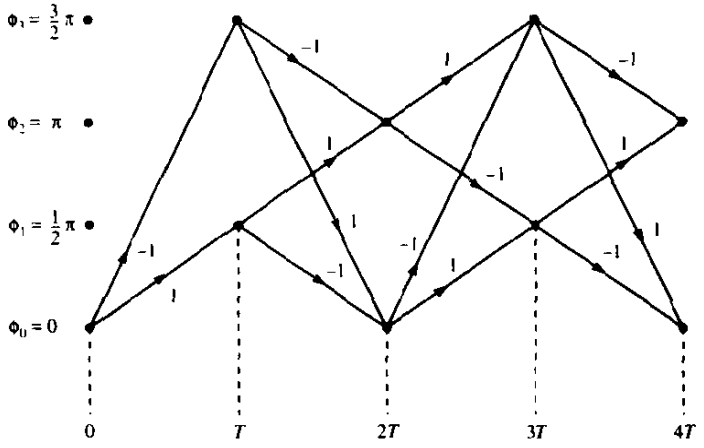


FIGURE 4-3-21 State trellis for binary CPFSK with $h = \frac{1}{2}$.

over L symbol intervals (partial response CPM), the number of phase states may increase up to a maximum of S_t, where

$$S_t = \begin{cases} pM^{L-1} & (\text{even } m) \\ 2pM^{L-1} & (\text{odd } m) \end{cases} \tag{4-3-62}$$

where M is the alphabet size. For example, the binary CPFSK signal (full response, rectangular pulse) with $h = \frac{1}{2}$, has $S_t = 4$ (terminal) phase states. The *state trellis* for this signal is illustrated in Fig. 4-3-21. We emphasize that the phase transitions from one state to another are not true phase trajectories. They represent phase transitions for the (terminal) states at the time instants $t = nT$.

An alternative representation to the state trellis is the state diagram, which also illustrates the state transitions at the time instants $t = nT$. This is an even more compact representation of the CPM signal characteristics. Only the possible (terminal) phase states and their transitions are displayed in the state diagram. Time does not appear explicitly as a variable. For example, the state diagram for the CPFSK signal with $h = \frac{1}{2}$ is shown in Fig. 4-3-22.

Minimum-Shift Keying (MSK) MSK is a special form of binary CPFSK

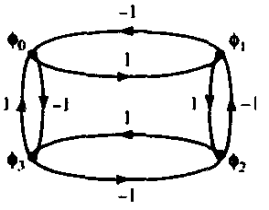


FIGURE 4-3-22 State diagram for binary CPFSK with $h = \frac{1}{2}$.

(and, therefore, CPM) in which the modulation index $h = \frac{1}{2}$. The phase of the carrier in the interval $nT \leq t \leq (n+1)T$ is

$$\begin{aligned}\phi(t; \mathbf{I}) &= \tfrac{1}{2}\pi \sum_{k=-\infty}^{n-1} I_k + \pi I_n q(t - nT) \\ &= \theta_n + \tfrac{1}{2}\pi I_n\left(\frac{t - nT}{T}\right), \quad nT \leq t \leq (n+1)T\end{aligned} \tag{4-3-63}$$

and the modulated carrier signal is

$$\begin{aligned}s(t) &= A\cos\left[2\pi f_c t + \theta_n + \tfrac{1}{2}\pi I_n\left(\frac{t - nT}{T}\right)\right] \\ &= A\cos\left[2\pi\left(f_c + \frac{1}{4T}I_n\right)t - \tfrac{1}{2}n\pi I_n + \theta_n\right], \quad nT \leq t \leq (n+1)T\end{aligned} \tag{4-3-64}$$

The expression (4-3-64) indicates that the binary CPFSK signal can be expressed as a sinusoid having one of two possible frequencies in the interval $nT \leq t \leq (n+1)T$. If we define these frequencies as

$$\begin{aligned}f_1 &= f_c - \frac{1}{4T} \\ f_2 &= f_c + \frac{1}{4T}\end{aligned} \tag{4-3-65}$$

then the binary CPFSK signal given by (4-3-64) may be written in the form

$$s_i(t) = A\cos[2\pi f_i t + \theta_n + \tfrac{1}{2}n\pi(-1)^{i-1}], \quad i = 1, 2 \tag{4-3-66}$$

The frequency separation $\Delta f = f_2 - f_1 = 1/2T$. Recall that $\Delta f = 1/2T$ is the minimum frequency separation that is necessary to ensure the orthogonality of the signals $s_1(t)$ and $s_2(t)$ over a signaling interval of length T. This explains why binary CPFSK with $h = \frac{1}{2}$ is called minimum-shift keying (MSK). The phase in the nth signaling interval is the phase state of the signal that results in phase continuity between adjacent intervals.

MSK may also be represented as a form of four-phase PSK. Specifically, we may express the equivalent lowpass digitally modulated signal in the form (see Problem 4-14)

$$v(t) = \sum_{n=-\infty}^{\infty} [I_{2n}g(t - 2nT) - jI_{2n+1}g(t - 2nT - T)] \tag{4-3-67}$$

where $g(t)$ is a sinusoidal pulse defined as

$$g(t) = \begin{cases} \sin\dfrac{\pi t}{2T} & (0 \leq t \leq 2T) \\ 0 & (\text{otherwise}) \end{cases} \tag{4-3-68}$$

Thus, this type of signal is viewed as a four-phase PSK signal in which the pulse shape is one-half cycle of a sinusoid. The even-numbered binary-valued (± 1) symbols $\{I_{2n}\}$ of the information sequence $\{I_n\}$ are transmitted via the cosine of the carrier, while the odd-numbered symbols $\{I_{2n+1}\}$ are transmitted via the sine of the carrier. The transmission rate on the two orthogonal carrier components is $1/2T$ bits per second so that the combined transmission rate is $1/T$ bits/s. Note that the bit transitions on the sine and cosine carrier components are staggered or offset in time by T seconds. For this reason, the signal

$$s(t) = A\left\{\left[\sum_{n=-\infty}^{\infty} I_{2n} g(t - 2nT)\right] \cos 2\pi f_c t + \left[\sum_{n=-\infty}^{\infty} I_{2n+1} g(t - 2nT - T)\right] \sin 2\pi f_c t\right\} \tag{4-3-69}$$

is called *offset quadrature PSK (OQPSK)* or *staggered quadrature PSK (SQPSK)*.

Figure 4-3-23 illustrates the representation of the MSK signals as two staggered quadrature-modulated binary PSK signals. The corresponding sum of the two quadrature signals is a constant amplitude, frequency-modulated signal.

It is also interesting to compare the waveforms for MSK with offset QPSK in which the pulse $g(t)$ is rectangular for $0 \leq t \leq 2T$, and with conventional

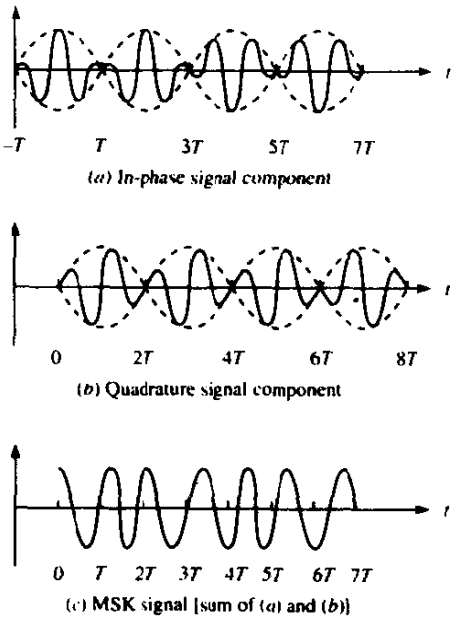


FIGURE 4-3-23 Representation of MSK signal as a form of two staggered binary PSK signals, each with a sinusoidal envelope.

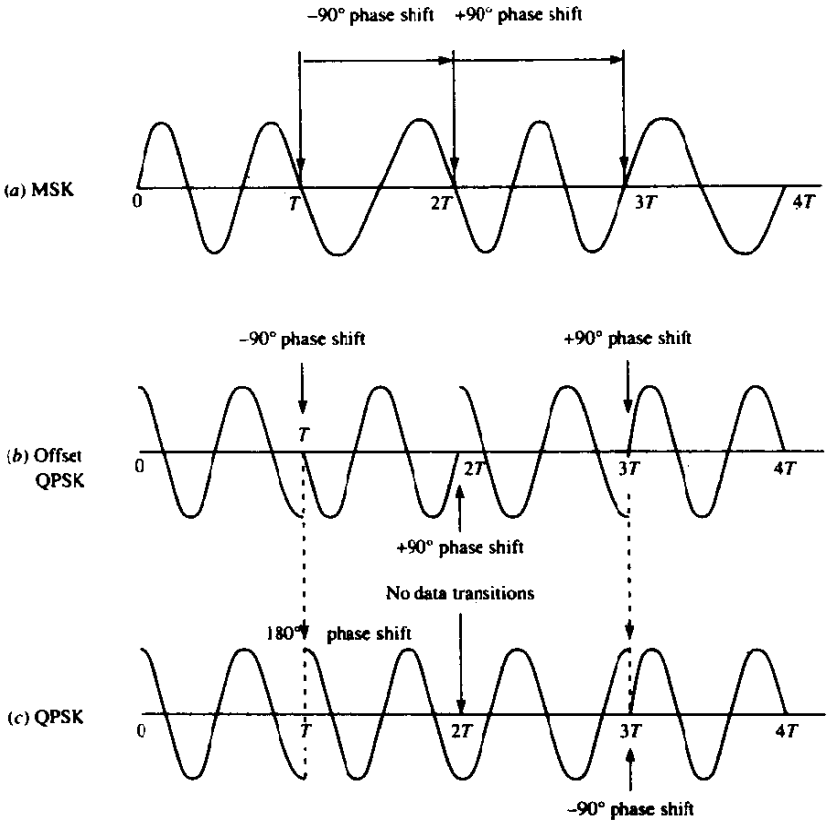


FIGURE 4-3-24 Signal waveforms for (*a*) MSK, (*b*) offset QPSK (rectangular pulse), and (*c*) conventional QPSK (rectangular pulse). [*From Gronemeyer and McBride (1976); © 1976 IEEE.*]

quadrature (four-phase) PSK (QPSK) in which the pulse $g(t)$ is rectangular for $0 \leq t \leq 2T$. Clearly, all three of the modulation methods result in identical data rates. The MSK signal has continuous phase. The offset QPSK signal with a rectangular pulse is basically two binary PSK signals for which the phase transitions are staggered in time by T seconds. Thus, the signal contains phase jumps of $\pm 90°$ that may occur as often as every T seconds. On the other hand, the conventional four-phase PSK signal with constant amplitude will contain phase jumps of $\pm 180°$ or $\pm 90°$ every $2T$ seconds. An illustration of these three signal types is given in Fig. 4-3-24.

Signal Space Diagrams for CPM In general, continuous-phase signals cannot be represented by discrete points in signal space as in the case of PAM, PSK, and QAM, because the phase of the carrier is time-variant. Instead, a continuous-phase signal is described by the various paths or trajectories from one phase state to another. For a constant-amplitude CPM signal, the various trajectories form a circle.

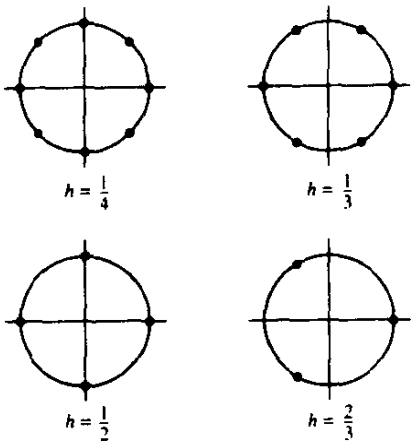


FIGURE 4-3-25 Signal space diagram for CPFSK.

For example, Fig. 4-3-25 illustrates the signal space (phase trajectory) diagram for CPFSK signals with $h = \frac{1}{4}$, $h = \frac{1}{3}$, $h = \frac{1}{2}$, and $h = \frac{2}{3}$. The beginning and ending points of these phase trajectories are marked in the figure by dots. Note that the length of the phase trajectory increases with an increase in h. An increase in h also results in an increase of the signal bandwidth, as demonstrated in the following section.

Multiamplitude CPM Multiamplitude CPM is a generalization of ordinary CPM in which the signal amplitude is allowed to vary over a set of amplitude values while the phase of the signal is constrained to be continuous. For example, let us consider a two-amplitude CPFSK signal, which may be represented as

$$s(t) = 2A \cos [2\pi f_c t + \phi_2(t; \mathbf{I})] + A \cos [2\pi f_c t + \phi_1(t; \mathbf{J})] \qquad (4\text{-}3\text{-}70)$$

where

$$\phi_2(t; \mathbf{I}) = \pi h \sum_{k=-\infty}^{n-1} I_k + \frac{\pi h I_n (t - nT)}{T}, \quad nT \leq t \leq (n+1)T \qquad (4\text{-}3\text{-}71)$$

$$\phi_1(t; \mathbf{J}) = \pi h \sum_{k=-\infty}^{n-1} J_k + \frac{\pi h J_n (t - nT)}{T}, \quad nT \leq t \leq (n+1)T \qquad (4\text{-}3\text{-}72)$$

The information is conveyed by the symbol sequences $\{I_n\}$ and $\{J_n\}$, which are related to two independent binary information sequences $\{a_n\}$ and $\{b_n\}$ that take values $\{0, 1\}$. We observe that the signal in (4-3-70) is a superposition of two CPFSK signals of different amplitude. However, the sequences $\{I_n\}$ and $\{J_n\}$ are not statistically independent, but are constrained in order to achieve phase continuity in the superposition of the two components.

To elaborate, let us consider the case where $h = \frac{1}{2}$, so that we have the superposition of two MSK signals. At the symbol transition points, the two

TABLE 4-3-1

a_n	b_n	I_n	J_n	Amplitude–phase relations
0	0	−1	−1	Amplitude is constant; phase decreases
0	1	−1	1	Amplitude changes; phase decreases
1	0	1	1	Amplitude is constant; phase increases
1	1	1	−1	Amplitude changes; phase increases

amplitude components are either in phase or 180° out of phase. The phase change in the signal is determined by the phase of the larger amplitude component, while the amplitude change is determined by the smaller component. Thus, the smaller component is constrained such that at the start and end of each symbol interval, it is either in phase or 180° out of phase with the larger component, independent of its phase. Under this constraint, the symbol sequences $\{I_n\}$ and $\{J_n\}$ may be expressed as

$$\begin{aligned} I_n &= 2a_n - 1 \\ J_n &= I_n(1 - 2b_n) = I_n\left(1 - \frac{b_n}{h}\right) \end{aligned} \tag{4-3-73}$$

These relationships are summarized in Table 4-3-1.

As a generalization, a multiamplitude CPFSK signal with n components may be expressed as

$$s(t) = 2^{N-1}\cos[2\pi f_c t + \phi_N(t; \mathbf{I})] + \sum_{m=1}^{N-1} 2^{m-1}\cos[2\pi f_c t + \phi_m(t; \mathbf{J}_m)] \tag{4-3-74}$$

where

$$\phi_N(t; \mathbf{I}) = \pi h I_n \frac{t - nT}{T} + \pi h \sum_{k=-\infty}^{n-1} I_k, \qquad nT \leq t \leq (n+1)T \tag{4-3-75}$$

and

$$\begin{aligned} \phi_m(t; \mathbf{J}_m) &= I_n \pi[h + \tfrac{1}{2}(J_{mn} + 1)]\frac{t - nT}{T} \\ &\quad + \sum_{k=-\infty}^{n-1} \pi I_k[h + \tfrac{1}{2}(J_{mk} + 1)], \qquad nT \leq t \leq (n+1)T \end{aligned} \tag{4-3-76}$$

The sequences $\{I_n\}$ and $\{J_{mn}\}$ are statistically independent, binary-valued sequences that take values from the set $\{1, -1\}$.

From (4-3-75) and (4-3-76), we observe that each component in the sum

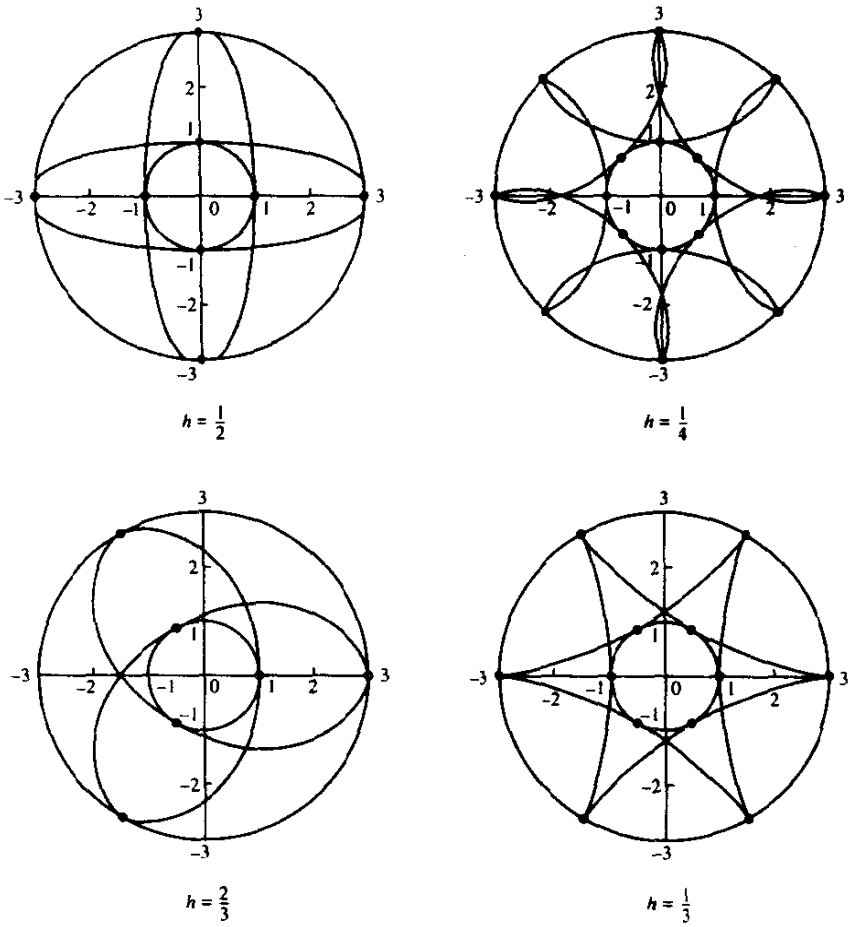


FIGURE 4-3-26 Signal space diagrams for two-component CPFSK.

will be either in phase or 180° out of phase with the largest component at the end of the nth symbol interval, i.e., at $t = (n + 1)T$. Thus, the signal states are specified by an amplitude level from the set of amplitudes $\{1, 3, 5, \ldots, 2^N - 1\}$ and a phase level from the set $\{0, \pi\theta, 2\pi\theta, \ldots, 2\pi - \pi h\}$. The phase constraint is required to maintain the phase continuity of the CPM signal.

Figure 4-3-26 illustrates the signal space diagrams for two-amplitude ($N = 2$) CPFSK with $h = \frac{1}{4}, \frac{1}{3}, \frac{1}{2}$, and $\frac{2}{3}$. The signal space diagrams for three-component ($N = 3$) CPFSK are shown in Fig. 4-3-27. In this case, there are *four* amplitude levels. The number of states depends on the modulation index h as well as N. Note that the beginning and ending points of the phase trajectories are marked by dots.

Additional multiamplitude CPM signal formats may be obtained by using

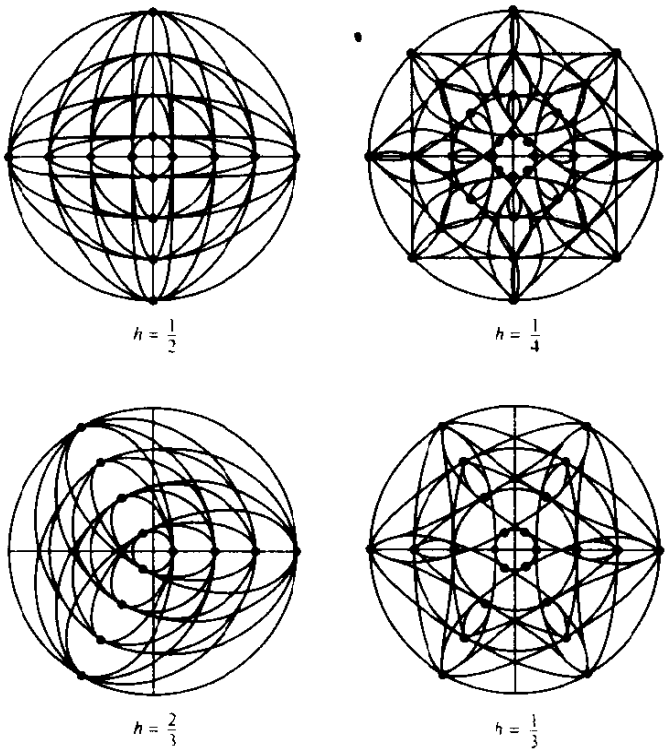


FIGURE 4-3-27 Signal space diagrams for three-component CPFSK.

pulse shapes other than rectangular, as well as signal pulses that span more than one symbol (partial response).

4-4 SPECTRAL CHARACTERISTICS OF DIGITALLY MODULATED SIGNALS

In most digital communications systems, the available channel bandwidth is limited. Consequently, the system designer must consider the constraints imposed by the channel bandwidth limitation in the selection of the modulation technique used to transmit the information. For this reason, it is important for us to determine the spectral content of the digitally modulated signals described in Section 4-3.

Since the information sequence is random, a digitally modulated signal is a stochastic process. We are interested in determining the power density spectrum of such a process. From the power density spectrum, we can determine the channel bandwidth required to transmit the information-bearing signal. Below, we first derive the spectral characteristics of the class of linearly

modulated signals. Then, we consider the nonlinear CPFSK, CPM, and baseband modulated signals with memory.

4-4-1 Power Spectra of Linearly Modulated Signals

Beginning with the form

$$s(t) = \text{Re}\,[v(t)e^{j2\pi f_c t}]$$

which relates the bandpass signal $s(t)$ to the equivalent lowpass signal $v(t)$, we may express the autocorrelation function of $s(t)$ as

$$\phi_{ss}(\tau) = \text{Re}\,[\phi_{vv}(\tau)e^{j2\pi f_c \tau}] \tag{4-4-1}$$

where $\phi_{vv}(\tau)$ is the autocorrelation function of the equivalent lowpass signal $v(t)$. The Fourier transform of (4-4-1) yields the desired expression for the power density spectrum $\Phi_{ss}(f)$ in the form

$$\Phi_{ss}(f) = \tfrac{1}{2}[\Phi_{vv}(f - f_c) + \Phi_{vv}(-f - f_c)] \tag{4-4-2}$$

where $\Phi_{vv}(f)$ is the power density spectrum of $v(t)$. It suffices to determine the autocorrelation function and the power density spectrum of the equivalent lowpass signal $v(t)$.

First we consider the linear digital modulation methods for which $v(t)$ is represented in the general form

$$v(t) = \sum_{n=-\infty}^{\infty} I_n g(t - nT) \tag{4-4-3}$$

where the transmission rate is $1/T = R/k$ symbols/s and $\{I_n\}$ represents the sequence of symbols that results from mapping k-bit blocks into corresponding signal points selected from the appropriate signal space diagram. Observe that in PAM, the sequence $\{I_n\}$ is real and corresponds to the amplitude values of the transmitted signal, but in PSK, QAM, and combined PAM–PSK, the sequence $\{I_n\}$ is complex-valued, since the signal points have a two-dimensional representation.

The autocorrelation function of $v(t)$ is

$$\begin{aligned}\phi_{vv}(t + \tau; t) &= \tfrac{1}{2}E[v^*(t)v(t + \tau)] \\ &= \tfrac{1}{2}\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} E[I_n^* I_m]g^*(t - nT)g(t + \tau - mT)\end{aligned} \tag{4-4-4}$$

We assume that the sequence of information symbols $\{I_n\}$ is wide-sense stationary with mean μ_i and autocorrelation function

$$\phi_{ii}(m) = \tfrac{1}{2}E[I_n^* I_{n+m}] \tag{4-4-5}$$

Hence (4-4-4) can be expressed as

$$\phi_{vv}(t+\tau;t) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \phi_{ii}(m-n)g^*(t-nT)g(t+\tau-mT)$$

$$= \sum_{m=-\infty}^{\infty} \phi_{ii}(m) \sum_{n=-\infty}^{\infty} g^*(t-nT)g(t+\tau-nT-mT) \quad (4\text{-}4\text{-}6)$$

The second summation in (4-4-6), namely,

$$\sum_{n=-\infty}^{\infty} g^*(t-nT)g(t+\tau-nT-mT)$$

is periodic in the t variable with period T. Consequently, $\phi_{vv}(t+\tau;t)$ is also periodic in the t variable with period T. That is,

$$\phi_{vv}(t+T+\tau;t+T) = \phi_{vv}(t+\tau;t) \quad (4\text{-}4\text{-}7)$$

In addition, the mean value of $v(t)$, which is

$$E[v(t)] = \mu_i \sum_{n=-\infty}^{\infty} g(t-nT) \quad (4\text{-}4\text{-}8)$$

is periodic with period T. Therefore $v(t)$ is a stochastic process having a periodic mean and autocorrelation function. Such a process is called a *cyclostationary process* or a *periodically stationary process in the wide sense*, as described in Section 2-2-6.

In order to compute the power density spectrum of a cyclostationary process, the dependence of $\phi_{vv}(t+\tau;t)$ on the t variable must be eliminated. This can be accomplished simply by averaging $\phi_{vv}(t+\tau;t)$ over a single period. Thus,

$$\bar{\phi}_{vv}(\tau) = \frac{1}{T}\int_{-T/2}^{T/2} \phi_{vv}(t+\tau;t)\,dt$$

$$= \sum_{m=-\infty}^{\infty} \phi_{ii}(m) \sum_{n=-\infty}^{\infty} \frac{1}{T}\int_{-T/2}^{T/2} g^*(t-nT)g(t+\tau-nT-mT)\,dt$$

$$= \sum_{m=-\infty}^{\infty} \phi_{ii}(m) \sum_{n=-\infty}^{\infty} \frac{1}{T}\int_{-T/2-nT}^{T/2-nT} g^*(t)g(t+\tau-mT)\,dt \quad (4\text{-}4\text{-}9)$$

We interpret the integral in (4-4-9) as the time-autocorrelation function of $g(t)$ and define it as

$$\phi_{gg}(\tau) = \int_{-\infty}^{\infty} g^*(t)g(t+\tau)\,dt \quad (4\text{-}4\text{-}10)$$

Consequently (4-4-9) can be expressed as

$$\bar{\phi}_{vv}(\tau) = \frac{1}{T} \sum_{m=-\infty}^{\infty} \phi_{ii}(m)\phi_{gg}(\tau - mT) \tag{4-4-11}$$

The Fourier transform of the relation in (4-4-11) yields the (average) power density spectrum of $v(t)$ in the form

$$\Phi_{vv}(f) = \frac{1}{T}|G(f)|^2 \Phi_{ii}(f) \tag{4-4-12}$$

where $G(f)$ is the Fourier transform of $g(t)$, and $\Phi_{ii}(f)$ denotes the power density spectrum of the information sequence, defined as

$$\Phi_{ii}(f) = \sum_{m=-\infty}^{\infty} \phi_{ii}(m)e^{-j2\pi fmT} \tag{4-4-13}$$

The result (4-4-12) illustrates the dependence of the power density spectrum of $v(t)$ on the spectral characteristics of the pulse $g(t)$ and the information sequence $\{I_n\}$. That is, the spectral characteristics of $v(t)$ can be controlled by design of the pulse shape $g(t)$ and by design of the correlation characteristics of the information sequence.

Whereas the dependence of $\Phi_{vv}(f)$ on $G(f)$ is easily understood upon observation of (4-4-12), the effect of the correlation properties of the information sequence is more subtle. First of all, we note that for an arbitrary autocorrelation $\phi_{ii}(m)$ the corresponding power density spectrum $\Phi_{ii}(f)$ is periodic in frequency with period $1/T$. In fact, the expression (4-4-13) relating the spectrum $\Phi_{ii}(f)$ to the autocorrelation $\phi_{ii}(m)$ is in the form of an exponential Fourier series with the $\{\phi_{ii}(m)\}$ as the Fourier coefficients. As a consequence, the autocorrelation sequence $\phi_{ii}(m)$ is given by

$$\phi_{ii}(m) = T \int_{-1/2T}^{1/2T} \Phi_{ii}(f)e^{j2\pi fmT}\, df \tag{4-4-14}$$

Second, let us consider the case in which the information symbols in the sequence are real and mutually uncorrelated. In this case, the autocorrelation function $\phi_{ii}(m)$ can be expressed as

$$\phi_{ii}(m) = \begin{cases} \sigma_i^2 + \mu_i^2 & (m = 0) \\ \mu_i^2 & (m \neq 0) \end{cases} \tag{4-4-15}$$

where σ_i^2 denotes the variance of an information symbol. When (4-4-15) is used to substitute for $\phi_{ii}(m)$ in (4-4-13), we obtain

$$\Phi_{ii}(f) = \sigma_i^2 + \mu_i^2 \sum_{m=-\infty}^{\infty} e^{-j2\pi fmT} \tag{4-4-16}$$

The summation in (4-4-16) is periodic with period $1/T$. It may be viewed as

the exponential Fourier series of a periodic train of impulses with each impulse having an area $1/T$. Therefore (4-4-16) can also be expressed in the form

$$\Phi_{ii}(f) = \sigma_i^2 + \frac{\mu_i^2}{T} \sum_{m=-\infty}^{\infty} \delta\left(f - \frac{m}{T}\right) \tag{4-4-17}$$

Substitution of (4-4-17) into (4-4-12) yields the desired result for the power density spectrum of $v(t)$ when the sequence of information symbols is uncorrelated. That is,

$$\Phi_{vv}(f) = \frac{\sigma_i^2}{T}|G(f)|^2 + \frac{\mu_i^2}{T^2} \sum_{m=-\infty}^{\infty} \left|G\left(\frac{m}{T}\right)\right|^2 \delta\left(f - \frac{m}{T}\right) \tag{4-4-18}$$

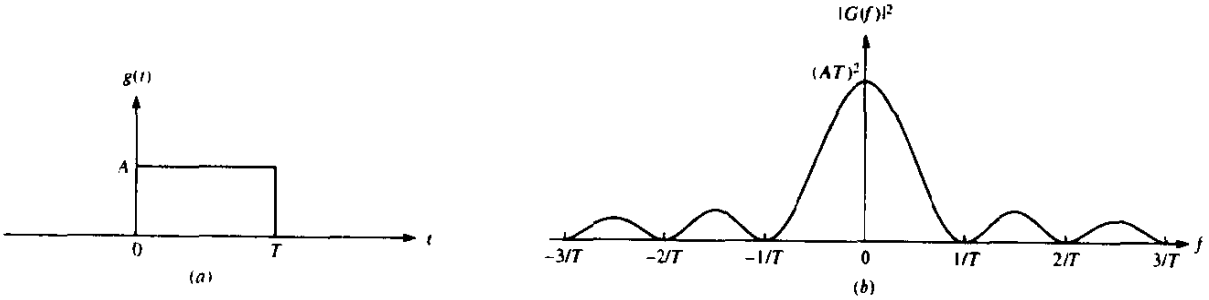
The expression (4-4-18) for the power density spectrum is purposely separated into two terms to emphasize the two different types of spectral components. The first term is the continuous spectrum, and its shape depends only on the spectral characteristic of the signal pulse $g(t)$. The second term consists of discrete frequency components spaced $1/T$ apart in frequency. Each spectral line has a power that is proportional to $|G(f)|^2$ evaluated at $f = m/T$. Note that the discrete frequency components vanish when the information symbols have zero mean, i.e., $\mu_i = 0$. This condition is usually desirable for the digital modulation techniques under consideration, and it is satisfied when the information symbols are equally likely and symmetrically positioned in the complex plane. Thus, the system designer can control the spectral characteristics of the digitally modulated signal by proper selection of the characteristics of the information sequence to be transmitted.

Example 4-4-1

To illustrate the spectral shaping resulting from $g(t)$, consider the rectangular pulse shown in Fig. 4-4-1(a). The Fourier transform of $g(t)$ is

$$G(f) = AT \frac{\sin \pi f T}{\pi f T} e^{-j\pi f T}$$

FIGURE 4-4-1 Rectangular pulse and its energy density spectrum $|G(f)|^2$.



Hence

$$|G(f)|^2 = (AT)^2\left(\frac{\sin \pi fT}{\pi fT}\right)^2 \tag{4-4-19}$$

This spectrum is illustrated in Fig. 4-4-1(*b*). Note that it contains zeros at multiples of $1/T$ in frequency and that it decays inversely as the square of the frequency variable. As a consequence of the spectral zeros in $G(f)$, all but one of the discrete spectral components in (4-4-18) vanish. Thus, upon substitution for $|G(f)|^2$ from (4-4-19), (4-4-18) reduces to

$$\Phi_{vv}(f) = \sigma_i^2 A^2 T\left(\frac{\sin \pi fT}{\pi fT}\right)^2 + A^2\mu_i^2\delta(f) \tag{4-4-20}$$

Example 4-4-2

As a second illustration of the spectral shaping resulting from $g(t)$, we consider the raised cosine pulse

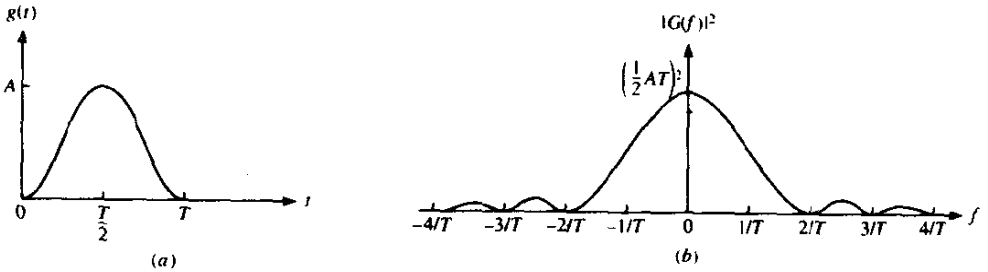
$$g(t) = \frac{A}{2}\left[1 + \cos\frac{2\pi}{T}\left(t - \frac{T}{2}\right)\right], \quad 0 \leq t \leq T \tag{4-2-21}$$

This pulse is graphically illustrated in Fig. 4-4-2(*a*). Its Fourier transform is easily derived and it may be expressed in the form

$$G(f) = \frac{AT}{2}\frac{\sin \pi fT}{\pi fT(1 - f^2T^2)}e^{-j\pi fT} \tag{4-4-22}$$

The square of the magnitude of $G(f)$ is shown in Fig. 4-4-2(*b*). It is interesting to note that the spectrum has zeros at $f = n/T$, $n = \pm 2, \pm 3, \pm 4, \ldots$. Consequently, all the discrete spectral components in (4-4-18), except the ones at $f = 0$ and $f = \pm 1/T$, vanish. When compared with the

FIGURE 4-4-2 Raised cosine pulse and its energy density spectrum $|G(f)|^2$.



spectrum of the rectangular pulse, the spectrum of the raised cosine pulse has a broader main lobe but the tails decay inversely as f^6.

Example 4-4-3

To illustrate that spectral shaping can also be accomplished by operations performed on the input information sequence, we consider a binary sequence $\{b_n\}$ from which we form the symbols

$$I_n = b_n + b_{n-1} \tag{4-4-23}$$

The $\{b_n\}$ are assumed to be uncorrelated random variables, each having zero mean and unit variance. Then the autocorrelation function of the sequence $\{I_n\}$ is

$$\begin{aligned}\phi_{ii}(m) &= E(I_n I_{n+m}) \\ &= \begin{cases} 2 & (m = 0) \\ 1 & (m = \pm 1) \\ 0 & \text{(otherwise)} \end{cases}\end{aligned} \tag{4-4-24}$$

Hence, the power density spectrum of the input sequence is

$$\begin{aligned}\Phi_{ii}(f) &= 2(1 + \cos 2\pi f T) \\ &= 4\cos^2 \pi f T\end{aligned} \tag{4-4-25}$$

and the corresponding power density spectrum for the (lowpass) modulated signal is

$$\Phi_{vv}(f) = \frac{4}{T}|G(f)|^2 \cos^2 \pi f T \tag{4-4-26}$$

4-4-2 Power Spectra of CPFSK and CPM Signals

In this section, we derive the power density spectrum for the class of constant amplitude CPM signals that were described in Section 4-3-3. We begin by computing the autocorrelation function and its Fourier transform, as was done in the case of linearly modulated signals.

The constant amplitude CPM signal is expressed as

$$s(t; \mathbf{I}) = A \cos [2\pi f_c t + \phi(t; \mathbf{I})] \tag{4-4-27}$$

where

$$\phi(t; \mathbf{I}) = 2\pi h \sum_{k=-\infty}^{\infty} I_k q(t - kT) \tag{4-4-28}$$

Each symbol in the sequence $\{I_n\}$ can take one of the M values $\{\pm 1, \pm 3, \ldots, \pm(M-1)\}$. These symbols are statistically independent and identically distributed with prior probabilities

$$P_n = P(I_k = n), \quad n = \pm 1, \pm 3, \ldots, \pm(M-1) \tag{4-4-29}$$

where $\sum_n P_n = 1$. The pulse $g(t) = q'(t)$ is zero outside of the interval $[0, LT]$, $q(t) = 0$, $t < 0$, and $q(t) = \frac{1}{2}$ for $t > LT$.

The autocorrelation function of the equivalent lowpass signal

$$v(t) = e^{j\phi(t;\mathbf{I})}$$

is

$$\phi_{vv}(t+\tau; t) = \tfrac{1}{2}E\left[\exp\left(j2\pi h \sum_{k=-\infty}^{\infty} I_k[q(t+\tau-kT) - q(t-kT)]\right)\right] \tag{4-4-30}$$

First we express the sum in the exponent as a product of exponents. The result is

$$\phi_{vv}(t+\tau; t) = \tfrac{1}{2}E\left(\prod_{k=-\infty}^{\infty} \exp\{j2\pi h I_k[q(t+\tau-kT) - q(t-kT)]\}\right) \tag{4-4-31}$$

Next, we perform the expectation over the data symbols $\{I_k\}$. Since these symbols are statistically independent, we obtain

$$\phi_{vv}(t+\tau; t) = \tfrac{1}{2}\prod_{k=-\infty}^{\infty}\left(\sum_{\substack{n=-(M-1)\\ n \text{ odd}}}^{M-1} P_n \exp\{j2\pi hn[q(t+\tau-kT) - q(t-kT)]\}\right) \tag{4-4-32}$$

Finally, the average autocorrelation function is

$$\bar{\phi}_{vv}(\tau) = \frac{1}{T}\int_0^T \phi_{vv}(t+\tau; t)\, dt \tag{4-4-33}$$

Although (4-4-32) implies that there are an infinite number of factors in the product, the pulse $g(t) = q'(t) = 0$ for $t < 0$ and $t > LT$, and $q(t) = 0$ for $t < 0$. Consequently only a finite number of terms in the product have nonzero exponents. Thus, (4-4-32) can be simplified considerably. In addition, if we let $\tau = \xi + mT$, where $0 \leq \xi < T$ and $m = 0, 1, \ldots$, the average autocorrelation in (4-4-33) reduces to

$$\bar{\phi}_{vv}(\xi + mT) = \frac{1}{2T}\int_0^T \prod_{k=1-L}^{m-1}\left(\sum_{\substack{n=-(M-1)\\ n \text{ odd}}}^{M-1} P_n \exp\{j2\pi hn[q(t+\xi-(k-m)T) - q(t-kT)]\}\right) \tag{4-4-34}$$

Let us focus on $\bar{\phi}_{vv}(\xi + mT)$ for $\xi + mT \geq LT$. In this case, (4-4-34) may be expressed as

$$\bar{\phi}_{vv}(\xi + mT) = [\psi(jh)]^{m-L}\lambda(\xi), \quad m \geq L, \quad 0 \leq \xi < T \tag{4-4-35}$$

where $\psi(jh)$ is the characteristic function of the random sequence $\{I_n\}$, defined as

$$\begin{aligned}\psi(jh) &= E(e^{j\pi h I_n}) \\ &= \sum_{\substack{n=-(M-1) \\ n \text{ odd}}}^{M-1} P_n e^{j\pi h n}\end{aligned} \tag{4-4-36}$$

and $\lambda(\xi)$ is the remaining part of the average autocorrelation function, which may be expressed as

$$\begin{aligned}\lambda(\xi) = \frac{1}{2T}\int_0^T \prod_{k=1-L}^{0} &\left(\sum_{\substack{n=-(M-1) \\ n \text{ odd}}}^{M-1} P_n \exp\{j2\pi hn[\tfrac{1}{2} - q(t - kT)]\} \right) \\ \times \prod_{k=1-L}^{1} &\left(\sum_{\substack{n=-(M-1) \\ n \text{ odd}}}^{M-1} P_n \exp[j2\pi hnq(t + \xi - kT)] \right) dt, \quad m \geq L\end{aligned} \tag{4-4-37}$$

Thus, $\bar{\phi}_{vv}(\tau)$ may be separated into a product of $\lambda(\xi)$ and $\psi(jh)$ as indicated in (4-4-35) for $\tau = \xi + mT \geq LT$ and $0 \leq \xi < T$. This property is used below.

The Fourier transform of $\bar{\phi}_{vv}(\tau)$ yields the average power density spectrum as

$$\begin{aligned}\Phi_{vv}(f) &= \int_{-\infty}^{\infty} \bar{\phi}_{vv}(\tau)e^{-j2\pi f\tau}\,d\tau \\ &= 2\,\mathrm{Re}\left[\int_0^{\infty} \bar{\phi}_{vv}(\tau)e^{-j2\pi f\tau}\,d\tau\right]\end{aligned} \tag{4-4-38}$$

But

$$\begin{aligned}\int_0^{\infty} \bar{\phi}_{vv}(\tau)e^{-j2\pi f\tau}\,d\tau &= \int_0^{LT} \bar{\phi}_{vv}(\tau)e^{-j2\pi f\tau}\,d\tau \\ &\quad + \int_{LT}^{\infty} \bar{\phi}_{vv}(\tau)e^{-j2\pi f\tau}\,d\tau\end{aligned} \tag{4-3-39}$$

With the aid of (4-4-35), the integral in the range $LT \leq \tau < \infty$ may be expressed as

$$\int_{LT}^{\infty} \bar{\phi}_{vv}(\tau)e^{-2\pi f\tau}\,d\tau = \sum_{m=L}^{\infty} \int_{mT}^{(m+1)T} \bar{\phi}_{vv}(\tau)e^{-j2\pi f\tau}\,d\tau \tag{4-4-40}$$

Now, let $\tau = \xi + mT$. Then (4-4-40) becomes

$$\begin{aligned}\int_{LT}^{\infty} \bar{\phi}_{vv}(\tau)e^{-j2\pi f\tau}\,dt &= \sum_{m=L}^{\infty} \int_0^T \bar{\phi}_{vv}(\xi + mT)e^{-j2\pi f(\xi+mT)}\,d\xi \\ &= \sum_{m=L}^{\infty} \int_0^T \lambda(\xi)[\psi(jh)]^{m-L}e^{-j2\pi f(\xi+mT)}\,d\xi \\ &= \sum_{n=0}^{\infty} \psi^n(jh)e^{-j2\pi fnT} \int_0^T \lambda(\xi)e^{-j2\pi f(\xi+LT)}\,d\xi \end{aligned} \tag{4-4-41}$$

A property of the characteristic function is $|\psi(jh)| \leq 1$. For values of h for which $|\psi(jh)| < 1$, the summation in (4-4-41) converges and yields

$$\sum_{n=0}^{\infty} \psi^n(jh)e^{-j2\pi fnT} = \frac{1}{1 - \psi(jh)e^{-j2\pi fT}} \tag{4-4-42}$$

In this case, (4-4-41) reduces to

$$\int_{LT}^{\infty} \bar{\phi}_{vv}(\tau)e^{-j2\pi f\tau}\,dt = \frac{1}{1 - \psi(jh)e^{-j2\pi fT}} \int_0^T \bar{\phi}_{vv}(\xi + LT)e^{-j2\pi f(\xi+LT)}\,d\xi \tag{4-4-43}$$

By combining (4-4-38), (4-4-39), and (4-4-43), we obtain the power density spectrum of the CPM signal in the form

$$\Phi_{vv}(f) = 2\,\mathrm{Re}\left[\int_0^{LT} \bar{\phi}_{vv}(\tau)e^{-j2\pi f\tau}\,d\tau + \frac{1}{1 - \psi(jh)e^{-j2\pi fT}} \int_{LT}^{(L+1)T} \bar{\phi}_{vv}(\tau)e^{-j2\pi f\tau}\,d\tau\right] \tag{4-4-44}$$

This is the desired result when $|\psi(jh)| < 1$. In general, the power density spectrum is evaluated numerically from (4-4-44). The average autocorrelation function $\bar{\phi}_{vv}(\tau)$ for the range $0 \leq \tau \leq (L+1)T$ may be computed numerically from (4-4-34).

For values of h for which $|\psi(jh)| = 1$, e.g., $h = K$, where K is an integer, we can set

$$\psi(jh) = e^{j2\pi\nu}, \qquad 0 \leq \nu < 1 \tag{4-4-45}$$

Then, the sum in (4-4-41) becomes

$$\sum_{n=0}^{\infty} e^{-j2\pi T(f-\nu/T)n} = \tfrac{1}{2} + \frac{1}{2T} \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{\nu}{T} - \frac{n}{T}\right) - j\,\tfrac{1}{2}\cot \pi T\left(f - \frac{\nu}{T}\right) \tag{4-4-46}$$

Thus, the power density spectrum now contains impulses located at frequencies

$$f_n = \frac{n+\nu}{T}, \quad 0 \leq \nu < 1, \quad n = 0, 1, 2, \ldots \tag{4-4-47}$$

The result (4-4-46) can be combined with (4-4-41) and (4-4-39) to obtain the entire power density spectrum, which includes both a continuous spectrum component and a discrete spectrum component.

Let us return to the case for which $|\psi(jh)| < 1$. When the symbols are equally probable, i.e.,

$$P_n = \frac{1}{M} \quad \text{for all } n$$

the characteristic function simplifies to the form

$$\begin{aligned} \psi(jh) &= \frac{1}{M} \sum_{\substack{n=-(M-1) \\ n \text{ odd}}}^{M-1} e^{j\pi hn} \\ &= \frac{1}{M} \frac{\sin M\pi h}{\sin \pi h} \end{aligned} \tag{4-4-48}$$

Note that in this case $\psi(jh)$ is real. The average autocorrelation function given by (4-4-34) also simplifies in this case to

$$\bar{\phi}_{vv}(\tau) = \frac{1}{2T} \int_0^T \prod_{k=1-L}^{\lfloor \tau/T \rfloor} \frac{1}{M} \frac{\sin 2\pi hM[q(t+\tau-kT) - q(t-kT)]}{\sin 2\pi h[q(t+\tau-kT) - q(t-kT)]} dt \tag{4-4-49}$$

The corresponding expression for the power density spectrum reduces to

$$\begin{aligned} \Phi_{vv}(f) = 2\Bigg[&\int_0^{LT} \bar{\phi}_{vv}(\tau) \cos 2\pi f\tau \, d\tau \\ &+ \frac{1 - \psi(jh)\cos 2\pi fT}{1 + \psi^2(jh) - 2\psi(jh)\cos 2\pi fT} \int_{LT}^{(L+1)T} \bar{\phi}_{vv}(\tau) \cos 2\pi f\tau \, d\tau \Bigg] \end{aligned} \tag{4-4-50}$$

Power Density Spectrum of CPFSK A closed-form expression for the power density spectrum can be obtained from (4-4-50) when the pulse shape $g(t)$ is rectangular and zero outside the interval $[0, T]$. In this case, $q(t)$ is linear for $0 \leq t \leq T$. The resulting power spectrum may be expressed as

$$\Phi_{vv}(f) = T\left[\frac{1}{M}\sum_{n=1}^{M} A_n^2(f) + \frac{2}{M^2}\sum_{n=1}^{M}\sum_{m=1}^{M} B_{nm}(f)A_n(f)A_m(f)\right] \tag{4-4-51}$$

where

$$A_n(f) = \frac{\sin \pi[fT - \frac{1}{2}(2n - 1 - M)h]}{\pi[fT - \frac{1}{2}(2n - 1 - M)h]}$$

$$B_{nm}(f) = \frac{\cos(2\pi fT - \alpha_{nm}) - \psi \cos \alpha_{nm}}{1 + \psi^2 - 2\psi \cos 2\pi fT} \tag{4-4-52}$$

$$\alpha_{nm} = \pi h(m + n - 1 - M)$$

$$\psi \equiv \psi(jh) = \frac{\sin M\pi h}{M \sin \pi h}$$

The power density spectrum of CPFSK for $M = 2$, 4, and 8 is plotted in Figs 4-4-3, 4-4-4, and 4-4-5 as a function of the normalized frequency fT, with the modulation index $h = 2f_dT$ as a parameter. Note that only one-half of the bandwidth occupancy is shown in these graphs. The origin corresponds to the carrier f_c. The graphs illustrate that the spectrum of CPFSK is relatively smooth and well confined for $h < 1$. As h approaches unity, the spectra become very peaked and, for $h = 1$ when $|\psi| = 1$, we find that impulses occur at M frequencies. When $h > 1$ the spectrum becomes much broader. In communication systems where CPFSK is used, the modulation index is designed to conserve bandwidth, so that $h < 1$.

The special case of binary CPFSK with $h = \frac{1}{2}$ (or $f_d = 1/4T$) and $\psi = 0$ corresponds to MSK. In this case, the spectrum of the signal is

$$\Phi_{vv}(f) = \frac{16A^2T}{\pi^2}\left(\frac{\cos 2\pi fT}{1 - 16f^2T^2}\right)^2 \tag{4-4-53}$$

where the signal amplitude $A = 1$ in (4-4-52). In contrast the spectrum of four-phase offset (quadrature) PSK (OQPSK) with a rectangular pulse $g(t)$ of duration T is

$$\Phi_{vv}(f) = A^2T\left(\frac{\sin \pi fT}{\pi fT}\right)^2 \tag{4-4-54}$$

If we compare these spectral characteristics, we should normalize the frequency variable by the bit rate or the bit interval T_b. Since MSK is binary FSK, it follows that $T = T_b$ in (4-4-53). On the other hand, in OQPSK, $T = 2T_b$ so that (4-4-54) becomes

$$\Phi_{vv}(f) = 2A^2T_b\left(\frac{\sin 2\pi fT_b}{2\pi fT_b}\right)^2 \tag{4-4-55}$$

The spectra of the MSK and OQPSK signals are illustrated in Fig. 4-4-6. Note that the main lobe of MSK is 50% wider than that for OQPSK. However, the side lobes in MSK fall off considerably faster. For example, if we compare the bandwidth W that contains 99% of the total power, we find that $W = 1.2/T_b$ for MSK and $W \approx 8/T_b$ for OQPSK. Consequently, MSK has a narrower spectral

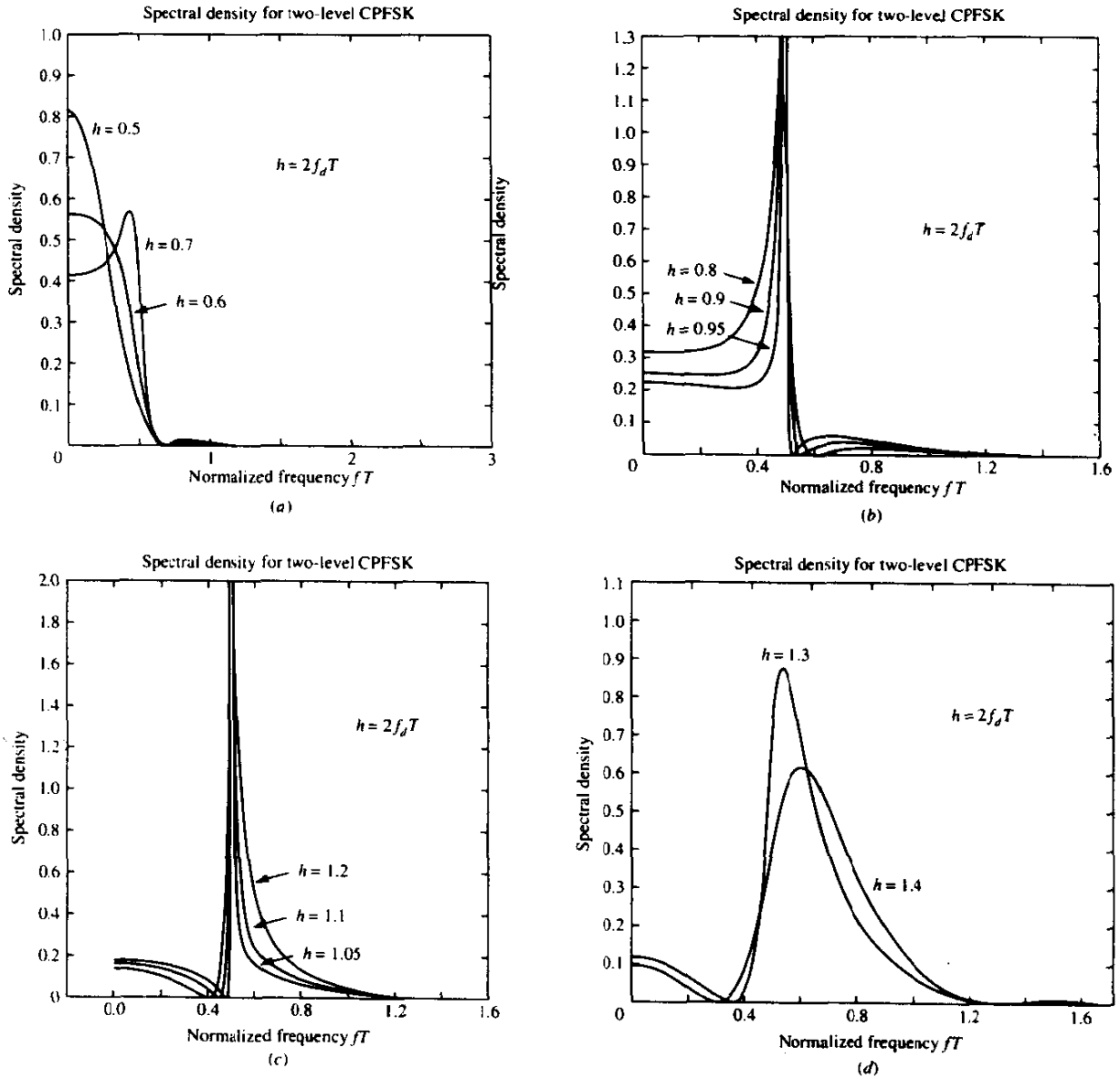


FIGURE 4-4-3 Power density spectrum of binary CPFSK.

occupancy when viewed in terms of fractional out-of-band power above $fT_b = 1$. Graphs for the fractional out-of-band power for OQPSK and MSK are shown in Fig. 4-4-7. Note that MSK is significantly more bandwidth-efficient than QPSK. This efficiency accounts for the popularity of MSK in many digital communications systems.

Even greater bandwidth efficiency than MSK can be achieved by reducing

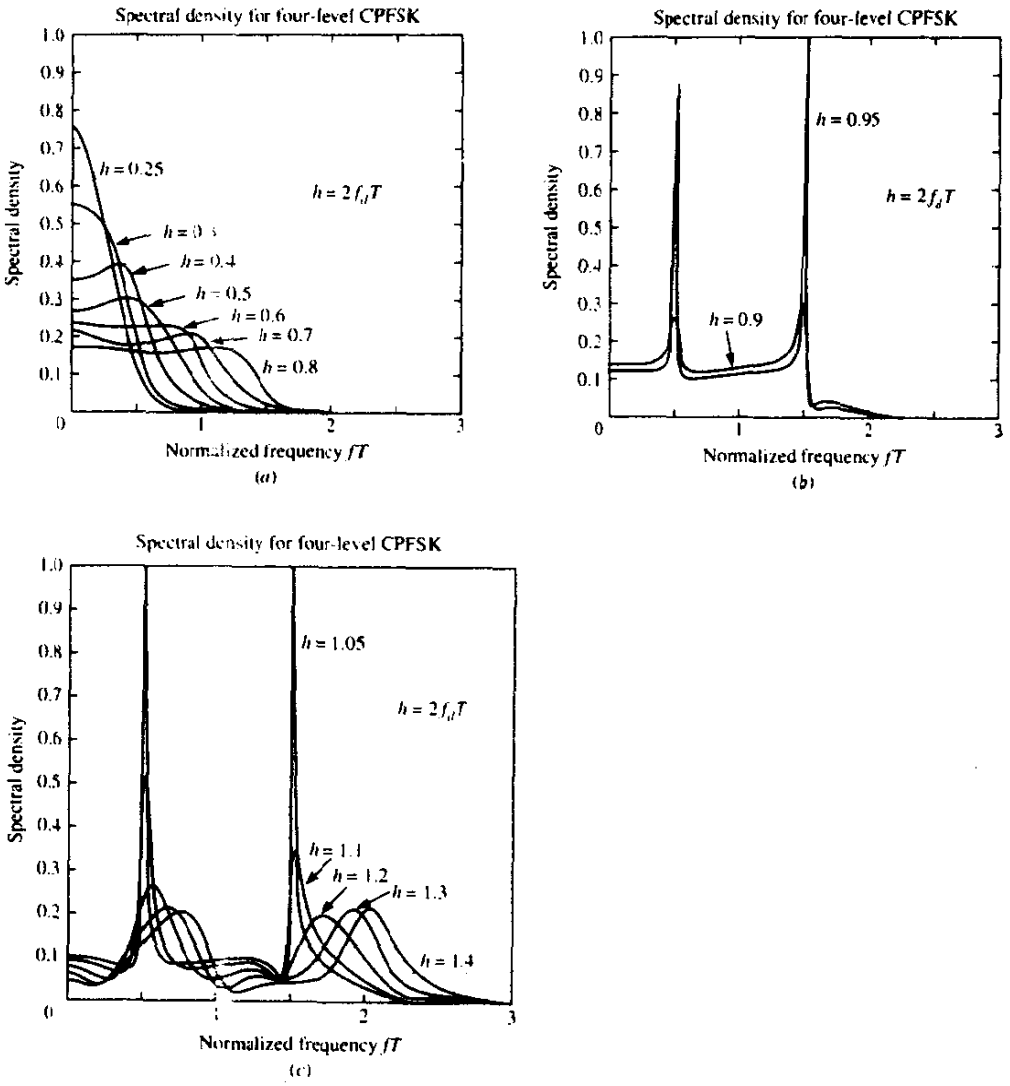


FIGURE 4-4-4 Power density spectrum of quaternary CPFSK.

the modulation index. However, the FSK signals will no longer be orthogonal and there will be an increase in the error probability.

Spectral Characteristics of CPM In general, the bandwidth occupancy of CPM depends on the choice of the modulation index h, the pulse shape $g(t)$, and the number of signals M. As we have observed for CPFSK, small values of h result in CPM signals with relatively small bandwidth occupancy, while large values of h result in signals with large bandwidth occupancy. This is also the case for the more general CPM signals.

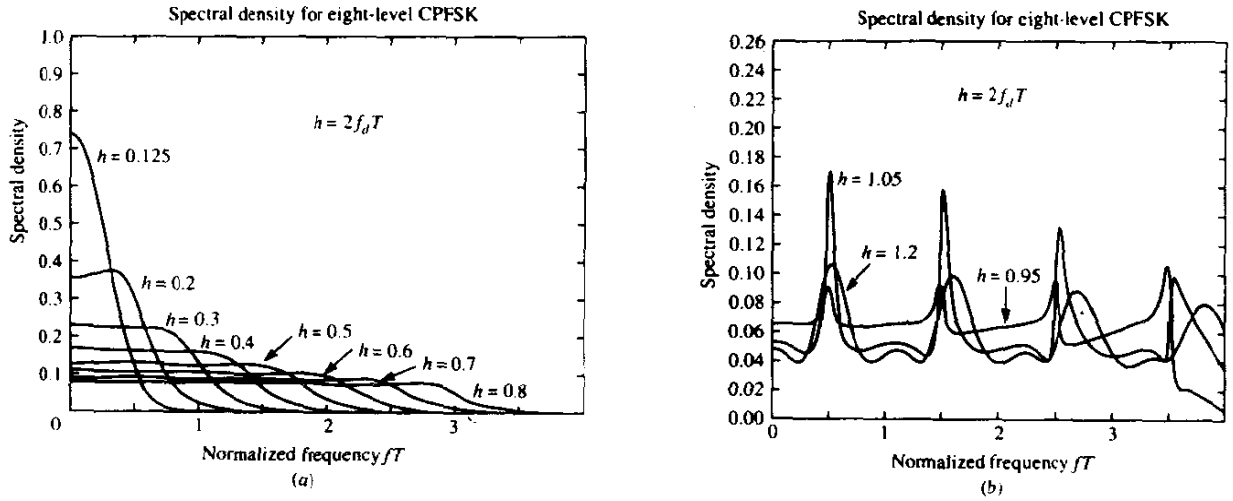
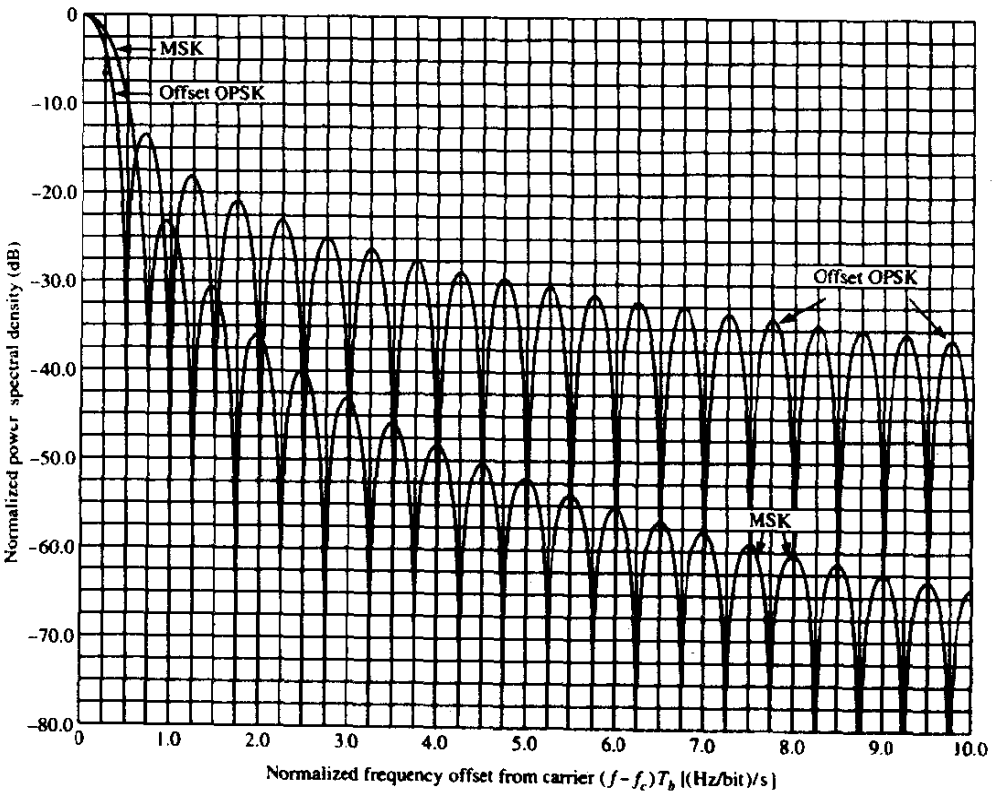


FIGURE 4-4-5 Power density spectrum of octal CPFSK.

FIGURE 4-4-6 Power density spectra of MSK and offset QPSK. [*From Gronemeyer and McBride (1976);* © *IEEE.*]



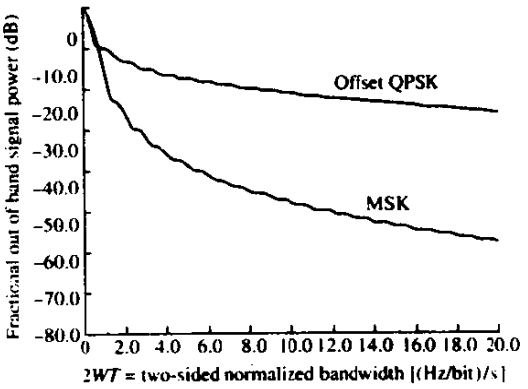


FIGURE 4-4-7 Fractional out-of-band power (normalized two-sided bandwidth = 2 BT). [*From Gronemeyer and McBride (1976); © 1976 IEEE.*]

The use of smooth pulses such as raised cosine pulses of the form

$$g(t)=\begin{cases}\dfrac{1}{2LT}\left(1-\cos\dfrac{2\pi t}{LT}\right) & (0\leq t\leq LT)\\ 0 & (\text{otherwise})\end{cases} \tag{4-4-56}$$

where $L = 1$ for full response and $L > 1$ for partial response, result in smaller bandwidth occupancy and, hence, greater bandwidth efficiency than the use of rectangular pulses. For example, Fig. 4-4-8 illustrates the power density spectrum for binary CPM with different partial response raised cosine (LRC) pulses when $h = \frac{1}{2}$. For comparison, the spectrum of binary CPFSK is also shown. Note that as L increases the pulse $g(t)$ becomes smoother and the corresponding spectral occupancy of the signal is reduced.

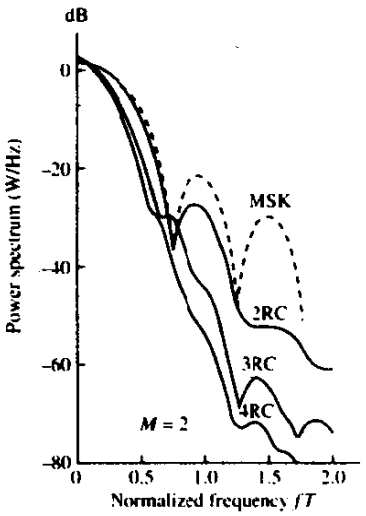


FIGURE 4-4-8 Power density spectrum for binary CPM with $h = \frac{1}{2}$ and different pulse shapes. [*From Aulin* et al. *(1981); © 1981 IEEE.*]

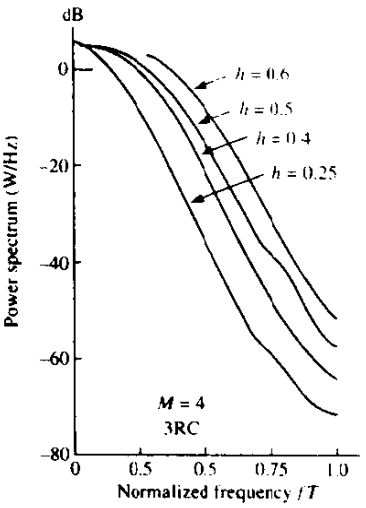
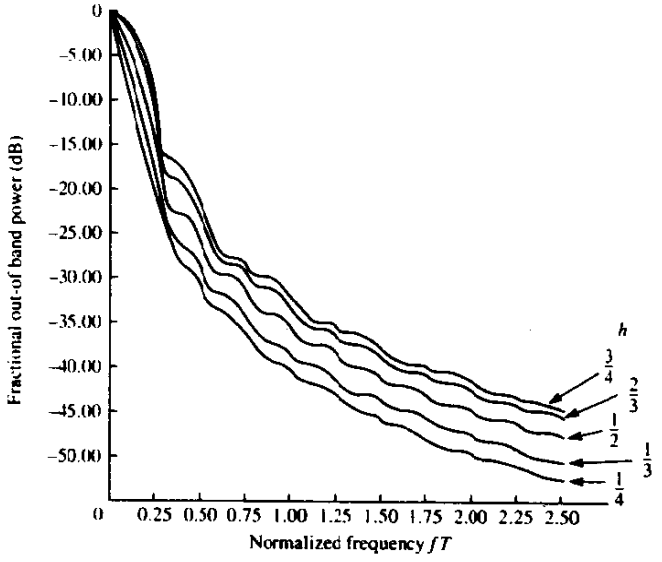


FIGURE 4-4-9 Power density spectrum for $M = 4$ CPM with $3RC$ and different modulation indices. [*From Aulin* et al. *(1981);* © *1981 IEEE.*]

The effect of varying the modulation index in a CPM signal is illustrated in Fig. 4-4-9 for the case of $M = 4$ and a raised cosine pulse of the form given in (4-4-56) with $L = 3$. Note that these spectral characteristics are similar to the ones illustrated previously for CPFSK, except that these spectra are narrower due to the use of a smoother pulse shape.

Finally, in Fig. 4-4-10, we illustrate the fractional out-of-band power for two-amplitude CPFSK with several different values of h.

FIGURE 4-4-10 Fractional out-of-band power for two-component CPFSK. (*Mulligan,1988.*)



4-4-3 Power Spectra of Modulated Signals with Memory

In the last two sections, we have determined the spectral characteristics for the class of linearly modulated signals without memory and for the class of angle-modulated signals such as CPFSK and CPM, which are nonlinear and possess memory. In this section, we consider the spectral characteristics of linearly modulated signals that have memory that can be modeled by a Markov chain. We have already encountered such signals in Section 4-3-2, where we described several types of baseband signals.

The power density spectrum of a digitally modulated signal that is generated by a Markov chain may be derived by following the basic procedure given in the previous section. Thus, we can determine the autocorrelation function and then evaluate its Fourier transform to obtain the power density spectrum. For signals that are generated by a Markov chain with transition probability matrix **P**, the power density spectrum of the modulated signal may be expressed in the general form (see Titsworth and Welch, 1961)

$$\Phi(f) = \frac{1}{T^2}\sum_{n=-\infty}^{\infty}\left|\sum_{i=1}^{K} p_i S_i\left(\frac{n}{T}\right)\right|^2 \delta\left(f-\frac{n}{T}\right) + \frac{1}{T}\sum_{i=1}^{K} p_i\,|S_i'(f)|^2 + \frac{2}{T}\mathrm{Re}\left[\sum_{i=1}^{K}\sum_{j=1}^{K} p_i S_i'^{*}(f) S_j'(f) P_{ij}(f)\right] \tag{4-4-57}$$

where $S_i(f)$ is the Fourier transform of the signal waveform $s_i(t)$,

$$s_i'(t) = s_i(t) - \sum_{k=1}^{K} p_k s_k(t)$$

$P_{ij}(f)$ is the Fourier transform of the discrete-time sequence $p_{ij}(n)$, defined as

$$P_{ij}(f) = \sum_{n=1}^{\infty} p_{ij}(n) e^{-j2\pi nfT} \tag{4-4-58}$$

and K is the number of states of the modulator. The term $p_{ij}(n)$ denotes the probability that the signal $s_j(t)$ is transmitted n signaling intervals after the transmission of $s_i(t)$. Hence, $\{p_{ij}(n)\}$ are the transition probabilities in the transition probability matrix $\mathbf{P}^n$. Note that $p_{ij}(1) = p_{ij}$.

When there is no memory in the modulation method, the signal waveform transmitted on each signaling interval is independent of the waveforms transmitted in previous signaling intervals. The power density spectrum of the resultant signal may still be expressed in the form of (4-4-57), if the transition probability matrix is replaced by

$$\mathbf{P} = \begin{bmatrix} p_1 & p_2 & \cdots & p_K \\ p_1 & p_2 & \cdots & p_K \\ \vdots & \vdots & & \vdots \\ p_1 & p_2 & \cdots & p_K \end{bmatrix} \tag{4-4-59}$$

and we impose the condition that $\mathbf{P}^n = \boldsymbol{P}$ for all $n \geq 1$. Under these conditions, the expression for the power density spectrum becomes a function of the stationary state probabilities $\{p_i\}$ only, and, hence, it reduces to the simpler form

$$\Phi(f) = \frac{1}{T^2} \sum_{n=-\infty}^{\infty} \left| \sum_{i=1}^{K} p_i S_i\left(\frac{n}{T}\right) \right|^2 \delta\left(f - \frac{n}{T}\right) + \frac{1}{T} \sum_{i=1}^{K} p_i(1 - p_i) |S_i(f)|^2 - \frac{2}{T} \sum_{i=1}^{K} \sum_{\substack{j=1 \\ i<j}}^{K} p_i p_j \operatorname{Re}[S_i(f) S_j^*(f)] \tag{4-4-60}$$

We observe that our previous result for the power density spectrum of memoryless linear modulation given by (4-4-18) may be viewed as a special case of (4-4-60) in which all waveforms are identical except for a set of scale factors that convey the digital information (Problem 4-30).

We also make the observation that the first term in the expression for the power density spectrum given by either (4-4-57) or (4-4-60) consists of discrete frequency components. This line spectrum vanishes when

$$\sum_{i=1}^{K} p_i S_i\left(\frac{n}{T}\right) = 0 \tag{4-4-61}$$

The condition (4-4-61) is usually imposed in the design of practical digital communications systems and is easily satisfied by an appropriate choice of signaling waveforms (Problem 4-31).

Now, let us determine the power density spectrum of the baseband-modulated signals described in Section 4-3-2. First, the NRZ signal is characterized by the two waveforms $s_1(t) = g(t)$ and $s_2(t) = -g(t)$, where $g(t)$ is a rectangular pulse of amplitude A. For $K = 2$, (4-4-60) reduces to

$$\Phi(f) = \frac{(2p-1)^2}{T^2} \sum_{n=-\infty}^{\infty} \left| G\left(\frac{n}{T}\right) \right|^2 \delta\left(f - \frac{n}{T}\right) + \frac{4p(1-p)}{T} |G(f)|^2 \tag{4-4-62}$$

where

$$|G(f)|^2 = (AT)^2 \left(\frac{\sin \pi f T}{\pi f T}\right)^2 \tag{4-4-63}$$

Observe that when $p = \frac{1}{2}$, the line spectrum vanishes and $\Phi(f)$ reduces to

$$\Phi(f) = \frac{1}{T} |G(f)|^2 \tag{4-4-64}$$

The NRZI signal is characterized by the transition probability matrix

$$\mathbf{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \tag{4-4-65}$$

Notice that in this case $\mathbf{P}^n = \mathbf{P}$ for all $n \geq 1$. Hence, the special form for the power density spectrum given by (4-4-62) applies to this modulation format as well. Consequently, the power density spectrum for the NRZI signal is identical to the spectrum of the NRZ signal.

Delay modulation has a transition probability matrix

$$\mathbf{P} = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix} \tag{4-4-66}$$

and stationary state probabilities $p_i = \frac{1}{4}$ for $i = 1, 2, 3, 4$. Powers of $\mathbf{P}$ may be obtained by use of the relation

$$\mathbf{P}^4\boldsymbol{\rho} = -\tfrac{1}{4}\boldsymbol{\rho} \tag{4-4-67}$$

where $\boldsymbol{\rho}$ is the signal correlation matrix with elements

$$\rho_{ij} = \frac{1}{T}\int_0^T s_i(t)s_j(t)\,dt \tag{4-4-68}$$

and where the four signals $\{s_i(t),\ i = 1, 2, 3, 4\}$ are shown in Fig. 4-3-15. It is easily seen that

$$\boldsymbol{\rho} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} \tag{4-4-69}$$

Consequently, powers of $\mathbf{P}$ can be generated from the relation

$$\mathbf{P}^{k+4}\boldsymbol{\rho} = -\tfrac{1}{4}\mathbf{P}^k\boldsymbol{\rho}, \qquad k > 1 \tag{4-4-70}$$

Use of (4-4-66), (4-4-69), and (4-4-70) in (4-4-57) yields the power density spectrum of delay modulation. It may be expressed in the form

$$\Phi(f) = \frac{1}{2\psi^2(17 + 8\cos 8\psi)}[23 - 2\cos\psi - 22\cos 2\psi - 12\cos 3\psi + 5\cos 4\psi + 12\cos 5\psi + 2\cos 6\psi - 8\cos 7\psi + 2\cos 8\psi] \tag{4-4-71}$$

where $\psi = \pi f T$.

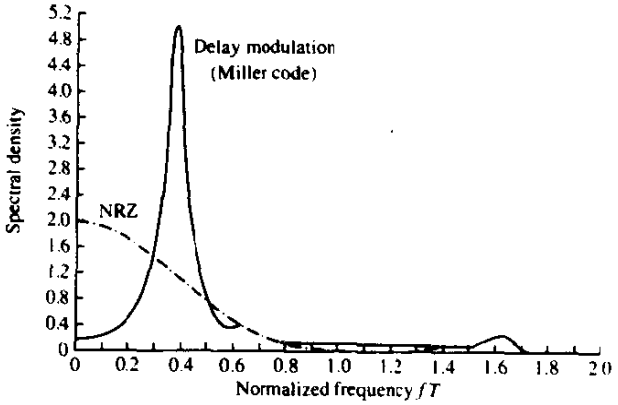


FIGURE 4-4-11 Power spectral density (one-sided) of Miller code (delay modulation) and NRZ/NRZI baseband signals. [*From Hecht and Guida (1969); © 1969 IEEE.*]

The spectra of these baseband signals are illustrated in Fig. 4-4-11. Observe that the spectra of the NRZ and NRZI signals peak at $f = 0$. Delay modulation has a narrower spectrum and a relatively small zero-frequency content. Its bandwidth occupancy is significantly smaller than that of the NRZ signal. These two characteristics make delay modulation an attractive choice for channels that do not pass dc, such as magnetic recording media.

4-5 BIBLIOGRAPHICAL NOTES AND REFERENCES

The characteristics of signals and systems given in this chapter are very useful in the design of optimum modulation/demodulation and coding/decoding techniques for a variety of channel models. In particular, the digital modulation methods introduced in this chapter are widely used in digital communication systems. The next chapter is concerned with optimum demodulation techniques for these signals and their performance in an additive, white gaussian noise channel. A general reference for signal characterization is the book by Franks (1969).

Of particular importance in the design of digital communications systems are the spectral characteristics of the digitally modulated signals, which are presented in this chapter in some depth. Of these modulation techniques, CPM is one of the most important due to its efficient use of bandwidth. For this reason, it has been widely investigated by many researchers, and a large number of papers have been published in the technical literature. The most comprehensive treatment of CPM, including its performance and its spectral characteristics, can be found in the book by Anderson *et al.* (1986). In addition to this text, the tutorial paper by Sundberg (1986) presents the basic concepts and an overview of the performance characteristics of various CPM techniques. This paper also contains over 100 references to published papers on this topic.

There are a large number of references dealing with the spectral characteristics of CPFSK and CPM. As a point of reference, we should mention that MSK was invented by Doelz and Heald in 1961. The early work on the power

spectral density of CPFSK and CPM was done by Bennett and Rice (1963), Anderson and Salz (1965), and Bennett and Davey (1965). The book by Lucky *et al.* (1968) also contains a treatment of the spectral characteristics of CPFSK. Most of the recent work is referenced in the paper by Sundberg (1986). We should also cite the special issue on bandwidth-efficient modulation and coding published by the *IEEE Transactions on Communications* (March 1981), which contains several papers on the spectral characteristics and performance of CPM.

The generalization of MSK to multiple amplitudes was investigated by Weber *et al.* (1978). The combination of multiple amplitudes with general CPM was proposed by Mulligan (1988) who investigated its spectral characteristics and its error probability performance in gaussian noise with and without coding.

4-1 Prove the following properties of Hilbert transforms:
- **a** If $x(t) = x(-t)$ then $\hat{x}(t) = -\hat{x}(-t)$;
- **b** If $x(t) = -x(-t)$ then $\hat{x}(t) = \hat{x}(-t)$;
- **c** If $x(t) = \cos \omega_0 t$ then $\hat{x}(t) = \sin \omega_0 t$;
- **d** If $x(t) = \sin \omega_0 t$ then $\hat{x}(t) = -\cos \omega_0 t$;
- **e** $\hat{\hat{x}}(t) = -x(t)$;
- **f** $\int_{-\infty}^{\infty} x^2(t)\,dt = \int_{-\infty}^{\infty} \hat{x}^2(t)\,dt$;
- **g** $\int_{-\infty}^{\infty} x(t)\hat{x}(t)\,dt = 0$.

4-2 If $x(t)$ is a stationary random process with autocorrelation function $\phi_{xx}(\tau) = E[x(t)x(t+\tau)]$ and spectral density $\Phi_{xx}(f)$ then show that $\phi_{\hat{x}\hat{x}}(\tau) = \phi_{xx}(\tau)$, $\phi_{x\hat{x}}(\tau) = -\hat{\phi}_{xx}(\tau)$, and $\Phi_{\hat{x}\hat{x}}(f) = \Phi_{xx}(f)$.

4-3 Suppose that $n(t)$ is a zero-mean stationary narrowband process represented by either (4-1-37), (4-1-38), or (4-1-39). The autocorrelation function of the equivalent lowpass process $z(t) = x(t) + jy(t)$ is defined as

$$\phi_{zz}(\tau) = \tfrac{1}{2}E[z^*(t)z(t+\tau)]$$

- **a** Show that

$$E[z(t)z(t+\tau)] = 0$$

- **b** Suppose $\phi_{zz}(\tau) = N_0\delta(\tau)$, and let

$$V = \int_0^T z(t)\,dt$$

Determine $E(V^2)$ and $E(VV^*) = E(|V|^2)$.

4-4 Determine the autocorrelation function of the stochastic process

$$x(t) = A \sin (2\pi f_c t + \theta)$$

where f_c is a constant and θ is a uniformly distributed phase, i.e.,

$$p(\theta) = \frac{1}{2\pi}, \quad 0 \le \theta \le 2\pi$$

4-5 Prove that $s_l(t)$ is generally a complex-valued signal and give the condition under which it is real. Assume that $s(t)$ is a real-valued bandpass signal.

4-6 Suppose that $s(t)$ is either a real- or complex-valued signal that is represented as a linear combination of orthonormal functions $\{f_n(t)\}$, i.e.,

$$\hat{s}(t) = \sum_{k=1}^{K} s_k f_k(t)$$

where

$$\int_{-\infty}^{\infty} f_n(t) f_m^*(t)\, dt = \begin{cases} 0 & (m \neq n) \\ 1 & (m = n) \end{cases}$$

Determine the expressions for the coefficients $\{s_k\}$ in the expansion $\hat{s}(t)$ that minimize the energy

$$\mathscr{E}_e = \int_{-\infty}^{\infty} |s(t) - \hat{s}(t)|^2\, dt$$

and the corresponding residual error $\mathscr{E}_e$.

4-7 Suppose that a set of M signal waveforms $\{s_{lm}(t)\}$ are complex-valued. Derive the equations for the Gram–Schmidt procedure that will result in a set of $N \leq M$ orthonormal signal waveforms.

4-8 Determine the correlation coefficients ρ_{km} among the four signal waveforms $\{s_i(t)\}$ shown in Fig. 4-2-1, and the corresponding Euclidean distances.

4-9 Consider a set of M orthogonal signal waveforms $s_m(t)$, $1 \leq m \leq M$, $0 \leq t \leq T$, all of which have the same energy $\mathscr{E}$. Define a new set of M waveforms as

$$s'_m(t) = s_m(t) - \frac{1}{M} \sum_{k=1}^{M} s_k(t), \qquad 1 \leq m \leq M, \quad 0 \leq t \leq T$$

Show that the M signal waveforms $\{s'_m(t)\}$ have equal energy, given by

$$\mathscr{E}' = (M-1)\mathscr{E}/M$$

and are equally correlated, with correlation coefficient

$$\rho_{mn} = \frac{1}{\mathscr{E}'} \int_0^T s'_m(t) s'_n(t)\, dt = -\frac{1}{M-1}$$

4-10 Consider the three waveforms $f_n(t)$ shown in Fig. P4-10.

a Show that these waveforms are orthonormal.

b Express the waveform $x(t)$ as a weighted linear combination of $f_n(t)$, $n = 1, 2, 3$, if

$$x(t) = \begin{cases} -1 & (0 \leq t < 1) \\ 1 & (1 \leq t < 3) \\ -1 & (3 \leq t < 4) \end{cases}$$

and determine the weighting coefficients.

4-11 Consider the four waveforms shown in Fig. P4-11.

a Determine the dimensionality of the waveforms and a set of basis functions.

b Use the basis functions to represent the four waveforms by vectors $\mathbf{s}_1$, $\mathbf{s}_2$, $\mathbf{s}_3$, and $\mathbf{s}_4$.

c Determine the minimum distance between any pair of vectors.

4-12 Determine a set of orthonormal functions for the four signals shown in Fig. P4-12.

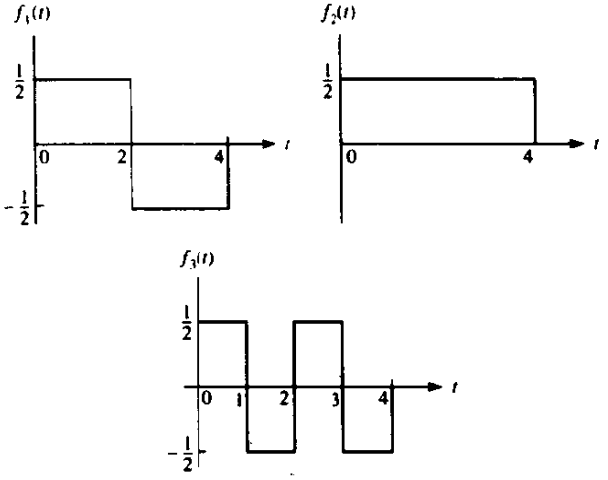


FIGURE P4-10

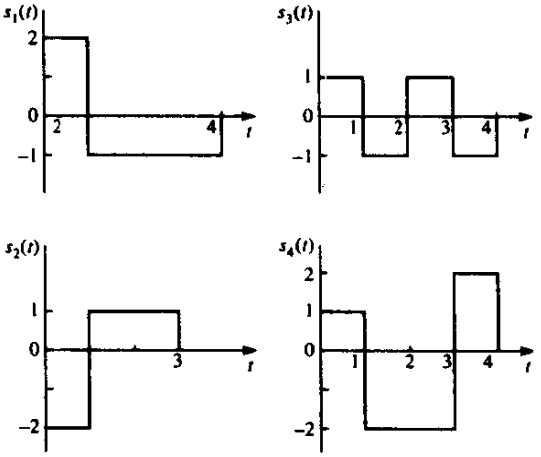


FIGURE P4-11

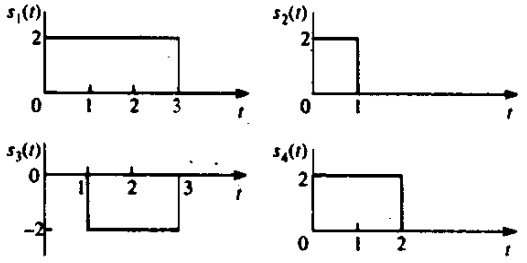


FIGURE P4-12

4-13 A lowpass gaussian stochastic process $x(t)$ has a power spectral density

$$\Phi(f) = \begin{cases} N_0 & (|f| < B) \\ 0 & (|f| > B) \end{cases}$$

Determine the power spectral density and the autocorrelation function of $y(t) = x^2(t)$.

4-14 Consider an equivalent lowpass digitally modulated signal of the form

$$u(t) = \sum_n [a_n g(t - 2nT) - jb_n g(t - 2nT - T)]$$

where $\{a_n\}$ and $\{b_n\}$ are two sequences of statistically independent binary digits and $g(t)$ is a sinusoidal pulse defined as

$$g(t) = \begin{cases} \sin(\pi t/2T) & (0 < t < 2T) \\ 0 & (\text{otherwise}) \end{cases}$$

This type of signal is viewed as a four-phase PSK signal in which the pulse shape is one-half cycle of a sinusoid. Each of the information sequences $\{a_n\}$ and $\{b_n\}$ is transmitted at a rate of $1/2T$ bits/s and, hence, the combined transmission rate is $1/T$ bits/s. The two sequences are staggered in time by T seconds in transmission. Consequently, the signal $u(t)$ is called *staggered four-phase PSK*.

a Show that the envelope $|u(t)|$ is a constant, independent of the information a_n on the in-phase component and information b_n on the quadrature component. In other words, the amplitude of the carrier used in transmitting the signal is constant.

b Determine the power density spectrum of $u(t)$.

c Compare the power density spectrum obtained from (b) with the power density spectrum of the MSK signal. What conclusion can you draw from this comparison?

4-15 Consider a four-phase PSK signal represented by the equivalent lowpass signal

$$u(t) = \sum_n I_n g(t - nT)$$

where I_n takes on one of the four possible values $\sqrt{\frac{1}{2}}(\pm 1 \pm j)$ with equal probability. The sequence of information symbols $\{I_n\}$ is statistically independent.

a Determine and sketch the power density spectrum of $u(t)$ when

$$g(t) = \begin{cases} A & (0 \leq t \leq T) \\ 0 & (\text{otherwise}) \end{cases}$$

b Repeat (a) when

$$g(t) = \begin{cases} A \sin(\pi t/T) & (0 \leq t \leq T) \\ 0 & (\text{otherwise}) \end{cases}$$

c Compare the spectra obtained in (a) and (b) in terms of the 3 dB bandwidth and the bandwidth to the first spectral zero.

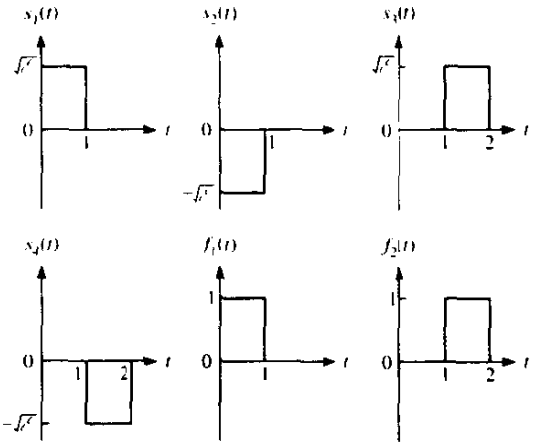


FIGURE P4-18

4-16 The random process $v(t)$ is defined as

$$v(t) = X \cos 2\pi f_c t - Y \sin 2\pi f_c t$$

where X and Y are random variables. Show that $v(t)$ is wide-sense stationary if and only if $E(X) = E(Y) = 0$, $E(X^2) = E(Y^2)$, and $E(XY) = 0$.

4-17 Carry out the Gram–Schmidt orthogonalization of the signals in Fig. 4-2-1(a) in the order $s_4(t)$, $s_3(t)$, $s_1(t)$, and, thus, obtain a set of orthonormal functions $\{f_m(t)\}$. Then, determine the vector representation of the signals $\{s_n(t)\}$ by using the orthonormal functions $\{f_m(t)\}$. Also, determine the signal energies.

4-18 Determine the signal space representation of the four signals $s_k(t)$, $k = 1, 2, 3, 4$, shown in Fig. P4-18, by using as basis functions the orthonormal functions $f_1(t)$ and $f_2(t)$. Plot the signal space diagram and show that this signal set is equivalent to that for a four-phase PSK signal.

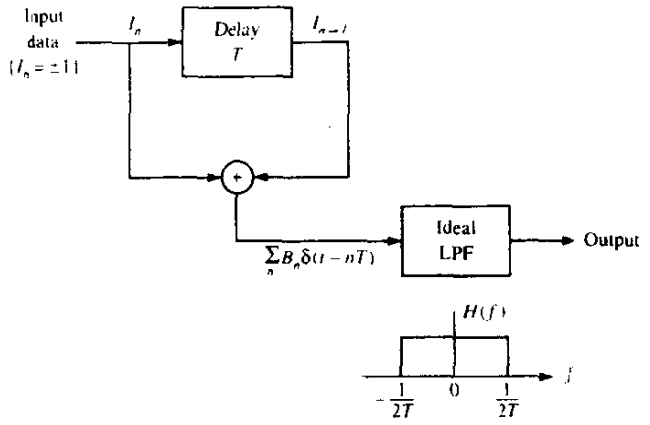
4-19 The power density spectrum of the cyclostationary process

$$v(t) = \sum_{n=-\infty}^{\infty} I_n g(t - nT)$$

was derived in Section 4-4-1 by averaging the autocorrelation function $\phi_{vv}(t + \tau, t)$ over the period T of the process and then evaluating the Fourier transform of the average autocorrelation function. An alternative approach is to change the cyclostationary process into a stationary process $v_\Delta(t)$ by adding a random variable Δ, uniformly distributed over $0 \leq \Delta < T$, so that

$$v_\Delta(t) = \sum_{n=-\infty}^{\infty} I_n g(t - nT - \Delta)$$

and defining the spectral density of $v(t)$ as the Fourier transform of the autocorrelation function of the stationary process $v_\Delta(t)$. Derive the result in (4-4-11), by evaluating the autocorrelation function of $v_\Delta(t)$ and its Fourier transform.



4-20 A PAM partial response signal (PRS) is generated as shown in Fig. P4-20 by exciting an ideal lowpass filter of bandwidth W by the sequence

$$B_n = I_n + I_{n-1}$$

at a rate $1/T = 2W$ symbols/s. The sequence $\{I_n\}$ consists of binary digits selected independently from the alphabet $\{1, -1\}$ with equal probability. Hence, the filtered signal has the form

$$v(t) = \sum_{n=-\infty}^{\infty} B_n g(t - nT), \qquad T = \frac{1}{2W}$$

a Sketch the signal space diagram for $v(t)$ and determine the probability of occurrence of each symbol.

b Determine the autocorrelation and power density spectrum of the three-level sequence $\{B_n\}$.

c The signal points of the sequence $\{B_n\}$ form a Markov chain. Sketch this Markov chain and indicate the transition probabilities among the states.

4-21 The lowpass equivalent representation of a PAM signal is

$$u(t) = \sum_n I_n g(t - nT)$$

Suppose $g(t)$ is a rectangular pulse and

$$I_n = a_n - a_{n-2}$$

where $\{a_n\}$ is a sequence of uncorrelated binary-valued $(1, -1)$ random variables that occur with equal probability.

a Determine the autocorrelation function of the sequence $\{I_n\}$.

b Determine the power density spectrum of $u(t)$.

c Repeat (b) if the possible values of the a_n are $(0, 1)$.

4-22 Show that $x(t) = s(t) \cos 2\pi f_c t \pm \hat{s}(t) \sin 2\pi f_c t$ is a single-sideband signal, where $s(t)$ is band-limited to $B \leq f_c$ Hz and $\hat{s}(t)$ is its Hilbert transform.

4-23 Use the results in Section 4-4-3 to determine the power density spectrum of the binary FSK signals in which the waveforms are

$$s_i(t) = \sin \omega_i t, \quad i = 1, 2, \quad 0 \leq t \leq T$$

where $\omega_1 = n\pi/T$ and $\omega_2 = m\pi/T$, $n \neq m$, and m and n are arbitrary positive integers. Assume that $p_1 = p_2 = \frac{1}{2}$. Sketch the spectrum and compare this result with the spectrum of the MSK signal.

4-24 Use the results in Section 4-4-3 to determine the power density spectrum of multitone FSK (MFSK) signals for which the signal waveforms are

$$s_n(t) = \sin \frac{2\pi n t}{T}, \quad n = 1, 2, \ldots, M, \quad 0 \leq t \leq T$$

Assume that the probabilities $p_i = 1/M$ for all i. Sketch the power spectral density.

4-25 A quadrature partial response signal (QPRS) is generated by two separate partial response signals of the type described in Problem 4-20 placed in phase quadrature. Hence, the QPRS is represented as

$$s(t) = \text{Re}\,[v(t)e^{j2\pi f_c t}]$$

where

$$\begin{aligned} v(t) &= v_c(t) + jv_s(t) \\ &= \sum_n B_n u(t - nT) + j \sum_n C_n u(t - nT) \end{aligned}$$

and $B_n = I_n + I_{n-1}$ and $C_n = J_n + J_{n-1}$. The sequences $\{B_n\}$ and $\{C_n\}$ are uncorrelated and $I_n = \pm 1$, $J_n = \pm 1$ with equal probability.

a Sketch the signal space diagram for the QPRS signal and determine the probability of occurrence of each symbol.

b Determine the autocorrelations and power spectra density of $v_c(t)$, $v_s(t)$, and $v(t)$.

c Sketch the Markov chain model and indicate the transition probabilities for the QPRS.

4-26 Determine the autocorrelation functions for the MSK and offset QPSK modulated signals based on the assumption that the information sequences for each of the two signals are uncorrelated and zero-mean.

4-27 Sketch the phase tree, the state trellis, and the state diagram for partial response CPM with $h = \frac{1}{2}$ and

$$u(t) = \begin{cases} 1/4T & (0 \leq t \leq 2T) \\ 0 & (\text{otherwise}) \end{cases}$$

4-28 Determine the number of terminal phase states in the state trellis diagram for

a a full response binary CPFSK with either $h = \frac{2}{3}$ or $\frac{3}{4}$;

b a partial response $L = 3$ binary CPFSK with either $h = \frac{2}{3}$ or $\frac{3}{4}$.

4-29 Show that 16 QAM can be represented as a superposition of two four-phase constant envelope signals where each component is amplified separately before summing, i.e.,

$$s(t) = G[A_n \cos 2\pi f_c t + B_n \sin 2\pi f_c t] + [C_n \cos 2\pi f_c t + D_n \sin 2\pi f_c t]$$

where $\{A_n\}$, $\{B_n\}$, $\{C_n\}$, and $\{D_n\}$ are statistically independent binary sequences

with elements from the set $\{+1, -1\}$ and G is the amplifier gain. Thus, show that the resulting signal is equivalent to

$$s(t) = I_n \cos 2\pi f_c t + Q_n \sin 2\pi f_c t$$

and determine I_n and Q_n in terms of A_n, B_n, C_n, and D_n.

4-30 Use the result in (4-4-60) to derive the expression for the power density spectrum of memoryless linear modulation given by (4-4-18) under the condition that

$$s_k(t) = I_k s(t), \quad k = 1, 2, \ldots, K$$

where I_k is one of the K possible transmitted symbols that occur with equal probability.

4-31 Show that a sufficient condition for the absence of the line spectrum component in (4-4-60) is

$$\sum_{i=1}^{K} p_i s_i(t) = 0$$

Is this condition necessary? Justify your answer.

4-32 The information sequence $\{a_n\}_{n=-\infty}^{\infty}$ is a sequence of iid random variables, each taking values +1 and −1 with equal probability. This sequence is to be transmitted at baseband by a biphase coding scheme, described by

$$s(t) = \sum_{n=-\infty}^{\infty} a_n g(t - nT)$$

where $g(t)$ is shown in Fig. P4-32.

a Find the power spectral density of $s(t)$.

b Assume that it is desirable to have a zero in the power spectrum at $f = 1/T$. To this end, we use a precoding scheme by introducing $b_n = a_n + ka_{n-1}$, where k is some constant, and then transmit the $\{b_n\}$ sequence using the same $g(t)$. Is it possible to choose k to produce a frequency null at $f = 1/T$? If yes, what are the appropriate value and the resulting power spectrum?

c Now assume we want to have zeros at all multiples of $f_0 = 1/4T$. Is it possible to have these zeros with an appropriate choice of k in the previous part? If not then what kind of precoding do you suggest to result in the desired nulls?

4-33 Starting with the definition of the transition probability matrix for delay modulation given in (4-4-66), demonstrate that the relation

$$\mathbf{P}^4\boldsymbol{\rho} = -\tfrac{1}{4}\boldsymbol{\rho}$$

holds, and, hence,

$$\mathbf{P}^{k+4}\boldsymbol{\rho} = -\tfrac{1}{4}\mathbf{P}^k\boldsymbol{\rho}, \quad k \geq 1$$

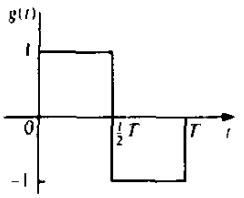


FIGURE P4-32

4-34 The two signal waveforms for binary FSK signal transmission with discontinuous phase are

$$s_0(t) = \sqrt{\frac{2\mathscr{E}_b}{T_b}} \cos\left[2\pi\left(f - \frac{\Delta f}{2}\right)t + \theta_0\right], \quad 0 \leq t < T$$

$$s_1(t) = \sqrt{\frac{2\mathscr{E}_b}{T_b}} \cos\left[2\pi\left(f + \frac{\Delta f}{2}\right)t + \theta_1\right], \quad 0 \leq t \leq T$$

where $\Delta f = 1/T \ll f_c$, and θ_0 and θ_1 are uniformly distributed random variables on the interval $(0, 2\pi)$. The signals $s_0(t)$ and $s_1(t)$ are equally probable.

a Determine the power spectral density of the FSK signal.

b Show that the power spectral density decays as $1/f^2$ for $f \gg f_c$.