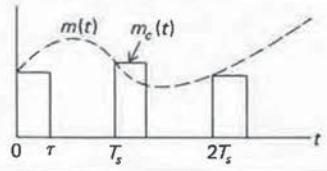
FIGURE 3.58 Pulse-amplitude modulation



operation with the sample values represented by the leading edge of each pulse.

The difference between pulse-amplitude modulation and sampling as developed in Chapter 2 is slight. We saw in Chapter 2 that a sampled signal can be represented as

$$m_\delta(t) = \sum_{n=-\infty}^{\infty} m(nT_s)\,\delta(t - nT_s) \tag{3.247}$$

From Figure 3.58, the PAM signal can be written as

$$m_c(t) = \sum_{n=-\infty}^{\infty} m(nT_s)\Pi\left[\frac{t - (nT_s + \frac{1}{2}\tau)}{\tau}\right] \tag{3.248}$$

where $\Pi[\cdot]$ is the time-domain unit pulse function defined in Chapter 2. The sampled signal $m_\delta(t)$ can be transformed into the PAM waveform defined by (3.248) by passing $m_\delta(t)$ through a network, as shown in Figure 3.59(a). The network simply holds the value of the sample for τ seconds. The $n = 0$ term in (3.247) and (3.248) illustrates that if the input to the holding network is $A\delta(t)$, the output must be $A\Pi[(t - \frac{1}{2}\tau)/\tau]$. Thus the impulse response of the holding network is

$$h(t) = \Pi\left[\frac{t - \frac{1}{2}\tau}{\tau}\right] \tag{3.249}$$

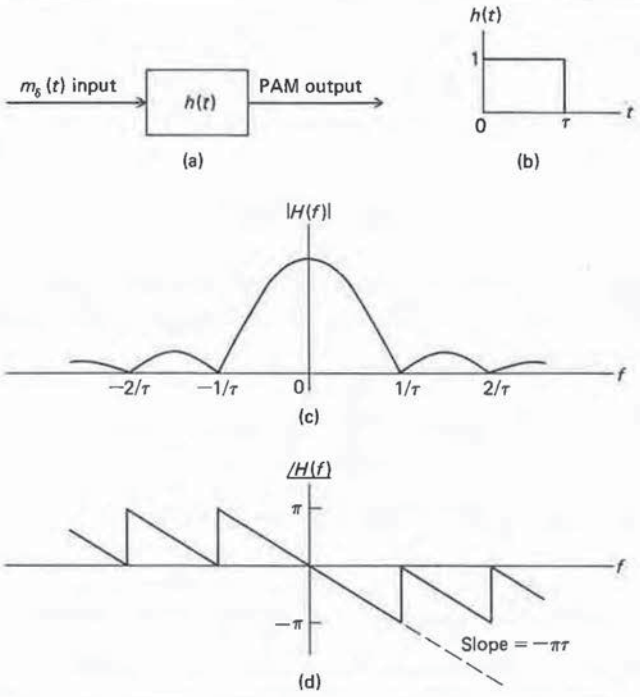
as shown in Figure 3.59(b). The transfer function $H(f)$ of the holding network is

$$H(f) = \tau\,\mathrm{sinc}\,f\tau\,e^{-j\pi f\tau} \tag{3.250}$$

as derived in Example 2.8 with $t_0 = \frac{1}{2}\tau$. The amplitude response and phase response are illustrated in Figures 3.59(c) and (d), respectively.

It should be remembered from Section 2.8 that $m(t)$ can be recovered from $m_\delta(t)$ by lowpass filtering. Thus a PAM signal can be demodulated using a two-step process. First $m_c(t)$ is passed through an equalization filter that

FIGURE 3.59 Generation of PAM. (a) Holding network. (b) Impulse response of holding network. (c) Amplitude response of holding network. (d) Phase response of holding network.

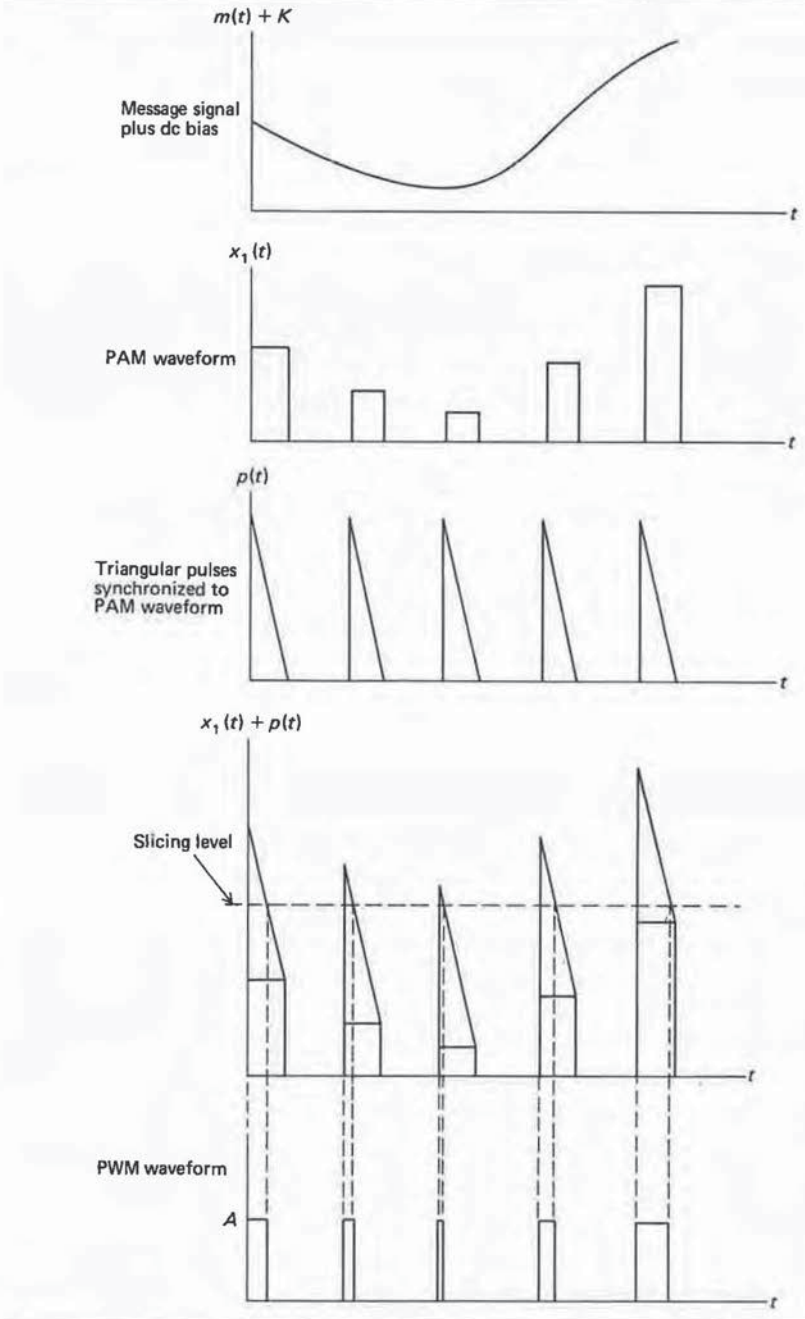


has the transfer function $H_e(f) = 1/H(f)$. If $\tau \ll T_s$, equalization is usually unnecessary. The message signal $m(t)$ can be recovered from the filter output $m_\delta(t)$ by lowpass filtering.

Pulse-Width Modulation (PWM) A PWM waveform consists of a sequence of pulses, each of which having a width proportional to the values of a message signal at the sampling instants. The generation of a PWM waveform is illustrated in Figure 3.60. Since the width of a pulse cannot be negative, a dc bias must be added to $m(t)$ prior to modulation. A PAM waveform $x_1(t)$ is then generated from the biased signal $m(t) + K$. To $x_1(t)$ is added a sequence of synchronized triangular pulses as illustrated. The signal, $x_1(t) + p(t)$, is level-sliced by a circuit that gives an output A when $x_1(t) + p(t)$ is above the slicing level. This signal is the desired PWM waveform.

A PWM waveform can be demodulated by lowpass filtering. The proof of

FIGURE 3.60 Generation of PWM



this statement requires derivation of the spectrum of the PWM signal, which is generally a rather difficult task. An excellent derivation is contained in the book by Schwartz, Bennett, and Stein (1966). The results given here are taken from that reference.

If we assume the message signal

$$m(t) = A \cos 2\pi f_m t \tag{3.251}$$

there are two fundamental frequencies present. These are the signal frequency f_m and the sampling frequency $f_s = 1/T_s$. If these two frequencies are incommensurable, which is certainly the general case, the resulting PWM waveform is nonperiodic, even though the message signal is assumed periodic and sampling is periodic. Despite this difficulty, Schwartz, Bennett, and Stein illustrate a technique that leads to a trigonometric series expression for the PWM signal. The result is

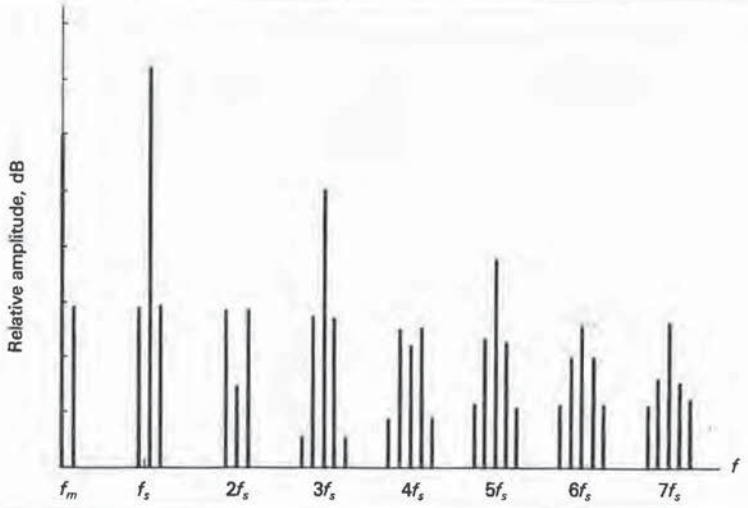
$$\begin{aligned} x(t) = {} & \tfrac{1}{2}h - \tfrac{1}{2}h\beta \cos 2\pi f_m t + \frac{1}{\pi} h \sum_{m=1}^{\infty} [(-1)^m - J_0(m\pi\beta)] \frac{1}{m} \sin 2\pi m f_s t \\ & - \frac{1}{\pi} h \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{1}{m} J_0(m\pi\beta) \left\{ \sin \left[2\pi(mf_s + nf_m)t + \frac{n\pi}{2} \right] \right. \\ & \left. + \sin \left[2\pi(mf_s - nf_m)t + \frac{n\pi}{2} \right] \right\} \end{aligned} \tag{3.252}$$

In this expression, h denotes the amplitude of the PWM pulses and β denotes the modulation index. The modulation index is specified by defining $\beta = 1$ as the value of β resulting in the maximum possible pulse width without causing interference with adjacent pulses. Note that the component at f_m is proportional to β.

The expression for $x(t)$ illustrates the character of the spectrum of a PWM signal with sinusoidal modulation. A dc component having value $\frac{1}{2}h$ is present. A component is also present at the message frequency f_m. The amplitude of this component is proportional to the modulation index β. Lowpass filtering allows the component at f_m, which represents the message $m(t)$, to be recovered from the PWM signal. Also present in the spectrum of $x(t)$ are spectral components at frequencies $mf_s + nf_m$ for all combinations of m and n. These spectral components are intermodulation products between the signal and sampling frequencies and their harmonics.

The spectrum is illustrated in Figure 3.61 for a modulation index $\beta = 0.1$. The sampling frequency f_s is $8f_m$. It can be seen that the spectrum consists of Fourier-Bessel spectra centered about the sampling frequency and harmonics of the sampling frequency. The Fourier-Bessel spectrum centered

FIGURE 3.61 Spectra of PWM signal with $\beta = 0.1$



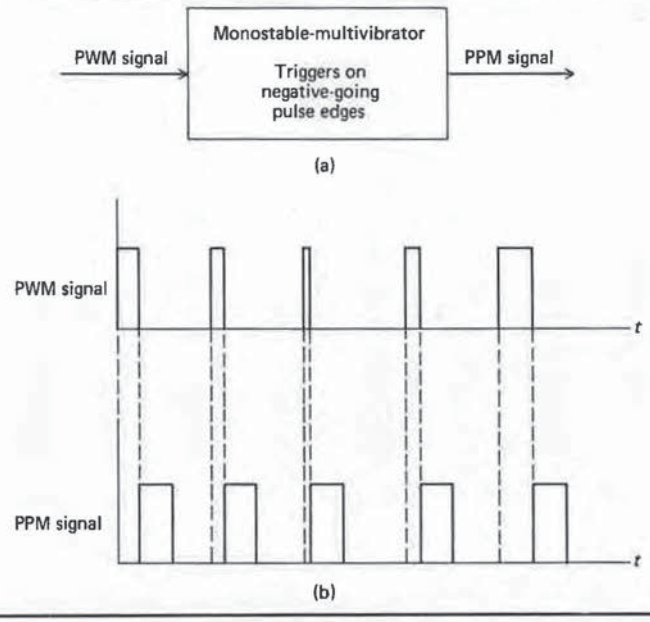
about $f = f_s$ consists of a single component for each sideband corresponding to narrowband modulation ($\beta = 0.1$). The Fourier-Bessel spectrum centered about $f = 7f_s$ consists of a pair of components in each sideband and therefore does not correspond to narrowband modulation. It can be seen from Figure 3.61 and from (3.252) that the effective index is a function of m, the harmonic of f_s. The process is very similar to that encountered in the narrowband to wideband conversion process.

Pulse-Position Modulation (PPM) A PPM signal consists of pulses in which the pulse displacement from a specified time reference is proportional to the sample values of the information-bearing signal. Just as for PWM, a dc bias must be added to the message signal prior to modulation so that the input to the PPM modulator is nonnegative for all values of time. As a matter of fact PPM is easily generated from PWM, as illustrated in Figure 3.62(a). The PWM signal is placed on the input of a monostable multivibrator, which is triggered on by negative-going transitions of the input signal. The on-time of the monostable multivibrator is fixed. The resulting PPM waveform for an assumed PWM waveform is illustrated in Figure 3.62(b).

The PPM signal can be represented by the expression

$$x(t) = \sum_{n=-\infty}^{\infty} g(t - t_n) \tag{3.253}$$

FIGURE 3.62 Generation of PPM from PWM. (a) Generation. (b) Waveforms.



where $g(t)$ represents the shape of the individual pulses and the occurrence times t_n are related to the values of the message signal $m(t)$ at the sampling instants nT_s, as discussed in the preceding paragraph. Determining the spectrum of a PPM signal is a process very similar to determine that of a PWM signal, which we just discussed. This problem is also considered by Schwartz, Bennett, and Stein (1966). If we assume the message signal

$$m(t) = A \cos 2\pi f_m t \tag{3.254}$$

we can write the PPM signal as

$$x(t) = f_s - \beta f_m \sin 2\pi f_m t + \sum_{m=1}^{\infty} \sum_{n=-\infty}^{\infty} \frac{1}{2m} (mf_s + nf_m) J_n(m\pi\beta) \cdot \cos\left[2\pi(mf_s + nf_m)t + \frac{n\pi}{2}\right] \tag{3.255}$$

assuming that $g(t)$ can be represented by an impulse function. If one uses a pulse shape $g(t)$ that cannot be approximated by an impulse function, one obtains the resulting expression for $x(t)$ by weighting each term in the pre-

ceding expression by the Fourier transform of $g(t)$ evaluated at the frequency of the respective term. For example, if $g(t)$ is an even pulse with real transform $G(f)$, the resulting expression for $x(t)$ is

$$x(t) = f_s G(0) - \beta f_m G(f_m) \sin 2\pi f_m t + \sum_{m=1}^{\infty} \sum_{n=-\infty}^{\infty} \frac{1}{2m} (mf_s + nf_m) J_n(m\pi\beta) G(mf_s + nf_m) \cdot \cos\left[2\pi(mf_s + nf_m)t + \frac{n\pi}{2}\right] \quad (3.256)$$

As in the case of PWM, β is the modulation index. A value of 1 for this index results in the largest allowable value of t_n without interference with adjacent pulse positions.

It can be seen from the expression for $x(t)$ that the frequency component at f_m is proportional to the derivative of the message signal $m(t)$. Thus, if the spectral component at f_m is extracted from the spectrum of $x(t)$ by lowpass filtering (the dc term can be removed by a blocking capacitor), the output of the filter is the derivative of $m(t)$ scaled by a constant. We can obtain the message signal $m(t)$ by integrating the filter output. In cases in which the sampling frequency is sufficiently high, the lowpass filter can be omitted, since an integrator also performs lowpass filtering.

The integration can be eliminated by converting the PPM signal to a PWM signal and then demodulating the PWM signal by lowpass filtering. PPM-to-PWM conversion is essentially the reverse of the operations shown in Figure 3.62. A train of pulses is generated at frequency f_s. The leading edges of these pulses are synchronized to the zero-displacement position of the PPM pulses. The equation for $x(t)$ illustrates that a frequency component is present at the sampling frequency if $J_0(\pi\beta) \neq 0$. This component can be used to generate the pulses at frequency f_s. These pulses are then turned off by the leading edge of the PPM pulses. It also can be seen from the equation for $x(t)$ that the frequencies present in a PPM signal are exactly the same as the frequencies present in a PWM signal.

Digital Pulse Modulation

In analog pulse modulation systems, the amplitude, width, or position of a pulse can vary over a continuous range in accordance with the message amplitude at the sampling instant. In systems utilizing digital pulse modulation, the transmitted samples take on only discrete values. We now examine two types of digital pulse modulation: delta modulation and pulse-code modulation.

Delta Modulation (DM) *Delta modulation (DM)* is a modulation technique in which the message signal is encoded into a sequence of binary symbols. These binary symbols are represented by the polarity of impulse functions at the modulator output. The electronic circuits to implement both the modulator and the demodulator are extremely simple. It is this electronic simplicity that makes DM an attractive technique.

A block diagram of a delta modulator is illustrated in Figure 3.63(a). The input to the pulse modulator portion of the circuit is

$$d(t) = m(t) - m_s(t) \tag{3.257}$$

where $m(t)$ is the message signal and $m_s(t)$ is a reference waveform. The signal $d(t)$ is hard-limited and multiplied by the pulse-generator output. This yields

$$x_c(t) = \Delta(t) \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \tag{3.258}$$

where $\Delta(t)$ is a hard-limited version of $d(t)$. The preceding expression can be written as

$$x_c(t) = \sum_{n=-\infty}^{\infty} \Delta(nT_s)\delta(t - nT_s) \tag{3.259}$$

Thus the output of the delta modulator is a series of impulses, each having positive or negative polarity depending on the sign of $d(t)$ at the sampling instants. In practical applications, the output of the pulse generator is not, of course, a sequence of impulse functions but rather a sequence of pulses that are narrow with respect to their periods. Impulse functions are assumed here because of the resulting mathematical simplicity.

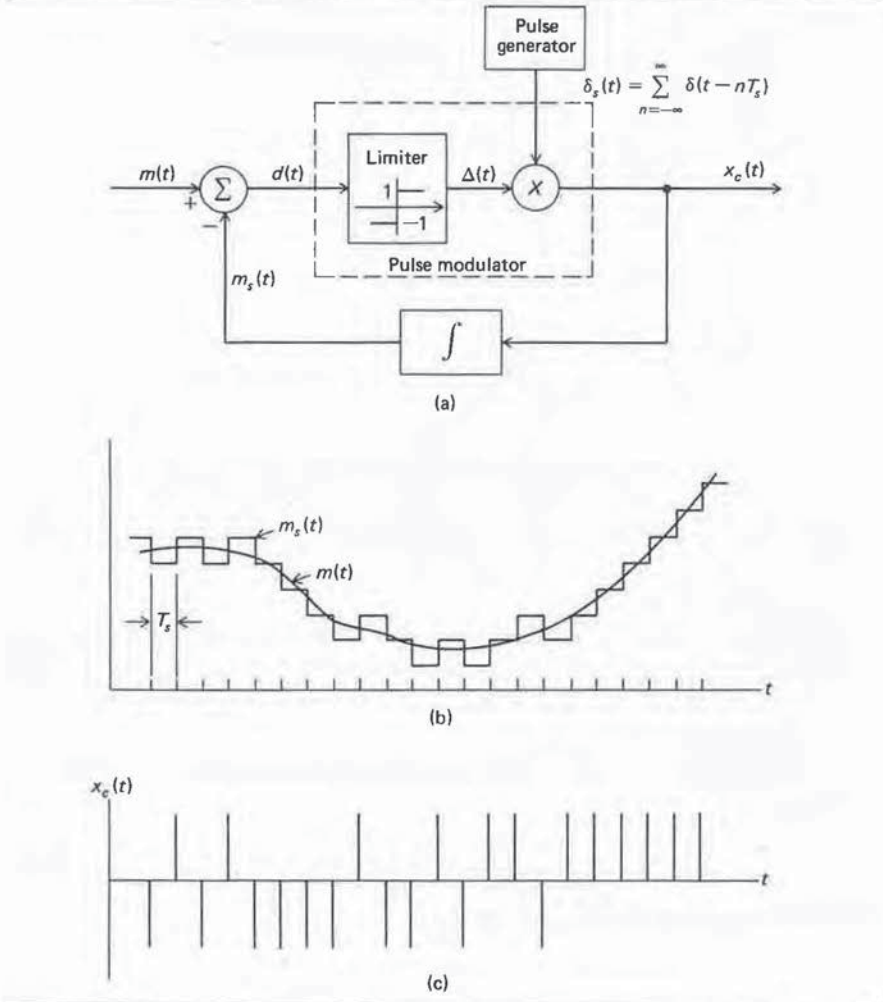
The reference signal $m_s(t)$ is generated by integrating $x_c(t)$. This yields

$$m_s(t) = \sum_{n=-\infty}^{\infty} \Delta(nT_s) \int^t \delta(\alpha - nT_s)\, d\alpha \tag{3.260}$$

which is a stairstep approximation of $m(t)$. The reference signal $m_s(t)$ is shown in Figure 3.63(b) for an assumed $m(t)$. The transmitted waveform $x_c(t)$ is illustrated in Figure 3.63(c).

Demodulation of DM is accomplished by integrating $x_c(t)$ to form the stairstep approximation $m_s(t)$. This signal can then be lowpass-filtered to suppress the discrete jumps in $m_s(t)$. Since a lowpass filter approximates an integrator, it is often possible to eliminate the integrator portion of the

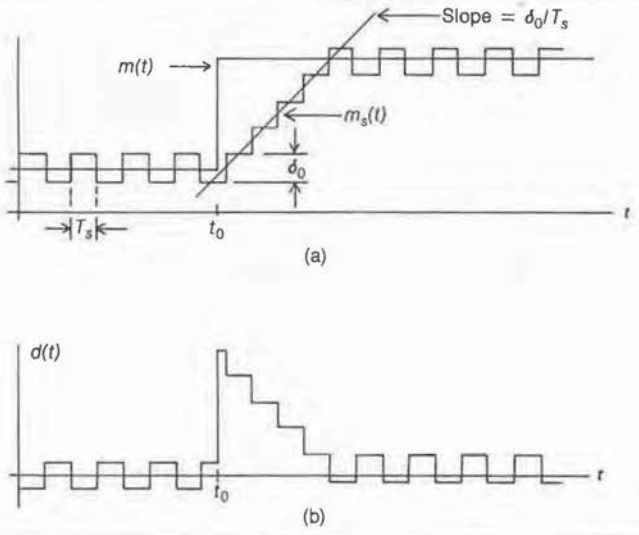
FIGURE 3.63 Delta modulation. (a) Delta modulator. (b) Modulation waveform and stairstep approximation. (c) Modulator output.



demodulator and to demodulate DM simply by lowpass filtering, as was done for PAM and PWM.

A difficulty with DM is the problem of slope overload. *Slope overload* occurs when the message signal $m(t)$ has a slope greater than can be followed by the stairstep approximation $m_s(t)$. This effect is illustrated in Figure 3.64(a), which shows a step change in $m(t)$ at time t_0. Assuming that each pulse in $x_c(t)$ has weight δ_0, the maximum slope that can be followed by $m_s(t)$ is δ_0/T_s, as shown. Figure 3.64(b) shows the resulting error signal due to a step

FIGURE 3.64 Illustration of slope overload. (a) Illustration of $m(t)$ and $m_s(t)$ with step change in $m(t)$. (b) Error between $m(t)$ and $m_s(t)$.



change in $m(t)$ at t_0. It can be seen that significant error exists for some time following the step change in $m(t)$. The duration of the error due to slope overload depends on the amplitude of the step, the impulse weights δ_0, and the sampling period T_s.

A simple analysis can be carried out assuming that the message signal $m(t)$ is a sinusoid:

$$m(t) = A \sin 2\pi f_1 t \tag{3.261}$$

The maximum slope that $m_s(t)$ can follow is

$$S_m = \frac{\delta_0}{T_s} \tag{3.262}$$

The derivative of $m(t)$ is

$$\frac{d}{dt} m(t) = 2\pi A f_1 \cos 2\pi f_1 t \tag{3.263}$$

It follows that $m_s(t)$ can follow $m(t)$ without slope overload if

$$\frac{\delta_0}{T_s} \geq 2\pi A f_1 \tag{3.264}$$

This example implies that a bandwidth constraint exists on $m(t)$ if slope overload is to be avoided.

One technique for overcoming the problem of slope overload is to modify the modulator, as shown in Figure 3.65. The result is known as *adaptive delta modulation*. The system is explained by recognizing that the weights δ_0, and consequently the step size of $m_s(t)$, can be very small if $m(t)$ is nearly constant. A rapidly changing $m(t)$ requires a larger value of δ_0 if slope overload is to be avoided. A lowpass filter is used as shown, with $x_c(t)$ as its input. If $m(t)$ is constant or nearly constant, the pulses constituting $x_c(t)$ will alternate in sign. Thus the dc value, determined over the time constant of the lowpass filter, is nearly zero. This small value controls the gain of the variable-gain amplifier such that it is very small under this condition. Thus δ_0 is made small at the integrator input. The square-law or magnitude device is used to ensure that the control voltage and amplifier gain $g(t)$ are always positive. If $m(t)$ is increasing or decreasing rapidly, the pulses $x_c(t)$ will have the same polarity over this period. Thus the magnitude of the output of the lowpass filter will be relatively large. The result is an increase in the gain of the variable-gain amplifier and consequently an increase in δ_0. This in turn reduces the time span of significant slope overload.

FIGURE 3.65 Adaptive delta modulator

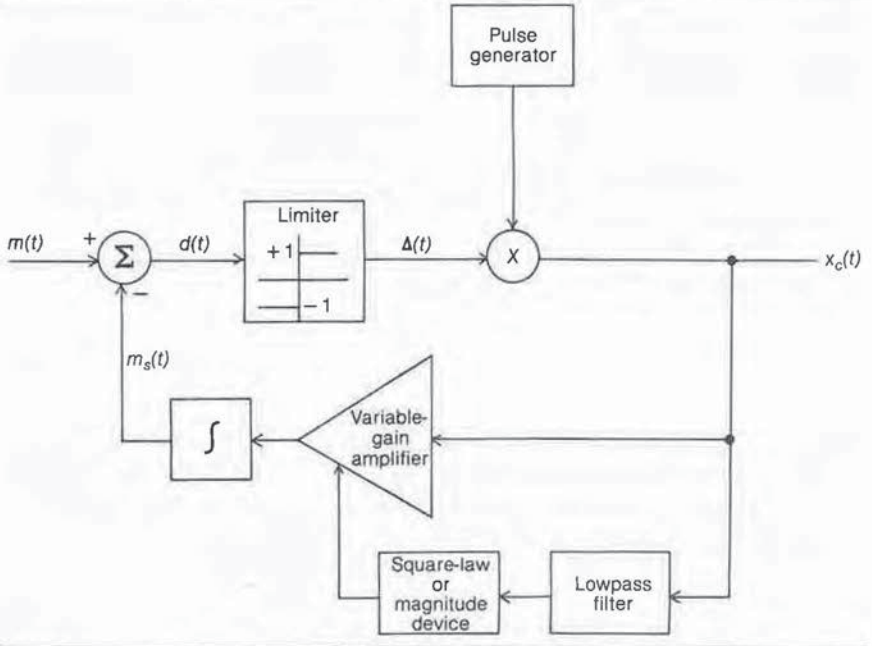
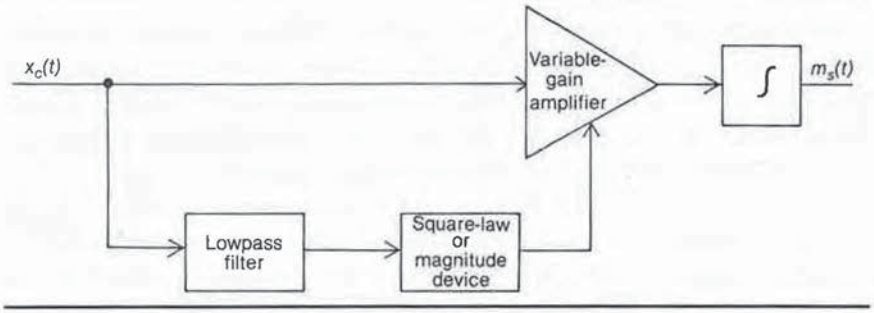


FIGURE 3.66 Adaptive DM receiver



The use of an adaptive delta modulator requires that the receiver be adaptive also, so that the step size at the receiver changes to match the changes in δ_0 at the modulator. This is illustrated in Figure 3.66.

Pulse-Code Modulation (PCM) The generation of PCM is a three-step process, as illustrated in Figure 3.67(a). The message signal $m(t)$ is first sampled. The sample values are then quantized. In PCM, the quantization level of each sample is the transmitted quantity instead of the sample value. Typically, the quantization level is encoded into a binary sequence, as shown in Figure 3.67(b). The modulator output is a pulse representation of the binary sequence, which is shown in Figure 3.67(c). A binary "one" is represented as a pulse, and a binary "zero" is represented as the absence of a pulse. This absence of a pulse is indicated by a dashed line in Figure 3.67(c). The PCM waveform of Figure 3.67(c) shows that a PCM system requires synchronization so that the starting points of the digital words can be determined at the demodulator.

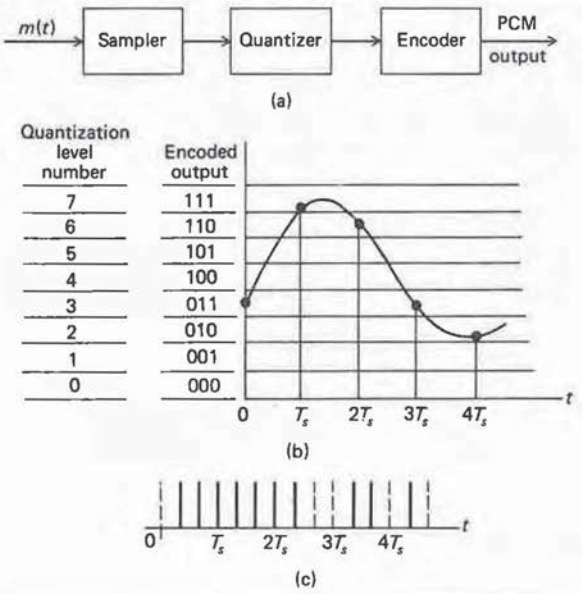
To consider the bandwidth requirements of a PCM system, suppose that q quantization levels are used, satisfying

$$q = 2^n \tag{3.265}$$

where n is an integer. For this case, $n = \log_2 q$ binary pulses must be transmitted for each sample of the message signal. If this signal has bandwidth W and the sampling rate is $2W$, then $2nW$ binary pulses must be transmitted per second. Thus the maximum width of each binary pulse is

$$(\Delta\tau)_{\max} = \frac{1}{2nW} \tag{3.266}$$

FIGURE 3.67 Generation of PCM. (a) PCM modulator. (b) Quantization and encoding. (c) Transmitted output.



We saw in Section 2.7 that the bandwidth required for transmission of a pulse is inversely proportional to the pulse width, so that

$$B = knW \tag{3.267}$$

where B is the required bandwidth of the PCM system and k is a constant of proportionality. This yields

$$B \cong 2Wk \log_2 q \tag{3.268}$$

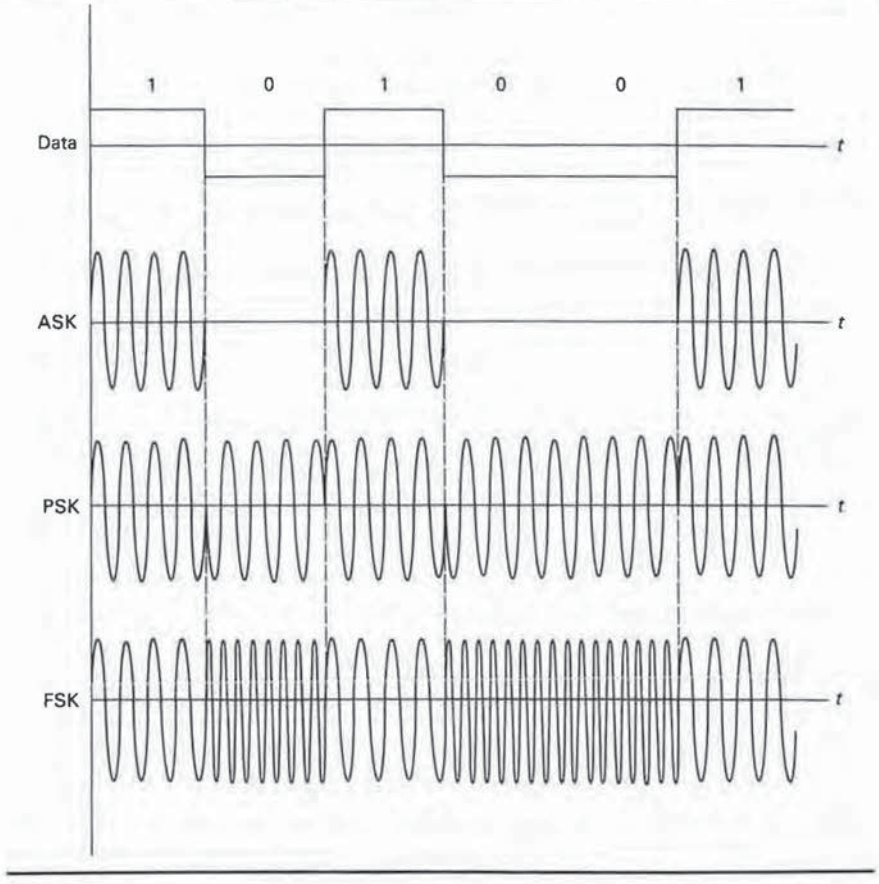
This represents a lower bound on bandwidth, since we have assumed both a minimum sampling rate and a minimum value of bandwidth for transmitting a pulse. Equation (3.268) shows that the PCM signal bandwidth is proportional to W and the logarithm of q.

If the major source of error in the system is quantization error, it follows that a small error requirement dictates large transmission bandwidth. Thus, in a PCM system, error can be exchanged for bandwidth. We shall see that this behavior is typical of many nonlinear systems operating in noisy environments. However, before noise effects can be analyzed, we must take a detour and develop the theory of probability and random processes. Knowl-

edge of this area enables one to accurately model realistic and practical communication systems operating in everyday, nonidealized environments.

The pulses representing the sample values of a PCM waveform can be transmitted on an RF carrier by the use of amplitude, phase, or frequency modulation. This is illustrated in Figure 3.68 for the case in which the data bits are represented by an NRZ (non–return to zero) waveform. Six bits are shown, corresponding to the data sequence 101001. For digital amplitude modulation, known as *amplitude shift keying (ASK)*, the carrier amplitude is determined by the data bit for that interval. For digital phase modulation, known as *phase-shift keying (PSK)*, the phase of the carrier is established by the data bit. The phase changes can be clearly seen in Figure 3.68. For digital

FIGURE 3.68 An example of digital modulation schemes



frequency modulation, known as *frequency-shift keying (FSK)*, the carrier frequency is established by the data bit. These techniques will be studied in detail in Chapter 7.

3.6 MULTIPLEXING

In many applications, a large number of data sources are located at a common point, and it is desirable to transmit these signals simultaneously using a single communication channel. This is accomplished using multiplexing. We will now examine several different types of multiplexing, each having advantages and disadvantages.

Frequency-Division Multiplexing (FDM)

Frequency-division multiplexing (FDM) is a technique whereby several message signals are translated, using modulation, to different spectral locations and added to form a baseband signal. The carriers used to form the baseband are usually referred to as *subcarriers*. Then, if desired, the baseband signal can be transmitted over a single channel using a single modulation process. Several different types of modulation can be used to form the baseband, as illustrated in Figure 3.69. In this example, there are N information signals contained in the baseband. Observation of the baseband spectrum in Figure 3.69(c) suggests that baseband modulator 1 is a DSB modulator with subcarrier frequency f_1. Modulator 2 is an upper-sideband SSB modulator, and modulator N is an angle modulator.

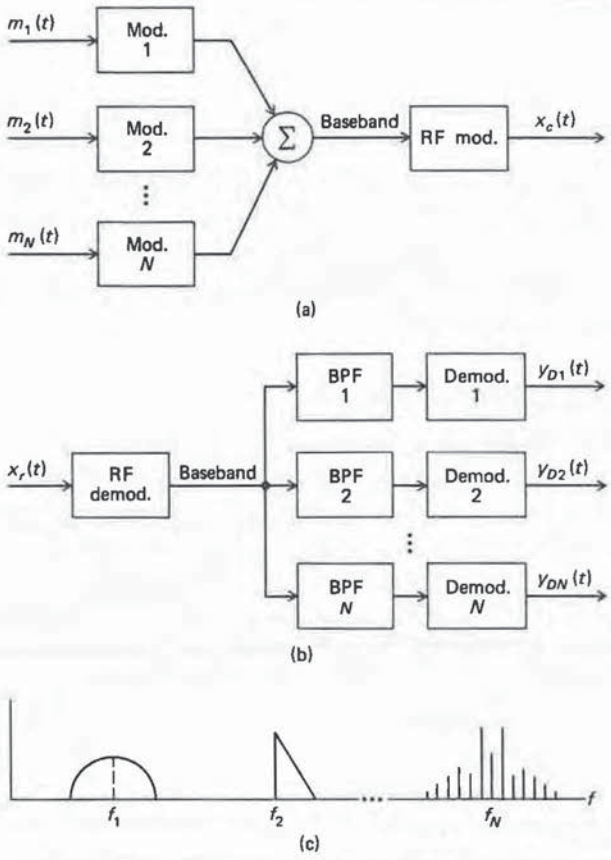
An FDM demodulator is shown in Figure 3.69(b). The RF demodulator output is ideally the baseband signal. The individual channels in the baseband are extracted using bandpass filters. The bandpass filter outputs are demodulated in the conventional manner.

Observation of the baseband spectrum illustrates that the baseband bandwidth is equal to the sum of the bandwidths of the modulated signals plus the sum of the *guardbands*, the empty spectral bands between the channels necessary for filtering. This bandwidth is lower-bounded by the sum of the bandwidths of the message signals. This bandwidth,

$$B = \sum_{i=1}^{N} W_i \tag{3.269}$$

where W_i is the bandwidth of $m_i(t)$, is achieved when all baseband modulators are SSB and all guardbands have zero width.

FIGURE 3.69 Frequency-division multiplexing. (a) FDM modulator. (b) FDM demodulator. (c) Baseband spectrum.

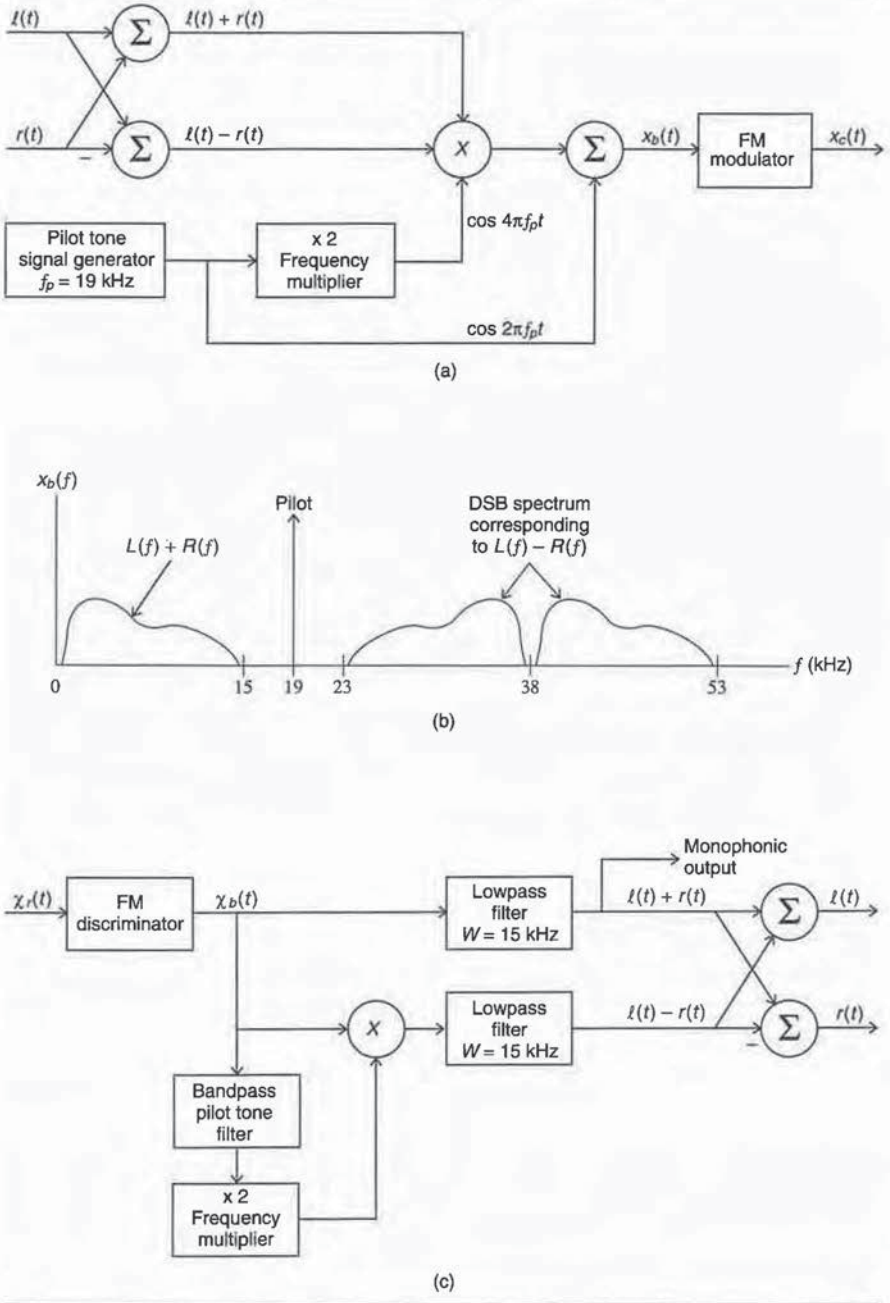


An Example of Frequency-Division Multiplexing: Stereophonic FM Broadcasting

As an example of frequency-division multiplexing we now consider stereophonic FM broadcasting. A necessary condition established in the development of stereophonic FM is that stereo FM be compatible with monophonic FM receivers. In other words, the output from a monophonic FM receiver must be the composite (left-channel plus right-channel) stereo signal.

The scheme adopted for stereophonic FM broadcasting is shown in Figure 3.70(a). As can be seen, the first step in the generation of a stereo FM signal is to first form the left-plus-right channel signal, $l(t) + r(t)$, and the left-minus-right channel signal, $l(t) - r(t)$. The left-minus-right channel signal is

FIGURE 3.70 Stereophonic FM transmitter and receiver. (a) Stereophonic FM transmitter. (b) Single-sided spectrum of FM baseband signal. (c) Stereophonic FM receiver.



then translated to 38 kHz using DSB modulation with a carrier derived from a 19-kHz oscillator. A frequency doubler is used to generate a 38-kHz carrier from a 19-kHz oscillator. We previously saw that a phase-lock loop could be used to implement a frequency doubler.

The baseband signal is found by adding the left-plus-right channel signal, the left-minus-right channel signal, and the 19-kHz pilot tone. The spectrum of the baseband signal is shown in Figure 3.70 for assumed left-channel and right-channel signals. The baseband signal is the input to the FM modulator. It is important to note that if a monophonic FM transmitter, having a message bandwidth of 15 kHz, and a stereophonic FM transmitter, having a message bandwidth of 53 kHz, both have the same constraint on the peak deviation, the deviation ratio, D, of the stereophonic FM transmitter is reduced by a factor of $53/15 = 3.53$. The impact of this reduction in the deviation ratio will be seen when we consider noise effects in Chapter 6.

The block diagram of a stereophonic FM receiver is shown in Figure 3.70(c). The output of the FM discriminator is the baseband signal $x_b(t)$ which, under ideal conditions, is identical to the baseband signal at the input to the FM modulator. As can be seen from the spectrum of the baseband signal, the left-plus-right channel signal can be generated by filtering the baseband signal with a lowpass filter having a bandwidth of 15 kHz. Note that this signal constitutes the monophonic output. The left-minus-right channel signal is obtained by coherently demodulating the DSB signal using a 38-kHz demodulation carrier. This coherent demodulation carrier is obtained by recovering the 19-kHz pilot using a bandpass filter and then using a frequency doubler as was done in the modulator. The left-plus-right channel signal and the left-minus-right channel signal are added and subtracted, as shown in Figure 3.70(c) to generate the left channel signal and the right channel signal.

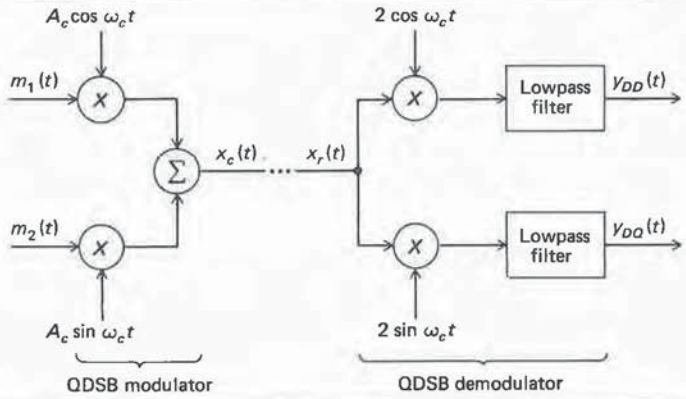
Quadrature Multiplexing (QM)

Another type of multiplexing is *quadrature multiplexing (QM)*, in which quadrature carriers are used for frequency translation. For the system shown in Figure 3.71, the signal

$$x_c(t) = A_c[m_1(t) \cos \omega_c t + m_2(t) \sin \omega_c t] \tag{3.270}$$

is a quadrature-multiplexed signal. By sketching the spectra of $m_1(t) \cos \omega_c t$ and $m_2(t) \sin \omega_c t$, we see these spectra overlap if the spectra of $m_1(t)$ and $m_2(t)$ overlap. Thus, strictly speaking, quadrature multiplexing is not a frequency-division technique.

FIGURE 3.71 Quadrature multiplexing



A QM signal is demodulated by using quadrature demodulation carriers. To show this, multiply $x_r(t)$ by 2 cos $(\omega_c t + \theta)$. This yields

$$2x_r(t) \cos(\omega_c t + \theta) = A_c[m_1(t) \cos\theta - m_2(t) \sin\theta + m_1(t) \cos(2\omega_c t + \theta) + m_2(t) \sin(2\omega_c t + \theta)] \quad (3.271)$$

The last two terms in the preceding expression can be removed by using a lowpass filter. The output of the lowpass filter is

$$y_{DD}(t) = A_c[m_1(t) \cos\theta - m_2(t) \sin\theta] \quad (3.272)$$

which yields $m_1(t)$, the desired output for $\theta = 0$. The quadrature channel is demodulated using a demodulation carrier of the form 2 sin $\omega_c t$.

The preceding result illustrates the effect of a demodulation phase error on QM. The result of this phase error is both an attenuation, which can be time-varying, of the desired signal and crosstalk from the quadrature channel. It should be noted that QM can be used to represent both DSB and SSB with appropriate definitions of $m_1(t)$ and $m_2(t)$.

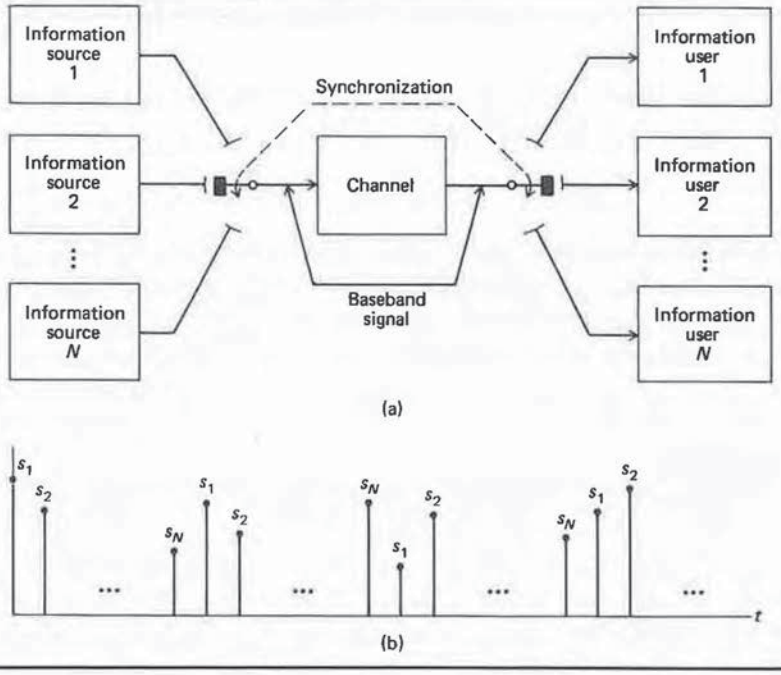
Frequency-division multiplexing can be used with quadrature multiplexing by translating pairs of signals, using quadrature carriers, to each subcarrier frequency. Each channel has bandwidth $2W$ and accommodates two message signals, each having bandwidth W. Thus, assuming zero-width guardbands, a baseband of bandwidth NW can accommodate N message signals, each of bandwidth W, and requires $\frac{1}{2}N$ separate subcarrier frequencies.

Time-Division Multiplexing (TDM)

Time-division multiplexing is best understood by considering Figure 3.72(a). The data sources are assumed to have been sampled at the Nyquist rate or higher. The commutator then interlaces the samples to form the baseband signal shown in Figure 3.72(b). At the channel output, the baseband signal is demultiplexed by using a second commutator as illustrated. Proper operation of this system depends on proper synchronization between the two commutators.

If all message signals have equal bandwidth, then the samples are transmitted sequentially, as shown in Figure 3.72(b). If the sampled data signals have unequal bandwidths, more samples must be transmitted per unit time from the wideband channels. This is easily accomplished if the bandwidths are harmonically related. For example, assume that a TDM system has four channels of data. Also assume that the bandwidth of the first and second data sources, $s_1(t)$ and $s_2(t)$, is W Hz, the bandwidth of $s_3(t)$ is $2W$ Hz, and the bandwidth of $s_4(t)$ is $4W$ Hz. It is easy to show that a permissible

FIGURE 3.72 Time-division multiplexing. (a) Time-division multiplexing system. (b) Baseband signal.



sequence of baseband samples is a periodic sequence, one period of which is $\ldots s_1 s_4 s_3 s_4 s_2 s_4 s_3 s_4 \ldots$.

The minimum bandwidth of a TDM baseband is easy to determine using the sampling theorem. Assuming Nyquist rate sampling, the baseband contains $2W_iT$ samples from the ith channel in each T-second interval, where W_i is the bandwidth of the ith channel. Thus the total number of baseband samples in a T-second interval is

$$n_s = \sum_{i=1}^{N} 2W_iT \tag{3.273}$$

Assuming that the baseband is a lowpass signal of bandwidth B, the required sampling rate is $2B$. In a T-second interval, we then have $2BT$ total samples. Thus

$$n_s = 2BT = \sum_{i=1}^{N} 2W_iT \tag{3.274}$$

or

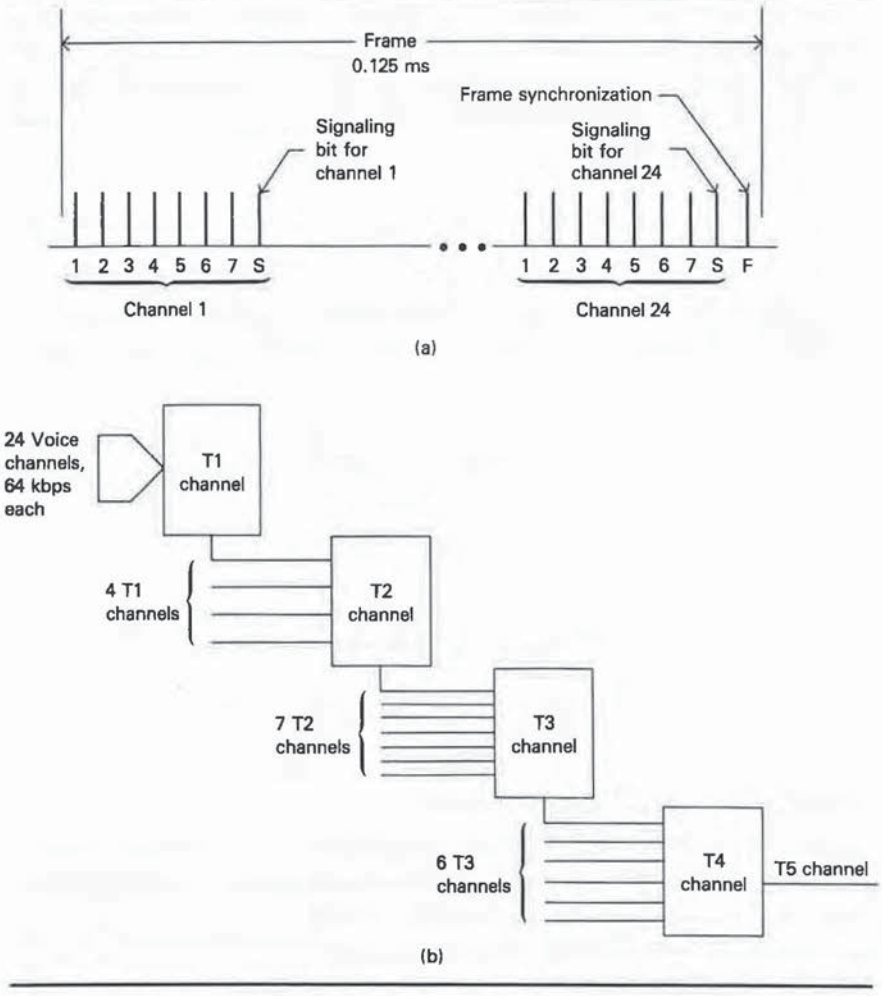
$$B = \sum_{i=1}^{N} W_i \tag{3.275}$$

which is the same as the minimum required bandwidth obtained for FDM.

An Example: The Digital Telephone System

As an example of a digital TDM system, we consider a multiplexing scheme common to many telephone systems. The sampling format is illustrated in Figure 3.73(a). A voice signal is sampled at 8000 samples per second, and each sample is quantized into seven binary digits. An additional binary digit, known as a *signaling bit,* is added to the basic seven bits that represent the sample value. The signaling bit is used in establishing calls and for synchronization. Thus eight bits are transmitted for each sample value, yielding a bit rate of 64,000 bits per second (64 kbps). Twenty-four of these 64 kbps voice channels are grouped together to yield a T1 carrier. The T1 frame consists of $24(8) + 1 = 193$ bits. The extra bit is used for frame synchronization. The frame duration is the reciprocal of the fundamental sampling frequency, or 0.125 ms. Since the frame rate is 8000 frames per second, with 193 bits per frame, the T1 data rate is 1.544 Mbps.

FIGURE 3.73 Digital multiplexing scheme for digital telephone. (a) T1 frame. (b) Digital multiplexing.



As shown in Figure 3.73(b), four T1 carriers can be multiplexed to yield a T2 carrier, which consists of 96 voice channels. Seven T2 carriers yield a T3 carrier, and 6 T3 carriers yield a T4 carrier. The bit rate of a T4 channel, consisting of 4032 voice channels with signaling bits and framing bits, is 274.176 Mbps. A T1 link is typically used for short transmission distances in areas of heavy usage. T4 and T5 channels are used for long transmission distances.

Comparison of Multiplexing Schemes

We have seen that for all three types of multiplexing studied, the baseband bandwidth is lower-bounded by the total information bandwidth. However, there are advantages and disadvantages to each multiplexing technique.

The basic advantage of FDM is simplicity of implementation, and if the channel is linear, disadvantages are difficult to identify. However, many channels have small, but nonnegligible nonlinearities. As we saw in Chapter 2, nonlinearities lead to intermodulation distortion. In FDM systems, the result of intermodulation distortion is crosstalk between channels in the baseband. This problem is avoided in TDM systems.

However, TDM also has inherent disadvantages. Samplers are required, and if continuous data are required by the data user, the continuous waveforms must be reconstructed from the samples. One of the biggest difficulties with TDM is maintaining synchronism between the multiplexing and demultiplexing commutators.

The basic advantage of QM is that QM allows simple DSB modulation to be used while at the same time making efficient use of baseband bandwidth. It also allows dc response, which SSB does not. The basic problem with QM is crosstalk between the quadrature channels, which results if perfectly coherent demodulation carriers are not available. Of course, if QM and FDM are used together, intermodulation becomes a problem.

Other advantages and disadvantages of FDM, QM, and TDM will become apparent when we study performance in the presence of noise in Chapter 6.

3.7 COMPARISON OF MODULATION SYSTEMS

In this chapter, a large number of different techniques for information transmission have been introduced. It is important that these techniques be compared so that logical choices can be made between the many available systems when various application needs arise. Unfortunately, at this point in our study, such a comparison cannot be accomplished with any rigor. The reason for this is that we have only studied systems in a highly idealized environment. Specifically, the assumption has been made that the received signal at the demodulator input $x_r(t)$ is *exactly* the transmitted signal $x_c(t)$.

In a practical environment, the transmitted signal is subjected to many undesirable perturbations prior to demodulation. The most important perturbation is usually *noise*, which is inadvertently added to the signal at several points in the system. Noise is always present in varying degrees in practical systems, and its effect must be considered in any meaningful comparison of systems.

The noise performance of modulation systems is often specified by comparing the signal-to-noise ratios (SNR) at the input and output of the demodulator. This is illustrated in Figure 3.74. The *predetection filter* is typically the combination of the radio-frequency (RF) and intermediate-frequency (IF) filters. As we shall see in Chapter 6, the parameter of interest at the output of the predetection filter is

$$(\text{SNR})_T = \frac{P_T}{\langle n_i^2(t) \rangle} \tag{3.277}$$

where P_T is the power of the received signal and $\langle n_i^2(t) \rangle$ is the noise power measured in the bandwidth of the message signal.

The signal at the output of the demodulator is

$$y_D(t) = m(t) + n(t) \tag{3.278}$$

where $n(t)$ is the output noise component. The signal-to-noise ratio at this point is

$$(\text{SNR})_D = \frac{\langle m^2(t) \rangle}{\langle n^2(t) \rangle} \tag{3.279}$$

For thermal noise sources, $\langle n_i^2(t) \rangle$ can be obtained from knowledge of the system noise figure. This system figure of merit is determined using techniques developed in Appendix A.

The noise performance of several important systems is illustrated in Figure 3.75. The curves shown give the performance of nonlinear systems only above threshold. The approximate point where threshold occurs is indicated by a heavy dot. Below threshold, the performance degrades rapidly.

Figure 3.75 is not intended to be complete. It is placed here only to illustrate the importance of considering noise effects and to illustrate the superior performance of nonlinear systems over certain ranges of signal-to-

FIGURE 3.74 Receiver block diagram

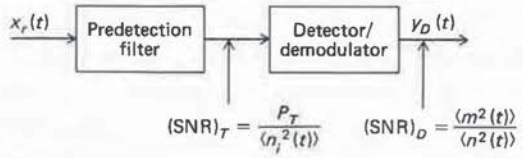
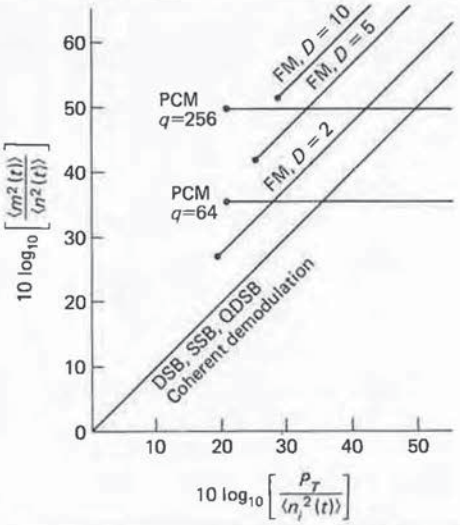


FIGURE 3.75 Noise performance curves



noise ratios. The noise performance of these systems is the subject of Chapter 6. At that time, the results shown in Figure 3.75 will be derived in detail.

For completeness, we include Table 3.4, which compares continuous-wave analog systems. The list of applications is by no means complete but does indicate typical uses.

SUMMARY

1. Modulation is the process by which a parameter of a carrier is varied in one-to-one correspondence with an information-bearing signal usually referred to as the message. Several uses of modulation are to achieve efficient transmission, to allocate channels, and for multiplexing.
2. If the carrier is continuous, the modulation is continuous-wave modulation. If the carrier is a sequence of pulses, the modulation is pulse modulation.
3. There are two basic types of continuous-wave modulation: linear modulation and angle modulation.
4. Assume that a general modulated carrier is given by

$$x_c(t) = A(t) \cos [\omega_c t + \phi(t)]$$

TABLE 3.4 A Comparison of Analog Systems

	Bandwidth	DC Response	Efficiency	Complexity	Typical Applications
SB	$2W$	Yes	100%	*Moderate.* Coherent demodulation is required	Low bandwidth communication systems
M	$2W$	No	<50%	*Minor.* Simple modulators and envelope detection.	Broadcast radio
SB	W	No	100%	*Major.* Phase shift modulators and coherent demodulators are required.	Voice communication systems
SB	$W+$	Yes	100%	*Major.* Symmetric filters and coherent demodulation are required.	Wideband systems
SB + rrier	$W+$	No	<50%	*Moderate.* Symmetric filter required, but envelope detection can be used.	TV video
M	$2(D+1)W$	Yes	Not applicable	*Moderate.* Simple phase-lock loop demodulators can be used.	High-fidelity broadcast radio
M	$2(D+1)W$	Yes (with calibration)	Not applicable	*Moderate.* Essentially the same as FM.	Data transmission; often used in the generation of FM

If $A(t)$ is proportional to the message signal, the result is linear modulation. If $\phi(t)$ is proportional to the message signal, the result is phase modulation (PM). If the time derivative of $\phi(t)$ is proportional to the message signal, the result is frequency modulation (FM). Both PM and FM are examples of angle modulation. Angle modulation is a nonlinear process.

5. The simplest example of linear modulation is double sideband (DSB). DSB is implemented as a simple product device, and coherent demodulation must be used.

6. If a carrier component is added to a DSB signal, the result is amplitude modulation (AM). This is a useful modulation technique because it allows simple envelope detection to be used.

7. The efficiency of a modulation process is defined as the percentage of total power that conveys information. For AM, this is given by

$$E = \frac{a^2\langle m_n^2(t)\rangle}{1 + a^2\langle m_n^2(t)\rangle}\,(100\%)$$

where the parameter a is known as the *modulation index* and $m_n(t)$ is $m(t)$ normalized so that the peak value is unity. If envelope demodulation is used, the index must be less than unity.

8. A single-sideband (SSB) signal is generated by transmitting only one of the sidebands in a DSB signal. SSB signals are generated either by sideband filtering a DSB signal or by using a phase-shift modulator. SSB signals can be written as

$$x_c(t) = \tfrac{1}{2}A_c m(t)\cos\omega_c t \pm \tfrac{1}{2}A_c\hat{m}(t)\sin\omega_c t$$

in which the plus sign is used for lower-sideband SSB and the minus sign is used for upper-sideband SSB. These signals can be demodulated either through the use of coherent demodulation or through the use of carrier reinsertion.

9. Vestigial sideband (VSB) results when a vestige of one sideband appears on an otherwise SSB signal. VSB is easier to generate than SSB. Demodulation can be coherent, or carrier reinsertion can be used.

10. Switching modulators, followed by filters, can be used as implementations of product devices.

11. Frequency translation is accomplished by multiplying a signal by a carrier and filtering. These systems are known as *mixers*. The concept of mixing is used in superheterodyne receivers. Mixing results in *image frequencies*, which can be troublesome.

12. The general expression for an angle-modulated signal is

$$x_c(t) = A_c\cos[\omega_c t + \phi(t)]$$

For a PM signal, $\phi(t)$ is given by

$$\phi(t) = k_p m(t)$$

and for an FM signal, it is

$$\phi(t) = 2\pi f_d\int^t m(\alpha)\,d\alpha$$

where k_p and f_d are the phase and frequency deviation constants, respectively.

13. Angle modulation results in an infinite number of sidebands. If only a single pair of sidebands is significant, the result is narrowband angle modulation. Narrowband angle modulation, with sinusoidal message, has approximately the same spectrum as a DSB signal except for a 180° phase shift of the lower sideband.

14. An angle-modulated carrier with a sinusoidal message signal can be expressed as

$$x_c(t) = A_c \sum_{n=-\infty}^{\infty} J_n(\beta) \cos(\omega_c + n\omega_m)t$$

The term $J_n(\beta)$ is the Bessel function of the first kind of order n and argument β. The parameter β is known as the *modulation index*. If $m(t) = A \sin \omega_m t$, then $\beta = k_p A$ for PM, and $\beta = f_d A/f_m$ for FM.

15. The power contained in an angle-modulated carrier is $\langle x_c^2(t) \rangle = \frac{1}{2}A_c^2$ if the carrier frequency is large compared to the bandwidth of the modulated carrier.

16. The bandwidth of an angle-modulated signal is, strictly speaking, infinite. However, a measure of the bandwidth can be obtained by defining the power ratio

$$P_r = J_0^2(\beta) + 2 \sum_{n=1}^{k} J_n^2(\beta)$$

which is the ratio of the total power $\frac{1}{2}A_c^2$ to the power in the bandwidth $B = 2kf_m$. A power ratio of 0.98 yields $B = 2(\beta + 1)f_m$.

17. The deviation ratio of an angle-modulated signal is

$$D = \frac{\text{peak frequency deviation}}{\text{bandwidth of } m(t)}$$

18. Carson's rule for estimating the bandwidth of an angle-modulated carrier with an arbitrary message signal is $B = 2(D + 1)W$.

19. Narrowband-to-wideband conversion is a technique whereby a wideband FM signal is generated from a narrowband FM signal. The system makes use of a frequency multiplier, which, unlike a mixer, multiplies the deviation as well as the carrier frequency.

20. Demodulation of an angle-modulated signal is accomplished through the use of a discriminator. This device yields an output signal propor-

tional to the frequency deviation of the input signal. Placing an integrator at the discriminator output allows PM signals to be demodulated.

21. An FM discriminator can be implemented as a differentiator followed by an envelope detector. Bandpass limiters are used at the differentiator input to eliminate amplitude variations.
22. *Interference,* the presence of undesired signal components, can be a problem in demodulation. Interference at the input of a demodulator results in undesired components at the demodulator output. If the interference is large and if the demodulator is nonlinear, thresholding can occur. The result of this is a complete loss of the signal component.
23. Interference is also a problem in angle modulation. In FM systems, the effect of interference is a function of both the amplitude and frequency of the interfering tone. In PM systems, the effect of interference is a function only of the amplitude of the interfering tone. For FM interference can be reduced by the use of preemphasis and deemphasis.
24. A phase-lock loop (PLL) is a simple and practical system for the demodulation of angle-modulated signals. It is a feedback control system and is analyzed as such. Phase-lock loops also provide simple implementations of frequency multipliers and frequency dividers.
25. Placing a bandpass filter and a discriminator within a phase-lock loop yields a frequency-compressive feedback system. This system is also useful for demodulation of angle-modulated signals.
26. The Costas phase-lock loop, which is a variation of the phase-lock loop, is a system for the demodulation of DSB signals.
27. Analog pulse modulation results when the message signal is sampled and a pulse train carrier is used. A parameter of each pulse is varied in one-to-one correspondence with the value of each sample.
28. Pulse-amplitude modulation (PAM) results when the *amplitude* of each carrier pulse is proportional to the value of the message signal at each sampling instant. PAM is essentially a sample-and-hold operation. Demodulation of PAM is accomplished by lowpass filtering.
29. Pulse-width modulation (PWM) results when the *width* of each carrier pulse is proportional to the value of the message signal at each sampling instant. Demodulation of PWM is also accomplished by lowpass filtering.
30. Pulse-position modulation (PPM) results when the *position* of each carrier pulse, as measured by the displacement of each pulse from a fixed ref-

erence, is proportional to the value of the message signal at each sampling instant.

31. Digital pulse modulation results when the sample values of the message signal are quantized and encoded prior to transmission.

32. Delta modulation (DM) is an easily implemented form of digital pulse modulation. In DM, the message signal is encoded into a sequence of binary symbols. The binary symbols are represented by the polarity of impulse functions at the modulator output. Demodulation is ideally accomplished by integration, but lowpass filtering is often a simple and satisfactory substitute.

33. Pulse-code modulation (PCM) results when the message signal is sampled and quantized, and each quantized sample value is encoded as a sequence of binary symbols. PCM differs from DM in that in PCM each quantized sample value is transmitted but in DM the transmitted quantity is the polarity of the change in the message signal from one sample to the next.

34. Multiplexing is a scheme allowing two or more message signals to be communicated simultaneously using a single system.

35. Frequency-division multiplexing (FDM) results when simultaneous transmission is accomplished by translating message spectra, using modulation, to *nonoverlapping* locations in a baseband spectrum. The baseband signal is then transmitted using any carrier modulation method.

36. Quadrature multiplexing (QM) results when two message signals are translated, using linear modulation with quadrature carriers, to the *same* spectral locations. Demodulation is accomplished coherently using quadrature demodulation carriers. A phase error in a demodulation carrier results in serious distortion of the demodulated signal. This distortion has two components: a time-varying attenuation of the desired output signal and crosstalk from the quadrature channel.

37. Time-division multiplexing (TDM) results when samples from two or more data sources are interlaced, using commutation, to form a baseband signal. Demultiplexing is accomplished by using a second commutator, which must be synchronous with the multiplexing commutator.

38. Any meaningful comparison of systems must take into consideration a large number of factors. Examples are cost or complexity of both transmitter and receiver, the bandwidth requirements, compatibility with existing systems, and performance, especially in the presence of noise.

FURTHER READING

One can find basic treatments of modulation theory at about the same technical level of this text in a wide variety of books. Examples are Carlson (1986), Stremler (1982), Haykin (1989), Lathi (1983), Roden (1985), Shanmugam (1979), and Couch (1990). Taub and Schilling (1986) have an excellent treatment of phase-lock loops. The performance of phase-lock loops in the absence of noise is discussed by Viterbi (1966, Chapters 2 and 3) and Gardner (1979). Also, as we mentioned in this chapter, Schwartz, Bennett, and Stein (1966) have an excellent section of PWM and PPM.

PROBLEMS

Section 3.1

3.1 Assume that a DSB signal

$$x_c(t) = A_c m(t) \cos(\omega_c t + \phi_0)$$

is demodulated using the demodulation carrier $2 \cos [\omega_c t + \theta(t)]$. Determine, in general, the demodulated output $y_D(t)$. Let $\theta(t) = \theta_0$, where θ_0 is a constant, and determine the mean-square error between $m(t)$ and the demodulated output as a function of ϕ_0 and θ_0. Now let $\theta(t) = \omega_0 t$ and compute the mean-square error between $m(t)$ and the demodulated output.

3.2 Show that an AM signal can be demodulated using coherent demodulation by assuming a demodulation carrier of the form

$$2 \cos [\omega_c t + \theta(t)]$$

where $\theta(t)$ is the demodulation phase error. Why is coherent demodulation of AM not used in practice?

3.3 Design an envelope detector that uses a full-wave rectifier rather than the half-wave rectifier shown in Figure 3.3. Sketch the resulting waveforms, as was done in Figure 3.3(b), for the full-wave rectifier. What are the advantages of the full-wave rectifier?

3.4 Three message signals are periodic with period T, as shown in Figure 3.76. Each of the three message signals is applied to an AM modulator. For each message signal, determine the modulation efficiency for $a = 0.7$ and $a = 1$.

3.5 The positive portion of the envelope of the output of an AM modulator is shown in Figure 3.77. The message signal is a waveform having

FIGURE 3.76

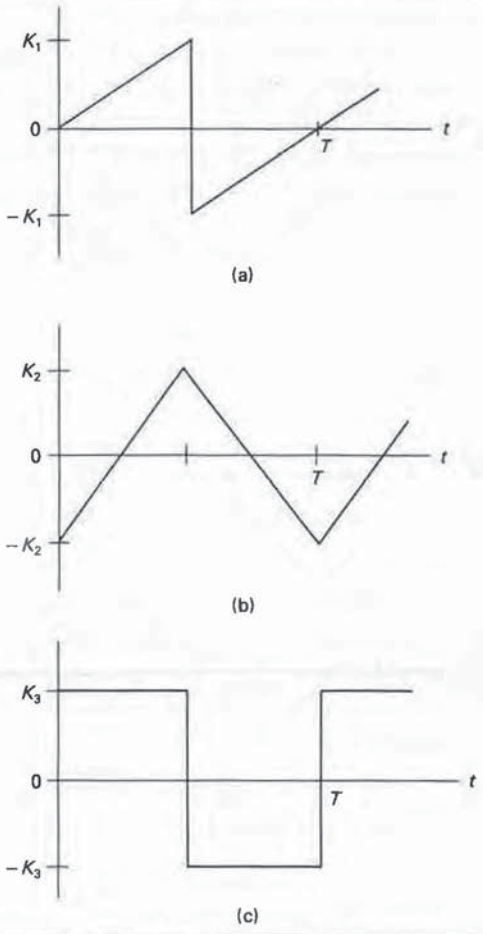


FIGURE 3.77

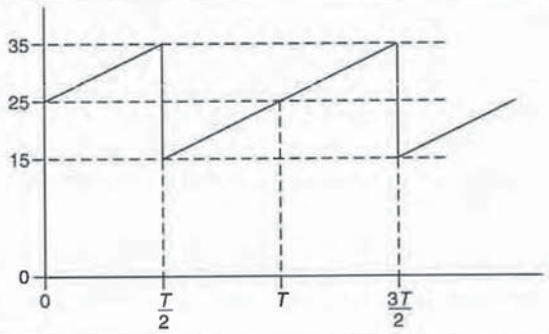
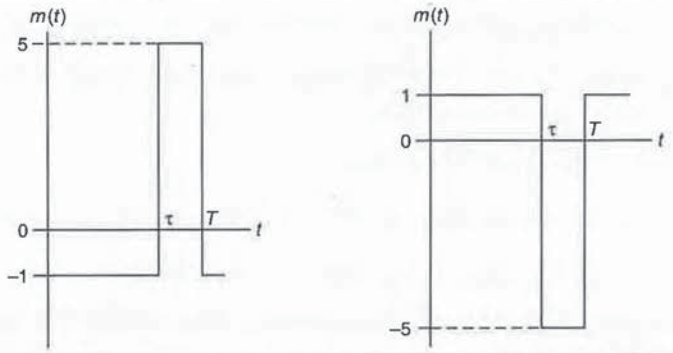


FIGURE 3.78



zero dc value. Determine the modulation index, the carrier power, the efficiency, and the power in the sidebands.

3.6 In this problem we examine the efficiency of AM modulation for the case in which the message signal does not have symmetrical maximum and minimum values. Two message signals are shown in Figure 3.78. Each is periodic with period T, and τ is chosen such that the dc value of $m(t)$ is zero. Calculate the efficiency for each $m(t)$ for $a = 1$.

3.7 An AM modulator operates with the message signal

$$m(t) = 7 \cos 20\pi t - 9 \cos 60\pi t$$

The unmodulated carrier is given by $100 \cos 200\pi t$, and the system operates with an index of $\frac{1}{2}$.

(a) Write the equation for $m_n(t)$, the normalized signal with a minimum value of -1.
(b) Determine $\langle m_n^2(t) \rangle$, the power in $m_n(t)$.
(c) Determine the efficiency of the modulator.
(d) Sketch the double-sided spectrum of $x_c(t)$, the modulator output, giving the weights and frequencies of all components.

3.8 Rework Problem 3.7 for the message signal

$$m(t) = 7 \cos 20\pi t + 9 \cos 60\pi t$$

3.9 An AM modulator has output

$$x_c(t) = 40 \cos 2\pi(200)t + 6 \cos 2\pi(180)t + 6 \cos 2\pi(220)t$$

Determine the modulation index and the efficiency.

3.10 An AM modulator has output

$$x_c(t) = A \cos 2\pi(200)t + B \cos 2\pi(180)t + B \cos 2\pi(220)t$$

The carrier power is 100 W, and the efficiency is 40%. Determine A, B, and the modulation index.

3.11 An AM modulator has output

$$x_c(t) = 20 \cos 2\pi(150)t + 6 \cos 2\pi(160)t + 6 \cos 2\pi(140)t$$

Determine the modulation index and the efficiency.

3.12 An AM modulator is operating with an index of 0.8. The modulation input is

$$m(t) = 2 \cos (2\pi f_m t) + \cos (4\pi f_m t) + 2 \cos (10\pi f_m t)$$

(a) Sketch the spectrum of the modulator output showing the weights of all impulse functions.
(b) What is the efficiency of the modulation process?

3.13 Consider the system shown in Figure 3.79. Assume that the average value of $m(t)$ is zero and that the maximum value of $|m(t)|$ is M. Also assume that the square-law device is defined by $y(t) = 4x(t) + 2x^2(t)$.

(a) Write the equation for $y(t)$.
(b) Describe the filter that yields an AM signal for $g(t)$. Give the necessary filter type and the frequencies of interest.
(c) What value of M yields a modulation index of 0.1?
(d) What is an advantage of this method of modulation?

3.14 Redraw Figure 3.7 to illustrate the generation of upper-sideband SSB. Give the equation defining the upper-sideband filter. Complete the analysis by deriving the expression for the output of an upper-sideband SSB modulator.

3.15 Show that both AM and DSB can be demodulated using the synchronous switching system shown in Figure 3.80.

FIGURE 3.79

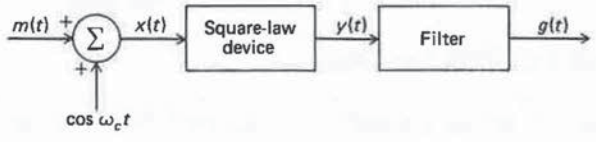
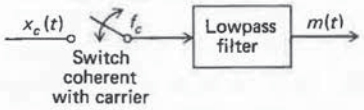


FIGURE 3.80



3.16 An upper-sideband SSB modulator has the message signal $m(t) = A_m \cos \omega_m t$. The unmodulated carrier is given by $c(t) = A_c \sin \omega_c t$. Sketch the modulator output $x_c(t)$. Repeat for a lower-sideband SSB modulator.

3.17 The VSB filter illustrated in Figure 3.13 is characterized by

$$H(f_c - f_1) = \epsilon e^{j\phi}$$
$$H(f_c + f_1) = (1 - \epsilon)e^{j\theta_1}$$
$$H(f_c + f_2) = 1 \cdot e^{j\theta_2}$$

The message signal is given by Equation (3.64) and is to be demodulated by multiplying by 4 cos $\omega_c t$ and lowpass filtering. It is required that the demodulated output be distortionless. Derive the expressions for θ_1 and θ_2 as functions of ϕ. Generalize your result by giving the ideal phase response of a VSB filter.

3.18 Prove that carrier reinsertion with envelope detection can be used for demodulation of VSB.

3.19 Sketch Figure 3.20 for the case where $\omega_{LO} = \omega_c - \omega_{IF}$.

3.20 A mixer is used in a short-wave superheterodyne receiver. The receiver is designed to receive transmitted signals between 5 and 25 MHz. High-side tuning is to be used. Determine the tuning range of the local oscillator for IF frequencies varying between 400 kHz and 2 MHz. Plot the ratio defined by the tuning range over this range of IF frequencies as in Table 3.1.

3.21 A superheterodyne receiver uses an IF frequency of 455 kHz. The receiver is tuned to a transmitter having a carrier frequency of 1120 kHz. Give two permissible frequencies of the local oscillator and the image frequency for each.

3.22 Repeat the preceding problem assuming that the IF frequency is 2500 kHz.

Section 3.2

3.23 Let the input to a phase modulator be $m(t) = u(t - t_0)$, as shown in Figure 3.22(a). Assume that the unmodulated carrier is $A_c \cos \omega_c t$ and that $\omega_c t_0 = 2\pi n$, where n is an integer. Sketch accurately the phase modulator output for $k_p = \pi$, $\frac{1}{4}\pi$, and $\frac{1}{8}\pi$, as was done in Figure 3.22(c) for $k_p = \frac{1}{2}\pi$.

3.24 The carrier power of an FM signal is 40 W, and the carrier frequency is $f_c = 40$ Hz. A sinusoidal message signal is used with index $\beta = 10$. The sinusoidal message signal has a frequency of 5 Hz. Determine the average value of $x_c(t)$. We usually assume $\langle x_c(t) \rangle = 0$. By drawing appropriate spectra, explain this apparent contradiction.

3.25 Given that $J_0(3) = -0.2601$ and that $J_1(3) = 0.3391$, determine $J_4(3)$.

3.26 Determine and sketch the spectrum (amplitude and phase) of an angle-modulated signal assuming that the instantaneous phase deviation is $\phi(t) = \beta \sin 2\pi f_m t$. Also assume $\beta = 8$, $f_m = 20$ Hz, and $f_c = 1000$ Hz.

3.27 Repeat the preceding problem assuming that the instantaneous phase deviation is $\phi(t) = \beta \cos 2\pi f_m t$.

3.28 An FM modulator has input $m(t) = 4 \cos 10\pi t$. The peak frequency deviation is 25 Hz. The modulator is followed by an ideal bandpass filter with a center frequency given by the carrier frequency and a bandwidth of 54 Hz. Determine the power at the filter output assuming that the power of the modulator output is 100 W. Sketch the amplitude and phase spectrum of the filter output.

3.29 A modulated signal is given by

$$x_c(t) = 5 \cos [2\pi(80)t] + 10 \cos [2\pi(100)t] + 3 \cos [2\pi(120)t]$$

Assuming a carrier frequency of 100 Hz, write this signal in the form of (3.1). Give equations for the envelope, $R(t)$, and the phase deviation, $\phi(t)$.

3.30 Repeat the preceding problem assuming that

$$x_c(t) = 5 \cos [2\pi(80)t] + 10 \cos [2\pi(1000)t] + 3 \sin [2\pi(120)t]$$

3.31 A transmitter uses a carrier frequency of 1000 Hz. Determine both the phase and frequency deviation for each of the following transmitter outputs:

(a) $x_c(t) = \cos [2\pi(1000)t + 10t^2]$

(b) $x_c(t) = \cos [2\pi(500)t^2]$
(c) $x_c(t) = \cos [2\pi(1100)t]$
(d) $x_c(t) = \cos [2\pi(1000)t + 10\sqrt{t}]$

3.32 An FM modulator has output

$$x_c(t) = 100 \cos \left[\omega_c t + 2\pi f_d \int^t m(\alpha)\, d\alpha \right]$$

where $f_d = 30$ Hz/V. Assume that $m(t)$ is the rectangular pulse $m(t) = 8\Pi[\frac{1}{4}(t - 2)]$.

(a) Sketch the phase deviation in radians.
(b) Sketch the frequency deviation in hertz.
(c) Determine the peak frequency deviation in hertz.
(d) Determine the peak phase deviation in radians.
(e) Determine the power at the modulator output.

3.33 Repeat the preceding problem assuming that $m(t)$ is the triangular pulse $4\Lambda[\frac{1}{2}(t - 6)]$.

3.34 An FM modulator with $f_d = 25$ Hz/V has the message signal shown in Figure 3.81. Plot the frequency deviation in hertz and the phase deviation in radians.

3.35 An FM modulator with $f_d = 15$ Hz/V has the message signal shown in Figure 3.82. Plot the frequency deviation in hertz and the phase deviation in radians.

3.36 An FM modulator with $f_d = 10$ Hz/V has the message signal shown in Figure 3.83. Plot the frequency deviation in hertz and the phase deviation in radians.

FIGURE 3.81

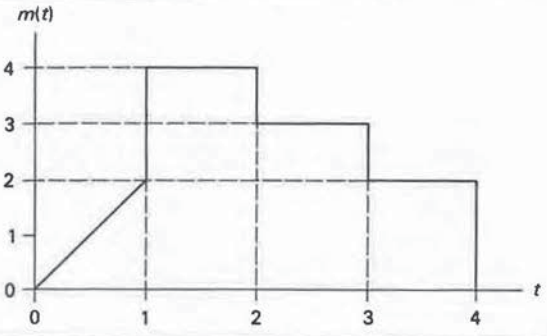
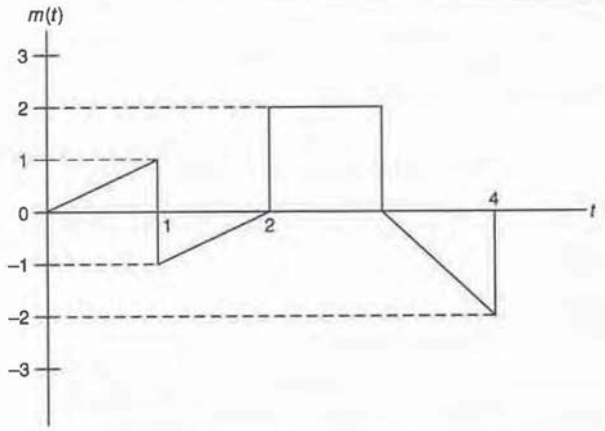


FIGURE 3.82



3.37 An FM modulator has $f_c = 1000$ Hz and $f_d = 12.5$. The modulator has input $m(t) = 4 \cos 2\pi(10)t$.

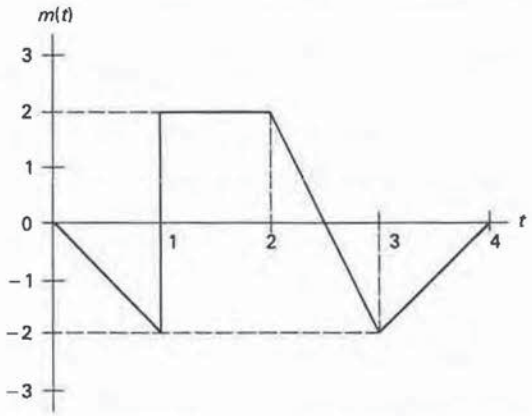
(a) What is the modulation index?

(b) Sketch, approximately to scale, the magnitude spectrum of the modulator output. Show all frequencies of interest.

(c) Is this narrowband FM? Why?

(d) If the same $m(t)$ is used for a phase modulator, what must k_p be to yield the index given in (a)?

FIGURE 3.83



3.38 An audio signal has a bandwidth of 25 kHz. The maximum value of $|m(t)|$ is 5 V. This signal frequency modulates a carrier. Estimate the bandwidth of the modulator output, assuming that the deviation constant of the modulator is (a) 20 Hz/V, (b) 200 Hz/V, (c) 2 kHz/V, and (d) 20 kHz/V.

3.39 By making use of Equations (3.104) and (3.112), show that

$$\sum_{n=-\infty}^{\infty} J_n^2(\beta) = 1$$

3.40 Prove that $J_n(\beta)$ can be expressed as

$$J_n(\beta) = \frac{1}{\pi} \int_0^{\pi} \cos(\beta \sin x - nx)\, dx$$

3.41 Using the result of Problem 3.40, show that

$$J_{-n}(\beta) = (-1)^n J_n(\beta)$$

3.42 An FM modulator is followed by an ideal bandpass filter having a center frequency of 500 Hz and a bandwidth of 70 Hz. The gain of the filter is 1 in the passband. The unmodulated carrier is given by $10 \cos(1000\pi t)$ and the message signal is $m(t) = 10 \cos(20\pi t)$. The transmitter constant f_d is 8 Hz/V.

(a) Determine the peak frequency deviation in hertz.
(b) Determine the peak phase deviation in radians.
(c) Determine the modulation index.
(d) Determine the power at the filter input and the filter output.
(e) Draw the single-sided spectrum of the signal at the filter input and the filter output. Label the amplitude and frequency of each spectral component.

3.43 An FM modulator operates with a peak deviation of 700 Hz. The message signal is given by

$$m(t) = 5 \cos 200\pi t$$

The power at the modulator ouput is 20 kW. The modulator is followed by a bandpass filter having a center frequency equal to the carrier frequency and a bandwidth of 1.3 kHz. Given that $J_0(7) = 0.3001$ and $J_1(7) = -0.0047$, compute the power at the filter output.

3.44 Reconstruct Figure 3.28 for the case where the five values of the modulation index are achieved by increasing the modulator deviation constant while holding f_m constant.

3.45 A sinusoidal message signal has a frequency of 150 Hz. This signal is the input to an FM modulator with an index of 10. Determine the bandwidth of the modulator output if a power ratio, P_r, of 0.8 is needed. Repeat for a power ratio of 0.9.

3.46 A narrowband FM signal has a carrier frequency of 110 kHz and a deviation ratio of 0.05. The modulation bandwidth is 10 kHz. This signal is used to generate a wideband FM signal with a deviation ratio of 20 and a carrier frequency of 100 MHz. The scheme utilized to accomplish this is illustrated in Figure 3.31. Give the required value of frequency multiplication, n. Also, fully define the mixer by giving *two* permissible frequencies for the local oscillator, and define the required bandpass filter (center frequency and bandwidth).

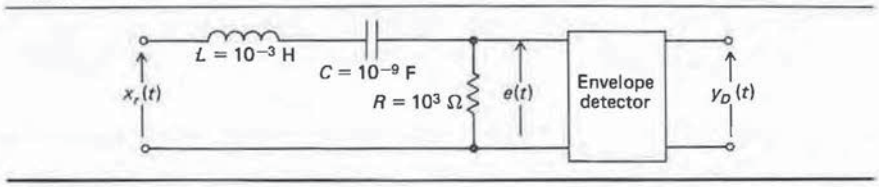
3.47 Consider the FM discriminator shown in Figure 3.84. The envelope detector can be considered ideal with an infinite input impedance. Plot the magnitude of the transfer function $E(f)/X_r(f)$. From this plot, determine a suitable carrier frequency and the discriminator constant, K_D, and estimate the allowable peak frequency deviation of the input signal.

3.48 By adjusting the values of R, L, and C in Problem 3.47, design a discriminator for a carrier frequency of 100 MHz, assuming that the peak frequency deviation is 6 MHz. What is the discriminator constant, K_D, for your design?

3.49 Solving system problems in which the input frequency is varying is often a difficult task. Approximate solutions can often be obtained by using sinusoidal steady-state techniques with the system-transfer function evaluated at the instantaneous input frequency. This is known as the *quasi-steady-state approximation*. In other words, if the input to a system is

$$x(t) = A \cos [\omega_c t + \phi(t)]$$

FIGURE 3.84



the system output is approximated by using

$$H\left(f_c + \frac{1}{2\pi}\frac{d\phi}{dt}\right)$$

for the transfer function. Rework Example 3.8 using this technique.

Section 3.3

3.50 Assume that an FM discriminator operates in the presence of sinusoidal interference. Show that the discriminator output is a nonzero constant (dc) for each of the following cases: $A_i = A_c$, $A_i = -A_c$, and $A_i \gg A_c$. Determine the discriminator output for each of these cases.

Section 3.4

3.51 Sketch the input phase deviation $\phi(t)$, the phase deviation of the VCO output $\theta(t)$, and the phase error for Example 3.9. Determine the relationship between the various parameters (A, k_f, K_t, K_v) that must exist for the maximum phase error to be less than 0.2 rad.

3.52 Rework Example 3.9 for $m(t) = A \cos \omega_m t$.

3.53 Using $x_r(t) = m(t) \cos \omega_c t$ and $e_0(t) = 2 \cos(\omega_c t + \theta)$ for the assumed Costas PLL input and VCO output, respectively, verify that all signals shown at the various points in Figure 3.57 are correct. Assuming that the VCO frequency deviation is defined by $d\theta/dt = -K_v e_v(t)$, where $e_v(t)$ is the VCO input and K_v is a positive constant, derive the phase plane. Using the phase plane, verify that the loop locks.

3.54 Using a single phase-lock loop, design a system that has an output frequency equal to $\frac{7}{3} f_0$, where f_0 is the input frequency. Describe fully, by sketching, the output of the VCO for your design. Draw the spectrum at the VCO output and at any other point in the system necessary to explain the operation of your design. Describe any filters used in your design by defining the center frequency and the appropriate bandwidth of each.

3.55 A first-order PLL is operating with zero frequency and phase error when a step in frequency of magnitude $\Delta\omega$ is applied. The loop gain K_t is $2\pi(100)$. Determine the steady-state phase error, in degrees, for $\Delta\omega = 2\pi(30)$, $2\pi(50)$, $2\pi(80)$, and $-2\pi(80)$ rad/s. What happens if $\Delta\omega = 2\pi(120)$ rad/s?

3.56 The imperfect second-order PLL is defined as a PLL with the loop filter

$$F(s) = \frac{s + a}{s + \epsilon}$$

Use the linear model of the PLL and derive the transfer function for $\Theta(s)/\Phi(s)$. Derive expressions for ω_n and ζ in terms of K_t, a, and ϵ.

3.57 Using the results of the previous problem, determine K_t, a, and ϵ for a second-order PLL having a natural frequency of 100 Hz and a damping factor of 0.707. Assume that $\epsilon = 0.1K_t$.

3.58 A frequency-compressive feedback system is used to demodulate an FM signal having a deviation ratio of 5. The VCO constant is 25 Hz/V. Determine the discriminator constant K_D so that the deviation ratio at the discriminator input is 0.4.

Section 3.5

3.59 Sketch a circuit using a capacitor that can convert PWM to PAM. The charge on the capacitor should be proportional to the area under the PWM pulse. Show typical signals at all points in your block diagram. How is the PAM signal demodulated?

3.60 A continuous data signal is quantized and transmitted using a PCM system. If each data sample at the receiving end of the system must be known to within $\pm 0.5\%$ of the peak-to-peak full-scale value, how many binary symbols must each transmitted digital word contain? Assume that the message signal is speech and has a bandwidth of 4 kHz. Estimate the bandwidth of the resulting PCM signal (choose k).

3.61 Extend Problem 3.60 by plotting the bandwidth of the PCM signal as a function of accuracy. Let the accuracy vary from 50% of the peak-to-peak full-scale value of the input signal to 0.1% of this value. Let the bandwidth of the message signal be W.

3.62 A delta modulator has the message signal

$$m(t) = 9 \sin 2\pi(10)t + 5 \sin 2\pi(20)t$$

Determine the minimum sampling frequency required to prevent slope overload, assuming that the impulse weights δ_0 are 0.05π.

3.67 Repeat Problem 3.62, assuming that

$$m(t) = 9 \cos 2\pi(10)t + 5 \cos 2\pi(20)t$$

Section 3.6

3.64 Five messages bandlimited to W, W, $2W$, $4W$, and $4W$ Hz, respectively, are to be time-division multiplexed. Devise a commutator configuration such that each signal is periodically sampled at its own minimum rate and the samples are properly interlaced. What is the *minimum* transmission bandwidth required for this TDM signal?

3.65 In an FDM communication system, the transmitted baseband signal is

$$x(t) = m_1(t) \cos \omega_1 t + m_2(t) \cos \omega_2 t$$

This system has a second-order nonlinearity between transmitter input and receiver output. Thus the received baseband signal $y(t)$ can be expressed as

$$y(t) = a_1 x(t) + a_2 x^2(t)$$

Assuming that the two message signals, $m_1(t)$ and $m_2(t)$, have the spectra

$$M_1(f) = M_2(f) = \Pi\left(\frac{f}{W}\right)$$

sketch the spectrum of $y(t)$. Discuss the difficulties encountered in demodulating the received baseband signal. In many FDM systems, the subcarrier frequencies ω_1 and ω_2 are harmonically related. Describe any additional problems this presents.



COMPUTER EXERCISES

The following exercises require that waveforms and the spectra of waveforms be plotted. Although these exercises can be developed using any high-level language, a special-purpose computer package, having significant computational and graphics capabilities, such as Mathcad or MATLAB, is recommended.

3.1 Develop a computer program to demonstrate DSB and AM modulation. Both the waveforms and the spectra corresponding to the waveforms at each point in the system are to be generated and plotted within the program. You may use coherent demodulation for AM. An interesting message signal, with a known spectrum, can easily be generated using either three-term or a four-term Fourier series.

3.2 The purpose of the exercise is to demonstrate the properties of SSB modulation. Develop a computer program to generate both upper-

sideband and lower-sideband SSB signals and display both the time-domain signals and the amplitude spectra of these signals. Plot the envelope of the SSB signals and show that both the upper-sideband and the lower-sideband SSB signals have the same envelope. It would also be interesting to look at the phase deviation and frequency deviation of the modulated carrier. Show that carrier reinsertion can be used to demodulate a SSB signal. Illustrate the effect of using a demodulation carrier with insufficient amplitude when using the carrier reinsertion technique.

3.3 In this computer exercise we investigate the properties of VSB modulation. Develop a computer program to generate and plot a VSB signal and the corresponding amplitude spectrum. Using the program, show that VSB can be demodulated using carrier reinsertion.

3.4 The purpose of this computer exercise is to study angle modulation using a sinusoidal message signal. Develop a program that allows you to generate an angle-modulated signal using a single sinusoid as the message signal. Using this message signal, plot the time-domain signal and both the amplitude and phase spectra of the signal for several values of the modulation index. Using several values of β show that the spectra shown in Figure 3.28 are correct.

3.5 Repeat the preceding problem using a message signal consisting of the sum of two sinusoids. It is not necessary to determine and plot the phase spectrum of $x_c(t)$. Using your program convince yourself that the result of Exercise 3.6 is correct.

3.6 Develop a computer program to generate the amplitude spectrum at the output of an FM modulator assuming a square-wave message signal. Plot the output for various values of the deviation ratio. When changing the deviation ratio, make sure that the peak deviation remains constant. Comment on your observations.

PROBABILITY AND RANDOM VARIABLES 4

The objective of this chapter is to provide the necessary background in probability theory to make possible the mathematical description or modeling of random signals. Our ultimate goal is to develop the mathematical tools necessary to deal with random signals, or *random processes,* in systems analysis, although this will not be accomplished until Chapter 5.

Because of their unpredictable nature, we cannot assign exact values to random signals or noise waveforms at a given instant in time but must describe them in terms of averages or the probability of a particular outcome. Thus we begin this chapter with an elementary consideration of probability theory. Next we introduce the idea of random variables and their statistical description. If we think of random variables as depending on a parameter, such as time, we are really speaking of random signals, the topic of Chapter 5.

4.1 WHAT IS PROBABILITY?

Each of us has an intuitive notion about probability from prior experience. Let us therefore begin by considering some of these intuitive approaches to probability.

Equally Likely Outcomes

Suppose you match coin tosses with someone. This is an example of a random, or chance, experiment for which there are two possible outcomes, or happenings. Either the coins will match when flipped or they will not match. Since there is no prior reason to believe that one of these outcomes is favored over the other, you would most likely guess, before flipping, that your chances of winning or losing are equally likely.

Stated more generally, if there are N possible *equally likely* and *mutally exclusive* outcomes (that is, the occurrence of one outcome precludes the occur-

rence of any of the others) to a random, or chance, experiment and if N_A of these outcomes correspond to an event A of interest, then the probability of event A, or $P(A)$, is

$$P(A) = \frac{N_A}{N} \tag{4.1}$$

However, there are practical difficulties with this definition of probability, which is often referred to as the *classical* (or *equally likely*) *definition of probability*. One must be able to break the chance experiment up into two or more equally likely outcomes and this is not always possible. The most obvious experiments fitting these conditions are card games, dice, and coin tossing.

Philosophically, there is difficulty with this definition in that use of the words "equally likely" really amounts to saying something about being equally probable, which means we are using probability to define probability.

To summarize, there are difficulties with the classical, or equally likely, definition of probability, but it is useful in engineering problems when it is reasonable to list N equally likely, *mutually exclusive** outcomes. The following example will illustrate its usefulness in a situation where it applies.

▼ EXAMPLE 4.1 Given a deck of 52 playing cards, (a) What is the probability of drawing the ace of spades? (b) What is the probability of drawing a spade?

SOLUTION

(a) Using the principle of equal likelihood, we have one favorable outcome in 52 possible outcomes. Therefore, $P(\text{ace of spades}) = \frac{1}{52}$.

(b) Again using the principle of equal likelihood, we have 13 favorable outcomes in 52, and $P(\text{spade}) = \frac{13}{52} = \frac{1}{4}$. ▲

Relative Frequency

Suppose we wish to assess the probability of an unborn childs being a boy. Using the classical definition, we predict a probability of $\frac{1}{2}$, since there are two possible mutually exclusive outcomes, which from outward appearances appear equally probable. However, yearly birth statistics for the United States consistently indicate that the ratio of males to total births is about 0.51. In the years 1987 to 1989, for example, the ratio of male births to

* *Mutually exclusive* means that if a given outcome occurs, it precludes the occurrence of any other possible outcomes.

female births in the United States was 1050 to 1000, respectively, or 51.22% of all births were male. This is an example of the relative-frequency approach to probability.

In the *relative-frequency approach,* we consider a random experiment, enumerate all possible outcomes, repeatedly perform the experiment (often conceptually), and take the ratio of the number of outcomes, N_A, favorable to an event of interest, A, to the total number of trials, N, as an approximation of the probability of A, $P(A)$. We define the limit of N_A/N, called the *relative frequency* of A, as $N \to \infty$, as $P(A)$:

$$P(A) \triangleq \lim_{N\to\infty} \frac{N_A}{N} \tag{4.2}$$

This definition of probability can be used to estimate $P(A)$. However, since the infinite number of experiments implied by (4.2) cannot be performed, only an approximation to $P(A)$ is obtained. Thus the relative-frequency notion of probability is useful for estimating a probability but is not satisfactory as a mathematical basis for probability.

The following example fixes these ideas and will be referred to later in this chapter.

▼ **EXAMPLE 4.2** Consider the simultaneous tossing of two fair coins. Thus, on any given trial, we have the possible outcomes HH, HT, TH, and TT, where, for example, HT denotes a head on the first coin and a tail on the second coin. (We imagine that numbers are painted on the coins so we can tell them apart.) What is the probability of two heads on any given trial?

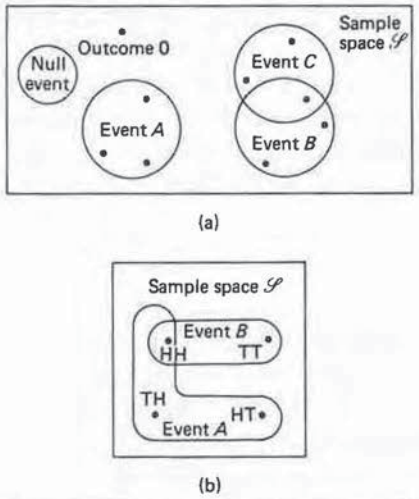
SOLUTION By distinguishing between the coins, the correct answer, using equal likelihood, is $\frac{1}{4}$. Similarly, it follows that $P(\text{HT}) = P(\text{TH}) = P(\text{TT}) = \frac{1}{4}$. ▲

Sample Spaces and the Axioms of Probability

Because of the difficulties mentioned for the preceding two definitions of probability, mathematicians prefer to approach probability on an axiomatic basis. We now briefly describe this axiomatic approach, which is general enough to encompass all the previously given definitions.

We can view a chance experiment geometrically by representing its possible outcomes as elements of a space referred to as a *sample space* $\mathcal{S}$. An *event* is defined as a collection of outcomes. An impossible collection of outcomes is referred to as the *null event.* Figure 4.1(a) shows a representation of

FIGURE 4.1 Sample spaces. (a) Pictorial representation of an arbitrary sample space. Points show outcomes; circles show events. (b) Sample space representation for the tossing of two coins.



a sample space. Three events of interest, A, B, and C, which do not encompass the entire sample space, are shown.

A specific example of a chance experiment might consist of measuring the dc voltage at the output terminals of a power supply. The sample space for this experiment would be the collection of all possible numerical values for this voltage. On the other hand, if the experiment is the tossing of two coins, as in Example 4.2, the sample space would consist of the four outcomes HH, HT, TH, and TT enumerated earlier. A sample-space representation for this experiment is shown in Figure 4.1(b). Two events of interest, A and B, are shown. Event A denotes at least one head, and event B consists of the coins matching. Note that A and B encompass all possible outcomes for this particular example.

In the axiomatic approach, a *measure,* called *probability,* is somehow assigned to the events of a sample space* such that this measure possesses the properties of probability. The properties or axioms of this probability measure are chosen to yield a satisfactory theory such that results from applying the theory will be consistent with experimentally observed phenomena. A set of satisfactory axioms is the following:

1. $P(A) \geq 0$ for all events A in the sample space $\mathcal{S}$.

* For example, by the relative-frequency or the equally likely approaches.

2. The probability of all posssible events occurring is unity, $P(\mathcal{S}) = 1$.
3. If the occurrence of A precludes the occurrence of B, and vice versa (that is, A and B are *mutually exclusive*), then $P(A \text{ or } B) = P(A) + P(B)$.*

It is emphasized that this approach to probability does not give us the number $P(A)$; it must be obtained by some other means.

Before we proceed further, it is convenient to introduce some useful notation. The event "A or B or both" will be denoted as $(A + B)$. The event "both A and B" will be denoted either as (AB) or as (A, B). We will sometimes be interested in the event "not A," which will be denoted $\bar{A}$. An event such as $(A + B)$, which is composed of two or more events, will be referred to as a *compound event*. Other notations often used for the events just defined are $A \cup B$ ("A union B"), $A \cap B$ ("A intersection B"), and A^c ("A complement"), respectively.

Some Useful Probability Relationships

We now employ concepts of relative frequency to show the reasonableness of some useful relationships among probabilities of other compound events. We could proceed directly from the axioms just stated, but the relative-frequency approach taken here is more intuitively satisfying.

Consider a chance experiment having, among other possible outcomes, two events of immediate interest, A and B, which may or may not happen together. For example, the experiment could be the simultaneous tossing of two fair coins, as in Example 4.2, with event A being the occurrence of *at least* one head (that is, HH, HT, or TH) and event B denoting that the outcomes on each coin match (that is, either HH or TT). Although, in this particular example, the compound event "A or B or both" encompasses the entire sample space, there may, in general, be other possible events.

To continue, suppose we repeat our experiment N times and record the information given in Table 4.1. According to the relative-frequency approach, as N becomes large without bound,

$$P(A) \cong \frac{N_{A\bar{B}} + N_{AB}}{N} \tag{4.3a}$$

$$P(B) \cong \frac{N_{\bar{A}B} + N_{AB}}{N} \tag{4.3b}$$

* To the third axiom is added the condition of countable additivity if the number of outcomes is countably infinite. We will not include it here in the interest of brevity.

TABLE 4.1

Case 1	$N_{A\bar{B}}$	= number of times A occurred alone.
Case 2	$N_{\bar{A}B}$	= number of times B occurred alone.
Case 3	N_{AB}	= number of times A and B occurred together.
Case 4	$N_{\bar{A}\bar{B}}$	= number of times neither A nor B occurred.
Thus	N	$= N_{\bar{A}B} + N_{A\bar{B}} + N_{AB} + N_{\bar{A}\bar{B}}$.

and

$$P(AB) \cong \frac{N_{AB}}{N} \tag{4.3c}$$

Consider now the probability of A or B or both, $P(A + B)$. Clearly, the number of occurrences of this compound event is $N_{A\bar{B}} + N_{\bar{A}B} + N_{AB}$. Thus, in a large number of trials,

$$\begin{aligned} P(A + B) &\cong \frac{N_{A\bar{B}} + N_{\bar{A}B} + N_{AB}}{N} \\ &= \frac{N_{A\bar{B}} + N_{AB}}{N} + \frac{N_{\bar{A}B} + N_{AB}}{N} - \frac{N_{AB}}{N} \end{aligned} \tag{4.4a}$$

Using (4.3) and taking the limit as $N \to \infty$, this becomes

$$P(A + B) = P(A) + P(B) - P(AB) \tag{4.4b}$$

Comparing this with axiom 3, we see that if A and B are not mutually exclusive, the probability $P(AB)$ must be subtracted from the right-hand side. Equation (4.4b) could have been derived directly from the axioms by defining appropriate compound events (see Problem 4.2).

Next consider the probability that A occurs, given that B occurred, $P(A \mid B)$, and the probability that B occurs, given that A occurred, $P(B \mid A)$. These probabilities are referred to as *conditional probabilities*. Considering $P(A \mid B)$ first, we see from Table 4.1 that given that event B occurred, we must consider only cases 2 and 3 for a total of $N_{\bar{A}B} + N_{AB}$ outcomes. Of these, only N_{AB} outcomes are favorable to event A occurring, so

$$P(A \mid B) = \frac{N_{AB}}{N_{\bar{A}B} + N_{AB}} = \frac{N_{AB}/N}{(N_{\bar{A}B} + N_{AB})/N} \tag{4.5}$$

In terms of previously defined probabilities, therefore,

$$P(A \mid B) = \frac{P(AB)}{P(B)} \tag{4.6a}$$

Similar reasoning results in the relationship

$$P(B \mid A) = \frac{P(AB)}{P(A)} \tag{4.6b}$$

In an axiomatic approach, conditional probabilities could be *defined* in this fashion, and we will take the preceding relations as definitions. However, the development given here shows their reasonableness.

Putting Equations (4.6a) and (4.6b) together, we obtain

$$P(A \mid B)P(B) = P(B \mid A)P(A) \tag{4.7a}$$

or

$$P(B \mid A) = \frac{P(B)P(A \mid B)}{P(A)} \tag{4.7b}$$

This is a special case of *Bayes' rule*.

Finally, suppose that the occurrence or nonoccurrence of B in no way influences the occurrence or nonoccurrence of A. If this is true, A and B are said to be *statistically independent*. Thus, if we are given B, this tells us nothing about A, and therefore, $P(A \mid B) = P(A)$. Similarly, $P(B \mid A) = P(B)$. From Equation (4.6a) or (4.6b) it follows that, for such events,

$$P(AB) = P(A)P(B) \tag{4.8}$$

Equation (4.8) will be taken as the definition of statistically independent events.

▼ **EXAMPLE 4.3** Referring to Example 4.2, suppose A denotes at least one head and B denotes a match. The sample space is shown in Figure 4.1(b). To find $P(A)$ and $P(B)$, we may proceed in several different ways.

SOLUTION First, if we use equal likelihood, there are three outcomes favorable to A (that is, HH, HT, and TH) among four possible outcomes, yielding $P(A) = \frac{3}{4}$. For B, there are two favorable outcomes in four possibilities, giving $P(B) = \frac{1}{2}$.

As a second approach, we note that the outcomes on separate coins are statistically independent with $P(H) = P(T) = \frac{1}{2}$. (Why?) Also, event A consists of any of the mutually exclusive outcomes HH, TH, and HT, giving

$$P(A) = (\tfrac{1}{2} \cdot \tfrac{1}{2}) + (\tfrac{1}{2} \cdot \tfrac{1}{2}) + (\tfrac{1}{2} \cdot \tfrac{1}{2}) = \tfrac{3}{4} \tag{4.9}$$

by Equation (4.8) and axiom 3. Similarly, since B consists of the mutually exclusive outcomes HH and TT,

$$P(B) = (\tfrac{1}{2} \cdot \tfrac{1}{2}) + (\tfrac{1}{2} \cdot \tfrac{1}{2}) = \tfrac{1}{2} \tag{4.10}$$

again through the use of (4.8) and axiom 3. Also, $P(AB) = P$(at least one head and a match) $= P(\text{HH}) = \frac{1}{4}$.

Next, consider the probability of at least one head given that we had a match, $P(A \mid B)$. Using Bayes' rule, we obtain

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2} \tag{4.11}$$

which is reasonable, since given B, the only outcomes under consideration are HH and TT, only one of which is favorable to event A. Next, finding $P(B \mid A)$, the probability of a match given at least one head, we obtain

$$P(B \mid A) = \frac{P(AB)}{P(A)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3} \tag{4.12}$$

Checking this result using the principle of equal likelihood, we have one favorable event among three candidate events (HH, TH, and HT), which yields a probability of $\frac{1}{3}$. We note that

$$P(AB) \neq P(A)P(B) \tag{4.13}$$

Thus events A and B are not statistically independent, although the events H and T on either coin are independent.

Finally, consider the joint probability $P(A + B)$. Using (4.4b), we obtain

$$P(A + B) = \tfrac{3}{4} + \tfrac{1}{2} - \tfrac{1}{4} = 1 \tag{4.14}$$

Remembering that $P(A + B)$ is the probability of at least one head, *or* a match, or both, we see that this includes all possible outcomes. Thus our result is correct. ▲

▼ EXAMPLE 4.4 This example illustrates the reasoning to be applied when trying to determine if two events are independent. A single card is drawn at random from a deck of cards. Which of the following pairs of events are independent? (a) The card is a club, and the card is black. (b) The card is a king, and the card is black.

SOLUTION We use the relationship $P(AB) = P(A \mid B)P(B)$ (always valid) and check it against the relation $P(AB) = P(A)P(B)$ (valid only for independent events). For part (a), we let A be the event that the card is a club and B be

the event that it is black. Since there are 26 black cards in an ordinary deck of cards, 13 of which are clubs, the conditional probability $P(A \mid B)$ is $\frac{13}{26}$ (given we are considering only black cards, we have 13 favorable outcomes for the card being a club). The probability that the card is black is $\frac{26}{52}$, because half the cards in the 52-card deck are black. The probability of a club (event A), on the other hand, is $\frac{13}{52}$ (13 cards in a 52-card deck are clubs). In this case,

$$P(A \mid B)P(B) = \frac{13}{26}\frac{26}{52} \neq P(A)P(B) = \frac{13}{52}\frac{26}{52} \tag{4.15}$$

and the events are not independent.

For part (b), we let A be the event that a king is drawn, and event B be that it is black. In this case, the probability of a king given that the card is black is $\frac{2}{26}$ (two cards of the 26 black cards are kings). The probability of a king is simply $\frac{4}{52}$ (4 kings in the 52-card deck). Hence,

$$P(A \mid B)P(B) = \frac{2}{26}\frac{26}{52} = P(A)P(B) = \frac{4}{52}\frac{26}{52} \tag{4.16}$$

▲ which shows that the events king and black are statistically independent.

▼ EXAMPLE 4.5 As an example more closely related to communications, consider the transmission of binary digits through a channel as might occur, for example, in computer systems. As is customary, we denote the two possible symbols as 0 and 1. Let the probability of receiving a zero, given a zero was sent, $P(0r \mid 0s)$, and the probability of receiving a 1, given a 1 was sent, $P(1r \mid 1s)$, be

$$P(0r \mid 0s) = P(1r \mid 1s) = 0.9 \tag{4.17}$$

Thus the probabilities $P(1r \mid 0s)$ and $P(0r \mid 1s)$ must be

$$P(1r \mid 0s) = 1 - P(0r \mid 0s) = 0.1 \tag{4.18a}$$

and

$$P(0r \mid 1s) = 1 - P(1r \mid 1s) = 0.1 \tag{4.18b}$$

respectively. These probabilities characterize the channel and would be obtained through experimental measurement or analysis. Techniques for calculating them for particular situations will be discussed in Chapter 7.

In addition to these probabilities, suppose that we have determined through measurement that the probability of sending a zero is

$$P(0s) = 0.8 \tag{4.19a}$$

and therefore the probability of sending a 1 is

$$P(1s) = 1 - P(0s) = 0.2 \tag{4.19b}$$

Note that once $P(0r \mid 0s)$, $P(1r \mid 1s)$, and $P(0s)$ are specified, the remaining probabilities are calculated using axioms 2 and 3.

The next question we ask is, "If a 1 was received, what is the probability, $P(1s \mid 1r)$, that a 1 was sent?" Applying Bayes' rule, we find that

$$P(1s \mid 1r) = \frac{P(1r \mid 1s)P(1s)}{P(1r)} \tag{4.20}$$

To find $P(1r)$, we note that

$$P(1r, 1s) = P(1r \mid 1s)P(1s) = 0.18 \tag{4.21a}$$

and

$$P(1r, 0s) = P(1r \mid 0s)P(0s) = 0.08 \tag{4.21b}$$

Thus

$$P(1r) = P(1r, 1s) + P(1r, 0s) = 0.18 + 0.08 = 0.26 \tag{4.22}$$

and

$$P(1s \mid 1r) = \frac{(0.9)(0.2)}{0.26} = 0.69 \tag{4.23}$$

Similarly, one can calculate $P(0s \mid 1r) = 0.31$, $P(0s \mid 0r) = 0.97$, and $P(1s \mid 0r) = 0.03$. For practice, you should go through the necessary calculations. ▲

Venn Diagrams and Tree Diagrams

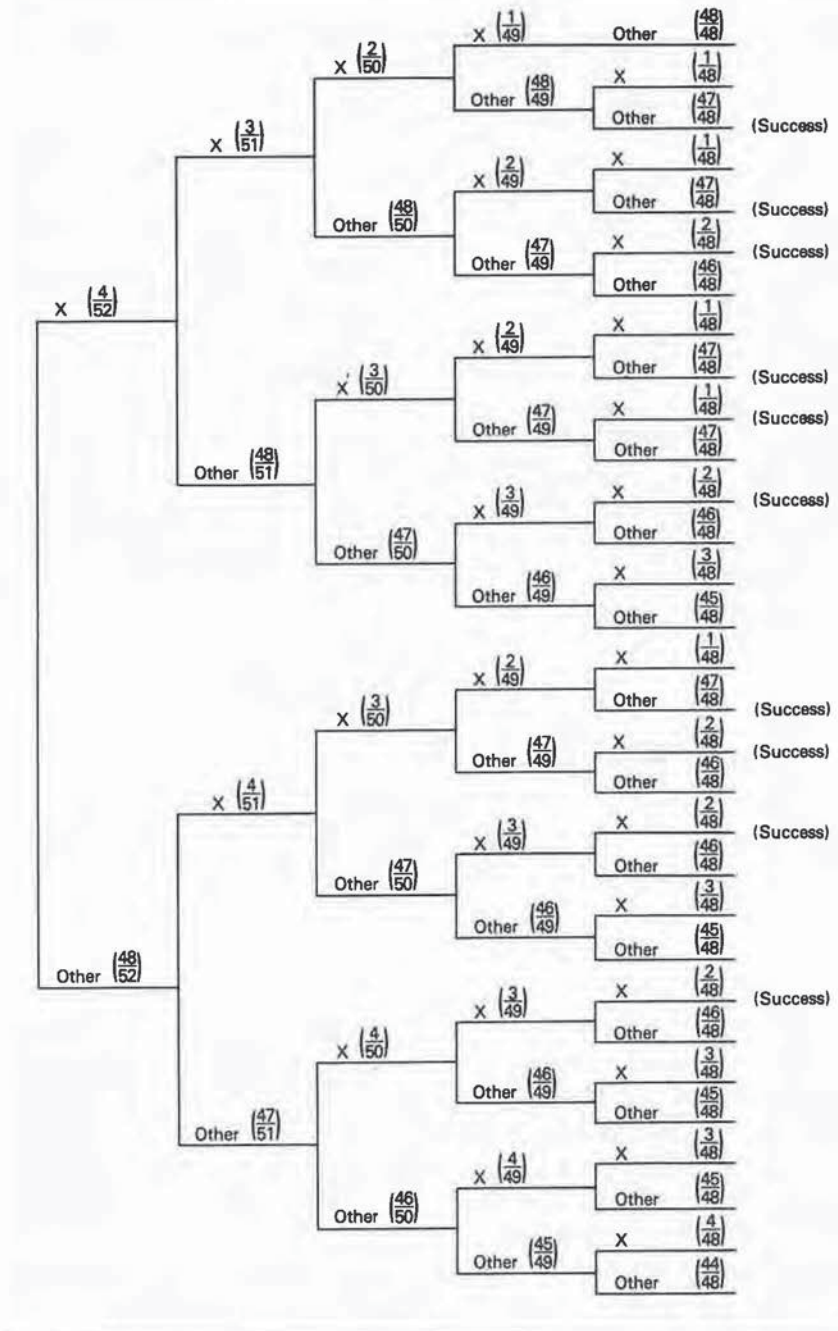
It is sometimes convenient to visualize the relationships between various events for a chance experiment in terms of a *Venn diagram*. In such diagrams, the sample space is indicated as a rectangle, with the various events indicated by circles or ellipses. Such a diagram looks exactly as shown in Figure 4.1(a), where it is seen that events B and C are not mutually exclusive, as indicated by the overlap between them, whereas event A is mutually exclusive of events B and C.

Another handy device for determining probabilities of compound events is a *tree diagram*, particularly if the compound event can be visualized as happening in time sequence. This device is illustrated by the following example.

▼ **EXAMPLE 4.6** Suppose five cards are drawn without replacement from a standard 52-card deck. What is the probability that three of a kind results?

SOLUTION The tree diagram for this chance experiment is shown in Figure 4.2. On the first draw, we focus on a particular card, denoted as X. The

FIGURE 4.2 A card-drawing problem illustrating the construction of a tree diagram



second draw results in four possible events of interest: a card is drawn that matches the first card with probability $\frac{3}{51}$, or a match is not obtained with probability $\frac{48}{51}$. If some card other than X was drawn on the first draw, then X results with probability $\frac{4}{51}$ on the second draw (lower half of Figure 4.2). At this point, 50 cards are left in the deck. If we follow the upper branch, which corresponds to a match of the first card, two events of interest are again possible: another match that will be referred to as a *triple* with probability of $\frac{2}{50}$ *on that draw,* or a card that does not match the first two with probability $\frac{48}{50}$. If a card other than X was obtained on the second draw, then X occurs with probability $\frac{4}{50}$ if X was obtained on the first draw, and probability $\frac{46}{50}$ if it was not. The remaining branches are filled in similarly. Each path through the tree will either result in success or failure, and the probability of drawing the cards along a particular path will be the product of the separate probabilities along each path. Since a particular sequence of draws resulting in success is mutually exclusive of the sequence of draws resulting in any other success, we simply add up all the products of probabilities along all paths that result in success. In addition to these sequences involving card X, there are 12 others involving other face values which result in three of a kind. Thus we multiply the result obtained from Figure 4.2 by 13. The probability of drawing three cards of the same value, in any order, is then given by

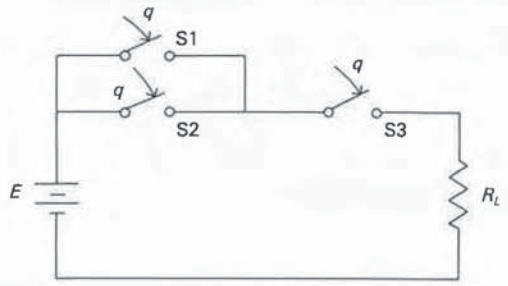
$$P(\text{3 of a kind}) = 13\frac{(10)(4)(3)(2)(48)(47)}{(52)(51)(50)(49)(48)}$$
$$= 0.02257 \tag{4.24}$$

(Note that drawing four of a kind is termed a failure, even though in a game of poker one would be very happy to have four of a kind. In this particular example, we are interested in the probability of obtaining three of a kind only.)

▲

▼ EXAMPLE 4.7 Another type of problem very closely related to those amenable to tree-diagram solution is a reliability problem. Reliability problems can result from considering the overall failure of a system composed of several components each of which may fail with a certain probability p. An example is shown in Figure 4.3, where a battery is connected to a load through the series-parallel combination of relay switches each of which may fail to close with probability p (or close with probability $q = 1 - p$). The problem is to find the probability that current flows in the load. From the diagram, it is clear that a circuit is completed if S1 or S2 and S3 are closed.

FIGURE 4.3 Circuit illustrating the calculation of reliability



Therefore,

$$\begin{aligned} P(\text{success}) &= P(\text{S1 or S2 and S3 closed}) \\ &= P(\text{S1 or S2 or both closed})P(\text{S3 closed}) \\ &= [1 - P(\text{both switches open})]P(\text{S3 closed}) \\ &= (1 - p^2)q \end{aligned} \tag{4.25}$$

where it is assumed that the separate switch actions are statistically independent. ▲

Some More General Relationships

We now obtain some useful formulas for a case somewhat more general than that given above. We consider an experiment, each mutually exclusive outcome of which is composed of a compound event (A_i, B_j). The totality of all these compound events, $i = 1, 2, \ldots, M$, $j = 1, 2, \ldots, N$, composes the entire sample space (that is, the events are exhaustive).

For example, the experiment might consist of rolling a pair of dice with (A_i, B_j) = (number of spots showing on die 1, number of spots showing on die 2).

Suppose the probability of the joint event (A_i, B_j) is $P(A_i, B_j)$. We can think of each compound event as a simple event, and if we sum the probabilities of all these mutually exclusive, exhaustive events, we will obtain a probability of 1, since we have included all possible outcomes. That is

$$\sum_{i=1}^{M} \sum_{j=1}^{N} P(A_i, B_j) = 1 \tag{4.26}$$

Now let us consider a particular event B_j. Associated with this particular event, we have M possible mutually exclusive, but not exhaustive outcomes

$(A_1, B_j), (A_2, B_j), \ldots, (A_M, B_j)$. If we sum over the corresponding probabilities, we will obtain the probability of B_j irrespective of the outcome on A. Thus

$$P(B_j) = \sum_{i=1}^{M} P(A_i, B_j) \tag{4.27}$$

Similar reasoning leads to the result

$$P(A_i) = \sum_{j=1}^{N} P(A_i, B_j) \tag{4.28}$$

$P(A_i)$ and $P(B_j)$ are referred to as *marginal probabilities*.

Suppose we wish to find the conditional probability of B_m given A_n, $P(B_m | A_n)$. In terms of the joint probabilities $P(A_i, B_j)$, we can write this conditional probability as

$$P(B_m | A_n) = \frac{P(A_n, B_m)}{\sum_{j=1}^{N} P(A_n, B_j)} \tag{4.29}$$

which is a more general form of Bayes' rule than that given by (4.7).

▼ EXAMPLE 4.8 A certain experiment has the joint and marginal probabilities shown in Table 4.2. Find the missing probabilities.

SOLUTION Using $P(B_1) = P(A_1, B_1) + P(A_2, B_1)$, we obtain $P(B_1) = 0.1 + 0.1 = 0.2$. Also, since $P(B_1) + P(B_2) + P(B_3) = 1$, we have $P(B_3) = 1 - 0.2 - 0.5 = 0.3$. Finally, using $P(A_1, B_3) + P(A_2, B_3) = P(B_3)$, we get $P(A_1, B_3) = 0.3 - 0.1 = 0.2$, and therefore, $P(A_1) = 0.1 + 0.4 + 0.2 = 0.7$. ▲

TABLE 4.2 $P(A_i, B_j)$

A_i \ B_j	B_1	B_2	B_3	$P(A_i)$
A_1	0.1	0.4		
A_2	0.1	0.1	0.1	0.3
$P(B_i)$		0.5		1

TABLE 4.3 Possible Random Variables

Outcome: S_i	R.V. No. 1: $X_1(S_i)$	R.V. No. 2: $X_2(S_i)$
S_1 = heads	$X_1(S_1) = 1$	$X_2(S_1) = \pi$
S_2 = tails	$X_1(S_2) = -1$	$X_2(S_2) = \sqrt{2}$

4.2 RANDOM VARIABLES, DISTRIBUTION FUNCTIONS, AND DENSITY FUNCTIONS

Random Variables

In the applications of probability it is often more convenient to work in terms of numerical outcomes (for example, the number of errors in a digital data message) rather than nonnumerical outcomes (for example, failure of a component). Because of this, we introduce the idea of a *random variable,* which is defined as a rule that assigns a numerical value to each possible outcome of a chance experiment. (The term *random variable* is a misnomer; a random variable is really a function, since it is a rule that assigns the members of one set to those of another.)

As an example, consider the tossing of a coin. Possible assignments of random variables are given in Table 4.3. These are examples of *discrete random variables* and are illustrated in Figure 4.4(a).

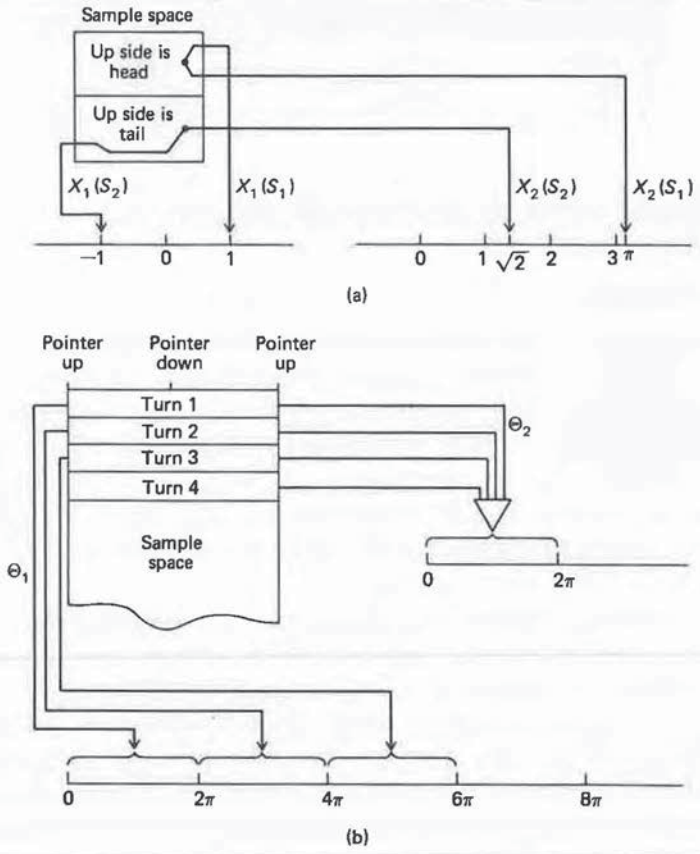
As an example of a *continuous random variable,* consider the spinning of a pointer, such as is typically found in children's games. A possible assignment of a random variable would be the angle Θ_1, in radians, that the pointer makes with the vertical when it stops. Defined in this fashion, Θ_1 has values that continuously increase with rotation of the pointer. A second possible random variable, Θ_2, would be Θ_1 minus integer multiples of 2π radians, such that $0 \leq \Theta_2 < 2\pi$, which is commonly denoted as Θ_1 modulo 2π. These random variables are illustrated in Figure 4.4(b).

At this point, we introduce a convention that will be adhered to throughout this book (although not in all books). Capital letters (X, Θ, and so on) denote random variables, and the corresponding lower-case letters (x, θ, and so on) *denote the values that the random variables take on.*

Probability (Cumulative) Distribution Functions

We need some way of probabilistically describing random variables that works equally well for discrete and continuous random variables. One way of accomplishing this is by means of the *cumulative distribution function* (*cdf*).

FIGURE 4.4 Pictorial representation of sample spaces and random variables. (a) Coin-tossing experiment. (b) Pointer-spinning experiment.



Consider a chance experiment with which we have associated a random variable X. We define the cdf $F_X(x)$ as

$$F_X(x) = \text{probability that } X \leq x = P(X \leq x) \tag{4.30}$$

We note that $F_X(x)$ is a *function* of x, not of the random variable X. But $F_X(x)$ also depends on the assignment of the random variable X, which accounts for the subscript.

The cdf has the following properties:

(a) $0 \leq F_X(x) \leq 1$, with $F_X(-\infty) = 0$ and $F_X(\infty) = 1$.
(b) $F_X(x)$ is continuous from the right; that is, $\lim_{x \to x_0+} F_X(x) = F_X(x_0)$.

(c) $F_X(x)$ is a nondecreasing function of x; that is, $F_X(x_1) \leq F_X(x_2)$ if $x_1 < x_2$.

The reasonableness of the preceding properties is shown by the following considerations.

Since $F_X(x)$ is a probability, it must, by our previous axioms, lie between 0 and 1, inclusive. Since $X = -\infty$ excludes all possible outcomes of our experiment, $F_X(-\infty) = 0$, and since $X = \infty$ includes all possible outcomes, $F_X(\infty) = 1$.

For $x_1 < x_2$, the events $X \leq x_1$ and $x_1 < X \leq x_2$ are mutually exclusive; furthermore, $X \leq x_2$ implies $X \leq x_1$ or $x_1 < X \leq x_2$. By axiom 3, therefore,

$$P(X \leq x_2) = P(X \leq x_1) + P(x_1 < X \leq x_2)$$

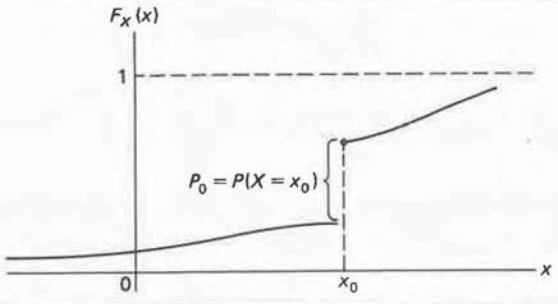
or

$$P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1) \tag{4.31}$$

Since probabilities are nonnegative, the left-hand side of (4.31) is nonnegative. Thus we see that property (c) holds.

The reasonableness of the right-continuity property is shown as follows. Suppose the random variable X takes on the value x_0 with probability P_0. Consider $P(X \leq x)$. If $x < x_0$, the event $X = x_0$ is not included, no matter how close x is to x_0. When $x = x_0$, we include the event $X = x_0$, which occurs with probability P_0. Since the events $X \leq x < x_0$ and $X = x_0$ are mutually exclusive, $P(X \leq x)$ must jump by an amount P_0 when $x = x_0$, as shown in Figure 4.5. Thus $F_X(x) = P(X \leq x)$ is continuous from the right. This is illustrated in Figure 4.5 by the dot on the curve to the right of the jump. What is more useful for our purposes, however, is that the *magnitude of any jump of $F_X(x)$, say at x_0, is equal to the probability that $X = x_0$.*

FIGURE 4.5 Illustration of the jump property of $F_X(x)$



Probability Density Function

From (4.31) we see that the cdf of a random variable is a complete and useful description for the computation of probabilities. However, we will also be interested in averages when we consider noise in communication systems. For the purpose of computing statistical averages, the *probability density function* (*pdf*) $f_X(x)$ of a random variable X is more convenient. The pdf of X is defined in terms of the cdf of X by

$$f_X(x) = \frac{dF_X(x)}{dx} \tag{4.32}$$

Since the cdf of a discrete random variable is discontinuous, its pdf, mathematically speaking, does not exist. By representing the derivative of a jump-discontinuous function at a point of discontinuity by a delta function of area equal to the magnitude of the jump, we can define pdf's for discrete random variables. In some books, this problem is avoided by defining a *probability mass function* for a discrete random variable, which consists simply of lines equal in magnitude to the probabilities that the random variable will take on its possible values.

Recalling that $F_X(-\infty) = 0$, we see from (4.32) that

$$F_X(x) = \int_{-\infty}^{x} f_X(x')\, dx' \tag{4.33}$$

That is, the *area* under the pdf from $-\infty$ to x is the probability that the observed value will be less than or equal to x.

From (4.32), (4.33), and the properties of $F_X(x)$, we see that the pdf has the following properties:

$$f_X(x) = \frac{dF_X(x)}{dx} \geq 0 \tag{4.34}$$

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1 \tag{4.35}$$

$$P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)\, dx \tag{4.36}$$

To obtain another enlightening and very useful interpretation of $f_X(x)$, we consider (4.36) with $x_1 = x - dx$ and $x_2 = x$. The integral then becomes $f_X(x)\, dx$, so

$$f_X(x)\, dx = P(x - dx < X \leq x) \tag{4.37}$$

TABLE 4.4

Outcome	X	$P(X = x_j)$
TT	$x_1 = 0$	$\frac{1}{4}$
TH, HT	$x_2 = 1$	$\frac{1}{2}$
HH	$x_3 = 2$	$\frac{1}{4}$

That is, the ordinate at any point x on the pdf curve multiplied by dx gives the probability of the random variable X lying in an infinitesimal range around the point x assuming that $f_X(x)$ is continuous at x.

The following two examples illustrate cdf's and pdf's for discrete and continuous cases, respectively.

▼ **EXAMPLE 4.9** Suppose two fair coins are tossed and X denotes the number of heads that turn up. The possible outcomes, the corresponding values of X, and the respective probabilities are summarized in Table 4.4. The cdf and pdf for this experiment and random variable definition are shown in Figure 4.6. The properties of the cdf and pdf for discrete random variables are demonstrated by this figure, as a careful examination will reveal. It is emphasized that the cdf and pdf change if the definition of the random variable or the probability assigned is changed. As an exercise, you should plot the cdf and pdf for the case where X denotes the number of heads divided by 2. ▲

FIGURE 4.6 The cdf and pdf for a coin-tossing experiment

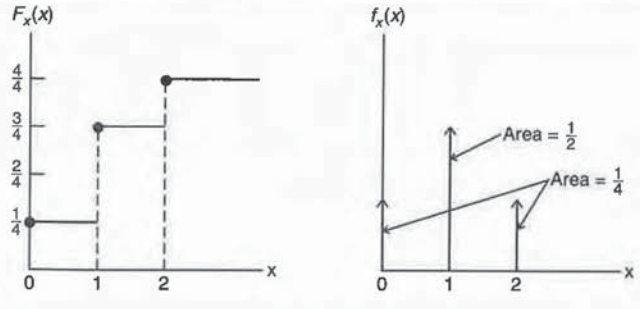
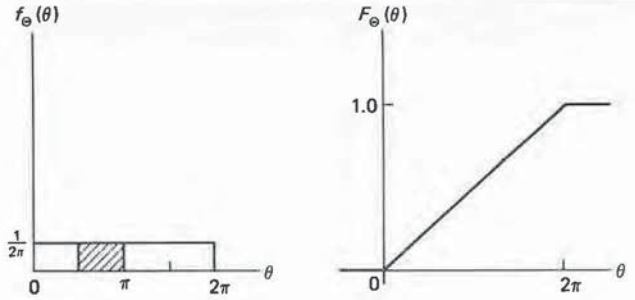


FIGURE 4.7 The pdf (a) and cdf (b) for a pointer-spinning experiment.



EXAMPLE 4.10 Consider the pointer-spinning experiment described earlier. We assume that any one stopping point is not favored over any other and that our random variable Θ is defined as the angle that the pointer makes with the vertical, modulo 2π. Thus Θ is limited to the range $(0, 2\pi)$, and for any two angles θ_1 and θ_2 in $(0, 2\pi)$, we have

$$P(\theta_1 - \Delta\theta < \Theta \leq \theta_1) = P(\theta_2 - \Delta\theta < \Theta \leq \theta_2) \tag{4.38}$$

by the assumption that the pointer is equally likely to stop at any angle in $(0, 2\pi)$. In terms of the pdf $f_\Theta(\theta)$, this can be written, using (4.37), as

$$f_\Theta(\theta_1) = f_\Theta(\theta_2), \qquad 0 \leq \theta_1, \theta_2 < 2\pi \tag{4.39}$$

Thus, in the interval $(0, 2\pi)$, $f_\Theta(\theta)$ is a constant, and outside $(0, 2\pi)$, $f_\Theta(\theta)$ is zero by the modulo 2π condition (this means that angles less than 0 or greater than 2π are impossible). By (4.35), it follows that $f_\Theta(\theta) = 1/(2\pi)$ in $(0, 2\pi)$; $f_\Theta(\theta)$ is shown graphically in Figure 4.7(a). The cdf $F_\Theta(\theta)$ is easily obtained by performing a graphical integration of $f_\Theta(\theta)$ and is shown in Figure 4.7(b).

To illustrate the use of these graphs, suppose we wish to find the probability of the pointer landing anyplace in the interval $(\frac{1}{2}\pi, \pi)$. The desired probability is given either as the area under the pdf curve from $\frac{1}{2}\pi$ to π, shaded in Figure 4.7(a), or as the value of the ordinate at $\theta = \pi$ minus the value of the ordinate at $\theta = \frac{1}{2}\pi$ on the cdf curve. The probability of the pointer's landing *exactly* at $\frac{1}{2}\pi$, however, is 0. ▲

Joint cdf's and pdf's

Some chance experiments must be characterized by two or more random variables. The cdf or pdf description is readily extended to such cases. For simplicity, we will consider only the case of two random variables.

To give a specific example, consider the chance experiment in which darts are repeatedly thrown at a target, as shown schematically in Figure 4.8. The point at which the dart lands on the target must be described in terms of two numbers. In this example, we denote the impact point by the two random variables X and Y, whose values are the xy coordinates of the point where the dart sticks, with the origin being fixed at the bull's eye.

The *joint cdf* of X and Y is defined as

$$F_{XY}(x, y) = P(X \leq x, \quad Y \leq y) \tag{4.40}$$

where the comma is interpreted as "and." The *joint pdf* of X and Y is defined as

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \, \partial y} \tag{4.41}$$

Just as we did in the case of single random variables, we can show that

$$P(x_1 < X \leq x_2, y_1 < Y \leq y_2) = \int_{y_1}^{y_2} \int_{x_1}^{x_2} f_{XY}(x, y) \, dx \, dy \tag{4.42}$$

which is the two-dimensional equivalent of (4.36). Letting $x_1 = y_1 = -\infty$ and $x_2 = y_2 = \infty$, we include the entire sample space. Thus

$$F_{XY}(\infty, \infty) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) \, dx \, dy = 1 \tag{4.43}$$

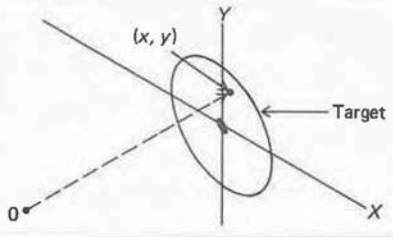
Letting $x_1 = x - dx$, $x_2 = x$, $y_1 = y - dy$, and $y_2 = y$, we obtain the following enlightening special case of (4.42):

$$f_{XY}(x, y) \, dx \, dy = P(x - dx < X \leq x, \quad y - dy < Y \leq y) \tag{4.44}$$

Thus the probability of finding X in an infinitesimal interval around x while simultaneously finding Y in an infinitesimal interval around y is $f_{XY}(x, y) \, dxdy$.

Given a joint cdf or pdf, we can obtain the cdf or pdf of one of the random

FIGURE 4.8 The dart-throwing experiment



variables using the following considerations. The cdf for X irrespective of the value Y takes on is simply

$$\begin{aligned} F_X(x) &= P(X \leq x, -\infty < Y < \infty) \\ &= F_{XY}(x, \infty) \end{aligned} \tag{4.45}$$

By similar reasoning, the cdf for Y alone is

$$F_Y(y) = F_{XY}(\infty, y) \tag{4.46}$$

$F_X(x)$ and $F_Y(y)$ are referred to as *marginal cdf's*. Using (4.40) and (4.42), we can express (4.45) and (4.46) as

$$F_X(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{x} f_{XY}(x', y')\, dx'\, dy' \tag{4.47}$$

and

$$F_Y(y) = \int_{-\infty}^{y} \int_{-\infty}^{\infty} f_{XY}(x', y')\, dx'\, dy' \tag{4.48}$$

respectively. Since

$$f_X(x) = \frac{dF_X(x)}{dx} \quad \text{and} \quad f_Y(y) = \frac{dF_Y(y)}{dy} \tag{4.49}$$

we obtain

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y')\, dy' \tag{4.50}$$

and

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x', y)\, dx' \tag{4.51}$$

from (4.47) and (4.48), respectively. Thus, to obtain the marginal pdf's $f_X(x)$ and $f_Y(y)$ from the joint pdf $f_{XY}(x, y)$, we simply integrate out the undesired variable (or variables for more than two random variables). Hence the joint cdf or pdf contains all the information possible about the joint random variables X and Y. Similar results hold for more than two random variables.

Two random variables are *statistically independent* (or simply independent) if the values each takes on do not influence the values of the other. Thus, for any x and y, it must be true that

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y) \tag{4.52a}$$

or, in terms of cdf's,

$$F_{XY}(x, y) = F_X(x)F_Y(y) \tag{4.52b}$$

That is, the joint cdf of independent random variables factors into the product of the separate marginal cdf's. Differentiating both sides of (4.52b) with respect to first x and then y, and using the definition of the pdf, we obtain

$$f_{XY}(x, y) = f_X(x)f_Y(y) \tag{4.53}$$

which shows that the joint pdf of independent random variables also factors.

If two random variables are not independent, we can write their joint pdf in terms of conditional pdf's $f_{X|Y}(x\,|\,y)$ and $f_{Y|X}(y\,|\,x)$ as

$$\begin{aligned} f_{XY}(x, y) &= f_X(x)f_{Y|X}(y\,|\,x) \\ &= f_Y(y)f_{X|Y}(x\,|\,y) \end{aligned} \tag{4.54}$$

These relations *define* the conditional pdf's of two random variables. An intuitively satisfying interpretation of $f_{X|Y}(x\,|\,y)$ is

$$f_{X|Y}(x\,|\,y)\,dx = P[x - dx < X \leq x \text{ given } Y = y] \tag{4.55}$$

There is a similar interpretation for $f_{Y|X}(y\,|\,x)$. Equation (4.55) is reasonable in that if X and Y are dependent, a given value of Y should influence the probability distribution for X. On the other hand, if X and Y are independent, information about one of the random variables tells us nothing about the other. Thus, for independent random variables,

$$f_{X|Y}(x\,|\,y) = f_X(x) \qquad \text{and} \qquad f_{Y|X}(y\,|\,x) = f_Y(y) \tag{4.56}$$

which could serve as an alternative definition of statistical independence. The following example illustrates the preceding ideas.

▼ **EXAMPLE 4.11** Two random variables X and Y have the joint pdf

$$f_{XY}(x, y) = \begin{cases} Ae^{-(2x+y)}, & x, y \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{4.57}$$

where A is a constant. We evaluate A from

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x, y)\,dx\,dy = 1 \tag{4.58}$$

Since

$$\int_{0}^{\infty}\int_{0}^{\infty} e^{-(2x+y)}\,dx\,dy = \tfrac{1}{2} \tag{4.59}$$

$A = 2$. We find the marginal pdf's from (4.50) and (4.51) as follows:

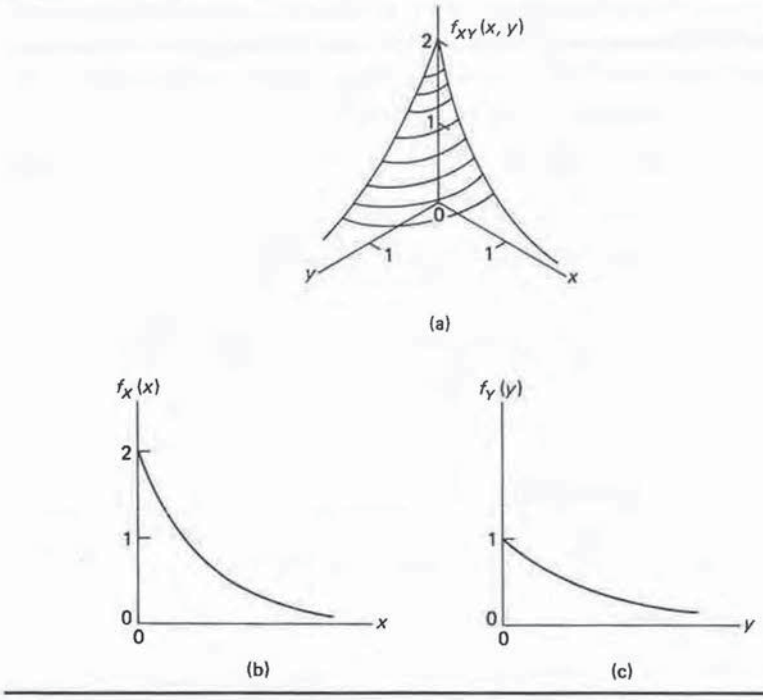
$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy = \begin{cases} \int_0^{\infty} 2e^{-(2x+y)}\, dy, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

$$= \begin{cases} 2e^{-2x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{4.60}$$

$$f_Y(y) = \begin{cases} e^{-y}, & y \geq 0 \\ 0, & y < 0 \end{cases} \tag{4.61}$$

These joint and marginal pdf's are shown in Figure 4.9. From these results, we note that X and Y are statistically independent since $f_{XY}(x, y) = f_X(x)f_Y(y)$.

FIGURE 4.9 Joint and marginal pdf's for two random variables. (a) Joint pdf. (b) Marginal pdf for X. (c) Marginal pdf for Y.



We find the joint cdf by integrating the joint pdf on both variables, using (4.42) and (4.40), which gives

$$F_{XY}(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{XY}(x', y')\, dx'\, dy' = \begin{cases} (1 - e^{-2x})(1 - e^{-y}), & x, y \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{4.62}$$

Dummy variables are used in the integration to avoid confusion. Note that $F_{XY}(-\infty, -\infty) = 0$ and $F_{XY}(\infty, \infty) = 1$, as they should, since in the first case we have the probability of an impossible event and in the latter we include all possible outcomes. We also can use the result for $F_{XY}(x, y)$ to obtain

$$F_X(x) = F_{XY}(x, \infty) = \begin{cases} (1 - e^{-2x}), & x \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{4.63a}$$

and

$$F_Y(y) = F_{XY}(\infty, y) = \begin{cases} (1 - e^{-y}), & y \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{4.63b}$$

Also note that the joint cdf factors into the product of the marginal cdf's, as it should be for statistically independent random variables.

The conditional pdf's are

$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)} = \begin{cases} 2e^{-2x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{4.64a}$$

and

$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_X(x)} = \begin{cases} e^{-y}, & y \geq 0 \\ 0, & y < 0 \end{cases} \tag{4.64b}$$

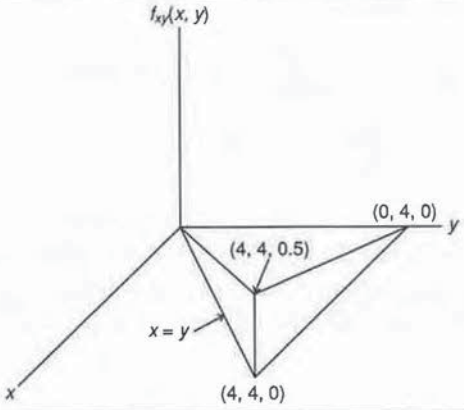
They are equal to the respective marginal pdf's, as they should be for independent random variables. ▲

▼ **EXAMPLE 4.12** To illustrate how to check a probability density function for statistical independence of the corresponding random variables, we consider the joint pdf

$$f_{XY}(x, y) = \begin{cases} \beta xy, & 0 \leq x \leq y,\ 0 \leq y \leq 4 \\ 0, & \text{otherwise} \end{cases} \tag{4.65}$$

For independence, the joint pdf should be the product of the marginal pdf's.

FIGURE 4.10 Probability density function for Example 4.12



SOLUTION This example is somewhat tricky because of the limits, so a diagram of the pdf is given in Figure 4.10. Although not absolutely necessary to answer the question of independence, we proceed by first finding the constant β which is chosen to normalize the pdf to unity. Integration over all x and y gives $\beta = \frac{1}{32}$. We next proceed to find the marginal pdf's. Integrating over x first and checking the figure to obtain the proper limits of integration, we obtain

$$f_Y(y) = \int_0^y \frac{xy}{32}\,dx, \qquad 0 \le y \le 4$$
$$= \begin{cases} y^3/64, & 0 \le y \le 4 \\ 0, & \text{otherwise} \end{cases} \tag{4.66}$$

The pdf on X is similarly obtained as

$$f_X(x) = \int_x^4 \frac{xy}{32}\,dy, \qquad 0 \le x \le 4$$
$$= \begin{cases} (x/4)[1 - (x/4)^2], & 0 \le x \le 4 \\ 0, & \text{otherwise} \end{cases} \tag{4.67}$$

It is clear that the product of the marginal pdf's is not equal to the joint pdf, so the random variables X and Y are not statistically independent. ▲

Transformation of Random Variables

Situations are often encountered in which we know the pdf (or cdf) of a random variable X and desire the pdf of a second random variable Y defined as a function of X, for example,

$$Y = g(X) \tag{4.68}$$

We initially consider the case where $g(X)$ is a monotonic function of its argument (for example, it is either nondecreasing or nonincreasing as the independent variable ranges from $-\infty$ to ∞), a restriction that may be relaxed if necessary.

A typical function is shown in Figure 4.11. The probability that X lies in the range $(x - dx, x)$ is the same as the probability that Y lies in the range $(y - dy, y)$, where $y = g(x)$. Using (4.37), we obtain

$$f_X(x)\, dx = f_Y(y)\, dy \tag{4.69a}$$

if $g(X)$ is monotonically increasing, and

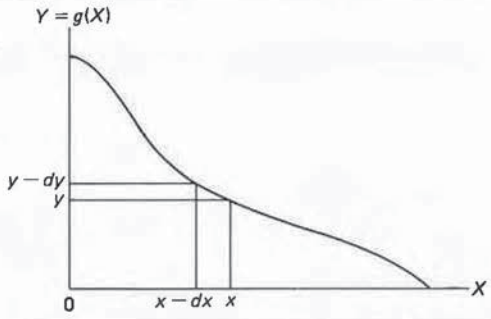
$$f_X(x)\, dx = -f_Y(y)\, dy \tag{4.69b}$$

if $g(X)$ is monotonically decreasing, since an *increase* in x results in a *decrease* in y. Both cases are taken into account by writing

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|_{x = g^{-1}(y)} \tag{4.70}$$

where $x = g^{-1}(y)$ denotes the inversion of (4.68) for x in terms of y.

FIGURE 4.11 A typical monotonic transformation of a random variable



EXAMPLE 4.13 To illustrate the use of (4.70), let us consider the pdf of Example 4.10, namely

$$f_\Theta(\theta) = \begin{cases} \dfrac{1}{2\pi}, & 0 \le \theta \le 2\pi \\ 0, & \text{otherwise} \end{cases} \tag{4.71}$$

and the transformation

$$Y = -\left(\frac{1}{\pi}\right)\Theta + 1 \tag{4.72}$$

Since $dy/d\theta = -1/\pi$, the pdf of Y is

$$f_Y(y) = f_\Theta(\theta = -\pi y + \pi)|-\pi| = \begin{cases} \frac{1}{2}, & -1 \le y \le 1 \\ 0, & \text{otherwise} \end{cases} \tag{4.73}$$

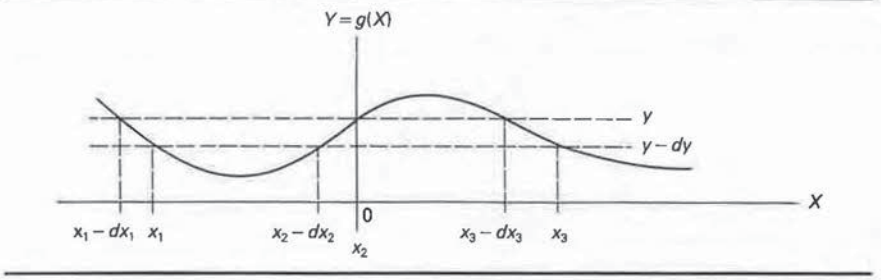
Consider next the case of $g(x)$ nonmonotonic. A typical case is shown in Figure 4.12. For the case shown, the infinitesimal interval $(y - dy, y)$ corresponds to three infinitesimal intervals on the x axis: $(x_1 - dx_1, x_1)$, $(x_2 - dx_2, x_2)$, and $(x_3 - dx_3, x_3)$. The probability that X lies in any one of these intervals is equal to the probability that Y lies in the interval $(y - dy, y)$. This can be generalized to the case of N disjoint intervals where it follows that

$$P(y - dy, y) = \sum_{i=1}^{N} P(x_i - dx_i, x_i) \tag{4.74}$$

where we have generalized to N intervals on the X axis corresponding to the interval $(y - dy, y)$ on the Y axis. Since

$$P(y - dy, y) = f_Y(y)\, dy \tag{4.75}$$

FIGURE 4.12 A nonmonotonic transformation of a random variable



and

$$P(x_i - dx_i, x_i) = f_X(x_i)\, dx_i \tag{4.76}$$

we have

$$f_Y(y) = \sum_{n=1}^{N} f_X(x_i) \left| \frac{dx_i}{dy} \right|_{x_i = g_i^{-1}(y)} \tag{4.77}$$

where the absolute value signs are used because a probability must be positive and $x_i = g_i^{-1}(y)$ is the ith solution to $g(y) = x$.

▼ **EXAMPLE 4.14** Consider the transformation

$$y = x^2 \tag{4.78}$$

If $f_X(x) = 0.5 \exp(-|x|)$, find $f_Y(y)$.

SOLUTION There are two solutions to $x^2 = y$; these are

$$x_1 = \sqrt{y} \quad \text{and} \quad x_2 = -\sqrt{y}, \quad y > 0 \tag{4.79}$$

Their derivatives are

$$\frac{dx_1}{dy} = 1/(2\sqrt{y}) \quad \text{and} \quad \frac{dx_2}{dy} = -1/(2\sqrt{y}), \quad y > 0 \tag{4.80}$$

Using these results in (4.77), we obtain $f_Y(y)$ to be

$$f_Y(y) = \frac{e^{-\sqrt{y}}}{2\sqrt{y}}, \quad y > 0 \tag{4.81}$$

▲ Since Y cannot be negative, $f_Y(y) = 0, y < 0$.

For two or more random variables, we consider only one-to-one transformations and the probability of the joint occurrence of random variables lying within infinitesimal areas (or volumes for more than two random variables). Thus, suppose two new random variables U and V are defined in terms of two old random variables X and Y by the relations

$$U = g_1(X, Y) \quad \text{and} \quad V = g_2(X, Y) \tag{4.82}$$

The new pdf $f_{UV}(u, v)$ is obtained from the old pdf $f_{XY}(x, y)$ by using (4.44) to write

$$P(u - du < U \leq u, \quad v - dv < V \leq v) = P(x - dx < X \leq x, \quad y - dy < Y \leq y)$$

or

$$f_{UV}(u, v)\, dA_{uv} = f_{XY}(x, y)\, dA_{xy} \tag{4.83}$$

where dA_{uv} is the infinitesimal area in the uv plane corresponding to the infinitesimal area dA_{xy} in the xy plane through the transformation (4.82).

The ratio of elementary area dA_{xy} to dA_{uv} is given by the Jacobian

$$\frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} \tag{4.84}$$

so that

$$f_{UV}(u, v) = f_{XY}(x, y) \left| \frac{\partial(x, y)}{\partial(u, v)} \right|_{\substack{x = g_1^{-1}(u, v) \\ y = g^{2-1}(u, v)}} \tag{4.85}$$

where the inverse functions $g_1^{-1}(u, v)$ and $g_2^{-1}(u, v)$ exist because the transformation (4.82) is assumed to be one-to-one. An example will help clarify this discussion.

▼ **EXAMPLE 4.15** Consider the dart-throwing game discussed in connection with joint cdf's and pdf's. We assume that the joint pdf in terms of rectangular coordinates for the impact point is

$$f_{XY}(x, y) = \frac{\exp[-(x^2 + y^2)/2\sigma^2]}{2\pi\sigma^2}, \qquad -\infty < x, y < \infty \tag{4.86}$$

where σ^2 is a constant. This is a special case of the *joint Gaussian pdf,* which we will discuss in more detail shortly.

Instead of rectangular coordinates, we wish to use polar coordinates R and Θ, defined by

$$R = \sqrt{X^2 + Y^2} \tag{4.87a}$$

and

$$\Theta = \tan^{-1}\left(\frac{Y}{X}\right) \tag{4.87b}$$

so that

$$X = R \cos \Theta = g_1^{-1}(R, \Theta) \tag{4.88a}$$

and

$$Y = R \sin \Theta = g_2^{-1}(R, \Theta) \tag{4.88b}$$

where

$$0 \leq \Theta < 2\pi, \qquad 0 \leq R < \infty$$

Under this transformation, the infinitesimal area $dx\,dy$ in the xy plane transforms to the area $r\,dr\,d\theta$ in the $r\theta$ plane, as determined by the Jacobian, which is

$$\frac{\partial(x, y)}{\partial(r, \theta)} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r \tag{4.89}$$

Thus the joint pdf of R and θ is

$$f_{R\Theta}(r, \theta) = \frac{re^{-r^2/2\sigma^2}}{2\pi\sigma^2}, \qquad \begin{matrix} 0 \leq \theta < 2\pi \\ 0 \leq r < \infty \end{matrix} \tag{4.90}$$

which follows from (4.85), which for this case takes the form

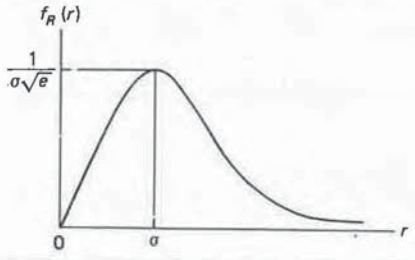
$$f_{R\Theta}(r, \theta) = rf_{XY}(x, y)\Big|_{\substack{x=r\cos\theta \\ y=r\sin\theta}} \tag{4.91}$$

If we integrate $f_{R\Theta}(r, \theta)$ over θ to get the pdf for R alone, we obtain

$$f_R(r) = \frac{re^{-r^2/2\sigma^2}}{\sigma^2}, \qquad 0 \leq r < \infty \tag{4.92}$$

which is referred to as the *Rayleigh pdf*. The probability that the dart lands in a ring of radius r from the bull's eye and of thickness dr is given by $f_R(r)\,dr$. From the sketch of the Rayleigh pdf given in Figure 4.13, we see that the most probable distance for the dart to land from the bull's eye is $R = \sigma$. By integrating (4.90) over r, it can be shown that the pdf of θ is uniform in $(0, 2\pi)$. ▲

FIGURE 4.13 The Rayleigh pdf



4.3 STATISTICAL AVERAGES

The probability functions (cdf and pdf) we have just discussed provide us with all the information possible about a random variable or a set of random variables. Often we do not need such a complete description, and in many cases, we are not able to obtain the cdf or pdf. A partial description of a random variable is given in terms of various statistical averages, or mean values.

Average of a Discrete Random Variable

To introduce the idea of a statistical average, we consider a discrete random variable X which takes on the possible values $x_1, x_2, \ldots, x_M$ with the respective probabilities $P_1, P_2, \ldots, P_M$. The statistical average, or *expectation,* of X is defined as

$$\bar{X} = E\{X\} = \sum_{j=1}^{M} x_j P_j \tag{4.93}$$

To show the reasonableness of this definition, we look at it in terms of relative frequency. If the underlying chance experiment is repeated a large number of times N and $X = x_1$ is observed n_1 times and $X = x_2$ is observed n_2 times, etc., the arithmetical average of the observed values is

$$\frac{n_1 x_1 + n_2 x_2 + \cdots + n_M x_M}{N} = \sum_{j=1}^{M} x_j \frac{n_j}{N} \tag{4.94}$$

But, by the relative-frequency interpretation of probability, n_j/N approaches P_j, the probability of the event $X = x_j$, as N becomes large. Thus, in the limit as N approaches infinity, (4.94) becomes (4.93).

▼ EXAMPLE 4.16 Suppose that a class of 100 students is given a test and The following scores result:

Score:	90	85	80	75	70	65	60
No. of students:	1	4	20	50	20	4	1

The average score, as we have known how to compute since our grammar school days, is

$$\frac{(90 \times 1) + (85 \times 4) + \cdots + (60 \times 1)}{100} = 75 \tag{4.95}$$

Now consider the same problem using (4.93). Defining the random variable X as taking on values numerically equal to the test scores, we have, using the relative-frequency approximation, the following probabilities:

$$\begin{aligned} P(X = 90) &\cong 0.01 \\ P(X = 85) &\cong 0.04 \\ P(X = 80) &\cong 0.2 \\ P(X = 75) &\cong 0.5 \\ P(X = 70) &\cong 0.2 \\ P(X = 65) &\cong 0.04 \\ P(X = 60) &\cong 0.01 \end{aligned} \tag{4.96}$$

Using (4.93), we obtain, for the expectation of X, the same result as before:

$$E\{X\} = (0.01 \times 90) + (0.04 \times 85) + \cdots + (0.01 \times 60) = 75 \tag{4.97}$$

▲

Average of a Continuous Random Variable

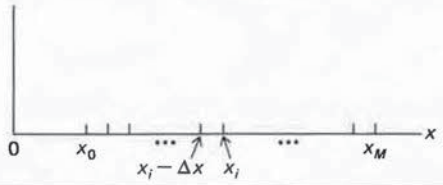
Now we look at the case where X is a continuous random variable with the pdf $f_X(x)$. In order to extend the definition given by (4.93) to this case, we consider the range of values that X may take on, say x_0 to x_M, to be broken up into a large number of small subintervals of length Δx, as shown in Figure 4.14.

The probability that X lies between $x_i - \Delta x$ and x_i is, from (4.37), given by

$$P(x_i - \Delta x < X \leq x_i) \cong f_X(x_i)\,\Delta x, \qquad i = 1, 2, \ldots, M \tag{4.98}$$

for Δx small. Thus we have approximated X by a discrete random variable that takes on the values $x_0, x_1, \ldots, x_M$ with probabilities $f_X(x_0)\Delta x, \ldots,$

FIGURE 4.14 A discrete approximation for a continuous random variable X



$f_X(x_M)\,\Delta x$, respectively. Using (4.93), the expectation of this random variable is

$$E\{X\} \cong \sum_{i=0}^{M} x_i f_X(x_i)\,\Delta x \tag{4.99a}$$

As Δx approaches zero, this becomes a better and better approximation for $E\{X\}$. In the limit, as Δx approaches dx, the sum becomes an integral, giving

$$E\{X\} = \int_{-\infty}^{\infty} x f_X(x)\,dx \tag{4.99b}$$

for the expectation of X.

Average of a Function of a Random Variable

We are interested not only in $E\{X\}$, which is referred to as the *mean* or *first moment* of X, but also in statistical averages of functions of X. Letting $Y = g(X)$, the statistical average or expectation of the new random variable Y could be obtained as

$$E\{Y\} = \int_{-\infty}^{\infty} y f_Y(y)\,dy \tag{4.100}$$

where $f_Y(y)$ is the pdf of Y, which can be found from $f_X(x)$ by application of (4.70). However, it is often more convenient simply to find the expectation of the function $g(X)$ as given by

$$\overline{g(X)} \triangleq E\{g(X)\} = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx \tag{4.101}$$

which is identical to $E\{Y\}$ as given by *(4.100)*. Two examples follow to illustrate the use of (4.100) and (4.101).

▼ EXAMPLE 4.17 Suppose the random variable Θ has the pdf

$$f_\Theta(\theta) = \begin{cases} \dfrac{1}{2\pi}, & |\theta| < \pi \\ 0, & \text{otherwise} \end{cases} \tag{4.102}$$

Then $E\{\Theta^n\}$ is referred to as the nth moment of Θ and is given by

$$E\{\Theta^n\} = \int_{-\infty}^{\infty} \theta^n f_\Theta(\theta)\,d\theta = \int_{-\pi}^{\pi} \theta^n \frac{d\theta}{2\pi} \tag{4.103}$$

Since the integrand is odd if n is odd, $E\{\Theta^n\} = 0$ for n odd. For n even,

$$E\{\Theta^n\} = \frac{1}{\pi}\int_0^{\pi} \theta^n \, d\theta = \frac{\pi^n}{n+1} \tag{4.104}$$

The first moment or mean of Θ, $E\{\Theta\}$, is a measure of the location of $f_\Theta(\theta)$ (that is, the "center of mass"). Since $f_\Theta(\theta)$ is symmetrically located about $\theta = 0$, it is not surprising that $\overline{\Theta} = 0$. ▲

▼ **EXAMPLE 4.18** Later we shall consider certain random waveforms that can be modeled as sinusoids with random phase angles having uniform pdf in $[-\pi, \pi]$. In this example, we consider a random variable X that is defined in terms of the uniform random variable Θ considered in Example 4.17 by

$$X = \cos \Theta \tag{4.105}$$

The density function of X, $f_X(x)$, is found as follows. First $-1 \leq \cos\theta \leq 1$, so $f_X(x) = 0$ for $|x| > 1$. Second, the transformation is not one-to-one, there being two values of Θ for each value of X, since $\cos\theta = \cos(-\theta)$. However, we can still apply (4.70) by noting that positive and negative angles have equal probabilities and writing

$$f_X(x) = 2f_\Theta(\theta)\left|\frac{d\theta}{dx}\right|, \qquad |x| < 1 \tag{4.106}$$

Now $\theta = \cos^{-1} x$ and $|d\theta/dx| = (1 - x^2)^{-1/2}$, which yields

$$f_X(x) = \begin{cases} \dfrac{1}{\pi\sqrt{1-x^2}}, & |x| \leq 1 \\ 0, & |x| > 1 \end{cases} \tag{4.107}$$

This pdf is illustrated in Figure 4.15.

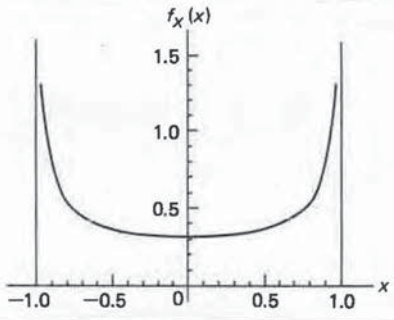
The mean and second moment of X can be calculated using either (4.100) or (4.101). Using (4.100), we obtain

$$\overline{X} = \int_{-1}^{1} \frac{x}{\pi\sqrt{1-x^2}}\,dx = 0 \tag{4.108}$$

because the integrand is odd, and

$$\overline{X^2} = \int_{-1}^{1} \frac{x^2\,dx}{\pi\sqrt{1-x^2}} = \frac{1}{2} \tag{4.109}$$

FIGURE 4.15 Probability density function of a sinusoid with uniform random phase



by a table of integrals. Using (4.101), we find that

$$\overline{X} = \int_{-\pi}^{\pi} \cos\theta \frac{d\theta}{2\pi} = 0 \tag{4.110}$$

and

$$\overline{X^2} = \int_{-\pi}^{\pi} \cos^2\theta \frac{d\theta}{2\pi} = \int_{-\pi}^{\pi} \frac{1}{2}(1 + \cos 2\theta) \frac{d\theta}{2\pi} = \frac{1}{2} \tag{4.111}$$

as obtained before. ▲

Average of a Function of More Than One Random Variable

The expectation of a function $g(X, Y)$ of two random variables X and Y is defined in a manner analogous to the case of a single random variable. If $f_{XY}(x, y)$ is the joint pdf of X and Y, the expectation of $g(X, Y)$ is

$$E\{g(X, Y)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y)\, dx\, dy \tag{4.112}$$

The generalization to more than two random variables should be obvious.

Equation (4.112) and its generalization to more than two random variables include the single-random-variable case, for suppose $g(X, Y)$ is replaced by a function of X alone, say $h(X)$. Then using (4.50), we obtain the following from (4.112):

$$\begin{aligned} E\{h(X)\} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x) f_{XY}(x, y)\, dx\, dy \\ &= \int_{-\infty}^{\infty} h(x) f_X(x)\, dx \end{aligned} \tag{4.113}$$

▼ **EXAMPLE 4.19** Consider the joint pdf of Example 4.11 and the expectation of $g(X, Y) = XY$. From (4.112), this expectation is

$$E\{XY\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{XY}(x, y)\,dx\,dy = \int_{0}^{\infty}\int_{0}^{\infty} 2xye^{-(2x+y)}\,dx\,dy$$
$$= 2\int_{0}^{\infty} xe^{-2x}\,dx\int_{0}^{\infty} ye^{-y}\,dy = \tfrac{1}{2} \tag{4.114}$$

We recall from Example 4.11 that X and Y are statistically independent. From the last line of the preceding equation for $E\{XY\}$, we see that

$$E\{XY\} = E\{X\}E\{Y\} \tag{4.115}$$

a result that holds in general for *statistically independent* random variables. In fact, for independent random variables, it readily follows that

$$E\{h(X)g(Y)\} = E\{h(X)\}E\{g(Y)\} \tag{4.116}$$

where $h(X)$ and $g(Y)$ are two functions of X and Y, respectively. In the special case where $h(X) = X^m$ and $g(Y) = Y^n$, the average $E\{h(X)g(Y)\} = E\{X^mY^n\}$. The expectations $E\{X^mY^n\}$ are referred to as the *joint moments* of order $m + n$ of X and Y. According to (4.116), the *joint moments of statistically independent random variables factor.* ▲

▼ **EXAMPLE 4.20** When finding the expectation of a function of more than one random variable, you may find it easier to use the concept of conditional expectation. Consider, for example, a function $g(X, Y)$ of two random variables X and Y, with the joint pdf $f_{XY}(x, y)$. The expectation of $g(X, Y)$ is

$$E\{g(X, Y)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)f_{XY}(x, y)\,dx\,dy$$
$$= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} g(x, y)f_{X|Y}(x\,|\,y)\,dx\right] f_Y(y)\,dy = E\{E[g(X, Y)\,|\,Y]\} \tag{4.117}$$

where $f_{X|Y}(x\,|\,y)$ is the conditional pdf of X, given Y, and $E[g(X, Y)\,|\,Y]$ denotes the conditional expectation of $g(X, Y)$, given Y.

As a specific application, let us consider the firing of projectiles at a target. Projectiles are fired until the target is hit for the first time, after which firing ceases. Assume that the probability of a projectile's hitting the target is p and that the firings are independent of one another. Find the average number of projectiles fired at the target.

SOLUTION To solve this problem, let N be a random variable denoting the number of projectiles fired at the target. Let the random variable H be 1 if the first projectile hits the target and 0 if it does not. Using the concept of conditional expectation, we find the average value of N is given by

$$\begin{aligned} E\{N\} = E\{E[N|H]\} &= pE[N|H = 1] + (1 - p)E[N|H = 0] \\ &= p \times 1 + (1 - p)(1 + E[N]) \end{aligned} \tag{4.118}$$

where $E[N|H = 0] = 1 + E[N]$ because $N \geq 1$ if a miss occurs on the first firing. By solving the last expression for $E[N]$, we obtain

$$E[N] = \frac{1}{p} \tag{4.119}$$

If $E[N]$ is evaluated directly, it is necessary to sum the series:

$$E[N] = 1 \times p + 2 \times (1 - p)p + 3 \times (1 - p)^2 p + \cdots \tag{4.120}$$

which is not too difficult in this instance. However, the conditional-expectation method clearly makes it easier to keep track of the bookkeeping. ▲

Variance of a Random Variable

The statistical average

$$\sigma_x^2 \triangleq (X - \bar{X})^2 = E\{[X - E(X)]^2\} \tag{4.121}$$

is called the *variance* of the random variable X; σ_x is called the *standard deviation* of X and is a measure of the concentration of the pdf of X, or $f_X(x)$, about the mean. The notation var $\{X\}$ for σ_x^2 is sometimes used. A useful relation for obtaining σ_x^2 is

$$\sigma_x^2 = E\{X^2\} - E^2\{X\} \tag{4.122}$$

which, in words, says that the variance of X is simply its second moment minus its mean, squared. To prove (4.122), we let $E\{X\} = m_x$. Then

$$\begin{aligned} \sigma_x^2 &= \int_{-\infty}^{\infty} (x - m_x)^2 f_X(x)\, dx = \int_{-\infty}^{\infty} (x^2 - 2xm_x + m_x^2) f_X(x)\, dx \\ &= E\{X^2\} - 2m_x^2 + m_x^2 = E\{X^2\} - E^2\{X\} \end{aligned} \tag{4.123}$$

which follows because $\int_{-\infty}^{\infty} x f_X(x)\, dx = m_x$.

EXAMPLE 4.21 Let X have the uniform pdf

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases} \tag{4.124}$$

Then

$$E\{X\} = \int_a^b x \frac{dx}{b-a} = \tfrac{1}{2}(a+b) \tag{4.125}$$

and

$$E\{X^2\} = \int_a^b x^2 \frac{dx}{b-a} = \tfrac{1}{3}(b^2 + ab + a^2) \tag{4.126}$$

Thus

$$\sigma_x^2 = \tfrac{1}{3}(b^2 + ab + a^2) - \tfrac{1}{4}(a^2 + 2ab + b^2) = \tfrac{1}{12}(a-b)^2 \tag{4.127}$$

Consider the following special cases:

1. $a = 1$ and $b = 2$, for which $\sigma_x^2 = \frac{1}{12}$
2. $a = 0$ and $b = 1$, for which $\sigma_x^2 = \frac{1}{12}$
3. $a = 0$ and $b = 2$, for which $\sigma_x^2 = \frac{1}{3}$

For cases 1 and 2, the pdf of X has the same width but is centered about different means; the variance is the same for both cases. In case 3, the pdf is wider than it is for cases 1 and 2, which is manifested by the larger variance. ▲

Average of a Linear Combination of N *Random Variables*

It is easily shown that the expected value, or average, of an arbitrary linear combination of random variables is the same as the linear combination of their respective means. That is,

$$E\left\{\sum_{i=1}^{N} a_i X_i\right\} = \sum_{i=1}^{N} a_i E\{X_i\} \tag{4.128}$$

where $X_1, X_2, \ldots, X_N$ are random variables and $a_1, a_2, \ldots, a_N$ are arbitrary constants. Equation (4.128) will be demonstrated for the special case $N = 2$; generalization to the case $N > 2$ is not difficult, but results in unwieldy notation.

Let $f_{X_1X_2}(x_1, x_2)$ be the joint pdf of X_1 and X_2. Then, using the definition of the expectation of a function of two random variables in (4.112), it follows that

$$\begin{aligned} E\{a_1X_1 + a_2X_2\} &\triangleq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (a_1x_1 + a_2x_2) f_{X_1X_2}(x_1, x_2)\, dx_1\, dx_2 \\ &= a_1 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 f_{X_1X_2}(x_1, x_2)\, dx_1\, dx_2 \\ &\quad + a_2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_2 f_{X_1X_2}(x_1, x_2)\, dx_1\, dx_2 \end{aligned} \tag{4.129}$$

Considering the first double integral and using (4.50) and (4.99), we find that

$$\begin{aligned} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 f_{X_1X_2}(x_1, x_2)\, dx_1\, dx_2 &= \int_{-\infty}^{\infty} x_1 \left\{ \int_{-\infty}^{\infty} f_{X_1X_2}(x_1, x_2)\, dx_2 \right\} dx_1 \\ &= \int_{-\infty}^{\infty} x_1 f_{X_1}(x_1)\, dx_1 \\ &\triangleq E\{X_1\} \end{aligned} \tag{4.130}$$

Similarly, it can be shown that the second double integral reduces to $E\{X_2\}$. Thus (4.128) has been proved for the case $N = 2$.

We emphasize that (4.128) holds regardless of whether or not the X_i terms are independent. Also, it should be noted that a similar result holds for a linear combination of functions of N random variables.

Variance of a Linear Combination of Independent Random Variables

If $X_1, X_2, \ldots, X_N$ are *statistically independent* random variables, then

$$\text{var}\left\{\sum_{i=1}^{N} a_iX_i\right\} = \sum_{i=1}^{N} a_i^2 \text{ var } \{X_i\} \tag{4.131}$$

where $a_1, a_2, \ldots, a_N$ are arbitrary constants and var $\{X_i\} \triangleq E\{(X_i - \bar{X}_i)^2\}$. This relation will be demonstrated for the case $N = 2$. Let $Z = a_1X_1 + a_2X_2$ and let $f_{X_i}(x_i)$ be the marginal pdf of X_i. Then the joint pdf of X_1 and X_2 is $f_{X_1}(x_1)f_{X_2}(x_2)$ by the assumption of statistical independence. Also, $\bar{Z} =$

$a_1\bar{X}_1 + a_2\bar{X}_2$ by (4.128). Also, var $\{Z\} = E\{(Z - \bar{Z})^2\}$. But, since $Z = a_1X_1 + a_2X_2$, we may write var $\{Z\}$ as

$$\begin{aligned}\text{var}\{Z\} &= E\{[(a_1X_1 + a_2X_2) - (a_1\bar{X}_1 + a_2\bar{X}_2)]^2\} \\ &= E\{[a_1(X_1 - \bar{X}_1) + a_2(X_2 - \bar{X}_2)]^2\} \\ &= a_1^2E\{(X_1 - \bar{X}_1)^2\} + 2a_1a_2E\{(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)\} \\ &\quad + a_2^2E\{(X_2 - \bar{X}_2)^2\} \end{aligned} \tag{4.132}$$

The first and last terms in the preceding equation are a_1^2 var $\{X_1\}$ and a_2^2 var $\{X_2\}$, respectively. The middle term is zero, since

$$\begin{aligned}&E\{(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)\} \\ &\quad = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_1 - \bar{X}_1)(x_2 - \bar{X}_2)f_{X_1}(x_1)f_{X_2}(x_2)\,dx_1\,dx_2 \\ &\quad = \int_{-\infty}^{\infty}(x_1 - \bar{X}_1)f_{X_1}(x_1)\,dx_1\int_{-\infty}^{\infty}(x_2 - \bar{X}_2)f_{X_2}(x_2)\,dx_2 \\ &\quad = (\bar{X}_1 - \bar{X}_1)(\bar{X}_2 - \bar{X}_2) = 0 \end{aligned} \tag{4.133}$$

We reiterate that the assumption of *statistical independence* was used to show that the middle term above is zero.

Another Special Average—The Characteristic Function

If we let $g(X) = e^{jvX}$ in (4.101), we obtain an average known as the *characteristic function* of X, or $M_X(jv)$, defined as

$$M_X(jv) \triangleq E\{e^{jvX}\} = \int_{-\infty}^{\infty} f_X(x)e^{+jvx}\,dx \tag{4.134}$$

It is seen that $M_X(jv)$ would be the *Fourier transform* of $f_X(x)$, as we have defined the Fourier transform in Chapter 2, provided a minus sign were used in the exponent instead of a plus sign. Thus, if we remember to replace $j\omega$ by $-jv$ in Fourier transform tables, we can use them to obtain characteristic functions from pdf's.

A pdf is obtained from the corresponding characteristic function by the inverse transform relationship

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} M_X(jv)e^{-jvx}\,dv \tag{4.135}$$

This illustrates one possible use of the characteristic function. It is sometimes easier to obtain the characteristic function than the pdf, and the latter is then obtained by inverse Fourier transformation.

Another use for the characteristic function is seen when we differentiate (4.134) with respect to v:

$$\frac{\partial M_X(jv)}{\partial v} = j \int_{-\infty}^{\infty} x f_X(x) e^{jvx}\, dx \tag{4.136}$$

Setting $v = 0$ after differentiation and dividing by j, we obtain

$$E\{X\} = (-j) \left.\frac{\partial M_X(jv)}{\partial v}\right|_{v=0} \tag{4.137}$$

For the nth moment, the relation

$$E\{X^n\} = (-j)^n \left.\frac{\partial^n M_X(jv)}{\partial v^n}\right|_{v=0} \tag{4.138}$$

can be proved by repeated differentiation.

▼ EXAMPLE 4.22 When we use a table of Fourier transforms, the one-sided exponential pdf

$$f_X(x) = \exp(-x)u(x) \tag{4.139}$$

has the corresponding characteristic function

$$M_X(jv) = \frac{1}{1 - jv} \tag{4.140}$$

When we use repeated differentiation or when we expand the characteristic function in a power series in jv, it follows from (4.138) that $E\{X^n\} = n!$ ▲

The pdf of the Sum of Two Independent Random Variables

The following problem is often encountered: given two *statistically independent* random variables X and Y with known pdf's $f_X(x)$ and $f_Y(y)$, respectively, what is the pdf of their sum $Z = X + Y$? To illustrate the use of the characteristic function, we will use it to find the pdf of Z, or $f_Z(z)$, even though we could find this pdf directly.

By definition of the characteristic function of Z, we can write

$$\begin{aligned} M_Z(jv) &= E\{e^{jvZ}\} = E\{e^{jv(X+Y)}\} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{jv(x+y)} f_X(x) f_Y(y)\, dx\, dy \end{aligned} \tag{4.141}$$

since the joint pdf of X and Y is $f_X(x)f_Y(y)$ by the assumption of statistical independence of X and Y. We can write (4.141) as the product of two integrals, since $e^{jv(x+y)} = e^{jvx}e^{jvy}$. This results in

$$\begin{aligned} M_Z(jv) &= \int_{-\infty}^{\infty} f_X(x)e^{jvx}\,dx \int_{-\infty}^{\infty} f_Y(y)e^{jvy}\,dy \\ &= E\{e^{jvX}\}E\{e^{jvY}\} \end{aligned} \tag{4.142}$$

But, from the definition of the characteristic function, Equation (4.134), we see that

$$M_Z(jv) = M_X(jv)M_Y(jv) \tag{4.143}$$

where $M_X(jv)$ and $M_Y(jv)$ are the characteristic functions of X and Y, respectively. Remembering that the characteristic function is the Fourier transform of the corresponding pdf and that a product in the frequency domain corresponds to convolution in the time domain (even with the minus sign in the exponent!), it follows that

$$f_Z(z) = f_X(x) * f_Y(y) = \int_{-\infty}^{\infty} f_X(z - u)f_Y(u)\,du \tag{4.144}$$

The following example illustrates the use of (4.144).

▼ **EXAMPLE 4.23** We consider the sum of four identically distributed, independent random variables:

$$Z = X_1 + X_2 + X_3 + X_4 \tag{4.145}$$

where the pdf of X_i is

$$f_{X_i}(x_i) = \Pi(x_i) = \begin{cases} 1, & |x_i| \leq \frac{1}{2} \\ 0, & \text{otherwise, } i = 1, 2, 3, 4 \end{cases} \tag{4.146}$$

where $\Pi(x_i)$ is the unit rectangular pulse function defined by Equation (2.2). We find $f_Z(z)$ by applying (4.144) twice. Thus we have

$$Z_1 = X_1 + X_2 \quad \text{and} \quad Z_2 = X_3 + X_4 \tag{4.147}$$

The pdf's of Z_1 and Z_2 are identical, both being the convolution of a uniform density with itself. From Table 2.2, we can immediately write down the result:

$$f_{Z_i}(z_i) = \Lambda(z_i) = \begin{cases} [1 - |z_i|], & |z_i| \leq 1 \\ 0, & \text{otherwise} \end{cases} \tag{4.148}$$

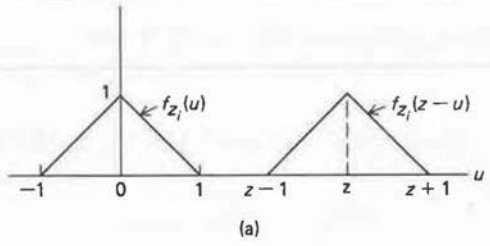
where $f_{Z_i}(z_i)$ is the pdf of Z_i, $i = 1, 2$. To find $f_Z(z)$, we simply convolve $f_{Z_i}(z_i)$ with itself. Thus

$$f_Z(z) = \int_{-\infty}^{\infty} f_{Z_i}(z - u) f_{Z_i}(u)\, du \tag{4.149}$$

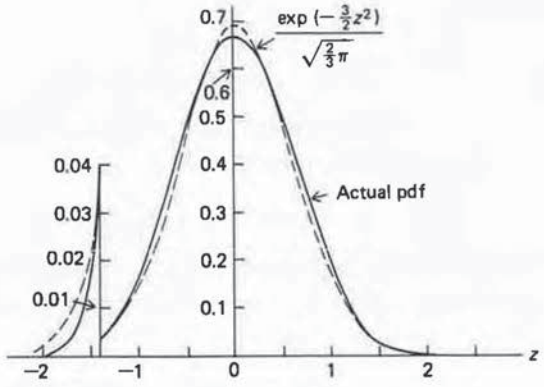
The factors in the integrand are sketched in Figure 4.16(a). Clearly, $f_Z(z) = 0$ for $z < -2$ or $z > 2$. Since $f_{Z_i}(z_i)$ is even, $f_Z(z)$ is also even. Thus we need not consider $f_Z(z)$ for $z < 0$. From Figure 4.16(a) it follows that for $1 \leq z \leq 2$,

$$f_Z(z) = \int_{z-1}^{1} (1 - u)(1 + u - z)\, du = \tfrac{1}{6}(2 - z)^3 \tag{4.150}$$

FIGURE 4.16 The pdf for the sum of four independent uniformly distributed random variables. (a) Convolution of two triangular pdf's. (b) Comparison of actual and Gaussian pdf's.



(a)



(b)

and for $0 \leq z \leq 1$, we obtain

$$f_Z(z) = \int_{z-1}^{0} (1 + u)(1 + u - z)\, du + \int_{0}^{z} (1 - u)(1 + u - z)\, du + \int_{z}^{1} (1 - u)(1 - u + z)\, du$$
$$= (1 - z) - \tfrac{1}{3}(1 - z)^3 + \tfrac{1}{6}z^3 \tag{4.151}$$

A graph of $f_Z(z)$ is shown in Figure 4.16(b) along with the graph of the function

$$\frac{\exp\left(-\frac{3}{2}z^2\right)}{\sqrt{\frac{2}{3}\pi}} \tag{4.152}$$

which represents a marginal Gaussian pdf of mean 0 and variance $\frac{1}{3}$, the same variance as $Z = X_1 + X_2 + X_3 + X_4$ [the results of Example (4.21) and Equation (4.131) can be used to obtain the variance of Z]. We will describe the Gaussian pdf more fully later.

The reason for the striking similarity of the two pdf's shown in Figure 4.16(b) will become apparent when we discuss the central-limit theorem. ▲

Covariance and the Correlation Coefficient

Two useful joint averages of a pair of random variables X and Y are their *covariance* μ_{XY}, defined as

$$\mu_{XY} = E\{(X - \bar{X})(Y - \bar{Y})\} \tag{4.153}$$

and their *correlation coefficient* ρ_{XY}, which is written in terms of the covariance as

$$\rho_{XY} = \frac{\mu_{XY}}{\sigma_X \sigma_Y} \tag{4.154}$$

Both are measures of the interdependence of X and Y. The correlation coefficient is more convenient because it is normalized such that $-1 \leq \rho_{XY} \leq 1$. If $\rho_{XY} = 0$, X and Y are said to be *uncorrelated*.

It is easily shown that $\rho_{XY} = 0$ for statistically independent random vari-

ables. If X and Y are independent, their joint pdf $f_{XY}(x, y)$ is the product of the respective marginal pdf's; that is, $f_{XY}(x, y) = f_X(x)f_Y(y)$. Thus

$$\begin{aligned}\mu_{XY} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - \bar{X})(y - \bar{Y})f_X(x)f_Y(y)\,dx\,dy \\ &= \int_{-\infty}^{\infty}(x - \bar{X})f_X(x)\,dx \int_{-\infty}^{\infty}(y - \bar{Y})f_Y(y)\,dy \\ &= (\bar{X} - \bar{X})(\bar{Y} - \bar{Y}) = 0\end{aligned} \tag{4.155}$$

Considering next the cases $X = \pm\alpha Y$, where α is a positive constant, we obtain

$$\begin{aligned}\mu_{XY} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\pm\alpha y \mp \alpha\bar{Y})(y - \bar{Y})f_{XY}(x, y)\,dx\,dy \\ &= \pm\alpha\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(y - \bar{Y})^2 f_{XY}(x, y)\,dx\,dy = \pm\alpha\sigma_y^2\end{aligned} \tag{4.156}$$

Using (4.131) with $N = 1$, we can write the variance of X as $\sigma_x^2 = \alpha^2\sigma_Y^2$. Thus the correlation coefficient is

$$\rho_{XY} = +1, \quad \text{for } X = +\alpha Y \qquad \text{and} \qquad \rho_{XY} = -1, \quad \text{for } X = -\alpha Y$$

To summarize, the correlation coefficient of two independent random variables is zero. When two random variables are linearly related, their correlation is $+1$ or -1 depending on whether one is a positive or a negative constant times the other.

4.4 SOME USEFUL pdf's*

We have already considered several often used probability distributions in the examples. These have included the Rayleigh pdf (Example 4.15), the pdf of a sinewave of random phase (Example 4.18), and the uniform pdf (Example 4.21). Some others, which will be useful in our future considerations, are given below.

Binomial Distribution

One of the most common discrete distributions in the application of probability to systems analysis is the binomial distribution. We consider a chance experiment with two mutually exclusive, exhaustive outcomes A and $\bar{A}$, with

* A number of useful probability distributions are given in a table in the summary at the end of this chapter.

probabilities $P(A) = p$ and $P(\bar{A}) = q = 1 - p$, respectively. Assigning the discrete random variable K to be numerically equal to the number of times event A occurs in N trials of our chance experiment, we seek the probability that exactly $k \leq n$ occurrences of the event A occur in n repetitions of the experiment. (Thus our actual chance experiment is the replication of the basic experiment n times.) The resulting distribution is called the *binomial distribution*.

Specific examples in which the binomial distribution is the result are the following: In n tosses of a coin, what is the probability of $k \leq n$ heads? In the transmission of n messages through a channel, what is the probability of $k \leq n$ errors?

Note that in all cases we are interested in *exactly* k occurrences of the event, not, for example, at least k of them, although we may find the latter probability if we have the former.

Although the problem we are considering is very general, let us solve it by visualizing the coin-tossing experiment. We wish to obtain the probability of k heads in n tosses of the coin if the probability of a head on a single toss is p and the probability of a tail is $1 - p = q$. One of the possible sequences of k heads in n tosses is

$$\underbrace{\mathrm{H}\,\mathrm{H}\cdots\mathrm{H}}_{k \text{ heads}}\ \underbrace{\mathrm{T}\,\mathrm{T}\cdots\mathrm{T}}_{n-k \text{ tails}}$$

Since the trials are independent, the probability of this particular sequence is

$$\underbrace{p \cdot p \cdots p}_{k \text{ factors}} \cdot \underbrace{q \cdot q \cdots q}_{n-k \text{ factors}} = p^k q^{n-k} \tag{4.157}$$

But the preceding sequence of k heads in n trials is only one of

$$\binom{n}{k} \triangleq \frac{n!}{k!(n-k)!} \tag{4.158}$$

possible sequences, where $\binom{n}{k}$ is the binomial coefficient. To see this, we consider the number of ways k *identifiable* heads can be arranged in n slots. The first head can fall in any of the n slots, the second in any of $n - 1$ slots (the first head already occupies one slot), the third in any of $n - 2$ slots, and so on for a total of

$$n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!} \tag{4.159}$$

possible arrangements in which we identify each head. However, we are not concerned about which head occupies which slot. For each possible identifiable arrangement, there are $k!$ arrangements for which we can switch the heads around and still have the same slots occupied. Thus the total number of arrangments, if we do not identify the particular coin occupying each slot, is

$$\frac{n(n-1)\cdots(n-k+1)}{k!} = \binom{n}{k} \tag{4.160}$$

Since the occurrence of any of these $\binom{n}{k}$ possible arrangements precludes the occurrence of any other [that is, the $\binom{n}{k}$ outcomes of our experiment are mutually exclusive], and since each occurs with probability $p^k q^{n-k}$, the probability of *exactly* k heads in n trials in *any* order is

$$P(K = k) \triangleq P_n(k) = \binom{n}{k} p^k q^{n-k}, \qquad k = 0, 1, \ldots, n \tag{4.161}$$

Equation (4.161), known as the *binomial probability distribution* (note that it is *not* a pdf or a cdf), is plotted in Figure 4.17(a)–(d) for four different values of p and n.

The mean of a binomially distributed random variable K, by Equation (4.93), is given by

$$E\{K\} = \sum_{k=0}^{n} k \frac{n!}{k!(n-k)!} p^k q^{n-k} \tag{4.162}$$

Noting that the sum can be started at $k = 1$ since the first term is zero, we can write

$$E\{K\} = \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k q^{n-k} \tag{4.163}$$

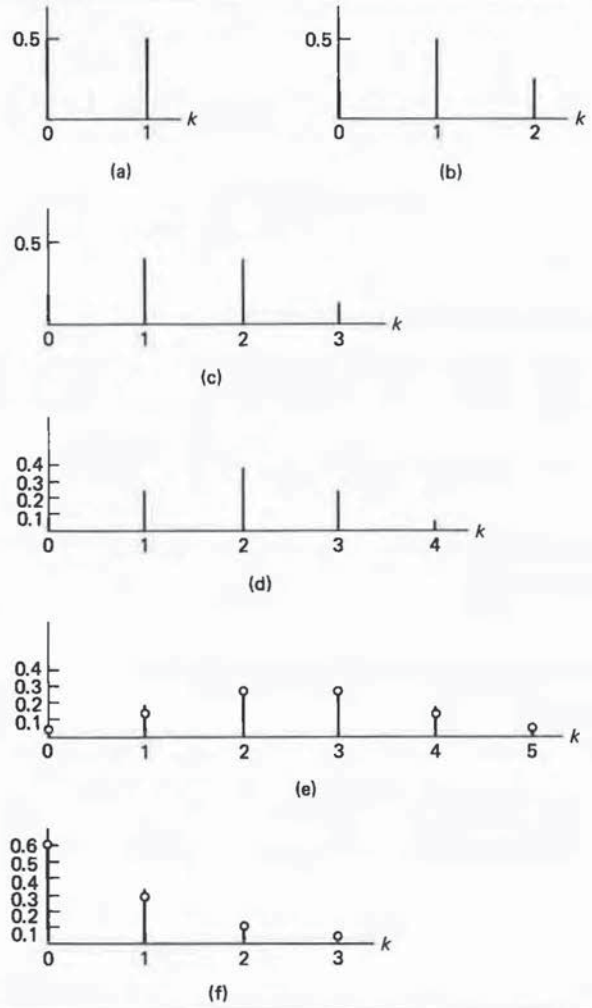
Letting $m = k - 1$ yields the sum

$$\begin{aligned} E\{K\} &= \sum_{m=0}^{n-1} \frac{n!}{m!(n-m-1)!} p^{m+1} q^{n-m-1} \\ &= np \sum_{m=0}^{n-1} \frac{(n-1)!}{m!(n-m-1)!} p^m q^{n-m-1} \end{aligned} \tag{4.164}$$

Finally, letting $\ell = n - 1$ and recalling that, by the binomial theorem,

$$(x+y)^\ell = \sum_{m=0}^{\ell} \binom{\ell}{m} x^m y^{\ell-m} \tag{4.165}$$

FIGURE 4.17 The binomial distribution with comparison to Laplace and Poisson approximations. (a) $n = 1$, $p = 0.5$. (b) $n = 2$, $p = 0.5$. (c) $n = 3$, $p = 0.5$. (d) $n = 4$, $p = 0.5$. (e) $n = 5$, $p = 0.5$. Circles are Laplace approximations. (f) $n = 5$, $p = \frac{1}{10}$. Circles are Poisson approximations.



we obtain

$$\overline{K} = E\{K\} = np(p + q)^{\ell} = np \tag{4.166}$$

since $p + q = 1$. The result is reasonable; in a long sequence of n tosses of a coin, we would expect about $np = \frac{1}{2}n$ heads.

We can go through a similar series of manipulations to show that $E\{K^2\} = np(np + q)$. Using this result, it follows that the variance of a binomially distributed random variable is

$$\sigma_K^2 = E\{K^2\} - E^2\{K\} = npq = \bar{K}(1 - p) \tag{4.167}$$

EXAMPLE 4.24 The probability of having two boys in a four-child family, assuming single births and the probability of a male $\cong 0.5$, from (4.161), is

$$P_4(2) = \binom{4}{2}\left(\frac{1}{2}\right)^4 = \frac{3}{8} \tag{4.168}$$

Laplace Approximation to the Binomial Distribution

When n becomes large, computations using (4.161) become unmanageable. In the limit as $n \to \infty$, it can be shown that for $|k - np| \leq \sqrt{npq}$

$$P_n(k) \cong \frac{1}{\sqrt{2\pi npq}} \exp\left[-\frac{(k - np)^2}{2npq}\right] \tag{4.169}$$

A comparison of the Laplace approximation with the actual distribution is given in Figure 4.17(e)

Poisson Distribution and Poisson Approximation to the Binomial Distribution

If we consider a chance experiment in which an event whose probability of occurrence in a very small time interval ΔT is $P = \alpha\Delta T$, where α is a constant of proportionality, and if successive occurrences are statistically independent, then the probability of k events in time T is

$$P_T(k) = \frac{(\alpha T)^k}{k!} e^{-\alpha T} \tag{4.170}$$

For example, the emission of electrons from a hot metal surface obeys this law, which is called the *Poisson distribution.*

The Poisson distribution can be used to approximate the binomial distribution when the number of trials n is large, the probability of each event p is small, and the product $np \cong npq$. The approximation is

$$P_n(k) \cong \frac{(\bar{K})^k}{k!} e^{-\bar{K}} \tag{4.171}$$

where, as calculated previously, $\overline{K} = E\{K\} = np$ and $\sigma_k^2 = E\{K\}q \cong E\{K\}$. This approximation is compared with the binomial distribution in Figure 4.17(f).

▼ **EXAMPLE 4.25** The probability of error on a single transmission in a digital communication system is $P_e = 10^{-4}$. What is the probability of more than three errors in 1000 transmissions?

SOLUTION We find the probability of three errors or less from (4.171):

$$P(K \leq 3) = \sum_{k=0}^{3} \frac{(\overline{K})^k}{k!} e^{-\overline{K}} \tag{4.172}$$

where $\overline{K} = (10^{-4})(1000) = 0.1$. Hence

$$P(K \leq 3) = e^{-0.1}\left[\frac{(0.1)^0}{0!} + \frac{(0.1)^1}{1!} + \frac{(0.1)^2}{2!} + \frac{(0.1)^3}{3!}\right] \cong 0.999996 \tag{4.173}$$

▲ Therefore, $P(K > 3) = 1 - P(K \leq 3) \cong 4 \times 10^{-6}$.

Geometric Distribution

Suppose we are interested in the probability of the first head in a series of coin tossings, or the first error in a long string of digital signal transmissions occurring on the kth trial. The distribution describing such experiments is called the *geometric distribution* and is

$$P(k) = pq^{k-1}, \qquad 1 \leq k < \infty \tag{4.174}$$

where p is the probability of the event of interest occurring (i.e., head, error, etc.) and q is the probability of it not occurring.

▼ **EXAMPLE 4.26** The probability of the first error occurring at the 1000th transmission in a digital data transmission system where the probability of error is $p = 10^{-6}$ is

▲ $$P(1000) = 10^{-6}(1 - 10^{-6})^{999} = 9.99 \times 10^{-7} \cong 10^{-6}$$

▼ **EXAMPLE 4.27** Find the probability of the first head in a sequence of 10 tossings of a fair coin occurring on the 1st, 2nd, . . . , 10th tossings.

SOLUTION The desired probabilities are

$$P(k) = 0.5(1 - 0.5)^{k-1} = 0.5^k, \qquad 1 \leq k \leq 10$$

The first five are listed below. The student should compute the rest of the values.

k	1	2	3	4	5
P_k	0.5	0.25	0.125	0.063	0.031

▲

Gaussian Distribution

In our future considerations, the Gaussian pdf will be used repeatedly. There are at least two reasons for this. One is that the assumption of Gaussian statistics for random phenomena often makes an intractable problem tractable. The other, and more fundamental reason, is that because of a remarkable phenomenon summarized by a theorem called the *central-limit theorem,* many naturally occurring random quantities, such as noise or measurement errors, are Gaussianly distributed. The following is a statement of the central-limit theorem.

▷**CENTRAL-LIMIT THEOREM**

Let $X_1, X_2, \ldots$ be independent, identically distributed random variables, each with finite mean m and finite variance σ^2. Let Z_n be a sequence of unit-variance, zero-mean random variables, defined as

$$Z_n \triangleq \frac{\sum_{i=1}^{n} X_i - nm}{\sigma\sqrt{n}} \tag{4.175}$$

Then

$$\lim_{n\to\infty} P(Z_n \leq z) = \int_{-\infty}^{z} \frac{e^{-t^2/2}}{\sqrt{2\pi}}\, dt \tag{4.176}$$

In other words, the cdf of the normalized sum (4.175) approaches a Gaussian cdf, no matter what the distribution of the component random variables. The only restriction is that they be independent and identically distributed and that their means and variances be finite. In some cases the independence and identically distributed assumptions can be relaxed. It is important, however, that no one of the component random variables or a finite combination of them dominate the sum.

We will not prove the central-limit theorem or use it in later work. We state it here simply to give partial justification for our almost exclusive assumption of Gaussian statistics from now on. For example, electrical noise is often the result of a superposition of voltages due to a large number of charge carriers. Turbulent boundary-layer pressure fluctuations on an aircraft skin are the superposition of minute pressures due to numerous eddies. Random errors in experimental measurements are due to many irregular fluctuating causes. In all these cases, the Gaussian approximation for the fluctuating quantity is useful and valid. Example 4.23 illustrates that surprisingly few terms in the sum are required to give a Gaussian-appearing pdf, even where the component pdf's are far from Gaussian.

The generalization of the joint Gaussian pdf first introduced in Example 4.15 is

$$f_{XY}(x, y) = \frac{\exp\left\{-\dfrac{\left(\dfrac{x - m_x}{\sigma_x}\right)^2 - 2\rho\left(\dfrac{x - m_x}{\sigma_x}\right)\left(\dfrac{y - m_y}{\sigma_y}\right) + \left(\dfrac{y - m_y}{\sigma_y}\right)^2}{2(1 - \rho^2)}\right\}}{2\pi\sigma_x\sigma_y\sqrt{1 - \rho^2}} \tag{4.177}$$

where, through straightforward but tedious integrations, it can be shown that

$$m_x = E\{X\} \tag{4.178a}$$

$$m_y = E\{Y\} \tag{4.178b}$$

$$\sigma_x^2 = \text{var}\{X\} \tag{4.178c}$$

$$\sigma_y^2 = \text{var}\{Y\} \tag{4.178d}$$

and

$$\rho = \frac{E\{(X - m_x)(Y - m_y)\}}{\sigma_x\sigma_y} \tag{4.178e}$$

The joint pdf for $N > 2$ Gaussian random variables may be written in a compact fashion through the use of matrix notation. The general form is given in Appendix B.

The marginal pdf for X (or Y) can be obtained by integrating (4.177) over

y (or x). Again, the integration is tedious and will be left to the problems. The marginal pdf for X is

$$n(m_x, \sigma_x) = \frac{\exp\left[-(x - m_x)^2/2\sigma_x^2\right]}{\sqrt{2\pi\sigma_x^2}} \tag{4.179}$$

where the notation $n(m_x, \sigma_x)$ has been introduced to denote a Gaussian pdf of mean m_x and standard deviation σ_x. This function is shown in Figure 4.18.

We will sometimes assume in the discussions to follow that $m_x = m_y = 0$ in (4.177) and (4.179), for if they are not zero, we can consider new random variables X' and Y' defined as $X' = X - m_x$ and $Y' = Y - m_y$ which do have zero means. Thus no generality is lost in assuming zero means.

For $\rho = 0$, that is, X and Y uncorrelated, the cross term in the exponent of (4.177) is zero, and $f_{XY}(x, y)$, with $m_x = m_y = 0$, can be written as

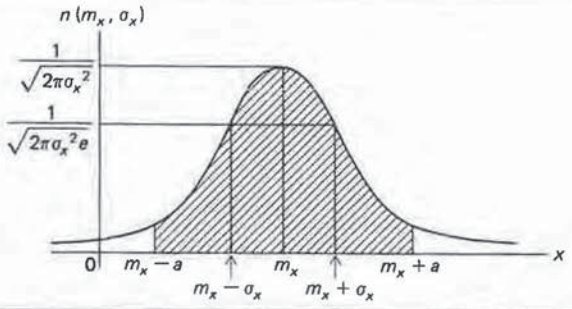
$$f_{XY}(x, y) = \frac{\exp\left(-x^2/2\sigma_x^2\right)}{\sqrt{2\pi\sigma_x^2}} \frac{\exp\left(-y^2/2\sigma_y^2\right)}{\sqrt{2\pi\sigma_y^2}} = f_X(x) f_Y(y) \tag{4.180}$$

Thus *uncorrelated Gaussian random variables are also statistically independent.* We emphasize that this does not hold for all pdf's, however.

It can be shown that the sum of any number of Gaussian random variables, independent or not, is Gaussian. The sum of two independent Gaussian random variables is easily shown to be Gaussian. Let $Z = X_1 + X_2$, where the pdf of X_i is $n(m_i, \sigma_i)$. Using a table of Fourier transforms or completing the square and integrating, we find that the characteristic function of X_i is

$$\begin{aligned} M_{X_i}(jv) &= \int_{-\infty}^{\infty} (2\pi\sigma_i^2)^{-1/2} \exp\left[\frac{-(x_i - m_i)^2}{2\sigma_i^2}\right] \exp(jvx_i)\, dx_i \\ &= \exp\left(jm_iv - \frac{\sigma_i^2 v^2}{2}\right) \end{aligned} \tag{4.181}$$

FIGURE 4.18 The Gaussian pdf with mean m_x and variance σ_x^2



Thus the characteristic function of Z is

$$M_Z(jv) = M_{X_1}(jv)M_{X_2}(jv) = \exp\left[j(m_1 + m_2)v - \frac{(\sigma_1^2 + \sigma_2^2)v^2}{2}\right] \quad (4.182)$$

which is the characteristic function of a Gaussian random variable of mean $m_1 + m_2$ and variance $\sigma_1^2 + \sigma_2^2$.

Gaussian Q-Function

As Figure 4.18 shows, $n(m_x, \sigma_x)$ describes a continuous random variable that may take on any value in $(-\infty, \infty)$ but is most likely to be found near $X = m_x$. The even symmetry of $n(m_x, \sigma_x)$ about $x = m_x$ leads to the conclusion that $P(X \leq m_x) = P(X \geq m_x) = \frac{1}{2}$.

Suppose we wish to find the probability that X lies in the interval $[m_x - a, m_x + a]$. Using (4.36), we can write this probability as

$$P[m_x - a \leq X \leq m_x + a] = \int_{m_x - a}^{m_x + a} \frac{\exp\left[-(x - m_x)^2/2\sigma_x^2\right]}{\sqrt{2\pi\sigma_x^2}}\, dx \quad (4.183)$$

which is the shaded area in Figure 4.18. With the change of variables $y = (x - m_x)/\sigma_x$, this gives

$$\begin{aligned} P[m_x - a \leq X \leq m_x + a] &= \int_{-a/\sigma_x}^{a/\sigma_x} \frac{e^{-y^2/2}}{\sqrt{2\pi}}\, dy \\ &= 2\int_0^{a/\sigma_x} \frac{e^{-y^2/2}}{\sqrt{2\pi}}\, dy \end{aligned} \quad (4.184)$$

where the last integral follows by virtue of the integrand being even. Unfortunately, this integral cannot be evaluated in closed form.

The Gaussian Q-function, or simply Q-function, is defined as

$$Q(u) = \int_u^\infty \frac{e^{-y^2/2}}{\sqrt{2\pi}}\, dy \quad (4.185)$$

It has been evaluated numerically, and rational and asympotic approximations are available to evaluate it for moderate and large arguments, respec-

tively.* Using this transcendental function definition, we may rewrite (4.184) as

$$P[m_x - a \leq X \leq m_x + a] = 2\left[\frac{1}{2} - \int_{a/\sigma_x}^{\infty} \frac{e^{-y^2/2}}{\sqrt{2\pi}}\, dy\right]$$

$$= 1 - 2Q\left(\frac{a}{\sigma_x}\right) \tag{4.186}$$

A useful approximation for the Q-function for large arguments is

$$Q(u) \cong \frac{e^{-u^2/2}}{u\sqrt{2\pi}}, \quad u \gg 1 \tag{4.187}$$

Numerical comparison of (4.185) and (4.187) shows that less than a 6% error results for $u \geq 3$ in using this approximation. This, and other results for the Q-function, are given in Appendix C.

A related integral is the error function, defined as

$$erf(u) = \frac{2}{\sqrt{\pi}} \int_0^u e^{-y^2}\, dy \tag{4.188}$$

It can be shown to be related to the Q-function by

$$erf(u) = 1 - 2Q(\sqrt{2u}) \tag{4.189}$$

Chebyshev's Inequality

The difficulties encountered above in evaluating (4.183) make an approximation to such probabilities desirable. *Chebyshev's inequality* gives us a lower bound, *regardless of the specific form of the pdf involved,* provided its second moment is finite. The probability of finding a random variable X within $\pm k$ standard deviations of its mean is at least $1 - 1/k^2$, according to Chebyshev's inequality. That is,

$$P[|X - m_x| \leq k\sigma_x] \geq 1 - \frac{1}{k^2}, \qquad k > 0 \tag{4.190}$$

Considering $k = 3$, we obtain

$$P[|X - m_x| \leq 3\sigma_x] \geq \tfrac{8}{9} \cong 0.889 \tag{4.191}$$

* These are provided in M. Abramowitz and I. Stegun (eds.), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* National Bureau of Standards, Applied Mathematics Series No. 55, Issued June 1964 (pp. 931 ff). Also New York: Dover, 1972.

Assuming X is Gaussian and using the Q-function, this probability can be computed exactly to be 0.9973. In words, the probability that a random variable deviates from its mean by more than ± 3 standard deviations is not greater than 0.111, regardless of its pdf. (There is the restriction that its second moment must be finite.) Note that the bound for this example is not very tight.

SUMMARY

1. The objective of probability theory is to attach real numbers between zero and one called *probabilities* to the *outcomes* of chance experiments—that is, experiments in which the outcomes are not uniquely determined by the causes but depend on chance—and to interrelate probabilities of *events,* which are defined to be combinations of outcomes.

2. Two events are *mutually exclusive* if the occurrence of one of them precludes the occurrence of the other. A set of events is said to be *exhaustive* if one of them must occur in the performance of a chance experiment. The *null event* cannot happen, and the *certain event* must happen in the performance of a chance experiment.

3. The *equally likely definition* of the probability $P(A)$ of an event A states that if a chance experiment can result in a number N of mutually exclusive, equally likely outcomes, then $P(A)$ is the ratio of the number of outcomes favorable to A, or N_A, to the total number. It is a circular definition in that probability is used to define probability, but it is nevertheless useful in many situations such as drawing cards from well-shuffled decks.

   The *relative-frequency definition* of the probability of an event A assumes that the chance experiment is replicated a large number of times N and

$$P(A) = \lim_{N \to \infty} \frac{N_A}{N}$$

   where N_A is the number of replications resulting in the occurrence of A. The *axiomatic approach* defines the probability $P(A)$ of an event A as a real number satisfying the following axioms:

   (a) $P(A) \geq 0$.
   (b) $P(\text{certain event}) = 1$.
   (c) If A and B are mutually exclusive events, $P(A \text{ or } B \text{ or both}) = P(A) + P(B)$.

   The axiomatic approach encompasses the first two definitions given.

4. Given two events A and B, the compound event "A or B or both" is denoted as $A + B$, the compound event "both A and B" is denoted as (AB) or as (A, B), and the event "not A" is denoted as $\bar{A}$. If A and B are not necessarily mutually exclusive, the axioms of probability may be used to show that $P(A + B) = P(A) + P(B) - P(AB)$. Letting $P(A \mid B)$ denote the probability of A occurring given that B occurred and $P(B \mid A)$ denote the probability of B given A, these probabilities are defined, respectively, as

$$P(A \mid B) = \frac{P(AB)}{P(B)} \quad \text{and} \quad P(B \mid A) = \frac{P(AB)}{P(A)}$$

A special case of *Bayes' rule* results by putting these two definitions together:

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}$$

Statistically independent events are events for which $P(AB) = P(A)P(B)$.

5. A random variable is a rule that assigns a real number to every outcome of a chance experiment. For example, in flipping a coin, assigning $X = +1$ to the occurrence of a head and $X = -1$ to the occurrence of a tail constitutes the assignment of a discrete-valued random variable.

6. The cumulative distribution function (cdf) $F_X(x)$ of a random variable X is defined as the probability that $X \leq x$ where x is a running variable. $F_X(x)$ lies between zero and 1 with $F_X(-\infty) = 0$ and $F_X(\infty) = 1$, is continuous from the right, and is a nondecreasing function of its argument. Discrete random variables have step-discontinuous cdf's, and continuous random variables have continuous cdf's.

7. The probability density function (pdf) $f_X(x)$ of a random variable X is defined to be the derivative of the cdf. Thus

$$F_X(x) = \int_{-\infty}^{x} f_X(x')\, dx'$$

The pdf is nonnegative and integrates over all x to unity. A useful interpretation of the pdf is that $f_X(x)\, dx$ is the probability of the random variable X lying in an infinitesimal range dx about x.

8. The joint cdf $F_{XY}(x, y)$ of two random variables X and Y is defined as the probability that $X \leq x$ and $Y \leq y$ where x and y are particular values of X and Y. Their joint pdf $f_{XY}(x, y)$ is the second partial derivative of the cdf first with respect to x and then with respect to y. The cdf of X (Y)

alone (that is, the marginal cdf) is found by setting y (x) to infinity in the argument of F_{XY}. The pdf of X (Y) alone (that is, the marginal pdf) is found by integrating f_{XY} over all y (x).

9. Two statistically independent random variables have joint cdf's and pdf's that factor into the respective marginal cdf's or pdf's.

10. The conditional pdf of X given Y is defined as

$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

with a similar definition for $f_{Y|X}(y|x)$. The expression $f_{X|Y}(x|y)\,dx$ can be interpreted as the probability that $x - dx < X \leq x$ given $Y = y$.

11. Given $Y = g(X)$ where $g(X)$ is a monotonic function,

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|_{x=g^{-1}(y)}$$

where $g^{-1}(y)$ is the inverse of $y = g(x)$. Joint pdf's of functions of more than one random variable can be transformed similarly.

12. Important probability functions defined in Chapter 4 are the Rayleigh pdf [Equation (4.92)], the pdf of a random-phased sinusoid (Example 4.18), the uniform pdf (Example 4.21), the binomial probability distribution [Equation (4.161)], the Laplace and Poisson approximations to the binomial distribution [Equations (4.169) and (4.171)], and the Gaussian pdf [Equations (4.177) and (4.179)].

13. The statistical average, or expectation, of a function $g(X)$ of a random variable X with pdf $f_X(x)$ is defined as

$$E\{g(X)\} = \overline{g(X)} = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx$$

The average of $g(X) = X^n$ is called the nth moment of X. The first moment is known as the *mean* of X. Averages of functions of more than one random variable are found by integrating the function times the joint pdf over all values of its arguments. The averages $\overline{g(X, Y)} = \overline{X^m Y^n}$ are called the *joint moments* of the order $m + n$. The variance of a random variable X is the average $\overline{(X - \bar{X})^2} = \overline{X^2} - \bar{X}^2$.

14. The average of $\overline{\Sigma a_i X_i}$ is $\Sigma a_i \bar{X}_i$; that is, summations and averaging can be interchanged. The variance of a sum of random variables is the sum of the respective variances *if the random variables are statistically independent.*

15. The characteristic function $M_X(jv)$ of a random variable X that has the pdf $f_X(x)$ is the expectation of exp (jvX) or, equivalently, the Fourier transform of $f_X(x)$ with a plus sign in the exponential of the Fourier transform integral. Thus the pdf is the inverse Fourier transform of the characteristic function.

    The nth moment of X can be found from $M_X(jv)$ by differentiating with respect to v for n times, multiplying by $(-j)^n$, and setting $v = 0$. The characteristic function of $Z = X + Y$, where X and Y are independent, is the product of the characteristic functions of X and Y. Thus, by the convolution theorem of Fourier transforms, the pdf of Z is the convolution of the pdf's of X and Y.

16. The covariance μ_{XY} of two random variables X and Y is the average

$$\mu_{XY} = E[(X - \bar{X})(Y - \bar{Y})]$$

    The correlation coefficient ρ_{XY} is

$$\frac{\mu_{XY}}{\sigma_X \sigma_Y}$$

    Both give a measure of the linear interdependence of X and Y, but ρ_{XY} is handier because it is bounded by ± 1.

17. The central-limit theorem states that, under suitable conditions, the sum of a large number N of random variables (not necessarily with the same pdf's) tends to a Gaussian pdf as N becomes large.

18. The Q-function can be used to compute probabilities of Gaussian random variables being in certain ranges. The Q-function is tabulated in Appendix C, and rational and asymptotic approximations are given for computing it.

19. Chebyshev's inequality gives the lower bound of the probability that a random variable is within k standard deviations of its mean as $1 - k^2$, *regardless of the pdf of the random variable* (its second moment must be finite).

20. Table 4.5 summarizes a number of useful probability distributions with their means and variances.

FURTHER READING

Several books are available that deal with probability theory from an engineering viewpoint. Among these are Beckman (1967), Breipohl (1970), Helstrom (1990), Leon-Garcia (1994), Peebles (1987), and Williams (1991).

TABLE 4.5 Probability Distributions of Some Random Variables and Their Means an Variances

Probability Density or Distribution Function	Mean	Variance
Uniform: $f_x(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & x < a \text{ or } x > b \end{cases}$	$\frac{1}{2}(b-a)$	$\frac{1}{12}(b-a)^2$
Gaussian: $f_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{(x-m)^2/2\sigma^2}$	m	σ^2
Rayleigh: $f_R(r) = \frac{re^{-r^2/2\sigma^2}}{\sigma^2}, \quad 0 \le r < \infty$	$\sqrt{\frac{\pi}{2}}\sigma$	$\frac{1}{2}(4-\pi)\sigma^2$
Laplacian: $f_x(x) = \frac{\alpha}{2} e^{-\alpha\lvert x\rvert}, \quad \alpha > 0$	0	$\frac{2}{\alpha^2}$
One-sided exponential: $f_x(x) = \alpha e^{-\alpha x} u(x)$	$\frac{1}{\alpha}$	$\frac{1}{\alpha^2}$
Hyperbolic: $f_x(x) = \frac{(m-1)h^{m-1}}{2(\lvert x\rvert + h)^m}, \; m > 3, h > 0$	0	$\frac{2h^2}{(m-3)(m-}$
Nakagami-m: $f_x(x) = \frac{2m^m}{\Gamma(m)} x^{2m-1} e^{-mx^2}, \quad x \ge 0$	$\frac{1\times3\times\ldots\times(2m-2)}{2^m\Gamma(m)}$	$\frac{\Gamma(m+1)}{\Gamma(m)\sqrt{m}}$
Lognormal:* $f_x(x) = \frac{\exp\left[-\frac{(\ln x - m_y)^2}{2\sigma_y^2}\right]}{x\sqrt{2\pi\sigma_y^2}}$	$e^{m_y + 2\sigma_y^2}$	$e^{2m_y+\sigma_y^2}(e^{\sigma_y^2} -$
Binomial: $P_n(k) = \binom{n}{k} p^k q^{n-k}, \; k = 0, 1, \ldots, n$ $p + q = 1$	np	npq
Poisson: $P(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad 0 \le k < \infty$	λ	λ
Geometric: $P(k) = pq^{k-1}, \quad 1 \le k < \infty$	$\frac{1}{p}$	$\frac{q}{p^2}$

* The lognormal random variable results from the transformation $y = \ln x$ where Y is a Gaussian r dom variable with mean m_y and variance σ_y^2.

More mathematical in nature are the following: Barr and Zehna (1983), Papoulis (1990), Papoulis (1984), and Roberts (1992). A good overview with many examples is Ash (1992).

PROBLEMS

Section 4.1

4.1 A circle is divided into 25 equal parts. A pointer is spun until it stops on one of the parts, which are numbered from 1 through 25. Describe the sample space and, assuming equally likely outcomes, find (a) P(an even number), (b) P(the number 25), (c) P(the numbers 4, 5, or 9), and (d) P(a number greater than 10).

4.2 (a) Prove (4.4) using the axioms of probability. (b) Repeat the derivation for three events A, B, and C, and show that

$$P(A \text{ or } B \text{ or } C) = P(A) + P(B) + P(C) - P(A, B) - P(B, C) - P(A, C) + P(A, B, C)$$

4.3 If five cards are drawn without replacement from an ordinary deck of cards, what is the probability that (a) three kings and two aces result, (b) four of a kind result, (c) all are of the same suit, (d) an ace, king, queen, jack, and ten of the same suit result, and (e) given that an ace, king, jack, and ten have been drawn, what is the probability that the next card drawn will be a queen (not all of the same suit)?

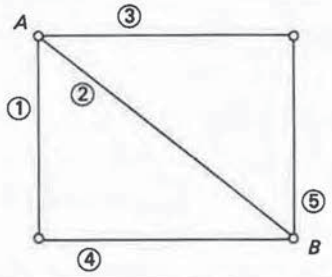
4.4 What equations must be satisifed in order for three events A, B, and C to be independent?

4.5 (a) Show that if two events are mutually exclusive and statistically independent, the probability of one, or both, is zero.
(b) Given three events A, B, and C, with A and B mutually exclusive and B and C independent, what can be said about the relationship between A and C?

4.6 Figure 4.19 is a graph that represents a communication network, where the nodes are receiver/repeater boxes and the edges (or links) represent communication channels which, if connected, convey the message perfectly. However, there is the probability p that a link will be broken and the probability $q = 1 - p$ that it will be whole. (*Hint:* Use a tree diagram like Fig. 4.2).

(a) What is the probability that at least one working path is available between the nodes labeled A and B?

FIGURE 4.19



(b) Remove link 4. Now what is the probability that at least one working path is available between nodes A and B?
(c) Remove link 2. What is the probability that at least one working path is available between nodes A and B?
(d) Which is the more serious situation, the removal of link 4 or link 2? Why?

4.7 Given a binary communication channel where A = input and B = output, let $P(A) = 0.4$, $P(B \mid A) = 0.9$, and $P(\bar{B} \mid \bar{A}) = 0.6$. Find $P(A \mid B)$ and $P(A \mid \bar{B})$.

4.8 (a) Find the probabilities omitted from Figure 4.20.
(b) Find the probabilities $P(A_3 \mid B_3)$, $P(B_2 \mid A_1)$, and $P(B_3 \mid A_2)$.

Section 4.2

4.9 Two dice are tossed.
(a) Let X_1 be a random variable that is numerically equal to the square of the total number of spots on the up faces of the dice. Construct a table that defines this random variable.

FIGURE 4.20 Table of joint and marginal probabilities $P(A_iB_j)$

	B_1	B_2	B_3	$P(A_i)$
A_1	0.05		0.45	0.55
A_2		0.15	0.10	
A_3	0.05	0.05		0.15
$P(B_j)$				1.0

(b) Let X_2 be a random variable that has the value of 1 if the sum of the number of spots up on both dice is less than 7 and the value zero if it is greater than or equal to 7. Repeat part (a) for this case.

4.10 Three fair coins are tossed simultaneously. Define a random variable as the total number of heads up on the coins. Plot the cumulative-distribution function and the probability-density function corresponding to this random variable.

4.11 A certain continuous random variable has the cumulative distribution function

$$F_X(x) = \begin{cases} 0, & x < 0 \\ Ax^3, & 0 \leq x \leq 10 \\ B, & x > 10 \end{cases}$$

(a) Find the proper values for A and B.
(b) Plot $F_X(x)$.
(c) Obtain and plot the pdf $f_X(x)$.
(d) Compute $P(X > 7)$.
(e) Compute $P(3 \leq X < 7)$.

4.12 The following functions can be pdf's if the constants are chosen properly. Find the proper conditions on the constants so that they are. (A, B, C, D, α, β, γ, and τ are positive constants.)

(a) $f(x) = Ae^{-\alpha x}u(x)$, where $u(x)$ is the unit step
(b) $f(x) = Be^{\beta x}u(-x)$
(c) $f(x) = Ce^{-\gamma|x|}$
(d) $f(x) = \dfrac{D}{\tau^2 + x^2}$

4.13 Test X and Y for independence if

(a) $f_{XY}(x, y) = Ae^{-|x|-2|y|}$

(b) $f_{XY}(x, y) = C(1 - x - y)$, $0 \leq x \leq 1 - y$ and $0 \leq y \leq 1$
Prove your answers.

4.14 The joint pdf of two random variables is

$$f_{XY}(x, y) = \begin{cases} C(1 + xy), & 0 \leq x \leq 6, 0 \leq y \leq 5 \\ 0, & \text{otherwise} \end{cases}$$

Find the following:

(a) The constant C.
(b) $F_{XY}(0.1, 1.5)$.
(c) $f_{XY}(x, 3)$.
(d) $f_{X|Y}(x \mid 3)$.

4.15 Repeat the previous problem for

$$f_{XY}(x, y) = \begin{cases} C(1 + xy), & 0 \leq x \leq 6,\ 0 \leq y \leq x \\ 0, & \text{otherwise} \end{cases}$$

4.16 The joint pdf of the random variables X and Y is

$$f_{XY}(x, y) = Axye^{-(x+y)}, \qquad x \geq 0 \quad \text{and} \quad y \geq 0$$

(a) Find the constant A.
(b) Find the marginal pdf's of X and Y, $f_X(x)$ and $f_Y(y)$.
(c) Are X and Y statistically independent? Justify your answer.

4.17 (a) For what value of α is the function

$$f(x) = \alpha x^{-2} u(x - \alpha)$$

a probability density function? [$u(x)$ is the unit step function.]
(b) Find the corresponding cumulative distribution function.
(c) Compute $P(X \geq 10)$.

4.18 Given the Gaussian random variable with the pdf

$$f_X(x) = \frac{e^{-x^2/2\sigma^2}}{\sqrt{2\pi}\,\sigma}$$

let $Y = X^2$. Find the pdf of Y. (*Hint:* Note that $Y = X^2$ is symmetrical about $X = 0$ and that it is impossible for Y to be less than zero.)

4.19 If X is Gaussian with the pdf given in Problem 4.18 and the random variable Y is given in terms of X as

$$Y = \begin{cases} aX, & X \geq 0 \\ 0, & X < 0 \end{cases}$$

find the pdf of Y. (*Hint:* When $X < 0$, what is Y? How is this manifested in the pdf for Y?)

4.20 A nonlinear system has input X and output Y. The pdf for the input is Gaussian as given in Problem 4.18. Determine the pdf of the output, assuming that the nonlinear system has the following input/output relationship: (a) $Y = |X|$; (b) $Y = X - X^3/3$.

Section 4.3

4.21 Let $f_X(x) = A\exp(-b|x|)$ for all x.

(a) Find the relationship between A and b such that this function is a pdf.
(b) Calculate $E(X)$ for this random variable.
(c) Find $E(X^2)$ for this random variable.
(d) What is the variance for this random variable?

4.22 Show that for any random variable it is true that $E(X^2) > E^2(X)$ unless the random variable is zero almost always.

4.23 Verify the entries in Table 4.5 for the mean and variances of the probability distributions given there.

4.24 A random variable X has the pdf

$$f_X(x) = Ae^{-bx}[u(x) - u(x - B)]$$

where $u(x)$ is the unit step function and A, B, and b are positive constants.

(a) Find the proper relationship between the constants A, b, and B.
(b) Determine and plot the cdf.
(c) Compute $E(X)$.
(d) Determine $E(X^2)$.
(e) What is the variance of X?

4.25 If

$$f_X(x) = (2\pi\sigma^2)^{-1/2} \exp\left(\frac{-x^2}{2\sigma^2}\right)$$

show that

(a) $E\{X^{2n}\} = 1 \cdot 3 \cdot 5 \cdots (2n - 1)\sigma^{2n}$, for $n = 1, 2, \ldots$
(b) $E\{X^{2n-1}\} = 0$, for $n = 1, 2, \ldots$

4.26 The random variable has pdf

$$f_x(x) = \tfrac{1}{2}\delta(x - 4) + \tfrac{1}{8}[u(x - 3) - u(x - 7)]$$

where $u(x)$ is the unit step. Determine the mean and the variance of the random variable thus defined.

4.27 Find the mean of $Z = X + R$, where X has the pdf given in Problem 4.21 and R has the Rayleigh pdf given in Equation (4.92).

4.28 Two random variables X and Y have means and variances given below:

$$m_x = 2 \quad \sigma_x^2 = 3 \quad m_y = 1 \quad \sigma_y^2 = 5$$

A new random variable Z is defined as

$$Z = 3X - 4Y$$

Determine the mean and variance of Z for each of the following cases of correlation between the random variables X and Y:

(a) $\rho_X iY = 0$; (b) $\rho_X iY = 0.25$; (c) $\rho_X iY = 0.6$; (d) $\rho_X iY = 1.0$.

4.29 Two Gaussian random variables X and Y, with zero mean and variance σ^2, between which there is a correlation coefficient ρ, have a joint probability density given by

$$f(x, y) = \frac{1}{2\pi\sigma^2\sqrt{1 - \rho^2}} \exp\left[- \frac{x^2 - 2\rho xy + y^2}{2\sigma^2(1 - \rho^2)} \right]$$

Verify that the symbol ρ in the expression for $f(x, y)$ is the correlation coefficient. That is, evalaute $E\{XY\}/\sigma^2$.

4.30 Using the definition of a conditional pdf given by Equation (4.54) and the expressions for the marginal and joint Gaussian pdf's, show that for two jointly Gaussian random variables X and Y, the conditional density function of X given Y has the form of a Gaussian density with conditional mean and the conditional variance given by

$$E(X \mid Y) = m_x + \frac{\rho\sigma_x}{\sigma_y}(Y - m_y)$$

$$\text{var}\,(X \mid Y) = \sigma_x^2 (1 - \rho^2)$$

respectively.

4.31 (a) Find the characteristic function corresponding to the pdf of Problem 4.21.
(b) Find $E\{X\}$ and $E\{X^2\}$ from the characteristic function.
(c) Do the results of (b) correspond with those of Problem 4.21?

4.32 The independent random variables X and Y have the probability densities

$$f(x) = e^{-x}u(x)$$
$$f(y) = \tfrac{1}{2}e^{-|y|}$$

Find and plot the probability density of the random variable $Z = X + Y$.

4.33 The random variable X has a probability density uniform in the range $-\frac{1}{2} \leq x \leq \frac{1}{2}$ and zero elsewhere. The independent variable Y has a density uniform in the range $1 \leq y \leq 3$ and zero elsewhere. Find the plot the density of $Z = X + Y$.

4.34 A random variable X is defined by

$$f_X(x) = 4e^{-8|x|}$$

The random variable Y is related to X by $Y = 2 + 3X$.

(a) Determine $E\{X\}$, $E\{X^2\}$, and σ_x^2.
(b) Determine $f_Y(y)$.
(c) Determine $E\{Y\}$, $E\{Y^2\}$, and σ_y^2.
(d) If you used $f_Y(y)$ in part (c), repeat that part using only $f_X(x)$.

4.35 A random variable X has the uniform probability density function

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$

(a) Determine the characteristic function $M_x(jv)$.
(b) Use the characteristic function to compute $E\{X\}$.
(c) Use the characteristic function to compute $E(X^2)$.
(d) What is the variance σ_x^2?

4.36 A random variable X has the probability density function

$$f_X(x) = \begin{cases} ae^{-ax}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

where a is an arbitrary positive constant.

(a) Determine the characteristic function $M_X(jv)$.
(b) Use the characteristic function to determine $E\{X\}$ and $E\{X^2\}$. Check your results by computing

$$\int_{-\infty}^{\infty} x^n f_X(x)\, dx$$

for $n = 1$ and 2.
(c) Compute σ_x^2.

Section 4.4

4.37 Compare the binomial, Laplace, and Poisson distributions for

(a) $n = 3$ and $p = \frac{1}{5}$ (b) $n = 3$ and $p = \frac{1}{10}$
(c) $n = 10$ and $p = \frac{1}{5}$ (d) $n = 10$ and $p = \frac{1}{10}$

4.38 An honest coin is flipped 20 times. (a) Determine the probability of the occurrence of 5 or 6 heads. (b) Determine the probability of the first head occurring at toss number 15.

4.39 An honest coin is flipped 100 times. Determine the probability of the occurrence of between 50 and 60 heads inclusive.

4.40 Passwords in a computer installation take the form $X_1X_2X_3X_4$, where each character X_i is one of the 26 letters of the alphabet. Determine the maximum possible number of different passwords available for assignment for each of the two following conditions:

(a) A given letter of the alphabet can be used only once in a password.
(b) Letters can be repeated if desired, so that each X_i is completely arbitrary.
(c) If selection of letters for a given password is completely random, what is the probability that your competitor could access, on a single try, your computer in part (a)? part (b)?

4.41 (a) By applying the binomial distribution, find the probability that there will be fewer than 3 heads when 10 honest coins are tossed.
(b) Do the same computation using the Laplace approximation.
(c) Compare the results of parts (a) and (b) by computing the percent error of the Laplace approximation.

4.42 A digital data transmission system has an error probability of 10^{-6} per digit. Find the probability of 3 or more errors in 10^6 digits.

4.43 (a) Using the expression for the joint Gaussian pdf [Equation (4.177)], obtain the marginal pdf's $f_X(x)$ and $f_Y(y)$. *Hint:* Complete the square in the exponent in x and use the integral

$$\int_{-\infty}^{\infty} \exp(-u^2/2)\, du = \sqrt{2\pi}$$

after making an appropriate substitution to obtain $f_X(x)$. Follow a similar procedure for $f_Y(y)$.

(b) From the results of (a), find $E\{X\}$, $E\{Y\}$, var $\{X\}$, and var $\{Y\}$.

4.44 Consider the Cauchy density function

$$f_X(x) = \frac{K}{1 + x^2}, \qquad -\infty \leq x \leq \infty$$

(a) Find K.
(b) Find $E\{X\}$ and show that var $\{X\}$ is not finite.
(c) Show that the characteristic function of a Cauchy random variable is $M_x(jv) = \pi K e^{-|v|}$.
(d) Now consider $Z = X_1 + \cdots + X_N$ where the X_i's are Cauchy and independent. Thus their characteristic function is

$$M_Z(jv) = (\pi K)^N e^{-N|v|}$$

Show that $f_Z(z)$ is Cauchy, *Comment:* $f_Z(z)$ is not Gaussian as $N \to \infty$ because var $\{X_i\}$ is not finite and the conditions of the central-limit theorem are violated.

4.45 (Chi-squared pdf) Consider the random variable $Y = \sum_{i=1}^{N} X_i^2$ where the X_i's are independent Gaussian random variables with pdf's $n(0, \sigma)$.

(a) Show that the characteristic function of X_i^2 is

$$M_{X_i^2}(jv) = (1 - 2jv\sigma^2)^{-1/2}$$

(b) Show that the pdf of Y is

$$f_Y(y) = \begin{cases} \dfrac{y^{N/2-1} e^{-y/2\sigma^2}}{2^{N/2}\sigma^N \Gamma(N/2)}, & y \geq 0 \\ 0, & y < 0 \end{cases}$$

where $\Gamma(x)$ is the gamma function, which, for $x = n$ an integer, is $\Gamma(n) = (n - 1)!$. This pdf is known as the χ^2 (chi-squared) pdf with N degrees of freedom. *Hint:* Use the Fourier transform pair

$$\frac{y^{N/2-1} e^{-y/\alpha}}{\alpha^{N/2}\Gamma(N/2)} \leftrightarrow (1 - j\alpha v)^{-N/2}$$

(c) Show that for N large, the χ^2 pdf can be approximated as

$$f_Y(y) \cong \exp\left[-\frac{1}{2}\left(\frac{y - N\sigma^2}{\sqrt{4N\sigma^4}}\right)^2\right] \Big/ \sqrt{4N\pi\sigma^4}, \qquad N \gg 1$$

Hint: Use the central-limit theorem. Since the X_i's are independent,

$$\overline{Y} = \sum_{i=1}^{N} \overline{X_i^2} = N\sigma^2$$

and

$$\text{var } (Y) = \sum_{i=1}^{N} \text{var } (X_i^2) = N \text{ var } (X_i^2)$$

(d) Compare the approximation obtained in part (c) with $f_Y(y)$ for $N = 2, 4, 8$.

(e) Let $R^2 = Y$. Show that the pdf of R for $N = 2$ is Rayleigh.

4.46 Compare the Q-function and the approximation to it for large arguments given by (4.187) by plotting both expressions on a log-log graph. (*Note:* A mathematics package, such as Mathcad is handy for this problem.)

4.47 Determine the cdf for a Gaussian random variable of mean m and variance σ^2. Express in terms of the Q-function. Plot the resulting cdf for $m = 0$, and $\sigma = 0.5$, 1, and 2.

4.48 Let X be uniformly distributed over $|x| \leq 1$. Plot $P(|X| \leq k\sigma_x)$ versus k and the corresponding bound given by Chebyshev's inequality.

4.49 A random variable X has the probability density function

$$f_X(x) = \frac{e^{-(x-12)^2/30}}{\sqrt{30\pi}}$$

Express the following probabilities in terms of the Q function:

(a) $P(X \leq 11)$ (b) $P(10 < X \leq 12)$

(c) $P(11 < X \leq 13)$ (d) $P(9 < X \leq 12)$

4.50 Prove Chebyshev's inequality. *Hint:* Let $Y = (X - m_X)/\sigma_X$ and find a bound for $P(|Y| < k)$ in terms of k.

4.51 If the random variable X is Gaussian, with zero mean and variance σ^2, obtain numerical values for the following probabilities:

(a) $P(|X| > \sigma)$ (b) $P(|X| > 2\sigma)$

(c) $P(|X| > 3\sigma)$

4.52 Speech is sometimes idealized as having a Laplacian-amplitude pdf. That is, the amplitude is distributed according to

$$f_X(x) = \left(\frac{a}{2}\right) \exp\left(-a|x|\right)$$

Express the variance of X in terms of a and find the three probabilities found in Problem 4.51. Compare your answers with those you obtained for Problem 4.51.



COMPUTER EXERCISES

4.1. (a) In preparation for this computer exercise, prove the following theorem: If X is a continuous random variable with cdf $F_X(x)$, then the random variable

$$Y = F_x(X)$$

is a uniformly distributed random variable in the interval (0, 1).

(b)Use the theorem proved in part (a) to generate a continuous random variable with desired pdf (or cdf) given that most programming languages (including the mathematics packages mentioned in the footnote to the computer exercises in Chapter 2) have a uniform random number generator.

(c) Design a random generator to generate a sequence of exponentially distributed random variables having the pdf

$$f_x(x) = \alpha e^{-\alpha x} u(x)$$

where $u(x)$ is the unit step. Plot histograms of the random numbers generated to check the validity of the random number generator you designed.

4.2 (a) Let U and V be two statistically independent random numbers uniformly distributed in [0, 1]. Show that the following transformation generates two statistically independent Gaussian random numbers with unit variance and zero mean:

$$X = R \cos(2\pi U)$$
$$Y = R \sin(2\pi U)$$

where

$$R = \sqrt{-2 \log_{10}(V)}$$

(*Hint:* First show that R is Rayleigh.)

(b) Generate 1000 random variable pairs according to the above algorithm. Plot histograms for each set (i.e., X and Y), and compare with Gaussian pdf's after properly scaling the histograms.

4.3. Using the results of Problem 4.30 and the Gaussian random number generator designed in Computer Exercise 4.2, design a Gaussian random number generator that will provide a specified correlation between adjacent samples. Let

$$\rho(\tau) = e^{-\alpha|\tau|}$$

and plot sequences of Gaussian random numbers for various choices of α. Show how stronger correlation between adjacent samples affects the variation from sample to sample. (*Note:* To get memory over more than adjacent samples, a digital filter should be used with independent Gaussian samples at the input.)

4.4. Design a computer program to repeatedly convolve two functions. You might wish to use the FFT to do this efficiently. Use it to verify that the pdf of a sum of independent random variables (with finite first and second moments) approaches a Gaussian pdf, thus giving a plausibility proof of the central-limit theorem.

4.5. Check the validity of the central-limit theorem by repeatedly generating n independent uniformly distributed random variables in the interval $(-0.5, 0.5)$, forming the sum given by (4.175), and plotting the histogram. Do this for $n = 5$, 10, and 20. Can you say anything qualitatively and quantitatively about the approach of the sums to Gaussian random numbers? Repeat for exponentially distributed component random variables (do Computer Exercise 4.1 first). Can you think of any drawbacks to this approach to generating Gaussian random variables?

5 RANDOM SIGNALS AND NOISE

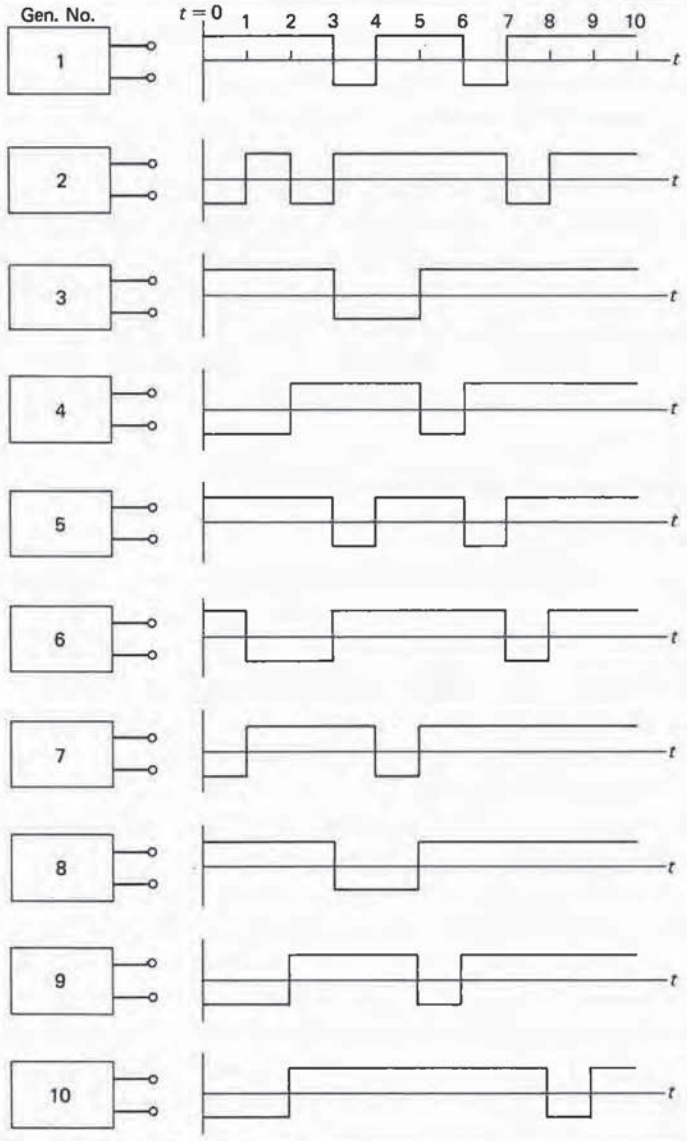
With the mathematical background developed in Chapter 4, we have finally arrived at the point where we can consider the statistical description of random waveforms. The importance of considering such waveforms, as pointed out in Chapter 1, lies in the fact that noise in communication systems is due to the random motion of charge carriers, and we observe macroscopic properties such as the current through a conductor that is due to the superposition of individual microscopic currents.

In the relative-frequency approach to probability, we imagined repeating the underlying chance experiment many times, the implication being that the replication process was carried out sequentially in time. In the study of random waveforms, however, the outcomes of our chance experiments are mapped into functions of time, or waveforms, rather than numbers, as in the case of random variables. The particular waveform is not predictable in advance of the experiment, just as the particular value of a random variable is not predictable before the chance experiment is performed. The problem we now address is the statistical description of chance experiments resulting in waveforms as outputs. To visualize how we may accomplish this, we again think in terms of relative frequency.

5.1 A RELATIVE-FREQUENCY DESCRIPTION OF RANDOM PROCESSES

For simplicity, we consider a binary digital waveform generator whose output randomly switches between $+1$ and -1 in T_0-second intervals, as shown in Figure 5.1. Let $X(t, \zeta_i)$ be the random waveform corresponding to the output of the ith generator. Suppose we use relative frequency to estimate $P(X = +1)$ by examining the outputs of all generators at a particular time. Since the outputs are functions of time, we must specify the time when

FIGURE 5.1 A statistically identical set of binary waveform generators with typical outputs



writing down the relative frequency. From an examination of the generator outputs we construct the following table:

Time Interval:	(0,1)	(1,2)	(2,3)	(3,4)	(4,5)	(5,6)	(6,7)	(7,8)	(8,9)	(9,10)
Relative Frequency:	$\frac{5}{10}$	$\frac{6}{10}$	$\frac{8}{10}$	$\frac{6}{10}$	$\frac{7}{10}$	$\frac{8}{10}$	$\frac{8}{10}$	$\frac{8}{10}$	$\frac{9}{10}$	$\frac{10}{10}$

From this table it is seen that the relative frequencies change with the time interval. Although this variation in relative frequency could be the result of *statistical irregularity,* we highly suspect that some phenomenon is making $X = +1$ more probable as time increases. To reduce the possibility that statistical irregularity is the culprit, we might repeat the experiment with 100 generators or 1000 generators. This is obviously a mental experiment in that it would be very difficult to obtain a set of identical generators and prepare them all in identical fashions. For some random processes, known as *ergodic,* it is possible to replace measurements on a set of generators at the same time with measurements across time on a single generator output.

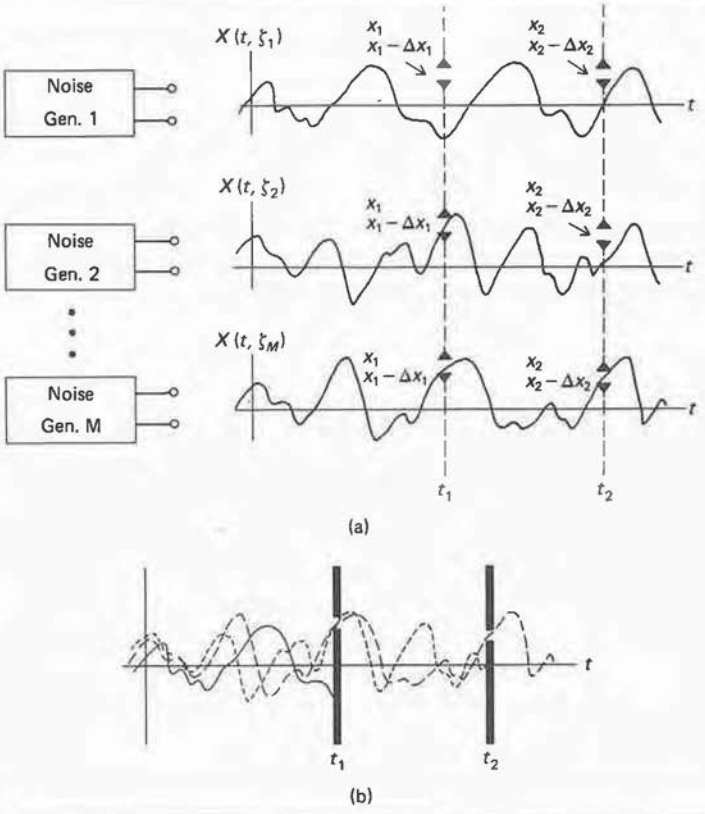
5.2 SOME TERMINOLOGY OF RANDOM PROCESSES

Sample Functions and Ensembles

In the same fashion as is illustrated in Figure 5.1, we could imagine performing any chance experiment many times simultaneously. If, for example, the random quantity of interest is the voltage at the terminals of a noise generator, the random variable X_1 may be assigned to represent the possible values of this voltage at time t_1 and the random variable X_2 the values at time t_2. As in the case of the digital waveform generator, we can imagine many noise generators all constructed in an identical fashion, insofar as we can make them, and run under identical conditions. Figure 5.2(a) shows typical waveforms generated in such an experiment. Each waveform $X(t, \zeta_i)$, is referred to as a *sample function,* where ζ_i is a member of a sample space $\mathcal{S}$. The totality of all sample functions is called an *ensemble.* The underlying chance experiment that gives rise to the ensemble of sample functions is called a *random,* or *stochastic, process.* Thus, to every outcome ζ we assign, according to a certain rule, a time function $X(t, \zeta)$. For a specific ζ, say ζ_i, $X(t, \zeta_i)$ signifies a single time function. For a specific time t_j, $X(t_j, \zeta)$ denotes a random variable. For fixed $t = t_j$ and fixed $\zeta = \zeta_i$, $X(t_j, \zeta_i)$ is a *number.* In what follows, we often suppress the ζ.

To summarize, the difference between a random variable and a random process is that for a random variable an outcome in the sample space is

FIGURE 5.2 Typical sample functions of a random process and illustration of the relative-frequency interpretation of its joint pdf. (a) Ensemble of sample functions. (b) Superposition of the sample functions shown in (a).



mapped into a number, whereas for a random process it is mapped into a function of time.

Description of Random Processes in Terms of Joint pdf's

A complete description of a random process $\{X(t, \zeta)\}$ is given by the N-fold joint pdf that probabilistically describes the possible values assumed by a typical sample function at times $t_N > t_{N-1} > \cdots > t_1$, where N is arbitrary. For $N = 1$, we can interpret this joint pdf $f_{X_1}(x_1, t_1)$ as

$$f_{X_1}(x_1, t_1)dx_1 = P(x_1 - dx_1 < X_1 \leq x_1 \text{ at time } t_1) \tag{5.1}$$

where $X_1 = X(t_1, \zeta)$. Similarly, for $N = 2$, we can interpret the joint pdf $f_{X_1X_2}(x_1, t_1; x_2, t_2)$ as

$$f_{X_1X_2}(x_1, t_1; x_2, t_2)dx_1dx_2 = P(x_1 - dx_1 < X_1 \leq x_1 \text{ at time } t_1 \text{ and } x_2 - dx_2 < X_2 \leq x_2 \text{ at time } t_2) \quad (5.2)$$

where $X_2 = X(t_2, \zeta)$.

To help visualize the interpretation of (5.2), Figure 5.2(b) shows the three sample functions of Figure 5.2(a) superimposed with barriers placed at $t = t_1$ and $t = t_2$. According to the relative-frequency interpretation, the joint probability given by (5.2) is the number of sample functions that pass through the slits in both barriers divided by the total number M of sample functions as M becomes large without bound.

Stationarity

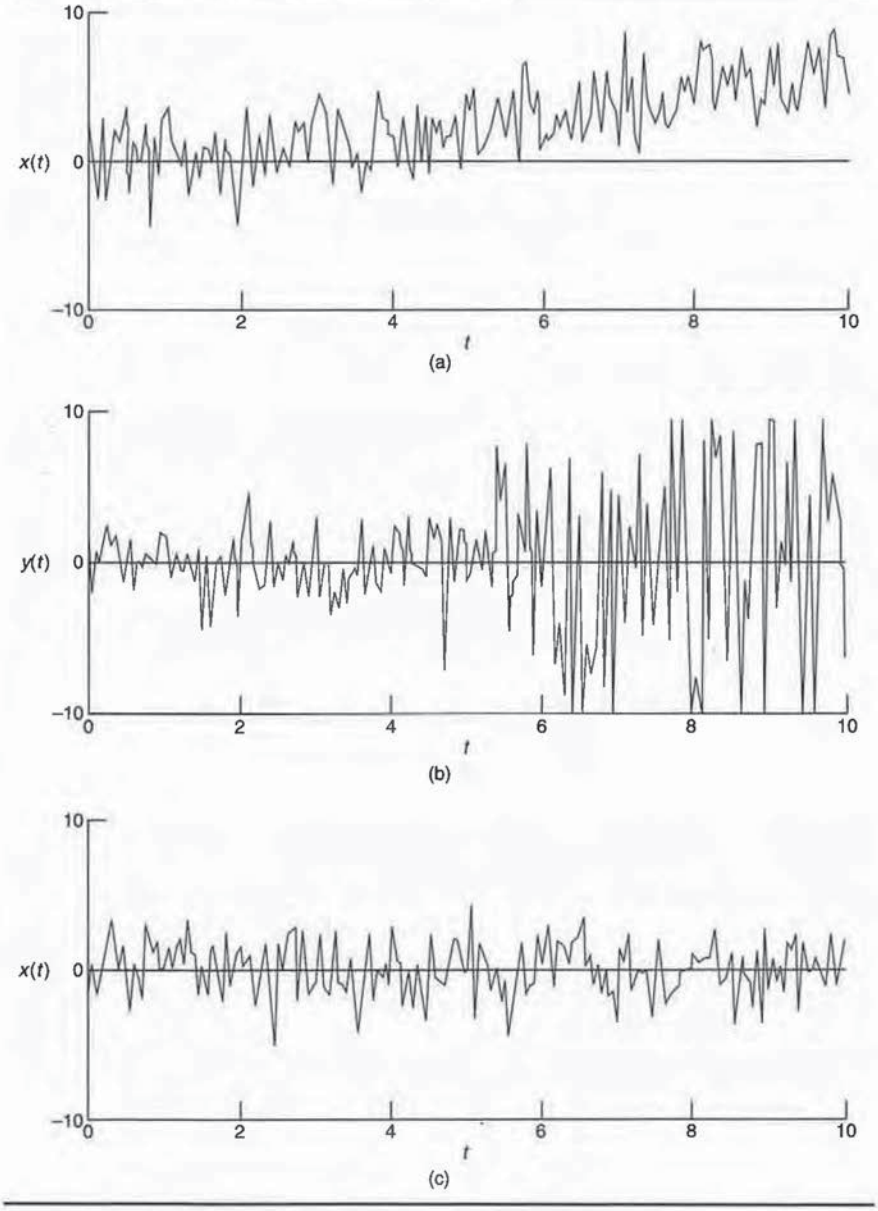
We have indicated the possible dependence of $f_{X_1X_2}$ on t_1 and t_2 by including them in its argument. If $\{X(t)\}$ were a Gaussian random process, for example, its values at time t_1 and t_2 would be described by Equation (4.177) where m_X, m_Y, σ_X^2, σ_Y^2, and ρ would, in general, depend on t_1 and t_2.* Note that we need a general N-fold pdf to completely describe the random process $\{X(t)\}$. In general, such a pdf depends on N time instants $t_1, t_2, \ldots, t_N$. In some cases, these joint pdf's depend only on the time differences $t_2 - t_1$, $t_3 - t_1, \ldots, t_N - t_1$; that is, the choice of time origin for the random process is immaterial. Such random processes are said to be *statistically stationary in the strict sense,* or simply *stationary.*

For stationary processes, means and variances are independent of time, and the correlation coefficient (or covariance) depends only on the time difference $t_2 - t_1$.† Figure 5.3 contrasts sample functions of stationary and nonstationary processes. It may happen that in some cases the mean and variance of a random process are time-independent and the covariance is a function only of the time difference, but the N-fold joint pdf depends on the time origin. Such random processes are called *wide-sense stationary processes* to distinguish them from strictly stationary processes (that is, processes whose N-fold pdf is independent of time origin). Strict-sense stationarity implies wide-sense stationarity, but the reverse is not necessarily true. An exception occurs for *Gaussian random processes for which wide-sense stationarity does imply strict-sense stationarity,* since the joint Gaussian pdf is completely

* For a stationary process, all joint moments are independent of time origin. We are interested primarily in the covariance, however.

† At N instants of time its values would be described by (B.1) of Appendix B.

FIGURE 5.3 Sample functions of nonstationary processes contrasted with a sample function of a stationary process. (a) Time-varying mean. (b) Time-varying variance. (c) Stationary.



specified in terms of the means, variances, and covariances of $X(t_1)$, $X(t_2)$, . . . , $X(t_N)$.

Partial Description of Random Processes: Ergodicity

As in the case of random variables, we may not always require a complete statistical description of a random process, or we may not be able to obtain the N-fold joint pdf even if desired. In such cases, we work with various moments, either by choice or by necessity. The most important averages are the mean,

$$m_X(t) = E\{X(t)\} = \overline{X(t)} \tag{5.3}$$

the variance,

$$\sigma_X^2(t) = E\{[X(t) - \overline{X(t)}]^2\} = \overline{X^2(t)} - \overline{[X(t)]}^2 \tag{5.4}$$

and the covariance,

$$\begin{aligned} \mu_X(t, t + \tau) &= E\{[X(t) - \overline{X(t)}][X(t + \tau) - \overline{X(t + \tau)}]\} \\ &= E\{X(t)X(t + \tau)\} - \overline{X(t)}\,\overline{X(t + \tau)} \end{aligned} \tag{5.5}$$

In (5.5), we let $t = t_1$ and $t + \tau = t_2$. The first term on the right-hand side is the *autocorrelation function* computed as a *statistical,* or *ensemble, average* (that is, the average is across the sample functions at times t and $t + \tau$). In terms of the joint pdf of the random process, the autocorrelation function is

$$R_X(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_{X_1 X_2}(x_1, t_1; x_2, t_2)\, dx_1\, dx_2 \tag{5.6}$$

where $X_1 = X(t_1)$ and $X_2 = X(t_2)$. If the process is stationary, $f_{X_1 X_2}$ does not depend on t but rather on the time difference, $\tau = t_2 - t_1$ and as a result, $R_X(t_1, t_2) = R_X(\tau)$ is a function only of τ. The following question arises: "If we compute the autocorrelation function using the definition of a time average given in Chapter 2, Equation (2.149), will the result be the same as the statistical average given by (5.6)?" For many processes, referred to as *ergodic,* the answer is affirmative. Ergodic processes are processes for which *time and ensemble averages are interchangeable.* Thus, if $X(t)$ is an ergodic process, *all time and corresponding ensemble averages are interchangeable.* In particular,

$$m_X = E\{X(t)\} = \langle X(t) \rangle$$

$$\sigma_X^2 = E\{[X(t) - \overline{X(t)}]^2\} = \langle [X(t) - \langle X(t) \rangle]^2 \rangle$$

and

$$R_X(\tau) = E\{X(t)X(t+\tau)\} = \langle X(t)X(t+\tau)\rangle$$

where

$$\langle v(t)\rangle \triangleq \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} v(t)\,dt$$

as defined in Chapter 2. We emphasize that for ergodic processes *all* time and ensemble averages are interchangeable, not just the mean, variance, and autocorrelation function.

▼ EXAMPLE 5.1 Consider the random process with sample functions*

$$n(t) = A\cos(\omega_0 t + \Theta)$$

where ω_0 is a constant and Θ is a random variable with the pdf

$$f_\Theta(\theta) = \begin{cases} \dfrac{1}{2\pi}, & |\theta| \le \pi \\ 0, & \text{otherwise} \end{cases} \tag{5.7}$$

Computed as statistical averages, the first and second moments are

$$\overline{n(t)} = \int_{-\infty}^{\infty} A\cos(\omega_0 t+\theta) f_\Theta(\theta)\,d\theta = \int_{-\pi}^{\pi} A\cos(\omega_0 t+\theta)\frac{d\theta}{2\pi} = 0 \tag{5.8a}$$

and

$$\overline{n^2(t)} = \int_{-\pi}^{\pi} A^2\cos^2(\omega_0 t+\theta)\frac{d\theta}{2\pi} = \frac{A^2}{2} \tag{5.8b}$$

respectively. The variance is equal to the second moment, since the mean is zero.

Computed as time averages, the first and second moments are

$$\langle n(t)\rangle = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} A\cos(\omega_0 t+\theta)\,dt = 0 \tag{5.9a}$$

* In this example we violate our earlier established convention that sample functions are denoted by capital letters. This is quite often done if confusion will not result.

and

$$\langle n^2(t)\rangle = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} A^2\cos^2(\omega_0 t+\theta)\,dt = \frac{A^2}{2} \tag{5.9b}$$

respectively. In general, the time average of some function of an ensemble member of a random process is a random variable. In this example, $\langle n(t)\rangle$ and $\langle n^2(t)\rangle$ are constants! We suspect that this random process is stationary and ergodic, even though the preceding results do not *prove* this. It turns out that this is indeed true.

To continue the example, consider the pdf

$$f_\Theta(\theta) = \begin{cases} \dfrac{2}{\pi}, & |\theta| \le \frac{1}{4}\pi \\ 0, & \text{otherwise} \end{cases} \tag{5.10}$$

For this case, the expected value, or mean, of the random process computed at an arbitrary time t is

$$\begin{aligned}\overline{n(t)} &= \int_{-\pi/4}^{\pi/4} A\cos(\omega_0 t+\theta)\frac{2}{\pi}\,d\theta \\ &= \frac{2}{\pi}A\sin(\omega_0 t+\theta)\Big|_{-\pi/4}^{\pi/4} = \frac{2\sqrt{2}A}{\pi}\cos\omega_0 t\end{aligned} \tag{5.11a}$$

The second moment, computed as a statistical average, is

$$\begin{aligned}\overline{n^2(t)} &= \int_{-\pi/4}^{\pi/4} A^2\cos^2(\omega_0 t+\theta)\frac{2}{\pi}\,d\theta \\ &= \int_{-\pi/4}^{\pi/4}\frac{A^2}{\pi}[1+\cos(2\omega_0 t+2\theta)]\,d\theta = \frac{A^2}{2}+\frac{A^2}{\pi}\cos 2\omega_0 t\end{aligned} \tag{5.11b}$$

Since stationarity of a random process implies that all moments are independent of time origin, the preceding results show that this process is not stationary. In order to comprehend the physical reason for this, you should sketch some typical sample functions. In addition, this process cannot be ergodic, since ergodicity requires stationarity. Indeed, the time-average first and second moments are still $\langle n(t)\rangle = 0$ and $\langle n^2(t)\rangle = \frac{1}{2}A^2$, respectively. Thus we have exhibited two time averages that are not equal to the corresponding statistical averages. ▲

Meanings of Various Averages for Ergodic Processes

It is useful to pause at this point and summarize the meanings of various averages for an ergodic process:

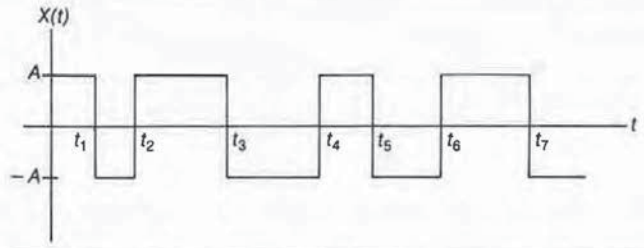
1. The mean $\overline{X(t)} = \langle X(t) \rangle$ is the dc component.
2. $\overline{X(t)}^2 = \langle X(t) \rangle^2$ is the dc power.
3. $\overline{X^2(t)} = \langle X^2(t) \rangle$ is the total power.
4. $\sigma_X^2 = \overline{X^2(t)} - \overline{X(t)}^2 = \langle X^2(t) \rangle - \langle X(t) \rangle^2$ is the power in the ac (time-varying) component.
5. The total power $\overline{X^2(t)} = \sigma_X^2 + \overline{X(t)}^2$ is the ac power plus the dc power.

Thus, in the case of ergodic processes, we see that these moments are measurable quantities in the sense that they can be replaced by the corresponding time averages and a finite-time approximation to these time averages can be measured in the laboratory.

▼ **EXAMPLE 5.2** To illustrate some of the definitions given above with regard to correlation functions, let us consider a random telegraph waveform $X(t)$, as illustrated in Figure 5.4. The sample functions of this random process have the following properties:

1. The values taken on at any time instant t_0 are either $X(t_0) = A$ or $X(t_0) = -A$ with equal probability.
2. The number k of switching instants in any time interval T obeys a Poisson distribution, as defined by Equation (4.170), with the attendant assumptions leading to this distribution. (That is, the probability of more than one switching instant occurring in an infinitesimal time interval dt is zero, with the probability of exactly one switching instant occurring in dt being $\alpha\, dt$, where α is a constant. Furthermore, successive switching occurrences are independent.)

FIGURE 5.4 Sample function of a random telegraph waveform



If τ is any positive time interval, the autocorrelation function of the random process defined by the preceding properties can be calculated as

$$\begin{aligned} R_X(\tau) &= E[X(t)X(t+\tau)] \\ &= A^2 P\{X(t) \text{ and } X(t+\tau) \text{ have the same sign}\} \\ &\quad + (-A^2)P\{X(t) \text{ and } X(t+\tau) \text{ have different signs}\} \\ &= A^2 P\{\text{even number of switching instants in } (t, t+\tau)\} \\ &\quad - A^2 P\{\text{odd number of switching instants in } (t, t+\tau)\} \\ &= A^2 \sum_{\substack{k=0 \\ k \text{ even}}}^{\infty} \frac{(\alpha\tau)^k}{k!} \exp(-\alpha\tau) - A^2 \sum_{\substack{k=0 \\ k \text{ odd}}}^{\infty} \frac{(\alpha\tau)^k}{k!} \exp(-\alpha\tau) \\ &= A^2 \exp(-\alpha\tau) \sum_{k=0}^{\infty} \frac{(-\alpha\tau)^k}{k!} \\ &= A^2 \exp(-\alpha\tau) \exp(-\alpha\tau) = A^2 \exp(-2\alpha\tau) \end{aligned} \tag{5.12}$$

The preceding development was carried out under the assumption that τ was positive. It could have been similarly carried out with τ negative, such that

$$R_X(\tau) = E\{X(t)X(t - |\tau|)\} = E\{X(t - |\tau|)X(t)\} = A^2 \exp(-2\alpha|\tau|) \tag{5.13}$$

This is a result that holds for all τ. That is, $R_X(\tau)$ is an even function of τ, which we will show in general shortly. ▲

5.3 CORRELATION FUNCTIONS AND POWER SPECTRA

The autocorrelation function, computed as a statistical average, has been defined by (5.6). If a process is ergodic, the autocorrelation function computed as a time average, as in Equation (2.149) of Chapter 2, is equal to the statistical average of (5.6). In Chapter 2, we defined the power spectral density $S(f)$ as the Fourier transform of the autocorrelation function $R(\tau)$. The *Weiner-Khinchine theorem* is a formal statement of this result for stationary random processes, for which $R(t_1, t_2) = R(t_2 - t_1) = R(\tau)$. For such processes, previously defined as wide-sense stationary, the power spectral density and autocorrelation function are Fourier transform pairs:

$$S(f) \leftrightarrow R(\tau) \tag{5.14}$$

If the process is ergodic, $R(\tau)$ can be calculated as either a time or an ensemble average.

Since $R_X(0) = \overline{X^2(t)}$ is the average power contained in the process, we have from the inverse Fourier transform that

$$\text{Average power} = R_X(0) = \int_{-\infty}^{\infty} S_X(f)\, df \tag{5.15}$$

which is reasonable, since the definition of $S_X(f)$ is that it is *power density* with respect to frequency.

Power Spectral Density

An intuitively satisfying, and in some cases computationally useful, expression for the power spectral density of a random process can be obtained by the following approach. Consider a particular sample function, $n(t, \zeta_i)$, of a random process. To obtain a function giving power density versus frequency using the Fourier transform, we consider a truncated version, $n_T(t, \zeta_i)$, defined as*

$$n_T(t, \zeta_i) = \begin{cases} n(t, \zeta_i), & |t| < \frac{1}{2}T \\ 0, & \text{otherwise} \end{cases} \tag{5.16}$$

Since sample functions of stationary random processes are power signals, the Fourier transform of $n(t, \zeta_i)$ does not exist, which necessitates defining $n_T(t, \zeta_i)$. The Fourier transform of a truncated sample function is

$$N_T(f, \zeta_i) = \int_{-T/2}^{T/2} n(t, \zeta_i) e^{-j2\pi ft}\, dt \tag{5.17}$$

and its energy spectral density, according to Equation (2.83), is $|N_T(f, \zeta_i)|^2$. The time-average power density over the interval $[-\frac{1}{2}T, \frac{1}{2}T]$ for this sample function is $|N_T(f, \zeta_i)|^2/T$. Since this time-average power density depends on the particular sample function chosen, we perform an ensemble average and take the limit as $T \to \infty$ to obtain the distribution of power density with frequency. This is defined as the power spectral density $S_n(f)$:

$$S_n(f) = \lim_{T\to\infty} \frac{\overline{|N_T(f, \zeta_i)|^2}}{T} \tag{5.18}$$

The operations of taking the limit and taking the ensemble average in (5.18) cannot be interchanged.

* Again, we use a lower-case letter to denote a random process for the simple reason that we need to denote the Fourier transform of $n(t)$ by an upper-case letter.

▼ **EXAMPLE 5.3** Let us find the power spectral density of the random process considered in Example 5.1 using (5.18). In this case,

$$n_T(t, \Theta) = A\Pi\left(\frac{t}{T}\right)\cos\left[\omega_0\left(t + \frac{\Theta}{\omega_0}\right)\right] \tag{5.19}$$

By the time-delay theorem of Fourier transforms and using the transform pair

$$\cos \omega_0 t \leftrightarrow \tfrac{1}{2}\delta(f - f_0) + \tfrac{1}{2}\delta(f + f_0) \tag{5.20}$$

we obtain

$$\mathcal{F}[\cos(\omega_0 t + \Theta)] = \tfrac{1}{2}\delta(f - f_0)e^{j\Theta} + \tfrac{1}{2}\delta(f + f_0)e^{-j\Theta} \tag{5.21}$$

We also recall from Chapter 2 (Example 2.8) that $\Pi(t/T) \leftrightarrow T \operatorname{sinc} Tf$, so, by the multiplication theorem of Fourier transforms,

$$\begin{aligned} N_T(f, \Theta) &= (AT \operatorname{sinc} Tf) * [\tfrac{1}{2}\delta(f - f_0)e^{j\Theta} + \tfrac{1}{2}\delta(f + f_0)e^{-j\Theta}] \\ &= \tfrac{1}{2}AT[e^{j\Theta} \operatorname{sinc}(f - f_0)T + e^{-j\Theta} \operatorname{sinc}(f + f_0)T] \end{aligned} \tag{5.22}$$

Therefore, the energy spectral density of the sample function is

$$\begin{aligned} |N_T(f, \Theta)|^2 = (\tfrac{1}{2}AT)^2[&\operatorname{sinc}^2 T(f - f_0) + e^{j2\Theta} \operatorname{sinc} T(f - f_0) \operatorname{sinc} T(f + f_0) \\ &+ e^{-j2\Theta} \operatorname{sinc} T(f - f_0) \operatorname{sinc} T(f + f_0) + \operatorname{sinc}^2 T(f + f_0)] \end{aligned} \tag{5.23}$$

In obtaining $\overline{[|N_T(f, \Theta)|^2]}$, we note that

$$\overline{\exp(\pm j2\Theta)} = \int_{-\pi}^{\pi} e^{\pm j2\theta}\frac{d\theta}{2\pi} = \int_{-\pi}^{\pi}(\cos 2\theta \pm j \sin 2\theta)\frac{d\theta}{2\pi} = 0 \tag{5.24}$$

Thus we obtain

$$\overline{|N_T(f, \Theta)|^2} = (\tfrac{1}{2}AT)^2[\operatorname{sinc}^2 T(f - f_0) + \operatorname{sinc}^2 T(f + f_0)] \tag{5.25}$$

and the power spectral density is

$$S_n(f) = \lim_{T\to\infty} \tfrac{1}{4}A^2[T \operatorname{sinc}^2 T(f - f_0) + T \operatorname{sinc}^2 T(f + f_0)] \tag{5.26}$$

However, a representation of the delta function is $\lim_{T\to\infty} T \operatorname{sinc}^2 Tu = \delta(u)$. See Figure 2.4(b). Thus

$$S_n(f) = \tfrac{1}{4}A^2\delta(f - f_0) + \tfrac{1}{4}A^2\delta(f + f_0) \tag{5.27}$$

The average power is $\int_{-\infty}^{\infty} S_n(f)\,df = \frac{1}{2}A^2$, the same as that obtained in Example 5.1. ▲

Wiener-Khinchine Theorem

To simplify the notation in the proof of the Wiener-Khinchine theorem, we rewrite (5.18) as

$$S_n(f) = \lim_{T\to\infty} \frac{E\{|\mathcal{F}[n_{2T}(t)]|^2\}}{2T} \tag{5.28}$$

where, for convenience, we have truncated over a $2T$-second interval and dropped ζ in the argument of $n_{2T}(t)$. Note that

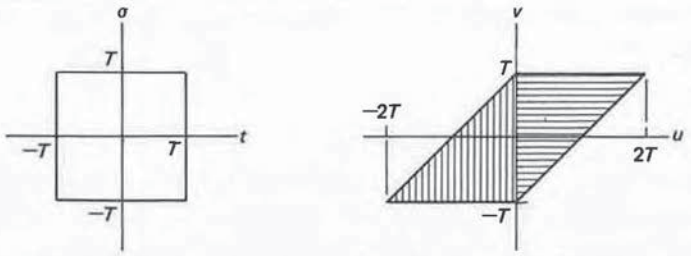
$$\begin{aligned} |\mathcal{F}[n_{2T}(t)]|^2 &= \left|\int_{-T}^{T} n(t)e^{-j\omega t}\,dt\right|^2, \qquad \omega = 2\pi f \\ &= \int_{-T}^{T}\int_{-T}^{T} n(t)n(\sigma)e^{-j\omega(t-\sigma)}\,dt\,d\sigma \end{aligned} \tag{5.29}$$

where the product of two integrals has been written as an iterated integral. Taking the ensemble average and interchanging the orders of averaging and integration, we obtain

$$\begin{aligned} E\{|\mathcal{F}[n_{2T}(t)]|^2\} &= \int_{-T}^{T}\int_{-T}^{T} E\{n(t)n(\sigma)\}e^{-j\omega(t-\sigma)}dt\,d\sigma \\ &= \int_{-T}^{T}\int_{-T}^{T} R_n(t-\sigma)e^{-j\omega(t-\sigma)}\,dt\,d\sigma \end{aligned} \tag{5.30}$$

by the definition of the autocorrelation function. The change of variables $u = t - \sigma$ and $v = t$ is now made with the aid of Figure 5.5. In the uv plane we integrate over v first and then over u by breaking the integration

FIGURE 5.5 Regions of integration for Equation (5.30)



over u up into two integrals, one for u negative and one for u postiive. Thus

$$E\{|\mathcal{F}[n_{2T}(t)]|^2\}$$

$$= \int_{u=-2T}^{0} R_n(u)e^{-j\omega u}\left(\int_{-T}^{u+T} dv\right) du + \int_{u=0}^{2T} R_n(u)e^{-j\omega u}\left(\int_{u-T}^{T} dv\right) du$$

$$= \int_{-2T}^{0} (2T + u)R_n(u)e^{-j\omega u}\, du + \int_{0}^{2T} (2T - u)R_n(u)e^{-j\omega u}\, du$$

$$= 2T\int_{-2T}^{2T}\left(1 - \frac{|u|}{2T}\right) R_n(u)e^{-j\omega u}\, du \tag{5.31}$$

The power spectral density is, by (5.28),

$$S_n(f) = \lim_{T\to\infty}\int_{-2T}^{2T}\left(1 - \frac{|u|}{2T}\right) R_n(u)e^{-j\omega u}\, du \tag{5.32}$$

which is the limit as $T \to \infty$ results in (5.14).

▼ EXAMPLE 5.4 Since the power spectral density and the autocorrelation function are transform pairs, the autocorrelation function of the random process defined in Example 5.1 is, from the result of Example 5.3, given by

$$R_n(\tau) = \mathcal{F}^{-1}[\tfrac{1}{4}A^2\delta(f - f_0) + \tfrac{1}{4}A^2\delta(f + f_0)] = \tfrac{1}{2}A^2\cos\omega_0\tau \tag{5.33}$$

Computing $R_n(\tau)$ as an ensemble average, we obtain

$$R_n(\tau) = E\{n(t)n(t+\tau)\} = \int_{-\pi}^{\pi} A^2\cos(\omega_0 t + \theta)\cos[\omega_0(t+\tau) + \theta]\frac{d\theta}{2\pi}$$

$$= \frac{A^2}{4\pi}\int_{-\pi}^{\pi}\{\cos\omega_0\tau + \cos[\omega_0(2t+\tau) + 2\theta]\}\, d\theta = \tfrac{1}{2}A^2\cos\omega_0\tau \tag{5.34}$$

which is the same result as that obtained using the Wiener-Khinchine theorem. ▲

Properties of the Autocorrelation Function

The properties of the autocorrelation function for a random process $X(t)$ were stated in Chapter 2, at the end of Section 2.6, and all time averages may now be replaced by statistical averages. These properties are now easily proved.

Property 1 states that $|R(\tau)| \leq R(0)$ for all τ. To show this, consider the nonnegative quantity

$$[X(t) \pm X(t+\tau)]^2 \geq 0 \tag{5.35}$$

where $\{X(t)\}$ is a stationary random process. Squaring and averaging term by term, we obtain

$$\overline{X^2(t)} \pm \overline{2X(t)X(t+\tau)} + \overline{X^2(t+\tau)} \geq 0 \tag{5.36a}$$

which reduces to

$$2R(0) \pm 2R(\tau) \geq 0 \qquad \text{or} \qquad -R(0) \leq R(\tau) \leq R(0) \tag{5.36b}$$

because $\overline{X^2(t)} = \overline{X^2(t+\tau)} = R(0)$ by the stationarity of $\{X(t)\}$.

Property 2 states that $R(-\tau) = R(\tau)$. This is easily proved by noting that

$$R(\tau) \triangleq \overline{X(t)X(t+\tau)} = \overline{X(t'-\tau)X(t')} \triangleq R(-\tau) \tag{5.37}$$

where the change of variables $t' = t + \tau$ has been made.

Property 3 states that $\lim_{|\tau|\to\infty} R(\tau) = \overline{X(t)}^2$ if $\{X(t)\}$ does not contain a periodic component. To show this, we note that

$$\begin{aligned} \lim_{|\tau|\to\infty} R(\tau) &\triangleq \lim_{|\tau|\to\infty} \overline{X(t)X(t+\tau)} \\ &\cong \overline{X(t)}\,\overline{X(t+\tau)}, \qquad \text{where } |\tau| \text{ is large} \\ &= \overline{X(t)}^2 \end{aligned} \tag{5.38}$$

where the second step follows intuitively because the interdependence between $X(t)$ and $X(t+\tau)$ becomes smaller as $|\tau| \to \infty$ (if no periodic components are present), and the last step results from the stationarity of $\{X(t)\}$.

Property 4, which states that $R(\tau)$ is periodic if $\{X(t)\}$ is periodic, follows by noting from the time-average definition of the autocorrelation function given by Equation (2.149) that periodicity of the integrand implies periodicity of $R(\tau)$.

Finally, property 5, which says that $\mathcal{F}[R(\tau)]$ is nonnegative, is a direct consequence of the Wiener-Khinchine theorem and (5.18).

▼ **EXAMPLE 5.5** Processes for which

$$S(f) = \begin{cases} \frac{1}{2}N_0, & |f| \leq B \\ 0, & \text{otherwise} \end{cases} \tag{5.39}$$

are commonly referred to as *bandlimited white noise,* since, as $B \to \infty$, all frequencies are present, in which case the process is simply called *white.* N_0

is the single-sided power spectral density of the nonbandlimited process. For a bandlimited white-noise process,

$$R(\tau) = BN_0 \operatorname{sinc} 2B\tau \tag{5.40}$$

As $B \to \infty$, $R(\tau) \to \frac{1}{2}N_0\delta(\tau)$. That is, no matter how close together we sample a white-noise process, the samples have zero correlation. If, in addition, the process is Gaussian, the samples are independent. A white-noise process has infinite power and is therefore a mathematical idealization, but it is nevertheless useful in systems analysis. ▲

▼ **EXAMPLE 5.6** As another example of calculating power spectra, let us consider a random process with sample functions that can be expressed as

$$X(t) = \sum_{k=-\infty}^{\infty} a_k p(t - kT - \Delta) \tag{5.41}$$

where $\ldots a_{-1}, a_0, a_1, \ldots, a_k \ldots$ is a sequence of random variables with

$$E[a_k] = 0 \qquad \text{and} \qquad E[a_k a_{k+m}] = R_m \tag{5.42}$$

The function $p(t)$ is a deterministic pulse-type waveform where T is the separation between pulses; Δ is a random variable that is independent of the value of a_k and uniformly distributed in the interval $(-T/2, T/2)$. The average value of this waveform is

$$E[X(t)] = \sum_{k=-\infty}^{\infty} E(a_k)E[p(t - kT - \Delta)] = 0 \tag{5.43}$$

The autocorrelation function is

$$\begin{aligned}
&R_X(\tau) \\
&= E[X(t)X(t + \tau)] \\
&= E\left\{\sum_{k=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} a_k a_{k+m} p(t - kT - \Delta)\, p[t + \tau - (k + m)T - \Delta]\right\} \\
&= \sum_{k=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} E[a_k a_{k+m}] E\{p(t - kT - \Delta)\, p[t + \tau - (k + m)T - \Delta]\} \\
&= \sum_{m=-\infty}^{\infty} R_m \sum_{k=-\infty}^{\infty} \int_{-T/2}^{T/2} p(t - kT - \Delta)\, p[t + \tau - (k + m)T - \Delta] \frac{d\Delta}{T} \\
&= \sum_{m=-\infty}^{\infty} R_m \sum_{k=-\infty}^{\infty} \int_{t-(k+1/2)T}^{t-(k-1/2)T} p(u)(u + \tau - mT) \frac{du}{T}
\end{aligned}$$

$$= \sum_{m=-\infty}^{\infty} R_m \left(\frac{1}{T}\right) \int_{-\infty}^{\infty} p(t + \tau + mT)p(t)\, dt$$

$$= \sum_{m=-\infty}^{\infty} R_m r(\tau + mT) \tag{5.44}$$

where

$$r(\tau) = \frac{1}{T} \int_{-\infty}^{\infty} p(t + \tau)p(t)\, dt \tag{5.45}$$

is the pulse-correlation function. We consider two cases as illustrations.

Case 1

Let $a_k = \pm A$, for all k, with $R_m = A^2$, $m = 0$, and $R_m = 0$, $m \neq 0$. Also, let $p(t) = \Pi(t/T)$, a unit pulse. This waveform is referred to as a *binary random waveform* and is typical of a baseband digital binary-modulated signal. It is useful in the computation of power spectra for some forms of modulation of binary digital data. The pulse-correlation function is

$$r(\tau) = \frac{1}{T}\int_{-\infty}^{\infty} \Pi\left(\frac{t + \tau}{T}\right) \Pi\left(\frac{t}{T}\right) dt$$

$$= \frac{1}{T}\int_{-T/2}^{T/2} \Pi\left(\frac{t + \tau}{T}\right) dt = \Lambda\left(\frac{\tau}{T}\right) \tag{5.46}$$

The autocorrelation function is

$$R_X(\tau) = A^2 \Lambda\left(\frac{\tau}{T}\right) \tag{5.47}$$

Using the Wiener-Khinchine theorem, we obtain the power spectral density:

$$S_X(f) = A^2 T \operatorname{sinc}^2(fT) \tag{5.48}$$

Case 2

Memory is built into the sequence of a_k's by the relationship

$$a_k = g_0 A_k + g_1 A_{k-1} \tag{5.49}$$

where g_0 and g_1 are constants and the A_k's are random variables such that $A_k = \pm A$ and

$$E[A_k A_{k+m}] = \begin{cases} A^2, & m = 0 \\ 0, & m \neq 0 \end{cases} \tag{5.50}$$

Again it is assumed that $p(t)$ is a unit-pulse function, although other pulse shapes could be assumed. It can be shown that

$$E[a_k a_{k+m}] = \begin{cases} A^2(g_0^2 + g_1^2), & m = 0 \\ g_0 g_1 A^2, & m = \pm 1 \\ 0, & \text{otherwise} \end{cases} \tag{5.51}$$

As before, the pulse-correlation function is

$$r(\tau) = \Lambda\left(\frac{\tau}{T}\right) \tag{5.52}$$

The autocorrelation function is

$$R_X(\tau) = A^2\left\{[g_0^2 + g_1^2]\Lambda\left(\frac{\tau}{T}\right) + g_0 g_1\left[\Lambda\left(\frac{\tau + T}{T}\right) + \Lambda\left(\frac{\tau - T}{T}\right)\right]\right\} \tag{5.53}$$

The power spectral density of $x(t)$ is

$$S_X(f) = A^2 T \operatorname{sinc}^2(fT)\{g_0^2 + g_1^2 + g_0 g_1[\exp(j2\pi fT) + \exp(-j2\pi fT)]\} \tag{5.54}$$

Let both g_0 and g_1 be unity. The resulting power spectral density is

$$S_X(f) = 4A^2 T \operatorname{sinc}^2(fT)\cos^2(\pi fT) \tag{5.55}$$

Figure 5.6 compares the power spectra for the two cases considered here. Note the effect of memory. Other values for g_0 and g_1 can be assumed, and memory between more than adjacent pulses also can be assumed.

Cross-Correlation Function and Cross-Power Spectral Density

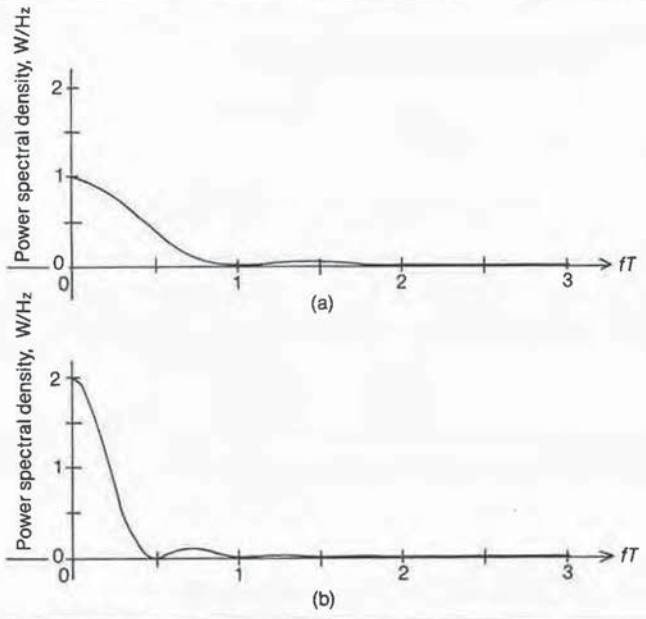
Suppose we wish to find the power in the sum of two noise voltages $X(t)$ and $Y(t)$. We might ask if we can simply add their separate powers. The answer is, in general, no. To see why, consider

$$n(t) = X(t) + Y(t) \tag{5.56}$$

where $X(t)$ and $Y(t)$ are two stationary random voltages that may be related (that is, that are not necessarily statistically independent). The power in the sum is

$$\begin{aligned} E\{n^2(t)\} &= E\{[X(t) + Y(t)]^2\} \\ &= E\{X^2(t)\} + 2E\{X(t)Y(t)\} + E\{Y^2(t)\} \\ &= P_X + 2P_{XY} + P_Y \end{aligned} \tag{5.57}$$

FIGURE 5.6 Power spectra of binary-valued waveforms. (a) Case in which there is no memory. (b) Case in which there is memory between adjacent pulses.



where P_X and P_Y are the powers of $X(t)$ and $Y(t)$, respectively, and P_{XY} is the cross power. More generally, we define the *cross-correlation function* as

$$R_{XY}(\tau) = E\{X(t)Y(t + \tau)\} \tag{5.58}$$

In terms of the cross-correlation function, $P_{XY} = R_{XY}(0)$. A *sufficient* condition for P_{XY} to be zero, so that we may simply add powers to obtain total power, is that

$$R_{XY}(0) = 0, \qquad \text{for all } \tau \tag{5.59}$$

Such processes are said to be *orthogonal*. If processes are statistically independent and at least one of them has zero mean, they are orthogonal. However, orthogonal processes are not necessarily statistically independent.

Cross-correlation functions can be defined for nonstationary processes also, in which case we have a function of two independent variables. We will not need to be this general in our considerations.

A useful symmetry property of the cross-correlation function for jointly stationary processes is

$$R_{XY}(\tau) = R_{YX}(-\tau) \tag{5.60}$$

which can be shown as follows. By definition,

$$R_{XY}(\tau) = E[X(t)Y(t + \tau)] \tag{5.61}$$

Defining $t' = t + \tau$, we obtain

$$R_{XY}(\tau) = E\{Y(t')X(t' - \tau)\} \triangleq R_{YX}(-\tau) \tag{5.62}$$

since the choice of time origin is immaterial for stationary processes.

The *cross-power spectral density* of two stationary random processes is defined as the Fourier transform of their cross-correlation function:

$$S_{XY}(f) = \mathcal{F}[R_{XY}(\tau)] \tag{5.63}$$

The cross-power spectral density provides, in the frequency domain, the same information about the random processes as does the cross-correlation function.

5.4 LINEAR SYSTEMS AND RANDOM PROCESSES*

Input-Output Relationships

In the consideration of the transmission of stationary random waveforms through fixed linear systems, a basic tool is the relationship of the output power spectral density to the input power spectral density, given as

$$S_y(f) = |H(f)|^2 S_x(f) \tag{5.64a}$$

The autocorrelation function of the output is the inverse Fourier transform of $S_y(f)$:

$$R_y(\tau) = \mathcal{F}^{-1}[S_y(f)] = \int_{-\infty}^{\infty} |H(f)|^2 S_x(f) e^{j2\pi f\tau}\, df \tag{5.64b}$$

$H(f)$ is the system's transfer function; $S_x(f)$ is the power spectral density of the input $x(t)$; $S_y(f)$ is the power spectral density of the output $y(t)$; and $R_y(\tau)$ is the autocorrelation function of the output. The analogous result for energy signals was proved in Chapter 2 [Equation (2.183)], and the result for power signals was simply stated.

* For the remainder of this chapter we use lower-case x and y to denote input and output random-process signals in keeping with Chapter 2 notation.

A proof of (5.64a) could be carried out by employing (5.18) in conjunction with Equation (2.166). We will take a somewhat longer route, however, and obtain several useful intermediate results. In addition, the proof provides practice in manipulating convolutions and expectations.

We begin by obtaining the cross-correlation function between input and output, $R_{xy}(\tau)$, defined as

$$R_{xy}(\tau) = E\{x(t)y(t + \tau)\} \tag{5.65}$$

Using the superposition integral, we have

$$y(t) = \int_{-\infty}^{\infty} h(u)x(t - u)\, du \tag{5.66}$$

where $h(t)$ is the system's impulse response. Equation (5.66) relates each sample function of the input and output processes, so we can write (5.65) as (transients are assumed to have died out)

$$R_{xy}(\tau) = E\left\{x(t) \int_{-\infty}^{\infty} h(u)x(t + \tau - u)\, du\right\} \tag{5.67}$$

Since the integral does not depend on t, we can take $x(t)$ inside and interchange the operations of expectation and convolution. (Both are simply integrals over different variables.) Since $h(u)$ is not random, (5.67) becomes

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} h(u)E\{x(t)x(t + \tau - u)\}\, du \tag{5.68}$$

By definition of the autocorrelation function of $x(t)$,

$$E\{x(t)x(t + \tau - u)\} = R_x(\tau - u) \tag{5.69}$$

Thus (5.68) can be written as

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} h(u)R_x(\tau - u)\, du \triangleq h(\tau) * R_x(\tau) \tag{5.70}$$

That is, the cross-correlation function of input with output is *the autocorrelation function of the input convolved with the impulse response,* an easily remembered result. Since (5.70) is a convolution, the Fourier transform of $R_{xy}(\tau)$, the cross-power spectral density of $x(t)$ with $y(t)$, is

$$S_{xy}(f) = H(f)S_x(f) \tag{5.71}$$

From the time-reversal theorem, item 3b of Table C.2, the cross-power spectral density $S_{yx}(f)$ is

$$S_{yx}(f) = \mathcal{F}[R_{yx}(\tau)] = \mathcal{F}[R_{xy}(-\tau)] = S^*_{xy}(f) \tag{5.72}$$

Employing (5.71) and using the relationships $H^*(f) = H(-f)$ and $S^*_x(f) = S_x(f)$ (where $S_x(f)$ is real), we obtain

$$S_{yx}(f) = H(-f)S_x(f) = H^*(f)S_x(f) \tag{5.73}$$

where the order of the subscripts is important. Taking the inverse Fourier transform of (5.73) with the aid of the convolution theorem, item 8 of Table C.2, and again using the time-reversal theorem, we obtain

$$R_{yx}(\tau) = h(-\tau) * R_x(\tau) \tag{5.74}$$

Let us pause to emphasize what we have obtained. By definition, $R_{xy}(\tau) \triangleq E\{x(t)y(t+\tau)\}$ can be written as

$$R_{xy}(\tau) \triangleq E\{x(t)\underbrace{[h(t) * x(t+\tau)]}_{y(t+\tau)}\} \tag{5.75}$$

Combining this with (5.70), we have

$$E\{x(t)[h(t) * x(t+\tau)]\} = h(\tau) * R_x(\tau) \triangleq h(\tau) * E\{x(t)x(t+\tau)\} \tag{5.76}$$

Similarly, (5.74) becomes

$$R_{yx}(\tau) \triangleq E\{\underbrace{[h(t) * x(t)]}_{y(t)}x(t+\tau)\} = h(-\tau) * R_x(\tau)$$
$$\triangleq h(-\tau) * E\{x(t)x(t+\tau)\} \tag{5.77}$$

Thus, bringing the convolution operation outside the expectation gives a convolution of $h(\tau)$ with the autocorrelation function if $h(t) * x(t+\tau)$ is inside the expectation or a convolution of $h(-\tau)$ with the autocorrelation function if $h(t) * x(t)$ is inside the expectation.

These results are combined to obtain the autocorrelation function of the output of a linear system in terms of the input autocorrelation function as follows:

$$R_y(\tau) \triangleq E\{y(t)y(t+\tau)\} = E\{y(t)[h(t) * x(t+\tau)]\} \tag{5.78}$$

which follows because $y(t+\tau) = h(t) * x(t+\tau)$. Using (5.76) with $x(t)$ replaced by $y(t)$, we obtain

$$\begin{aligned} R_y(\tau) &= h(\tau) * E\{y(t)x(t+\tau)\} \\ &= h(\tau) * R_{yx}(\tau) \\ &= h(\tau) * \{h(-\tau) * R_x(\tau)\} \end{aligned} \tag{5.79}$$

where the last line follows by substituting from (5.74). Written in terms of integrals, (5.79) is

$$R_y(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(u)h(v)R_x(\tau + v - u)\, dv\, du \tag{5.80}$$

The Fourier transform of (5.79) is the output power spectral density and is easily obtained as follows:

$$\begin{aligned} S_y(f) \triangleq \mathcal{F}[R_y(\tau)] &= \mathcal{F}[h(\tau) * R_{yx}(\tau)] = H(f)S_{yx}(f) \\ &= |H(f)|^2 S_x(f) \end{aligned} \tag{5.81}$$

where (5.73) has been substituted to obtain the last line.

▼ EXAMPLE 5.7 The input to a filter with impulse response $h(t)$ and transfer function $H(f)$ is a white-noise process with power spectral density

$$S_x(f) = \tfrac{1}{2}N_0, \qquad -\infty < f < \infty \tag{5.82}$$

The cross-power spectral density between input and output is

$$S_{xy}(f) = \tfrac{1}{2}N_0 H(f) \tag{5.83}$$

and the cross-correlation function is

$$R_{xy}(\tau) = \tfrac{1}{2}N_0 h(\tau) \tag{5.84}$$

Hence we could measure the impulse response of a filter by driving it with white noise and determining the cross-correlation function of input with output. Applications include system identification and channel measurement. ▲

Filtered Gaussian Processes

Suppose the input to a linear system is a random process. What can we say about the output statistics? For general inputs and systems, this is usually a difficult question to answer. However, *if the input to a linear system is Gaussian, the output is also Gaussian.*

A nonrigorous demonstration of this is carried out as follows. The sum of two independent Gaussian random variables has already been shown to be Gaussian. By repeated application of this result, we can find that the sum of

any number of independent Gaussian random variables is Gaussian.* For a fixed linear system, the output $y(t)$ in terms of the input $x(t)$ is given by

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

$$= \lim_{\Delta\tau \to 0} \sum_{k=-\infty}^{\infty} x(k\,\Delta\tau)h(t - k\,\Delta\tau)\,\Delta\tau \tag{5.85}$$

where $h(t)$ is the impulse response. By writing the integral as a sum, we have demonstrated that if $x(t)$ is a white Gaussian process, the output is also Gaussian (but not white) because, at any time t, the right-hand side of (5.85) is simply a linear combination of independent Gaussian random variables. (Recall Example 5.5, where the correlation function of white noise was shown to be an impulse. Also recall that uncorrelated Gaussian random variables are independent.)

If the input is not white, we can still show that the output is Gaussian by considering the cascade of two linear systems, as shown in Figure 5.7. The system in question is the one with the impulse response $h(t)$. To show that its output is Gaussian, we note that the cascade of $h_1(t)$ with $h(t)$ is a linear system with the impulse response

$$h_2(t) = h_1(t) * h(t) \tag{5.86}$$

This system's input, $z(t)$, is Gaussian and white. Therefore, its output, $y(t)$, is also Gaussian by application of the theorem just proved. However, the output of the system with impulse response $h_1(t)$ is Gaussian by application of the same theorem, but not white. Hence the output of a linear system with nonwhite Gaussian input is Gaussian.

▼ **EXAMPLE 5.8** The input to the lowpass RC filter shown in Figure 5.8 is white Gaussian noise with the power spectral density $S_{n_i}(f) = \frac{1}{2}N_0$, $-\infty < f < \infty$. The power spectral density of the output is

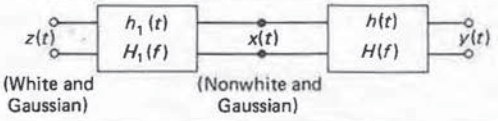
$$S_{n_0}(f) = S_{n_i}(f)|H(f)|^2 = \frac{\frac{1}{2}N_0}{1 + (f/f_3)^2} \tag{5.87}$$

where $f_3 = (2\pi RC)^{-1}$ is the 3-dB cutoff frequency. Inverse Fourier transforming $S_{n_0}(f)$, we obtain $R_{n_0}(\tau)$, the output autocorrelation function:

$$R_{n_0}(\tau) = \frac{N_0}{4RC}\, e^{-|\tau|/RC} \tag{5.88}$$

* This also follows from Appendix B, (B.11).

FIGURE 5.7 Cascade of two linear systems with Gaussian input



The square of the mean of $n_0(t)$ is

$$\overline{n_0(t)}^2 = \lim_{|\tau| \to \infty} R_{n_0}(\tau) = 0 \tag{5.89}$$

and the mean-squared value, which is also equal to the variance since the mean is zero, is

$$\overline{n_0^2(t)} = \sigma_{n_0}^2 = R_{n_0}(0) = \frac{N_0}{4RC} \tag{5.90}$$

Alternatively, we can find the average power at the filter output by integrating the power spectral density of $n_0(t)$. The result is as above:

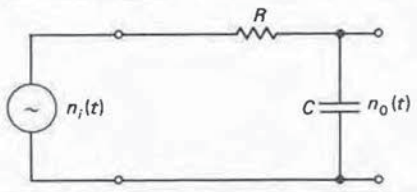
$$\overline{n_0^2(t)} = \int_{-\infty}^{\infty} \frac{\frac{1}{2}N_0}{1 + (f/f_3)^2}\, df = \frac{N_0}{2\pi RC} \int_0^{\infty} \frac{dx}{1 + x^2} = \frac{N_0}{4RC} \tag{5.91}$$

Since the input is Gaussian, the output is Gaussian as well. The first-order pdf is

$$f_{n_0}(y, t) = f_{n_0}(y) = \frac{e^{-2RCy^2/N_0}}{\sqrt{\pi N_0/2RC}} \tag{5.92}$$

by employing Equation (4.179). The second-order pdf at time t and $t + \tau$ is found by substitution into Equation (4.177). Letting X be a random variable that refers to the values the output takes on at time t and Y be a random

FIGURE 5.8 A lowpass RC filter with a white-noise input



variable that refers to the values the output takes on at time $t + \tau$, we have, from the preceding results,

$$m_x = m_y = 0 \tag{5.93a}$$

$$\sigma_x^2 = \sigma_y^2 = \frac{N_0}{4RC} \tag{5.93b}$$

and the correlation coefficient is

$$\rho(\tau) = \frac{R_{n_0}(\tau)}{R_{n_0}(0)} = e^{-|\tau|/RC} \tag{5.93c}$$

Referring to Example 5.2, one can see that the random telegraph waveform has the same autocorrelation function as that of the output of the lowpass RC filter of Example 5.8. This demonstrates that processes with drastically different sample functions can have the same second-order averages. ▲

Noise-Equivalent Bandwidth

If we pass white noise through a filter that has the transfer function $H(f)$, the average power at the output, by (5.64), is

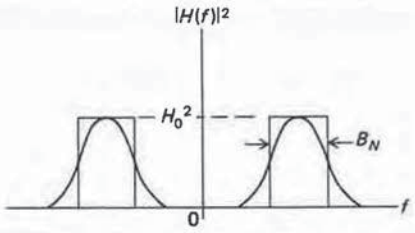
$$P_{n_0} = \int_{-\infty}^{\infty} \tfrac{1}{2}N_0 \, |H(f)|^2 \, df = N_0 \int_0^{\infty} |H(f)|^2 \, df \tag{5.94}$$

where $\frac{1}{2}N_0$ is the two-sided power spectral density of the input. If the filter were ideal with bandwidth B_N and midband gain H_0, as shown in Figure 5.9, the noise power at the output would be

$$P_{n_0} = H_0^2(\tfrac{1}{2}N_0)(2B_N) = N_0 B_N H_0^2 \tag{5.95}$$

The question we now ask is the following: What is the bandwidth of an ideal, fictitious filter that has the same midband gain as $H(f)$ and that passes the

FIGURE 5.9 Comparison between $|H(f)|^2$ and an idealized approximation



same noise power? If the midband gain of $H(f)$ is H_0, the answer is obtained by equating the preceding two results. Thus

$$B_N = \frac{1}{H_0^2} \int_0^\infty |H(f)|^2 \, df \tag{5.96}$$

is the single-side bandwidth of the fictitious filter. B_N is called the *noise-equivalent bandwidth* of $H(f)$.

It is sometimes useful to determine the noise-equivalent bandwidth of a system using time-domain integration. By Rayleigh's energy theorem [see (2.82)], we have

$$\int_{-\infty}^\infty |H(f)|^2 \, df = \int_{-\infty}^\infty |h(t)|^2 \, dt \tag{5.97}$$

Thus, (5.96) can be written as

$$B_N = \frac{1}{2H_0^2} \int_0^\infty |h(t)|^2 \, dt = \frac{\displaystyle\int_{-\infty}^\infty |h(t)|^2 \, dt}{2\left[\displaystyle\int_{-\infty}^\infty h(t) \, dt\right]^2} \tag{5.98}$$

where it is noted that

$$H_0 = H(f)|_{f=0} = \int_{-\infty}^\infty h(t) e^{-j2\pi ft} \, dt|_{f=0} = \int_{-\infty}^\infty h(t) \, dt \tag{5.99}$$

For some systems, (5.98) is easier to evaluate than (5.96).

▼ EXAMPLE 5.9 The noise-equivalent bandwidth of an nth-order Butterworth filter for which

$$|H_n(f)|^2 = \frac{1}{1 + (f/f_3)^{2n}} \tag{5.100}$$

is

$$\begin{aligned} B_N(n) &= \int_0^\infty \frac{1}{1 + (f/f_3)^{2n}} \, df = f_3 \int_0^\infty \frac{1}{1 + x^{2n}} \, dx \\ &= \frac{\pi f_3/2n}{\sin(\pi/2n)}, \qquad n = 1, 2, \ldots \end{aligned} \tag{5.101}$$

where f_3 is the 3-dB frequency of the filter. For $n = 1$, (5.101) gives the result for a lowpass RC filter. As n approaches infinity, $H_n(f)$ approaches

the transfer function of an ideal lowpass filter of single-sided bandwidth f_3. The noise-equivalent bandwidth is

$$\lim_{n\to\infty} B_N(n) = f_3 \tag{5.102}$$

as it should be by its definition. As the cutoff of a filter becomes sharper, its noise-equivalent bandwidth approaches its 3-dB bandwidth. ▲

▼ EXAMPLE 5.10 To illustrate the application of (5.98), consider the computation of the noise-equivalent bandwidth of a first-order Butterworth filter in the time domain. Its impulse response is

$$h(t) = \mathcal{F}^{-1}\left[\frac{1}{1 + j\dfrac{f}{f_3}}\right] = 2\pi f_3 e^{-2\pi f_3 t}u(t) \tag{5.103}$$

According to (5.98), the noise-equivalent bandwidth of this filter is

$$B_N = \frac{\displaystyle\int_0^\infty (2\pi f_3)^2 e^{-4\pi f_3 t}\,dt}{2\left[\displaystyle\int_0^\infty 2\pi f_3 e^{-2\pi f_3 t}\,dt\right]^2} = \frac{\pi f_3}{2}\left[-e^{4\pi f_3 t}\right]_0^\infty = \frac{\pi f_3}{2} \tag{5.104}$$

which checks with (5.101) if $n = 1$ is substituted. ▲

5.5 NARROWBAND NOISE

Quadrature-Component and Envelope-Phase Representation

In most communication systems operating at a carrier frequency f_c, the bandwidth of the channel, B, is small compared with f_c. In such situations, it is convenient to represent the noise in terms of quadrature components as

$$n(t) = n_c(t)\cos(\omega_0 t + \theta) - n_s(t)\sin(\omega_0 t + \theta) \tag{5.105}$$

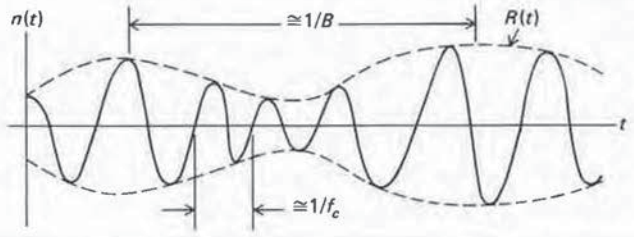
where $\omega_0 = 2\pi f_0$ and θ is an arbitrary phase angle. In terms of envelope and phase components, $n(t)$ can be written as

$$n(t) = R(t)\cos[\omega_0 t + \phi(t) + \theta] \tag{5.106}$$

where

$$R(t) = \sqrt{n_c^2 + n_s^2} \tag{5.107a}$$

FIGURE 5.10 A typical narrowband noise waveform



and

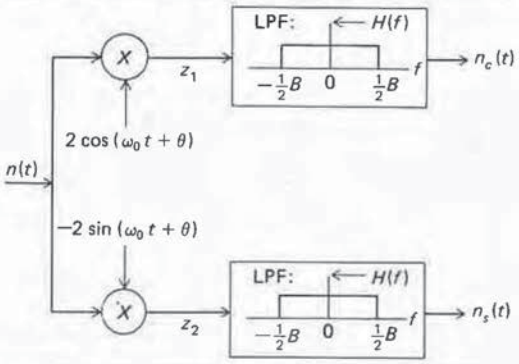
$$\phi(t) = \tan^{-1}\left[\frac{n_s(t)}{n_c(t)}\right] \tag{5.107b}$$

Actually, any random process can be represented in either of these forms, but if a process is narrowband, $R(t)$ and $\phi(t)$ can be interpreted as the slowly varying envelope and phase, respectively, as sketched in Figure 5.10.

Figure 5.11 shows the block diagram of a system for producing $n_c(t)$ and $n_s(t)$ where θ is, as yet, an arbitrary phase angle. Note that the composite operations used in producing $n_c(t)$ and $n_s(t)$ constitute linear systems (superposition holds from input to output). Thus, if $n(t)$ is a Gaussian process, so are $n_c(t)$ and $n_s(t)$. (The system of Figure 5.11 is to be interpreted as relating input and output processes sample function by sample function.)

We will prove several properties of $n_c(t)$ and $n_s(t)$. Most important, of

FIGURE 5.11 The operations involved in producing $n_c(t)$ and $n_s(t)$



course, is whether equality really holds in (5.105) and in what sense. We will show that

$$E\{[n(t) - [n_c(t)\cos(\omega_0 t + \theta) - n_s(t)\sin(\omega_0 t + \theta)]]^2\} = 0 \quad (5.108)$$

That is, the mean-squared error between a sample function of the actual noise process and the right-hand side of (5.105) is zero (averaged over the ensemble of sample functions). More useful when using the representation in (5.105), however, are the following properties:

Mean and Variances

$$\overline{n(t)} = \overline{n_c(t)} = \overline{n_s(t)} = 0 \quad (5.109a)$$

$$\overline{n^2(t)} = \overline{n_c^2(t)} = \overline{n_s^2(t)} \triangleq N \quad (5.109b)$$

Power Spectral Densities

$$S_{n_c}(f) = S_{n_s}(f) = L_p[S_n(f - f_0) + S_n(f + f_0)] \quad (5.110a)$$

$$S_{n_c n_s}(f) = jL_p[S_n(f - f_0) - S_n(f + f_0)] \quad (5.110b)$$

where $L_p[\ ]$ denotes the lowpass part of the quantity in brackets; $S_n(f)$, $S_{n_c}(f)$, and $S_{n_s}(f)$ are the power spectral densities of $n(t)$, $n_c(t)$, and $n_s(t)$, respectively; and $S_{n_c n_s}(f)$ is the cross-power spectral density of $n_c(t)$ and $n_s(t)$.

From (5.110b), we see that

$$R_{n_c n_s}(\tau) \equiv 0, \qquad \text{for all } \tau, \text{ if } L_p[S_n(f - f_0) - S_n(f + f_0)] = 0 \quad (5.110c)$$

This is an especially useful property in that it tells us that $n_c(t)$ and $n_s(t)$ are uncorrelated if the power spectral density of $n(t)$ is symmetrical about $f = f_0$ where $f > 0$. If, in addition, $n(t)$ is Gaussian, $n_c(t)$ and $n_s(t)$ will be *independent* Gaussian processes because they are uncorrelated, and the joint pdf of $n_c(t)$ and $n_s(t + \tau)$, for *any* delay τ, will simply be of the form

$$f(n_c, t; n_s, t + \tau) = \frac{e^{-(n_c^2 + n_s^2)/2N}}{2\pi N} \quad (5.111)$$

If $S_n(f)$ is not symmetrical about $f = f_0$, where $f > 0$, then (5.111) holds only for $\tau = 0$ or those values of τ for which $R_{n_c n_s}(\tau) = 0$.

Using the results of Example 4.15, the envelope and phase functions of (5.106) have the joint pdf

$$f(r, \phi) = \frac{r}{2\pi N} e^{-r^2/2N}, \qquad \text{for } r > 0 \text{ and } |\phi| \leq \pi \quad (5.112)$$

which holds for the same conditions as for (5.111). The properties given by (5.110) will be proved first, followed by (5.108). Note that the results expressed by (5.109) follow from (5.110a).

The Power Spectral Density Function of $n_c(t)$ and $n_s(t)$

To prove (5.110a), we first find the power spectral density of $z_1(t)$, as defined in Figure 5.11, by computing its autocorrelation function and Fourier transforming the result. To simplify the derivation, it is assumed that θ is a uniformly distributed random variable in $(0, 2\pi)$ and is statistically independent of $n(t)$.

The autocorrelation function of $z_1(t) = 2n(t)\cos(\omega_0 t + \theta)$ is

$$\begin{aligned} R_{z_1}(\tau) &= E\{4n(t)n(t+\tau)\cos(\omega_0 t + \theta)\cos[\omega_0(t+\tau) + \theta]\} \\ &= 2E\{n(t)n(t+\tau)\}\cos\omega_0\tau \\ &\quad + 2E\{n(t)n(t+\tau)\cos(2\omega_0 t + \omega_0\tau + 2\theta)\} \\ &= 2R_n(\tau)\cos\omega_0\tau \end{aligned} \tag{5.113}$$

where $R_n(\tau)$ is the autocorrelation function of $n(t)$. In obtaining (5.113), we used appropriate trigonometric identities in addition to the independence of $n(t)$ and θ. Thus, by the multiplication theorem of Fourier transforms, the power spectral density of $z_1(t)$ is

$$\begin{aligned} S_{z_1}(f) &= S_n(f) * [\delta(f - f_0) + \delta(f + f_0)] \\ &= S_n(f - f_0) + S_n(f + f_0) \end{aligned} \tag{5.114}$$

of which only the lowpass part is passed by $H(f)$. Thus the result for $S_{n_c}(f)$ expressed by (5.110a) follows. A similar proof can be carried out for $S_{n_s}(f)$. Equation (5.109b) follows by integrating (5.110a) over all f.

Next, let us consider (5.110b). To prove it, we need an expression for $R_{z_1z_2}(\tau)$, the cross-correlation function of $z_1(t)$ and $z_2(t)$. (See Figure 5.11.) By definition, and from Figure 5.11,

$$\begin{aligned} R_{z_1z_2}(\tau) &= E\{z_1(t)z_2(t+\tau)\} \\ &= -E\{4n(t)n(t+\tau)\cos(\omega_0 t + \theta)\sin[\omega_0(t+\tau) + \theta]\} \\ &= -2R_n(\tau)\sin\omega_0\tau \end{aligned} \tag{5.115}$$

where we again used appropriate trigonometric identities and the independence of $n(t)$ and θ. Letting $h(t)$ be the impulse response of the lowpass filters in Figure 5.11 and employing (5.76) and (5.77), the cross-correlation function of $n_c(t)$ and $n_s(t)$ can be written as

$$\begin{aligned} R_{n_cn_s}(\tau) &\triangleq E\{n_c(t)n_s(t+\tau)\} = E\{[h(t) * z_1(t)]n_s(t+\tau)\} \\ &= h(-\tau) * E\{z_1(t)n_s(t+\tau)\} = h(-\tau) * E\{z_1(t)[h(t) * z_2(t+\tau)]\} \\ &= h(-\tau) * h(\tau) * E\{z_1(t)z_2(t+\tau)\} = h(-\tau) * [h(\tau) * R_{z_1z_2}(\tau)] \end{aligned} \tag{5.116}$$

The Fourier transform of $R_{n_c n_s}(\tau)$ is the cross-power spectral density, $S_{n_c n_s}(f)$, which, from the convolution theorem, is given by

$$\begin{aligned} S_{n_c n_s}(f) &= H(f)\mathcal{F}[h(-\tau) * R_{z_1 z_2}(\tau)] = H(f)H^*(f)S_{z_1 z_2}(f) \\ &= |H(f)|^2 S_{z_1 z_2}(f) \end{aligned} \tag{5.117}$$

From (5.115) and the frequency-translation theorem, it follows that

$$\begin{aligned} S_{z_1 z_2}(f) &= \mathcal{F}[\, jR_n(\tau)(e^{j\omega_0\tau} - e^{-j\omega_0\tau})] \\ &= j[S_n(f - f_0) - S_n(f + f_0)] \end{aligned} \tag{5.118}$$

Thus, from (5.117),

$$\begin{aligned} S_{n_c n_s}(f) &= j|H(f)|^2[S_n(f - f_0) - S_n(f + f_0)] \\ &= jL_p[S_n(f - f_0) - S_n(f + f_0)] \end{aligned} \tag{5.119}$$

which proves (5.110b). Note that since the cross-power spectral density $S_{n_c n_s}(f)$ is imaginary, the cross-correlation function $R_{n_c n_s}(\tau)$ is odd. Thus $R_{n_c n_s}(0)$ is zero if the cross-correlation function is continuous at $\tau = 0$, which is the case for bandlimited signals.

▼ **EXAMPLE 5.11** Let us consider a bandpass random process with the power spectral density shown in Figure 5.12(a). Choosing the center frequency of $f_0 = 7$ Hz results in $n_c(t)$ and $n_s(t)$ being uncorrelated. Figure 5.12(b) shows $S_{z_1}(f)$ [or $S_{z_2}(f)$] for $f_0 = 7$ Hz with $S_{n_c}(f)$ [or $S_{n_s}(f)$], that is, the lowpass part of $S_{z_1}(f)$, shaded. The integral of $S_n(f)$ is 2(6)(2) = 24 W, which is the same result obtained from integrating the shaded portion of Figure 5.12(b).

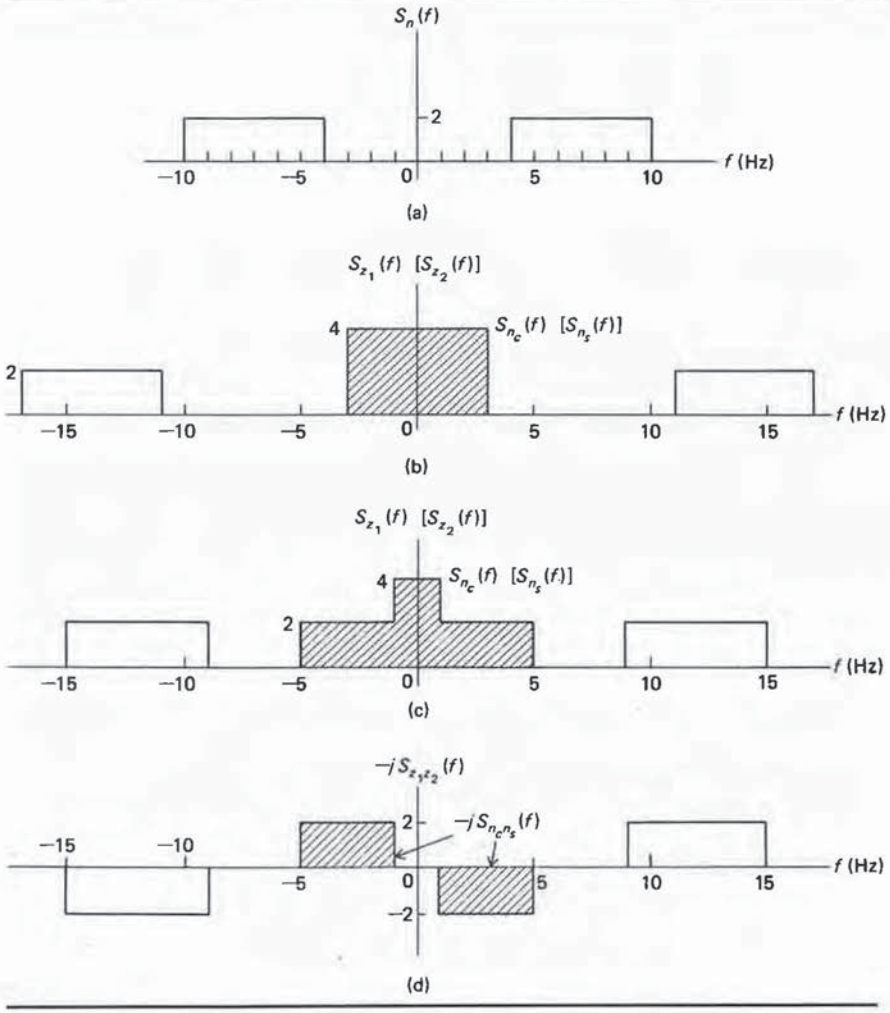
Now suppose f_0 is chosen as 5 Hz. Then $S_{z_1}(f)$ and $S_{z_2}(f)$ are as shown in Figure 5.12(c), with $S_{n_c}(f)$ shown shaded. From Equation (5.110b), it follows that $-jS_{n_c n_s}(f)$ is the shaded portion of Figure 5.12(d). Because of the asymmetry that results from the choice of f_0, $n_c(t)$ and $n_s(t)$ are not uncorrelated. As a matter of interest, we can calculate $R_{n_c n_s}(\tau)$ easily by using the transform pair

$$2AW \operatorname{sinc} 2W\tau \leftrightarrow A\Pi\left(\frac{f}{2W}\right) \tag{5.120}$$

and the frequency-translation theorem. From Figure 5.12(d), it follows that

$$S_{n_c n_s}(f) = 2j\{-\Pi[\tfrac{1}{4}(f - 3)] + \Pi[\tfrac{1}{4}(f + 3)]\} \tag{5.121}$$

FIGURE 5.12 Spectra for Example 5.11. (a) Bandpass spectrum. (b) Lowpass spectra for $f_0 = 7$ Hz. (c) Lowpass spectra for $f_0 = 5$ Hz. (d) Cross spectra for $f_0 = 5$ Hz.

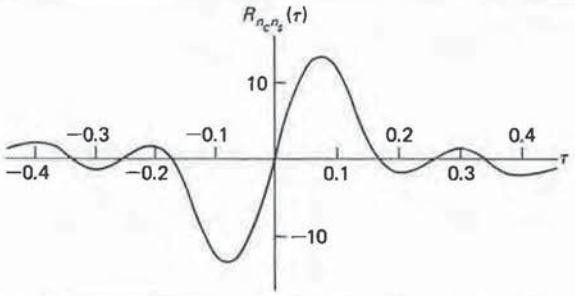


which results in the cross-correlation function

$$R_{n_c n_s}(\tau) = 2j(-4 \text{ sinc } 4\tau e^{j6\pi\tau} + 4 \text{ sinc } 4\tau e^{-j6\pi\tau}) = 16 \text{ sinc } 4\tau \sin 6\pi\tau \tag{5.122}$$

This cross-correlation function is shown in Figure 5.13. Although $n_c(t)$ and $n_s(t)$ are not uncorrelated, we see that τ may be chosen such that $R_{n_c n_s}(\tau) = 0$ for particular values of τ. ▲

FIGURE 5.13 Cross-correlation function of $n_c(t)$ and $n_s(t)$ for Example 5.11



Proof That Equality Holds in Equation (5.105) in the Sense of Zero Mean-Square Error

We now show that (5.105) holds. To simplify notation, let

$$\hat{n}(t) = n_c(t) \cos(\omega_0 t + \theta) - n_s(t) \sin(\omega_0 t + \theta) \tag{5.123}$$

where $\hat{n}(t)$ is *not* to be confused with the Hilbert transform. Thus we must show that

$$E\{[n(t) - \hat{n}(t)]^2\} = 0 \tag{5.124}$$

Expanding and taking the expectation term by term, we obtain

$$E\{(n - \hat{n})^2\} = \overline{n^2} - 2\overline{n\hat{n}} + \overline{\hat{n}^2} \tag{5.125}$$

where the arguments have been dropped to simplify notation.

Let us consider the last term in (5.125) first. By the definition of $\hat{n}(t)$,

$$\begin{aligned}
\overline{\hat{n}^2} &= E\{[n_c(t) \cos \omega_0 t + \theta) - n_s(t) \sin(\omega_0 t + \theta)]^2\} \\
&= \overline{n_c^2}\ \overline{\cos^2(\omega_0 t + \theta)} + \overline{n_s^2}\ \overline{\sin^2(\omega_0 t + \theta)} \\
&\quad - 2\overline{n_c n_s}\ \overline{\cos(\omega_0 t + \theta) \sin(\omega_0 t + \theta)} \\
&= \tfrac{1}{2}\overline{n_c^2} + \tfrac{1}{2}\overline{n_s^2} \\
&= \overline{n^2}
\end{aligned} \tag{5.126}$$

where we have employed (5.109b) along with the averages

$$\overline{\cos^2(\omega_0 t + \theta)} = \tfrac{1}{2} + \tfrac{1}{2}\overline{\cos 2(\omega_0 t + \theta)} = \tfrac{1}{2}$$
$$\overline{\sin^2(\omega_0 t + \theta)} = \tfrac{1}{2} - \tfrac{1}{2}\overline{\cos 2(\omega_0 t + \theta)} = \tfrac{1}{2}$$

and

$$\overline{\cos(\omega_0 t + \theta) \sin(\omega_0 t + \theta)} = \tfrac{1}{2}\overline{\sin 2(\omega_0 t + \theta)} = 0$$

Next, we consider $\overline{n\hat{n}}$. By definition of $\hat{n}(t)$, it can be written as

$$\overline{n\hat{n}} = E\{n(t)[n_c(t)\cos(\omega_0 t + \theta) - n_s(t)\sin(\omega_0 t + \theta)]\} \tag{5.127}$$

But, from Figure 5.11,

$$n_c(t) = h(t') * [2n(t')\cos(\omega_0 t' + \theta)] \tag{5.128a}$$

and

$$n_s(t) = h(t') * [2n(t')\sin(\omega_0 t' + \theta)] \tag{5.128b}$$

where $h(t')$ is the impulse response of the lowpass filter in Figure 5.11. The argument t' has been used in (5.128) to remind us that the variable of integration in the convolution is different from the variable t in (5.127). Substituting (5.128) into (5.127), we obtain

$$\begin{aligned}
\overline{n\hat{n}} &= E\{n(t)[h(t') * [2n(t')\cos(\omega_0 t' + \theta)]\cos(\omega_0 t + \theta) \\
&\quad + h(t') * [2n(t')\sin(\omega_0 t' + \theta)]\sin(\omega_0 t + \theta)]\} \\
&= E\{2n(t)h(t') * n(t')[\cos(\omega_0 t' + \theta)\cos(\omega_0 t + \theta) \\
&\quad + \sin(\omega_0 t' + \theta)\sin(\omega_0 t + \theta)]\} \\
&= E\{2n(t)h(t') * [n(t')\cos\omega_0(t - t')]\} \\
&= 2h(t') * [E\{n(t)n(t')\}\cos\omega_0(t - t')] \\
&= 2h(t') * [R_n(t - t')\cos\omega_0(t - t')] \\
&\triangleq 2\int_{-\infty}^{\infty} h(t - t')R_n(t - t')\cos\omega_0(t - t')\,dt'
\end{aligned} \tag{5.129}$$

Letting $u = t - t'$, we obtain

$$\overline{n\hat{n}} = 2\int_{-\infty}^{\infty} h(u)\cos(\omega_0 u)R_n(u)\,du \tag{5.130}$$

Now, a general case of Parseval's theorem is

$$\int_{-\infty}^{\infty} x(t)y(t)\,dt = \int_{-\infty}^{\infty} X(f)Y^*(f)\,df \tag{5.131}$$

where $x(t) \leftrightarrow X(f)$ and $y(t) \leftrightarrow Y(f)$. In (5.130) we note that

$$h(u)\cos\omega_0 u \leftrightarrow \tfrac{1}{2}H(f - f_0) + \tfrac{1}{2}H(f + f_0) \tag{5.132}$$

and

$$R_n(u) \leftrightarrow S_n(f) \tag{5.133}$$

Thus, by (5.131), we may write (5.130) as

$$\overline{n\hat{n}} = \int_{-\infty}^{\infty} [H(f - f_0) + H(f + f_0)]S_n(f)\, df \tag{5.134}$$

which follows because $S_n(f)$ is real. However, $S_n(f)$ is nonzero only where $H(f - f_0) + H(f + f_0) = 1$ because it was assumed narrowband. Thus (5.130) reduces to

$$\overline{n\hat{n}} = \int_{-\infty}^{\infty} S_n(f)\, df = \overline{n^2(t)} \tag{5.135}$$

Substituting (5.135) and (5.126) into (5.125), we obtain

$$E\{(n - \hat{n})^2\} = \overline{n^2} - 2\overline{n^2} + \overline{n^2} \equiv 0 \tag{5.136}$$

which shows that the mean-square error between $n(t)$ and $\hat{n}(t)$ is zero.

5.6 DISTRIBUTIONS ENCOUNTERED IN FM DEMODULATION OF SIGNALS IN GAUSSIAN NOISE

We now derive two probability distributions that arise in the consideration of demodulation of signals in Gaussian noise. These derivations make use of the quadrature-component and envelope-phase representations for noise described in the preceding section. They are the probability of a zero crossing of a bandlimited Gaussian process and the average rate of origin encirclement of a constant-amplitude sinusoid plus narrowband Gaussian noise.

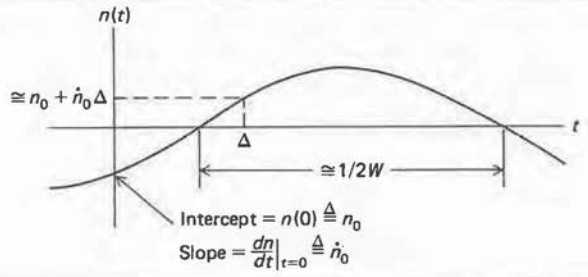
The Zero-Crossing Problem

Consider a sample function of a lowpass, zero-mean Gaussian process $n(t)$, as illustrated in Figure 5.14. Let its effective noise bandwidth be W, its power spectral density be $S_n(f)$, and its autocorrelation function be $R_n(\tau)$.

Consider the probability of a zero crossing in a small time interval Δ seconds in duration. For Δ sufficiently small, so that more than one zero crossing is unlikely, the probability $P_{\Delta-}$ of a minus-to-plus zero crossing in a time interval $\Delta \ll 1/(2W)$ is the probability that $n_0 < 0$ and $n_0 + \dot{n}_0\Delta > 0$. That is,

$$\begin{aligned} P_{\Delta-} &= P(n_0 < 0 \text{ and } n_0 + \dot{n}_0\Delta > 0) \\ &= P(n_0 < 0 \text{ and } n_0 > -\dot{n}_0\Delta, \text{ all } \dot{n}_0 \geq 0) \\ &= P(-\dot{n}_0\Delta < n_0 < 0, \text{ all } \dot{n}_0 \geq 0) \end{aligned} \tag{5.137}$$

FIGURE 5.14 Sample function of a lowpass Gaussian process of bandwidth W



This can be written in terms of the joint pdf of n_0 and $\dot{n}_0$, $f_{n_0\dot{n}_0}(y, z)$, as

$$P_{\Delta-} = \int_0^\infty \left[\int_{-z\Delta}^0 f_{n_0\dot{n}_0}(y, z)\, dy \right] dz \tag{5.138}$$

where y and z are running variables for n_0 and $\dot{n}_0$, respectively. Now $\dot{n}_0$ is a Gaussian random variable, since it involves a linear operation on $n(t)$, which is Gaussian by assumption. In Problem 5.39, it is shown that

$$E\{n_0\dot{n}_0\} = \frac{dR_n(\tau)}{d\tau}\bigg|_{\tau=0} \tag{5.139}$$

Thus, if the derivative of $R_n(\tau)$ exists at $\tau = 0$, it is zero because $R_n(\tau)$ must be even. It follows that

$$E\{n_0\dot{n}_0\} = 0 \tag{5.140}$$

Therefore, n_0 and $\dot{n}_0$, which are samples of $n_0(t)$ and $dn_0(t)/dt$, respectively, are statistically independent, since uncorrelated Gaussian processes are independent. Thus, letting var $\{n_0\} = \overline{n_0^2}$ and var $\{\dot{n}_0\} = \overline{\dot{n}_0^2}$, the joint pdf of n_0 and $\dot{n}_0$ is

$$f_{n_0\dot{n}_0}(y, z) = \frac{\exp\left(-y^2/2\overline{n_0^2}\right)}{\sqrt{2\pi\overline{n_0^2}}} \frac{\exp\left(-z^2/2\overline{\dot{n}_0^2}\right)}{\sqrt{2\pi\overline{\dot{n}_0^2}}} \tag{5.141}$$

which, when substituted into (5.138), yields

$$P_{\Delta-} = \int_0^\infty \frac{\exp\left(-z^2/2\overline{\dot{n}_0^2}\right)}{\sqrt{2\pi\overline{\dot{n}_0^2}}} \left[\int_{-z\Delta}^0 \frac{\exp\left(-y^2/2\overline{n_0^2}\right)}{\sqrt{2\pi\overline{n_0^2}}}\, dy \right] dz \tag{5.142}$$

For Δ small, the inner integral of (5.142) can be approximated as $z\Delta/\sqrt{2\pi\overline{n_0^2}}$, which allows (5.142) to be simplified to

$$P_{\Delta-} \cong \frac{\Delta}{\sqrt{2\pi\overline{n_0^2}}} \int_0^\infty z \frac{\exp\left(-z^2/2\overline{\dot{n}_0^2}\right)}{\sqrt{2\pi\overline{\dot{n}_0^2}}}\, dz \tag{5.143}$$

Letting $\zeta = z^2/2\overline{\dot{n}_0^2}$ yields

$$\begin{aligned} P_{\Delta-} &\cong \frac{\Delta}{2\pi\sqrt{\overline{n_0^2}\,\overline{\dot{n}_0^2}}} \int_0^\infty \overline{\dot{n}_0^2} e^{-\zeta}\, d\zeta \\ &= \frac{\Delta}{2\pi} \sqrt{\frac{\overline{\dot{n}_0^2}}{\overline{n_0^2}}} \end{aligned} \tag{5.144}$$

for the probability of a minus-to-plus zero crossing in Δ seconds. By symmetry, the probability of a plus-to-minus zero crossing is the same. Thus, the probability of zero crossing in Δ seconds is

$$P_\Delta \cong \frac{\Delta}{\pi} \sqrt{\frac{\overline{\dot{n}_0^2}}{\overline{n_0^2}}} \tag{5.145}$$

For example, suppose that $n(t)$ is an ideal lowpass process with the power spectral density

$$S_n(f) = \begin{cases} \frac{1}{2}N_0, & |f| \leq W \\ 0, & \text{otherwise} \end{cases} \tag{5.146}$$

Thus

$$R_n(\tau) = N_0 W \operatorname{sinc} 2W\tau \tag{5.147}$$

which possesses a derivative at zero. Therefore, n_0 and $\dot{n}_0$ are independent. It follows that

$$\overline{n_0^2} = \operatorname{var}\{n_0\} = \int_{-\infty}^{\infty} S_n(f)\, df = R_n(0) = N_0 W \tag{5.148}$$

and, since the transfer function of a differentiator is $H_d(f) = j2\pi f$, that

$$\begin{aligned} \overline{\dot{n}_0^2} = \operatorname{var}\{\dot{n}_0\} &= \int_{-\infty}^{\infty} |H_d(f)|^2 S_n(f)\, df = \int_{-W}^{W} (2\pi f)^2 \tfrac{1}{2}N_0\, df \\ &= \tfrac{1}{3}(2\pi W)^2 (N_0 W) \end{aligned} \tag{5.149}$$

Substitution of these results into (5.145) gives

$$2P_{\Delta-} = 2P_{\Delta+} = P_{\Delta} = \frac{\Delta}{\pi}\frac{2\pi W}{\sqrt{3}} = \frac{2W\Delta}{\sqrt{3}} \tag{5.150}$$

for the probability of a zero crossing in a small time interval Δ seconds in duration for a random process with an ideal rectangular lowpass spectrum.

Average Rate of Origin Encirclement for Sinusoid Plus Narrowband Gaussian Noise

Consider next the sum of a sinusoid plus narrowband Gaussian noise:

$$\begin{aligned} z(t) &= A\cos\omega_0 t + n(t) \\ &= A\cos\omega_0 t + n_c(t)\cos\omega_0 t - n_s(t)\sin\omega_0 t \end{aligned} \tag{5.151}$$

where $n_c(t)$ and $n_s(t)$ are lowpass processes with statistical properties as described in Section 5.5. We may write $z(t)$ in terms of envelope $R(t)$ and phase $\theta(t)$ as

$$z(t) = R(t)\cos[\omega_0 t + \theta(t)] \tag{5.152}$$

where

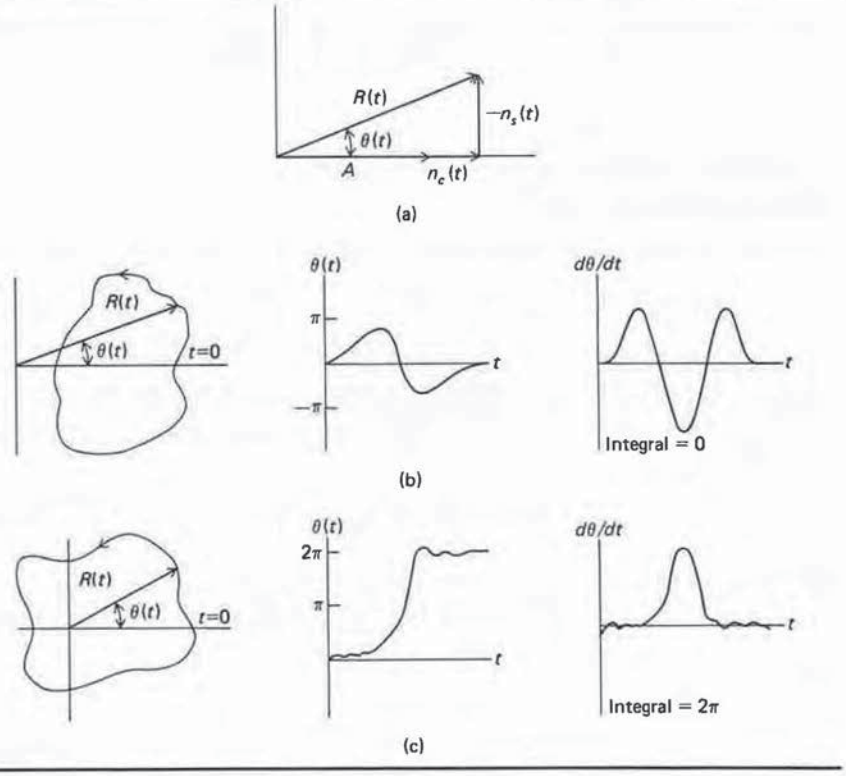
$$R(t) = \sqrt{[A + n_c(t)]^2 + n_s^2(t)} \tag{5.153}$$

and

$$\theta(t) = \tan^{-1}\left[\frac{n_s(t)}{A + n_c(t)}\right] \tag{5.154}$$

A phasor representation for this process is shown in Figure 5.15(a). In Figure 5.15(b), a possible trajectory for the tip of $R(t)$ that does not encircle the origin is shown along with $\theta(t)$ and $d\theta(t)/dt$. In Figure 5.15(c), a trajectory that encircles the origin is shown along with $\theta(t)$ and $d\theta(t)/dt$ for this case. For this latter case, the area under $d\theta/dt$ must be 2π radians. Recalling the definition of an ideal FM discriminator in Chapter 3, we see that the sketches for $d\theta/dt$ shown in Figure 5.15 represent the output of a discriminator in response to input of an unmodulated signal plus noise. For a high signal-to-noise ratio, the phasor will randomly fluctuate near the horizontal axis. Occasionally, however, it will encircle the origin as shown in Figure 5.15(c). Intuitively, these encirclements become more probable as the signal-to-noise ratio decreases. Because of its nonzero area, the impulsive type of output illustrated in Figure 5.15(c), caused by an encirclement of the origin, has a much more serious effect on the noise level of the discriminator output

FIGURE 5.15 Phasor diagrams showing possible trajectories for a sinusoid plus Gaussian noise. (a) Phasor representation of a sinusoid plus narrowband noise. (b) Trajectory that does not encircle origin. (c) Trajectory that does encircle origin.



than does the noise excursion illustrated in Figure 5.15(b), which has zero area. We now derive an expression for the average number of noise spikes per second of the type illustrated in Figure 5.15(c). Only positive spikes, caused by counterclockwise origin encirclements, will be considered, since the average rate for negative spikes, which result from clockwise origin encirclements, is the same by symmetry.

Assume that if $R(t)$ crosses the horizontal axis when it is in the second quadrant, the origin encirclement will be completed. With this assumption, and considering a small interval Δ seconds in duration, the probability of a counterclockwise encirclement $P_{cc\Delta}$ in the interval $(0, \Delta)$ is

$$\begin{aligned} P_{cc\Delta} &= P[A + n_c(t) < 0 \text{ and } n_s(t) \text{ makes } + \text{ to } - \text{ zero crossing in } (0, \Delta)] \\ &= P[n_c(t) < -A]P_{\Delta-} \end{aligned} \tag{5.155}$$

where $P_{\Delta-}$ is the probability of a minus-to-plus zero crossing in $(0, \Delta)$ as given by (5.150) with $n(t)$ replaced by $n_s(t)$, and the statistical independence of $n_c(t)$ and $n_s(t)$ has been used. Now by (5.109b), $\overline{n_c^2(t)} = \overline{n_s^2(t)} = \overline{n^2(t)}$. If $n(t)$ is an ideal bandpass process with single-sided bandwidth B and power spectral density N_0, then $\overline{n^2(t)} = N_0B$, and

$$P[n_c(t) < -A] = \int_{-\infty}^{-A} \frac{e^{-n_c^2/2N_0B}}{\sqrt{2\pi N_0 B}}\, dn_c$$
$$= Q(\sqrt{A^2/N_0B}) \tag{5.156}$$

where $Q(\cdot)$ is the Q-function defined by (4.185). From (5.150) with $W = \frac{1}{2}B$, which is the bandwidth of $n_s(t)$, we have

$$P_{\Delta-} = \frac{\Delta B}{2\sqrt{3}} \tag{5.157}$$

Substituting (5.156) and (5.157) into (5.155), we obtain

$$P_{cc\Delta} = \frac{\Delta B}{2\sqrt{3}} Q\left(\sqrt{\frac{A^2}{N_0B}}\right) \tag{5.158}$$

The probability of a clockwise encirclement $P_{c\Delta}$ is the same by symmetry. Thus the expected number of encirclements per second, clockwise and counterclockwise, is

$$\nu = \frac{1}{\Delta}(P_{c\Delta} + P_{cc\Delta})$$
$$= \frac{B}{\sqrt{3}} Q\left(\sqrt{\frac{A^2}{N_0B}}\right) \tag{5.159}$$

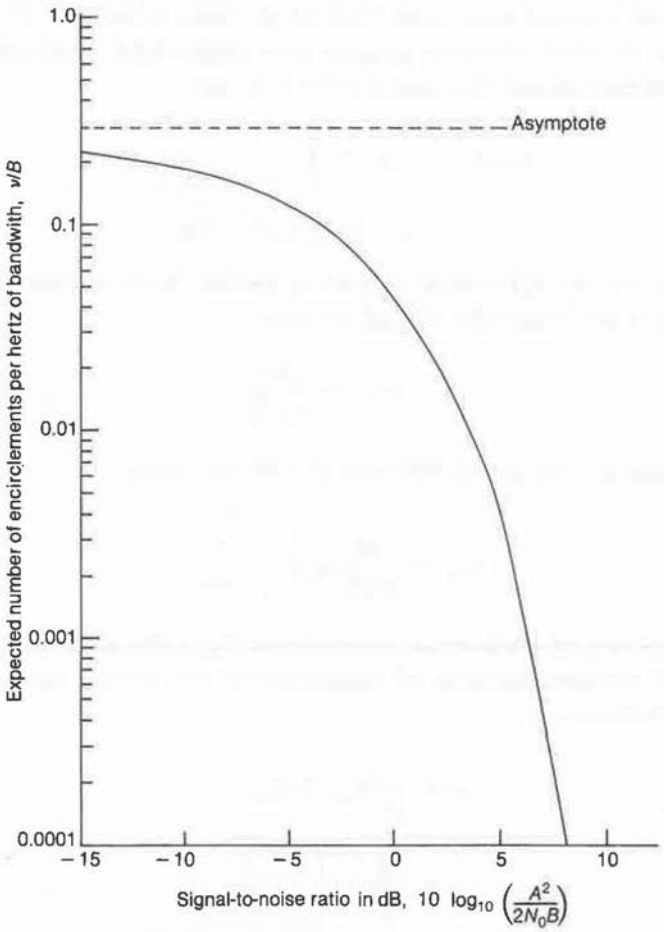
We note that the average number of encirclements per second increases in direct proportion to the bandwidth and decreases essentially exponentially with an increasing signal-to-noise ratio $A^2/2N_0B$.

We can see this in Figure 5.16, which illustrates ν/B as a function of signal-to-noise ratio. Figure 5.16 also shows the asymptote $\nu/B = 1/2\sqrt{3} = 0.2887$.

The results derived above say nothing about the statistics of the number of impulses, N, in a time interval, T. In Problem 5.36, however, it is shown that the power spectral density of periodic impulse noise processes is given by

$$S_I(f) = \nu\overline{a^2} \tag{5.160}$$

FIGURE 5.16 Rate of origin encirclements as a function of signal-to-noise ratio



where ν is the average number of impulses per second ($\nu = f_s$ for a periodic impulse train) and $\overline{a^2}$ is the mean-squared value of the impulse weights a_k. A similar result can be shown for impulses which have exponentially distributed intervals between them (i.e., Poisson impulse noise). Approximating the impulse portion of $d\theta/dt$ as a Poisson impulse noise process with sample functions of the form

$$x(t) \triangleq \frac{d\theta(t)}{dt}\bigg|_{\text{impulse}} = \sum_{k=-\infty}^{\infty} \pm 2\pi\delta(t - t_k) \tag{5.161}$$

where t_k is a Poisson point process with average rate ν given by (5.159), we may approximate the power spectral density of this impulse noise process as white with a spectral level given by

$$S_x(f) = \nu(2\pi)^2$$

$$= \frac{4\pi^2 B}{\sqrt{3}} Q\left(\sqrt{\frac{A^2}{N_0 B}}\right), \qquad -\infty < f < \infty \tag{5.162}$$

If the sinusoidal signal component in (5.151) is FM modulated, the average number of impulses per second is increased over that obtained for no modulation. Intuitively, the reason may be explained as follows. Consider a carrier that is FM modulated by a unit step. Thus

$$z(t) = A \cos 2\pi[f_c + f_d u(t)]t + n(t) \tag{5.163}$$

where $f_d \leq \frac{1}{2}B$ is the frequency-deviation constant in hertz per volt. Because of this frequency step, the carrier phasor shown in Figure 5.15(a) rotates counterclockwise at f_d Hz for $t > 0$. Since the noise is bandlimited to B Hz with center frequency f_c Hz, its average frequency is *less* than the instantaneous frequency of the modulated carrier when $t > 0$. Hence there will be a greater probability for a 2π clockwise rotation of $R(t)$ relative to the carrier phasor if it is frequency offset by f_d Hz (that is, modulated) than if it is not. In other words, the average rate for negative spikes will increase for $t > 0$ and that for positive spikes will decrease. Conversely, for a negative frequency step, the average rate for positive spikes will increase and that for negative spikes will decrease. It can be shown that the result is a net increase $\delta\nu$ in the spike rate over the case for no modulation, with the average increase approximated by (see Problems 5.34 and 5.35)

$$\overline{\delta\nu} = \overline{|\delta f|} \exp\left(\frac{-A^2}{2N_0 B}\right) \tag{5.164}$$

where $\overline{|\delta f|}$ is the average of the magnitude of the frequency deviation. For the case just considered, $\overline{|\delta f|} = f_d$. The total average spike rate is then $\nu + \overline{\delta\nu}$. The power spectral density of the spike noise for modulated signals is obtained by substituting $\nu + \overline{\delta\nu}$ for ν in (5.162).

SUMMARY

1. A random process is completely described by the N-fold joint pdf of its amplitudes at the arbitrary times $t_1, t_2, \ldots, t_N$. If this pdf is invariant under a shift of the time origin, the process is said to be *statistically stationary in the strict sense.*

2. The autocorrelation function of a random process, computed as a statistical average, is defined as

$$R(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_{X_1 X_2}(x_1, t_1; x_2, t_2)\, dx_1\, dx_2$$

where $f_{X_1X_2}$ is the joint amplitude pdf of the process at times t_1 and t_2. *If the process is stationary,*

$$R(t_1, t_2) = R(t_2 - t_1) = R(\tau)$$

where $\tau \triangleq t_2 - t_1$.

3. A process whose statistical average mean and variance are time-independent and whose autocorrelation function is a function only of $t_2 - t_1 = \tau$ is termed *wide-sense stationary. Strict-sense stationary processes are also wide-sense stationary. The converse is true only for special cases, one of which being Gaussian processes.*

4. A process for which statistical averages and time averages are equal is called *ergodic. Ergodicity implies stationarity, but the reverse is not necessarily true.*

5. The Wiener-Khinchine theorem states that the autocorrelation function and the power spectral density of a stationary random process are a Fourier transform pair.

   An expression for the power spectral density of a random process that is often useful is

$$S_n(f) = \lim_{T \to \infty} \frac{1}{T} E\{|\mathcal{F}[n_T(t)]|^2\}$$

where $n_T(t)$ is a sample function truncated to T seconds, centered about $t = 0$.

6. The autocorrelation function of a random process is a real, even function of the delay variable τ with an absolute maximum at $\tau = 0$. It is periodic for periodic random processes, and its Fourier transform is nonnegative for all frequencies. As $\tau \to \pm\infty$, the autocorrelation function approaches the square of the mean of the random process unless the random process is periodic. $R(0)$ gives the total average power in a process.

7. White noise has a constant power spectral density $\frac{1}{2}N_0$ for all f. Its autocorrelation function is $\frac{1}{2}N_0\delta(\tau)$. For this reason, it is sometimes called delta-correlated noise. It has infinite power and is therefore a mathematical idealization. However, it is, nevertheless, a useful approximation in many cases.

8. The cross-correlation function of two stationary random processes $X(t)$ and $Y(t)$ is defined as

$$R_{XY}(\tau) = E\{X(t)Y(t+\tau)\}$$

Their cross-power spectral density is

$$S_{XY}(f) = \mathcal{F}[R_{XY}(\tau)]$$

They are said to be *orthogonal* if $R_{XY}(\tau) = 0$ for all τ.

9. Consider a linear system with the impulse response $h(t)$ and the transfer function $H(f)$ with random input $x(t)$ and output $y(t)$. Then

$$S_Y(f) = |H(f)|^2 S_X(f)$$

$$R_Y(\tau) = \mathcal{F}^{-1}[S_Y(f)] = \int_{-\infty}^{\infty} |H(f)|^2 S_X(f) e^{j2\pi f\tau}\, df$$

$$R_{XY}(\tau) = h(\tau) * R_X(\tau)$$

$$S_{XY}(f) = H(f) S_X(f)$$

$$R_{YX}(\tau) = h(-\tau) * R_X(\tau)$$

$$S_{YX}(f) = H^*(f) S_X(f)$$

where $S(f)$ denotes the spectral density and $R(\tau)$ denotes the correlation function.

10. The output of a linear system with Gaussian input is Gaussian.

11. The noise-equivalent bandwidth of a linear system with a transfer function $H(f)$ is defined as

$$B_N = \frac{1}{H_0^2} \int_0^{\infty} |H(f)|^2\, df$$

where H_0 = maximum of $|H(f)|$. If the input is white noise with the single-sided power spectral density N_0, the output power is

$$P_0 = H_0^2 N_0 B_N$$

An equivalent expression for the noise-equivalent bandwidth written in terms of the impulse response of the filter is

$$B_N = \frac{\displaystyle\int_{-\infty}^{\infty} |h(t)|^2\, dt}{2\left[\displaystyle\int_{-\infty}^{\infty} h(t)\, dt\right]^2}$$

12. The quadrature-component representation of a bandlimited random process $n(t)$ is

$$n(t) = n_c(t) \cos(\omega_0 t + \theta) - n_s(t) \sin(\omega_0 t + \theta)$$

where θ is an arbitrary phase angle. The envelope-phase representation is

$$n(t) = R(t) \cos(\omega_0 t + \phi(t) + \theta)$$

where $R^2(t) = n_c^2(t) + n_s^2(t)$ and $\tan[\phi(t)] = n_s(t)/n_c(t)$. If the process is narrowband, n_c, n_s, R, and ϕ vary slowly with respect to $\cos \omega_0 t$ and $\sin \omega_0 t$. If the power spectral density of $n(t)$ is $S_n(f)$, the power spectral densities of $n_c(t)$ and $n_s(t)$ are

$$S_{n_c}(f) = S_{n_s}(f) = L_p[S_n(f - f_0) + S_n(f + f_0)]$$

where $L_p[\ ]$ denotes the low-frequency part of the quantity in the brackets. If $L_p[S_n(f + f_0) - S_n(f - f_0)] = 0$, then $n_c(t)$ and $n_s(t)$ are orthogonal. The average powers of $n_c(t)$, $n_s(t)$, and $n(t)$ are equal. The processes $n_c(t)$ and $n_s(t)$ are given by

$$n_c(t) = L_p[2n(t) \cos(\omega_0 t + \theta)]$$

and

$$n_s(t) = -L_p[2n(t) \sin(\omega_0 t + \theta)]$$

Since these operations are linear, $n_c(t)$ and $n_s(t)$ will be Gaussian if $n(t)$ is Gaussian. Thus, $n_c(t)$ and $n_s(t)$ are independent if $n(t)$ is zero-mean Gaussian with a power spectral density that is symmetrical about $f = f_0$ for $f > 0$.

FURTHER READING

The references given in Chapter 4 also provide further reading on the subject matter of this chapter.

PROBLEMS

Section 5.1

5.1 A fair die is thrown. Depending on the number of spots on the up face, the following random processes are generated. Sketch several examples of sample functions for each case.

$$\text{(a) } X(t, \zeta) = \begin{cases} A, & \text{1 or 2 spots up} \\ 0, & \text{3 or 4 spots up} \\ -A, & \text{5 or 6 spots up} \end{cases}$$

$$\text{(b) } X(t, \zeta) = \begin{cases} 5A, & \text{1 spot up} \\ 3A & \text{2 spots up} \\ A, & \text{3 spots up} \\ -A, & \text{4 spots up} \\ -3A, & \text{5 spots up} \\ -5A, & \text{6 spots up} \end{cases}$$

$$\text{(c) } X(t, \zeta) = \begin{cases} 4A, & \text{1 spot up} \\ 2A, & \text{2 spots up} \\ At, & \text{3 spots up} \\ -At, & \text{4 spots up} \\ -2A, & \text{5 spots up} \\ -4A, & \text{6 spots up} \end{cases}$$

Section 5.2

5.2 Referring to Problem 5.1, what are the following probabilities for each case?

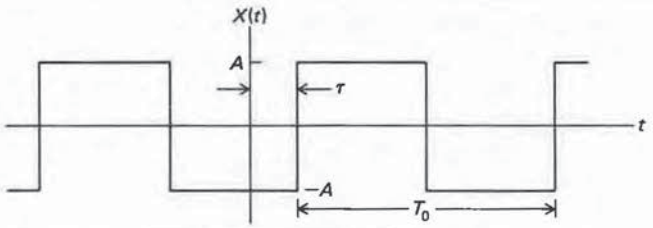
(a) $F_X(X \leq 2A, t = 4)$
(b) $F_X(X \leq 0, t = 4)$
(c) $F_X(X \leq 2A, t = 2)$

5.3 A random process is composed of sample functions that are square-waves, each with constant amplitude A, period T_0, and random delay τ as sketched in Figure 5.17. The pdf of τ is

$$f(\tau) = \begin{cases} 1/T_0, & |\tau| \leq T_0/2 \\ 0, & \text{otherwise} \end{cases}$$

(a) Sketch several typical sample functions.
(b) Write the first-order pdf for this random process at some arbitrary time t_0. (*Hint:* Because of the random delay τ, the pdf is independent of t_0. Also, it might be easier to deduce the cdf and differentiate it to get the pdf.)

FIGURE 5.17



5.4 Let the sample functions of a random process be given by

$$X(t) = A \cos \omega_0 t$$

where ω_0 is fixed and A has the pdf

$$f_A(a) = \frac{e^{-a^2/2\sigma_a^2}}{\sqrt{2\pi}\sigma_a}$$

This random process is passed through an ideal integrator to give a random process $Y(t)$.

(a) Find an expression for the sample functions of the output process $Y(t)$.
(b) Write down an expression for the pdf of $Y(t)$ at time t_0. (*Hint:* Note that $\sin \omega_0 t_0$ is just a constant.)
(c) Is $Y(t)$ stationary? Is it ergodic?

5.5 (a) Find the time-average mean and the autocorrelation function for the random process of Problem 5.3.
(b) Find the ensemble-average mean and the autocorrelation function.
(c) Is this process wide-sense stationary? Why?

5.6 Consider the random process of Example 5.1 with the pdf of θ given by

$$p(\theta) = \begin{cases} 4/\pi, & \pi/4 \leq \theta \leq \pi/2 \\ 0, & \text{otherwise} \end{cases}$$

(a) Find the statistical-average and time-average mean and variance.
(b) Find the statistical-average and time-average autocorrelation functions.
(c) Is this process ergodic?

5.7 Let a random process be given as

$$Z(t) = X(t) \cos(\omega_0 t + \theta)$$

where $X(t)$ is a stationary random process with $E\{X(t)\} = 0$ and $E\{X^2(t)\} = \sigma_X^2$.

(a) If $\theta = 0$, find $E\{Z(t)\}$ and $E\{Z^2\}$. Is $Z(t)$ stationary?

(b) If θ is a random variable indpendent of $X(t)$ and uniformly distributed in the interval $(-\pi, \pi)$, show that

$$E\{Z(t)\} = 0 \qquad \text{and} \qquad E\{Z^2(t)\} = \sigma_X^2/2$$

(c) Is $Z(t)$ stationary?

5.8 The voltage of the output of a noise generator whose statistics are known to be closely Gaussian and stationary is measured with a dc voltmeter and a true root-mean-square (rms) voltmeter that is ac coupled. The dc meter reads 5 V, and the true rms meter reads 10 V. Write down an expression for the first-order pdf of the voltage at any time $t = t_0$. Sketch and dimension the pdf.

Section 5.3

5.9 Which of the following functions are suitable autocorrelation functions? Tell why or why not. (ω_0, τ_0, τ_1, A, B, and C are positive constants.)

(a) $A \cos \omega_0 \tau$

(b) $A\Lambda(\tau/\tau_0)$, where $\Lambda(x)$ is the unit-area triangular function defined in Chapter 2

(c) $A\Pi(\tau/\tau_0)$, where $\Pi(x)$ is the unit-area pulse function defined in Chapter 2

(d) $B\Pi(\tau/\tau_1) + C\Pi(\tau/2\tau_1)$

(e) $A \sin \omega_0 \tau$

5.10 A bandlimited white-noise process has a double-sided power spectral density of 2×10^{-10} W/Hz in the frequency range $|f| \leq 1$ MHz. Find its autocorrelation function and sketch and fully dimension it.

5.11 Consider a random binary pulse waveform as analyzed in Example 5.6, but with half-cosine pulses given by $p(t) = \cos(2\pi t/2T)\ \Pi(t/T)$. Obtain and sketch the autocorrelation function for the two cases considered in Example 5.6, namely,

(a) $a_k = \pm A$ for all k, where A is a constant, with $R_m = A^2$, $m = 0$, and $R_m = 0$ otherwise.

(b) $a_k = A_k + A_{k-1}$ with $A_k = \pm A$ and $E\{A_k A_{k+m}\} = A^2$, $m = 0$, and zero otherwise.

(c) Find and sketch the power spectral density for each preceding case.

5.12 Two random processes are given by

$$X(t) = n(t) + A \cos(\omega_0 t + \theta)$$

and

$$Y(t) = n(t) + A \sin(\omega_0 t + \theta)$$

where A and ω_0 are constants and θ is a random variable uniformly distributed in the interval $(-\pi, \pi)$. The first term, $n(t)$, represents a stationary random noise process with autocorrelation function $R_n(\tau) = B\Lambda(\tau/\tau_0)$, where B and τ_0 are nonnegative constants. Find and sketch the cross-correlation function of these two random processes.

5.13 Given two independent, wide-sense stationary random processes $X(t)$ and $Y(t)$ with autocorrelation functions $R_X(\tau)$ and $R_Y(\tau)$, respectively.

(a) Show that the autocorrelation function $R_Z(\tau)$ of their product $Z(t) = X(t)Y(t)$ is given by

$$R_Z(\tau) = R_X(\tau)R_Y(\tau)$$

(b) Express the power spectral density of $Z(t)$ in terms of the power spectral densities of $X(t)$ and $Y(t)$, denoted as $S_X(f)$ and $S_Y(f)$, respectively.

(c) Let $X(t)$ be a bandlimited stationary noise process with power spectral density $S_X(f) = 10\Pi(f/200)$, and let $Y(t)$ be the process defined by sample functions of the form

$$Y(t) = 5 \cos(50\pi t + \theta)$$

where θ is a uniformly distributed random variable in the interval $(0, 2\pi)$. Using the results derived in parts (a) and (b), obtain the autocorrelation function and power spectral density of $Z(t) = X(t)Y(t)$.

5.14 A random signal has the autocorrelation function

$$R(\tau) = 5 + 2\Lambda(\tau/10)$$

where $\Lambda(x)$ is the unit-area triangular function defined in Chapter 2. Determine the following:

(a) The ac power.

(b) The dc power.

(c) The total power.
(d) The power spectral density. Sketch it and label carefully.

5.15 A random process is defined as $Y(t) = X(t) + X(t - T)$, where $X(t)$ is a wide-sense stationary random process with autocorrelation function $R_X(\tau)$ and power spectral density $S_X(f)$.

(a) Show that $R_Y(\tau) = 2R_X(\tau) + R_X(\tau + T) + R_X(\tau - T)$.
(b) Show that $S_Y(f) = 4S_X(f) \cos^2(\pi fT)$.
(c) If $X(t)$ has autocorrelation function $R_X(\tau) = 5\Lambda(\tau)$, where $\Lambda(\tau)$ is the unit-area triangular function, and $T = 0.5$, find and sketch the power spectral density of $Y(t)$ as defined in the problem statement.

5.16 The power spectral density of a wide-sense stationary random process is given by

$$S_X(f) = 10\delta(\tau) + 25 \operatorname{sinc}^2(5f)$$

(a) Sketch and fully dimension this power spectral density function.
(b) Find the power in the dc component of the random process.
(c) Find the total power.
(d) Given that the area under the main lobe of the sinc-squared function is approximately 0.9 of the total area, which is unity if it has unity amplitude, find the fraction of the total power contained in this process for frequencies between 0 and 0.2 Hz.

5.17 Given the following autocorrelation functions $R_X(\tau)$ and power spectral densities $S_X(f)$ (α, K, K_1, K_2, and b are all positive constants):

$$R_X(\tau) = K \exp(-\alpha|\tau|) \cos 2\pi b\tau \tag{1}$$

$$R_X(\tau) = K_1[\exp(-\alpha|\tau|) + 1] + K_2 \cos 2\pi b\tau \tag{2}$$

$$S_X(f) = K \exp(-\alpha f^2) + K_1\delta(f - 10) + K_1\delta(f + 10) + K_2\delta(\tau) \tag{3}$$

(a) Sketch each function and fully dimension.
(b) Determine the value of the dc power, if any, for each one.
(c) Determine the total power for each.
(d) Determine the frequency of the periodic component, if any, for each.

Section 5.4

5.18 A stationary random process $n(t)$ has a power spectral density of 10^{-6} W/Hz. It is passed through an ideal lowpass filter with transfer func-

tion $H(f) = \Pi(f/1\text{ MHz})$, where $\Pi(x)$ is the unit-area pulse function defined in Chapter 2.

(a) What is the power spectral density of the output? Sketch.
(b) Obtain the autocorrelation function of the output. Sketch and label carefully.
(c) What is the power of the output process? Find it two different ways.

5.19 An ideal finite-time integrator is characterized by the input-output relationship

$$Y(t) = \frac{1}{T}\int_{t-T}^{t} X(\alpha)\, d\alpha$$

(a) What is its impulse response?
(b) What is its transfer function?
(c) The input is white noise with two-sided power spectral density $N_0/2$. Find the power spectral density of the output of the filter.
(d) Show that the autocorrelation function of the output is

$$R_0(\tau) = \frac{N_0}{2T}\Lambda(\tau/T)$$

where $\Lambda(x)$ is the unit-area triangular function defined in Chapter 2.
(e) What is the equivalent noise bandwidth of the integrator?
(f) Show that the result for the output noise power obtained using the equivalent noise bandwidth found in part (e) coincides with the result found from the autocorrelation function of the output found in part (d).

5.20 White noise with two-sided power spectral density $N_0/2$ drives a second-order Butterworth filter with transfer function magnitude

$$|H_{2bu}(f)| = \frac{1}{\sqrt{1 + (f/f_3)^4}}$$

(a) What is the power spectral density of the output?
(b) Show that the autocorrelation function of the output is

$$R_0(\tau) = \frac{\pi f_3 N_0}{2}\exp\left(-\sqrt{2}\pi f_3|\tau|\right)\cos\left[\sqrt{2}\pi f_3|\tau| - \pi/4\right]$$

Plot as a function of $f_3\tau$. (*Hint:* Use the integral.)

$$\int_0^\infty \frac{\cos(ax)}{b^4 + x^4}\,dx = \frac{\sqrt{2}\pi}{4b^3}\exp(-ab/\sqrt{2})[\cos(ab/\sqrt{2}) + \sin(ab/\sqrt{2})], \qquad a, b > 0$$

(c) Does the output power obtained by taking the proper limit of $R_0(\tau)$ check with that calculated using the equivalent noise bandwidth for a Butterworth filter as given by (5.101)?

5.21 A power spectral density given by

$$S_Y(f) = \frac{(2\pi f)^2}{(2\pi f)^4 + 10{,}000}$$

is desired. A white-noise source of two-sided power spectral density 1 W/Hz is available. What is the transfer function of the filter to be placed at the noise-source output to produce the desired power spectral density?

5.22 Obtain the autocorrelation functions and power spectral densities of the outputs of the following systems with the input autocorrelation functions or power spectral densities given.

(a) $H(f) = \Pi(f/2B)$

$R_X(\tau) = \dfrac{N_0}{2}\delta(\tau)$

N_0 and B are positive constants.

(b) $h(t) = A\exp(-\alpha t)\,u(t)$

$S_X(f) = \dfrac{B}{1 + (2\pi\beta f)^2}$

A, α, B, and β are positive constants.

5.23 The input to a lowpass filter with impulse response

$$h(t) = \exp(-20\pi t)\,u(t)$$

is white, Gaussian noise with single-sided power spectral density of 2 W/Hz. Obtain the following:

(a) The mean of the output.
(b) The power spectral density of the output.
(c) The autocorrelation function of the output.
(d) The probability density function of the output at an arbitrary time t_1.

(e) The joint probability density function of the output at time instants t_1 and $t_1 + 0.03$ s.

5.24 Find the noise-equivalent bandwidths for the following first- and second-order lowpass filters in terms of their 3-dB bandwidths. Refer to Chapter 2 to determine their transfer function magnitudes.

(a) Chebyshev
(b) Bessel

5.25 A second-order Butterworth filter has 3-dB bandwidth of 500 Hz. Determine the unit impulse response of the filter and use it to compute the noise-equivalent bandwidth of the filter. Check your result against the appropriate special case of Example 5.9.

5.26 Determine the noise-equivalent bandwidths for the two filters having transfer functions given below:

(a) $H_a(f) = \Pi(f/6) + \Pi(f/2)$
(b) $H_b(f) = 2\Lambda(f/100)$

5.27 A filter has transfer function

$$H(f) = H_0(f - 500) + H_0(f + 500)$$

where

$$H_0(f) = 2\Lambda(f/100)$$

Find its noise-equivalent bandwidth.

5.28 Determine the noise-equivalent bandwidths of the systems having the following transfer functions. (*Hint:* Use the time domain approach.)

(a) $H_a(f) = \dfrac{10}{(j2\pi f + 1)(j2\pi f + 20)}$

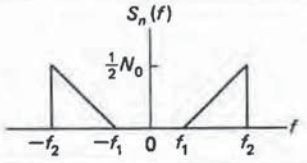
(b) $H_b(f) = \dfrac{25}{(j2\pi f + 5)^2}$

Section 5.5

5.29 Noise $n(t)$ has the power spectral density shown in Figure 5.18. We write

$$n(t) = n_c(t) \cos(2\pi f_0 t + \theta) - n_s(t) \sin(2\pi f_0 t + \theta)$$

FIGURE 5.18



Make plots of the power spectral densities of $n_c(t)$ and $n_s(t)$ for the following cases:

(a) $f_0 = f_1$ (b) $f_0 = f_2$ (c) $f_0 = \frac{1}{2}(f_2 + f_1)$

(d) For which of these cases are $n_c(t)$ and $n_s(t)$ uncorrelated?

5.30 (a) If $S_n(f) = \alpha^2/(\alpha^2 + \omega^2)$, show that $R_n(\tau) = Ke^{-\alpha|\tau|}$. Find K.

(b) Find $R_n(\tau)$ if

$$S_n(f) = \frac{\frac{1}{2}\alpha^2}{\alpha^2 + (\omega - \omega_0)^2} + \frac{\frac{1}{2}\alpha^2}{\alpha^2 + (\omega + \omega_0)^2}, \qquad \begin{aligned} \omega &= 2\pi f \\ \omega_0 &= 2\pi f_0 \end{aligned}$$

(c) If $n(t) = n_c(t)\cos(2\pi f_0 t f\theta) - n_s(t)\sin(2\pi f_0 t f\theta)$ find $S_{n_c}(f)$, and $S_{n_c n_s}(f)$, where $S_n(f)$ is as given in part (b). Sketch.

5.31 The double-sided power spectral density of noise $n(t)$ is shown in Figure 5.19. If $n(t) = n_c(t)\cos(2\pi f_0 t + \theta) - n_s(t)\sin(2\pi f_0 t + \theta)$, find and plot $S_{n_c}(f)$, $S_{n_s}(f)$, and $S_{n_c n_s}(f)$ for the following cases:

(a) $f_0 = \frac{1}{2}(f_1 + f_2)$ (b) $f_0 = f_1$ (c) $f_0 = f_2$

(d) Find $R_{n_c n_s}(\tau)$ for each case for which $S_{n_c n_s}(f)$ is not zero. Plot.

5.32 A noise waveform $n_1(t)$ has the bandlimited power spectral density shown in Figure 5.20. Find and plot the power spectral density of $n_2(t) = n_1(t)\cos(\omega_0 t + \theta) - n_1(t)\sin(\omega_0 t + \theta)$, where θ is a uniformly distributed random variable in $(0, 2\pi)$.

FIGURE 5.19

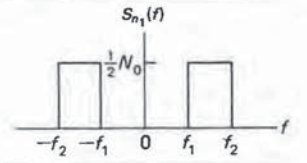
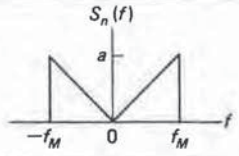


FIGURE 5.20



5.33 (Rice-Nakagami pdf) Consider the sum of a constant-amplitude sinusoid and narrowband Gaussian noise:

$$y(t) = [A + n_c(t)] \cos(\omega_0 t + \theta) - n_s(t) \sin(\omega_0 t + \theta)$$

Let the variance of n_c and n_s be σ^2. Show that the pdf of

$$r = \sqrt{(A + n_c)^2 + n_s^2}$$

is

$$f_R(r) = \left(\frac{r}{\sigma^2}\right) \exp\left[-\frac{(r^2 + A^2)}{2\sigma^2}\right] I_0\left(\frac{rA}{\sigma^2}\right), \qquad r \geq 0$$

where $I_0(x) \triangleq (2\pi)^{-1} \int_0^{2\pi} e^{x\cos\theta}\, d\theta$ is the modified Bessel function of the first kind and of zero order.

Section 5.6

5.34 Consider a signal-plus-noise process of the form

$$z(t) = A \cos 2\pi(f_0 + f_d)t + n(t)$$

where $\omega_0 = 2\pi f_0$, with

$$n(t) = n_c(t) \cos \omega_0 t - n_s(t) \sin \omega_0 t$$

an ideal bandlimited white-noise process with double-sided power spectral density equal to $\frac{1}{2}N_0$, for $-\frac{1}{2}B \leq f \pm f_0 \leq \frac{1}{2}B$, and zero otherwise. Write $z(t)$ as

$$z(t) = A \cos 2\pi(f_0 + f_d)t + n_c'(t) \cos 2\pi(f_0 + f_d)t - n_s'(t) \sin 2\pi(f_0 + f_d)t$$

(a) Express $n_c'(t)$ and $n_s'(t)$ in terms of $n_c(t)$ and $n_s(t)$. Using the techniques developed in Section 5.5, find the power spectral densities of $n_c'(t)$ and $n_s'(t)$, $S_{n_c'}(f)$ and $S_{n_s'}(f)$.

(b) Find the cross-spectral density of $n_c'(t)$ and $n_s'(t)$, $S_{n_c'n_s'}(f)$, and the cross-correlation function, $R_{n_c'n_s'}(\tau)$. Are $n_c'(t)$ and $n_s'(t)$

correlated? Are $n_c'(t)$ and $n_s'(t)$, sampled at the same instant, independent?

5.35 (a) Using the results of Problem (5.34), derive Equation (5.164) with $\overline{|\delta f|} = f_d$.

(b) Compare Equations (5.164) and (5.159) for a squarewave-modulated FM signal with deviation f_d by letting $\overline{|\delta f|} = f_d$ and $B = 2f_d$ for $f_d = 5$ and 10 for signal-to-noise ratios of $A^2/N_0B = 1$, 10, 100, 1000. Plot ν and $\overline{\delta\nu}$ versus A^2/N_0B.

Problems Extending Text Material

5.36 A random process is composed of sample functions of the form

$$x(t) = n(t) \sum_{k=-\infty}^{\infty} \delta(t - kT_s) = \sum_{k=-\infty}^{\infty} n_k \delta(t - kT_s)$$

where $n(t)$ is a wide-sense stationary random process with the autocorrelation function $R_n(\tau)$, and $n_k = n(kT_s)$. If T_s is chosen such that

$$R_n(kT_s) = 0, \qquad k = 1, 2, \ldots$$

[that is, the samples $n_k = n(kT_s)$ are orthogonal], use Equation (5.28) to show that the power spectral density of $x(t)$ is

$$S_x(f) = \frac{R_n(0)}{T_s} = f_s R_n(0) = f_s \overline{n^2(t)}, \qquad -\infty < f < \infty$$

5.37 A random pulse-position-modulated waveform is as shown in Figure 5.21, where the delays Δt_i of the pulses are independent uniform random variables with identical pdf's of the form

$$f_{\Delta t_i}(x) = \begin{cases} \dfrac{4}{T}, & 0 \le x \le \frac{1}{4}T,\ \tau_0 < \frac{3}{4}T \\ 0, & \text{otherwise} \end{cases}$$

Find the power spectral density of $y(t)$. (*Hint:* Use Equation (5.18).)

FIGURE 5.21

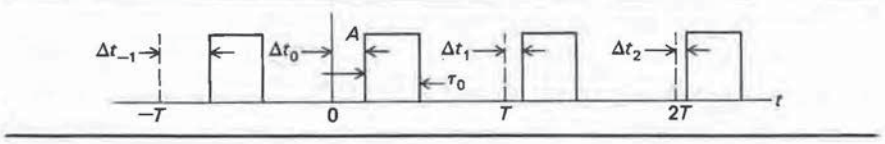
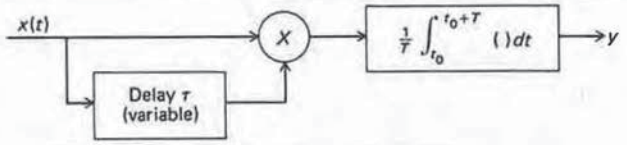


FIGURE 5.22



5.38 Consider the system shown in Figure 5.22 as a means of approximately measuring $R_x(\tau)$ where $x(t)$ is stationary.

(a) Show that $E\{y\} = R_X(\tau)$.

(b) Find an expression for σ_y^2 if $x(t)$ is Gaussian and has zero mean. *Hint:* If x_1, x_2, x_3, and x_4 are Gaussian with zero mean, it can be shown that

$$E\{x_1x_2x_3x_4\} = E\{x_1x_2\}E\{x_3x_4\} + E\{x_1x_3\}E\{x_2x_4\} + E\{x_1x_4\}E\{x_2x_3\}$$

5.39 A useful average in the consideration of noise in FM demodulation is the cross-correlation

$$R_{y\dot{y}}(\tau) \triangleq E\left\{y(t)\frac{dy\,(t+\tau)}{dt}\right\}$$

where $y(t)$ is assumed stationary.

(a) Show that

$$R_{y\dot{y}}(\tau) = \frac{dR_y(\tau)}{d\tau}$$

where $R_y(\tau)$ is the autocorrelation function of $y(t)$. (*Hint:* The transfer function of a differentiator is $H(f) = j2\pi f$.)

(b) If $y(t)$ is Gaussian, write down the joint pdf of

$$Y \triangleq y(t) \qquad \text{and} \qquad Z \triangleq \frac{dy(t)}{dt}$$

at any time t, assuming the ideal lowpass power spectral density

$$S_y(f) = \tfrac{1}{2}N_0\Pi\left(\frac{f}{2B}\right)$$

Express your answer in terms of N_0 and B.

(c) Can one obtain a result for the joint pdf of y and $\dot{y}$ if $y(t)$ is obtained by passing white noise through a lowpass RC filter? Why or why not?



COMPUTER EXERCISES

5.1 Check the correlation between the random variable X and Y generated by the random number generator of Computer Exercise 4.2 by computing the sample correlation coefficient of 1000 pairs according to the definition

$$\rho(X,Y) = \frac{1}{(N-1)\hat{\sigma}_1\hat{\sigma}_2}\sum_{n=1}^{N}(X_n - \hat{\mu}_X)(Y_n - \hat{\mu}_Y)$$

where

$$\hat{\mu}_X = \frac{1}{N}\sum_{n=1}^{N} X_n$$

$$\hat{\mu}_Y = \frac{1}{N}\sum_{n=1}^{N} Y_n$$

$$\hat{\sigma}_X^2 = \frac{1}{N-1}\sum_{n=1}^{N}(X_n - \hat{\mu}_X)^2$$

and

$$\hat{\sigma}_Y^2 = \frac{1}{N-1}\sum_{n=1}^{N}(Y_n - \hat{\mu}_X)^2$$

5.2 Using the correlated Gaussian random-number generator designed in Computer Exercise 4.3, write a simulation for the origin encirclement problem described by Equations (5.151) and following. Note that the $\rho(\tau)$ given in that computer exercise corresponds to passing white noise through a first-order lowpass filter. Thus, you should be able to relate the filter bandwidth, and therefore of the noise, to α. You should design your simulation to check out the curve given in Figure 5.16.

6 NOISE IN MODULATION SYSTEMS

In Chapters 4 and 5 we investigated the subjects of probability and random processes. These concepts led to a representation for bandlimited noise, which will now be used in the analysis of analog communication systems operating in the presence of noise.

Noise is present in varying degrees in all electrical systems and in communication systems noise can be especially damaging. As illustrated in Appendix A, under ordinary conditions, whenever the temperature of a conductor is above 0 K, random motion of charge carriers results in thermal noise. This noise is often low-level and can usually be neglected in those portions of a system where the signal level is high. Often, however, systems are encountered in which the signal levels are low, and in such systems the effects of even low-level noise can seriously degrade overall system performance. Examples of this are found in the predetection portions of an AM or FM radio receiver that is tuned to a distant transmitter and in a radar receiver that is processing a return pulse from a distant target. There are many other examples.

As pointed out in Chapter 1, system noise results from sources external to the system as well as from sources internal to the system. Since noise is unavoidable in any practical system, techniques for minimizing the effects of noise must often be used if high-performance communications are desired. In the present chapter, appropriate performance criteria for system evaluation will be developed. After this, a number of systems will be analyzed to determine the effect of noise on system operation. It is especially important to note the differences between linear and nonlinear systems. We will find that the use of nonlinear modulation, such as FM, allows *improved performance* to be obtained at the expense of *increased transmission bandwidth*. Such tradeoffs do not exist when linear modulation is used.

Our concern in this chapter will be with systems using analog modulation, both continuous wave and pulse. For completeness, we briefly consider pulse code modulation in this chapter. In the following chapters, our study will

be extended to include digital systems and the optimization of system performance in the presence of noise.

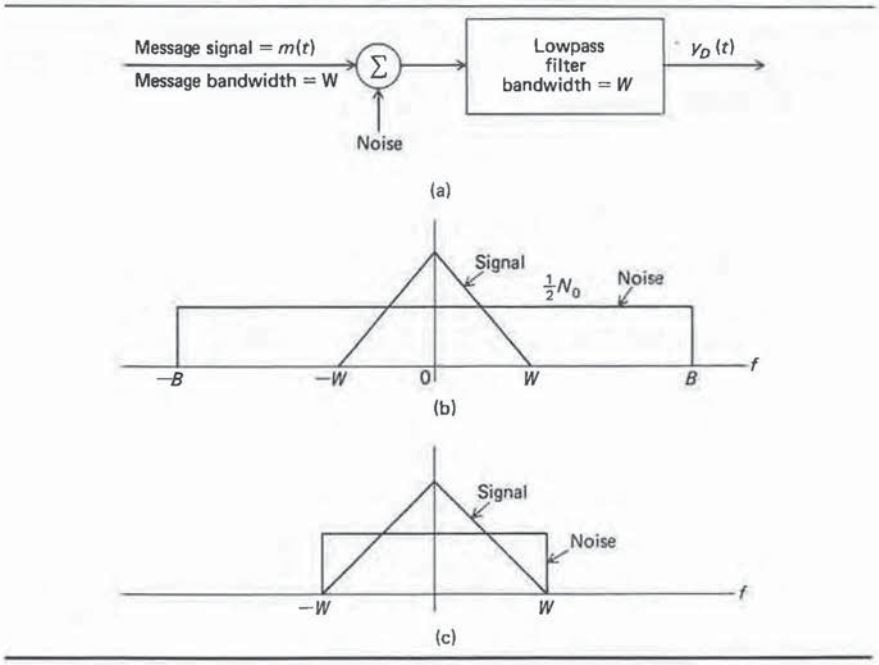
6.1 SIGNAL-TO-NOISE RATIOS

In Chapter 3, systems that involve the operations of modulation and demodulation were studied. In this section we extend that study to the performance of linear demodulators in the presence of noise. We concentrate our efforts on the calculation of signal-to-noise ratios. The signal-to-noise ratio is often a useful figure of merit for evaluating system performance.

Baseband Systems

In order to have a basis for comparing system performance, let us determine the signal-to-noise ratio at the output of a baseband system involving no modulation or demodulation. Such a system is illustrated in Figure 6.1(a). Assume that the signal power is finite at P_T W and that the noise added to

FIGURE 6.1 Baseband system. (a) Diagram. (b) Spectra at filter input. (c) Spectra at filter output.



the signal has the double-sided power spectral density $\frac{1}{2}N_0$ W/Hz over a bandwidth B, which exceeds W. This is illustrated in Figure 6.1(b).

If the message signal $m(t)$ is assumed to be bandlimited, a lowpass filter can be used to enhance the signal-to-noise ratio, as shown in Figure 6.1(c). If the filter bandwidth equals W, the signal is totally passed and the signal power at the lowpass filter output is P_T. The power of the output noise is

$$\int_{-W}^{W} \tfrac{1}{2}N_0 \, df = N_0 W \tag{6.1}$$

Thus the signal-to-noise ratio (SNR) at the filter output is

$$\text{SNR} = \frac{P_T}{N_0 W} \tag{6.2}$$

The filter enhances the SNR by a factor of B/W. Since (6.2) describes the SNR achieved with a simple baseband system using no modulation, it is a reasonable standard for making comparisons of system performance.

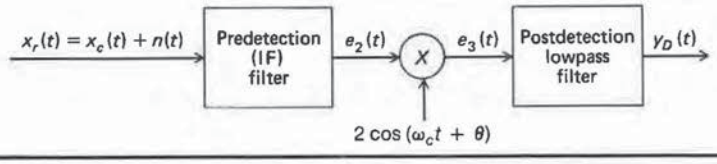
DSB Systems

We can now compute the noise performance of a coherent double-sideband (DSB) demodulator. Consider the block diagram in Figure 6.2, which illustrates a coherent demodulator preceded by a predetection filter. Typically, the predetection filter is the intermediate-frequency (IF) filter discussed in Chapter 3. The input to this filter is the modulated signal plus white Gaussian noise of double-sided power spectral density $\frac{1}{2}N_0$ W/Hz. Since the transmitted signal $x_c(t)$ is assumed to be a DSB signal, the received signal $x_r(t)$ can be written as

$$x_r(t) = A_c m(t) \cos(\omega_c t + \theta) + n(t) \tag{6.3}$$

where $m(t)$ is the message and θ is used to denote our uncertainty of the carrier phase or, equivalently, the time origin. If the predetection filter bandwidth is $2W$, the DSB signal is completely passed by the filter. Using

FIGURE 6.2 DSB demodulator



the technique developed in Chapter 5, the noise at the filter output $n_0(t)$ can be expanded into its direct and quadrature components as follows:

$$e_2(t) = A_c m(t) \cos(\omega_c t + \theta) + n_c(t) \cos(\omega_c t + \theta) - n_s(t) \sin(\omega_c t + \theta) \tag{6.4}$$

where the total noise power is $\overline{n_0^2(t)} = \frac{1}{2}\overline{n_c^2(t)} + \frac{1}{2}\overline{n_s^2(t)}$ and is equal to $2N_0W$.

The predetection SNR, measured at the input to the multiplier, is easily determined. The signal power is $\frac{1}{2}A_c^2\overline{m^2}$, where m is understood to be a function of t and the noise power is $2N_0W$ as shown in Figure 6.3(a). This yields the predetection SNR

$$(\text{SNR})_T = \frac{A_c^2\overline{m^2}}{4WN_0} \tag{6.5}$$

In order to compute the postdetection SNR, $e_3(t)$ is first computed:

$$e_3(t) = A_c m(t) + A_c m(t) \cos 2(\omega_c t + \theta) + n_c(t) + n_c(t) \cos 2(\omega_c t + \theta) - n_s(t) \sin 2(\omega_c t + \theta) \tag{6.6}$$

The double-frequency terms are removed by the postdetection filter so that

$$y_D(t) = A_c m(t) + n_c(t) \tag{6.7}$$

for the demodulated output. The postdetection signal power is $A_c^2\overline{m^2}$, and the postdetection noise power is $\overline{n_c^2}$ or $2N_0W$, as shown on Figure 6.3(b). The postdetection SNR is

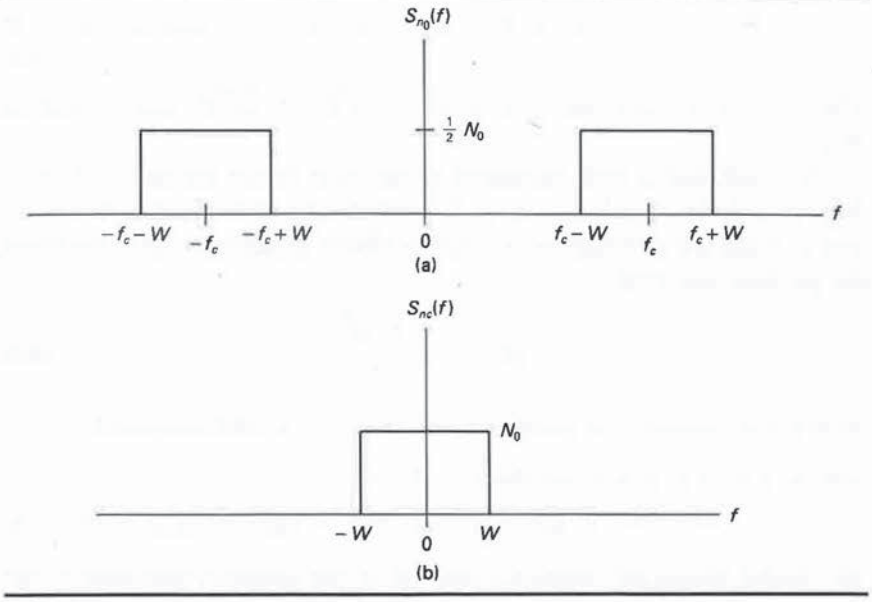
$$(\text{SNR})_D = \frac{A_c^2\overline{m^2}}{2N_0W}$$

The ratio of $(\text{SNR})_D$ to $(\text{SNR})_T$ is referred to as the *detection gain* and is often used as a figure of merit for a demodulator. Thus, for the coherent DSB demodulator, the detection gain is

$$\frac{(\text{SNR})_D}{(\text{SNR})_T} = \frac{A_c^2\overline{m^2}}{2N_0W} \cdot \frac{4N_0W}{A_c^2\overline{m^2}} = 2 \tag{6.8}$$

At first sight, this result is somewhat misleading, for it appears that we have gained 3 dB. This is true for the demodulator because it effectively suppresses the quadrature noise component. However, a comparison with the baseband system reveals that nothing is gained, insofar as the SNR at the system output is concerned, by the use of modulation, because the predetection filter bandwidth must be $2W$ if DSB modulation is used. Thus the channel bandwidth is doubled, resulting in double the noise bandwidth and,

FIGURE 6.3 Predetection (a) and postdetection (b) filter output noise spectra for DSB demodulation



consequently, double the noise power at the demodulator input. The 3-dB detection gain is exactly enough to overcome this effect.

In order to see this mathematically, let us return to the expression for the postdetection SNR. The term $\frac{1}{2}A_c^2\overline{m^2}$ is the transmitted power P_T. Thus

$$(\text{SNR})_D = \frac{P_T}{N_0 W} \tag{6.9}$$

which is the output SNR of the equivalent baseband system.

SSB Systems

Similar calculations are easily carried out for SSB systems. For SSB, the predetection filter input can be written as

$$x_r(t) = A_c[m(t)\cos(\omega_c t + \theta) \pm \hat{m}(t)\sin(\omega_c t + \theta)] + n(t) \tag{6.10}$$

where $\hat{m}(t)$ denotes the Hilbert transform of $m(t)$. Recall from Chapter 3 that the plus sign is used for lower-sideband SSB and the minus sign is used for upper-sideband SSB. Since the minimum bandwidth of the predetection bandpass filter is W for single-sideband, the center frequency of the predetection filter is $\omega_x = \omega_c \pm \frac{1}{2}(2\pi W)$, where the sign depends on the choice of sideband.

We could expand the noise about the center frequency ω_x. However, as we saw in Chapter 5, we are free to expand the noise about any frequency we choose. It is slightly more convenient, however, to expand the noise about the carrier frequency ω_c. For this case, the predetection filter output can be written as $x_c(t) + n_0(t)$, which is

$$e_2(t) = A_c[m(t)\cos(\omega_c t + \theta) \pm \hat{m}(t)\sin(\omega_c t + \theta)] + n_c(t)\cos(\omega_c t + \theta) - n_s(t)\sin(\omega_c t + \theta) \quad (6.11)$$

where, as can be seen from Figure 6.4(a),

$$N_T = \overline{n_0^2} = \overline{n_c^2} = \overline{n_s^2} = N_0 W \quad (6.12)$$

Equation (6.11) can be written

$$e_2(t) = [A_c m(t) + n_c(t)]\cos(\omega_c t + \theta) \pm [A_c\hat{m}(t) \mp n_s(t)]\sin(\omega_c t + \theta) \quad (6.13)$$

As discussed in Chapter 3, demodulation is accomplished by multiplying $e_2(t)$ by the demodulation carrier $2\cos(\omega_c t + \theta)$ and lowpass filtering. Thus the DSB demodulator illustrated in Figure 6.2 also applies to the SSB demodulation problem. It follows that

$$y_D(t) = A_c m(t) + n_c(t) \quad (6.14)$$

The power spectral density of $n_c(t)$ is illustrated in Figure 6.4(b) for the case of lower-sideband SSB. Note that the postdetection filter passes $n_c(t)$, so that the postdetection noise power is

$$N_D = \overline{n_c^2} = N_0 W \quad (6.15)$$

From (6.14) it follows that the postdetection signal power is

$$S_D = A_c^2\overline{m^2} \quad (6.16)$$

We now turn our attention to the predetection terms.

The predetection signal power is

$$S_T = \overline{\{A_c[m(t)\cos(\omega_c t + \theta) - \hat{m}(t)\sin(\omega_c t + \theta)]\}^2} \quad (6.17)$$

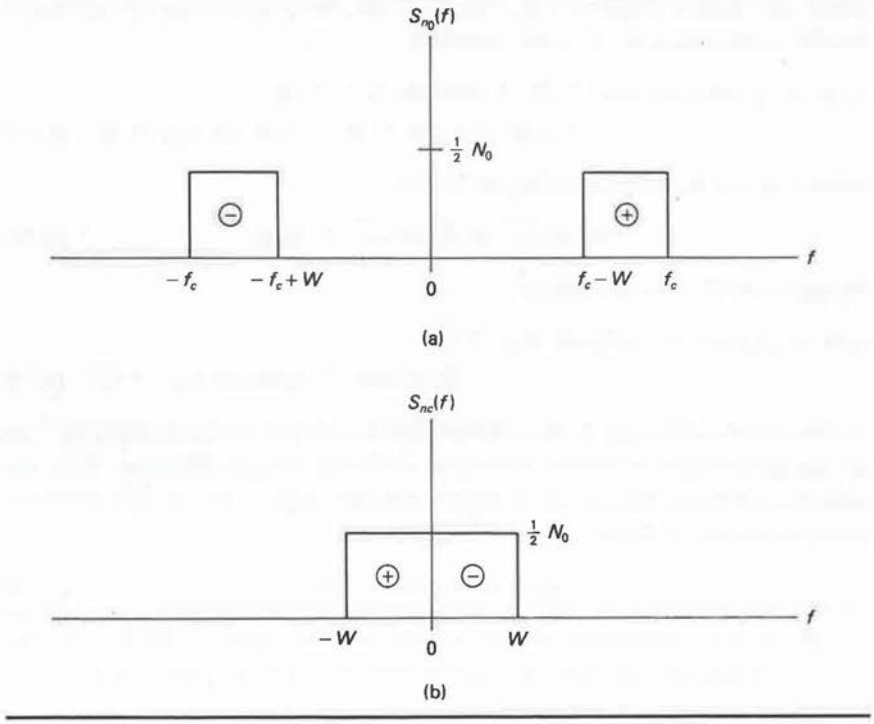
In Chapter 2 we pointed out that a function and its Hilbert transform are orthogonal. If $\overline{m(t)} = 0$, it follows that $\overline{m(t)\hat{m}(t)} = \langle m(t)\hat{m}(t)\rangle = 0$. Thus the preceding expression becomes

$$S_T = A_c^2\{[\tfrac{1}{2}\overline{m^2(t)}] + [\tfrac{1}{2}\overline{\hat{m}^2(t)}]\} \quad (6.18)$$

It was also shown in Chapter 2 that a function and its Hilbert transform have equal power. Applying this to (6.18) yields

$$S_T = A_c^2\overline{m^2} \quad (6.19)$$

FIGURE 6.4 Predetection (a) and postdetection (b) filter output spectra for lower-sideband SSB (+ and − signs denote spectral translation of positive and negative portions of $S_{n_0}(f)$ due to demodulation, respectively).



The same reasoning applied to the predetection and postdetection noise shows that they have equal power and that this power is

$$N_T = N_D = \tfrac{1}{2}\overline{n_c^2} + \tfrac{1}{2}\overline{n_s^2} = N_0 W \tag{6.20}$$

Therefore, the detection gain is

$$\frac{(\text{SNR})_D}{(\text{SNR})_T} = \frac{A_c^2\overline{m^2}}{N_0 W} \cdot \frac{N_0 W}{A_c^2\overline{m^2}} = 1 \tag{6.21}$$

The SSB system lacks the 3-dB detection gain of the DSB system. However, the predetection noise power of the SSB system is 3 dB less than that for the DSB system if the predetection filters have minimum bandwidth. This results in equal performance, given by

$$(\text{SNR})_D = \frac{A_c^2\overline{m^2}}{N_0 W} = \frac{P_T}{N_0 W} \tag{6.22}$$

Thus coherent demodulation of both DSB and SSB results in performance equivalent to baseband.

AM Systems: Coherent Detection

We saw in Chapter 3 that the representation for an AM signal is

$$x_c(t) = A_c[1 + am_n(t)] \cos(\omega_c t + \theta) \tag{6.23}$$

where $m_n(t)$ is the modulation signal normalized so that the maximum value of $|m_n(t)|$ is unity and a is the modulation index. It is easily shown, by using a development parallel to that for DSB systems, that the demodulated output in the presence of noise is

$$y_D(t) = A_c a m_n(t) + n_c(t) \tag{6.24}$$

The dc term that mathematically results in $y_D(t)$ is not included in (6.24) for two reasons. First, this term is not considered part of the signal since it contains no information. [Recall that we have assumed $m(t) = 0$.] Second, most practical AM demodulators are not dc-coupled, so a dc term does not appear on the output of a practical system. In addition, the dc term is frequently used for AGC (automatic gain control) and has to be held constant at the transmitter.

From (6.24) it follows that the signal power in $y_D(t)$ is

$$S_D = A_c^2 a^2 \overline{m_n^2} \tag{6.25}$$

and the noise power is

$$N_D = \overline{n_c^2} = 2N_0W \tag{6.26}$$

For the predetection case, the signal power is

$$S_T = \tfrac{1}{2}A_c^2 + \tfrac{1}{2}A_c^2 a^2 \overline{m_n^2} \tag{6.27}$$

and the noise power is

$$N_T = 2N_0W \tag{6.28}$$

since the required transmission bandwidth for AM is $2W$. Thus the detection gain is

$$\frac{(\text{SNR})_D}{(\text{SNR})_T} = \frac{A_c^2 a^2 \overline{m_n^2}/2N_0W}{(A_c^2 + A_c^2 a^2 \overline{m_n^2})/4N_0W} = \frac{2a^2 \overline{m_n^2}}{1 + a^2 \overline{m_n^2}} \tag{6.29}$$

Recall that when we studied AM in Chapter 3 the efficiency of an AM transmission system was defined as the ratio of sideband power to total

power in the transmitted signal $x_c(t)$. This resulted in the efficiency E being expressed as

$$E = \frac{a^2\overline{m_n^2}}{1 + a^2\overline{m_n^2}} \tag{6.30}$$

where the overbar, denoting a statistical average, has been substituted for the time-average notation $\langle \cdot \rangle$ used in Chapter 3. It follows from (6.29) and (6.30) that the detection gain can be expressed as

$$\frac{(\text{SNR})_D}{(\text{SNR})_T} = 2E \tag{6.31}$$

Since the predetection signal-to-noise ratio can be written as

$$(\text{SNR})_T = \frac{S_T}{2N_0W} = \frac{P_T}{2N_0W} \tag{6.32}$$

it follows that the signal-to-noise ratio at the demodulator output can be written as

$$(\text{SNR})_D = \frac{EP_T}{N_0W} \tag{6.33}$$

If the efficiency *could* be 100%, then AM would have the same postdetection signal-to-noise ratio as the ideal DSB and SSB systems. Of course, as we saw in Chapter 3, the efficiency of AM is typically much less than 100% and the postdetection signal-to-noise ratio is correspondingly lower.

▼ EXAMPLE 6.1 An AM system operates with a modulation index of 0.5 and the power in the normalized message signal is 0.1. The efficiency is

$$E = \frac{(0.5)^2(0.1)}{1 + (0.5)^2(0.1)} = 0.0244 \tag{6.34}$$

and the postdetection signal-to-noise ratio is

$$(SNR)_D = 0.0244\,\frac{P_T}{N_0W} \tag{6.35}$$

The detection gain is

$$\frac{(SNR)_D}{(SNR)_T} = 2E = 0.0488 \tag{6.36}$$

This is more than 16 dB inferior to the ideal system requiring the same bandwidth. It should be remembered, however, that the motivation for using AM is not noise performance but rather that AM allows the use of simple envelope detectors for demodulation. The reason, of course, for the poor efficiency of AM is that a large fraction of the total transmitted power is devoted to the carrier component, which conveys no information since it is not a function of the message signal. ▲

AM Systems: Envelope Detection

Since envelope detection is the usual method of demodulating an AM signal, it is important to understand how envelope demodulation differs from coherent demodulation in the presence of noise. The received signal, at the input to the envelope detector, is assumed to be $x_c(t)$ plus narrowband noise. Thus

$$x_r(t) = A_c[1 + am_n(t)] \cos(\omega_c t + \theta) + n_c(t) \cos(\omega_c t + \theta) - n_s(t) \sin(\omega_c t + \theta) \quad (6.37)$$

where, as before, $\overline{n_c^2} = \overline{n_s^2} = 2N_0W$. The output of the envelope detector can be determined with the aid of the phasor diagram in Figure 6.5. The signal $x_r(t)$ can be written in terms of envelope and phase as

$$x_r(t) = r(t) \cos[\omega_c t + \theta + \phi(t)] \quad (6.38)$$

where

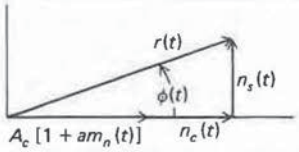
$$r(t) = \sqrt{\{A_c[1 + am_n(t)] + n_c(t)\}^2 + n_s^2(t)} \quad (6.39)$$

and

$$\phi(t) = \tan^{-1} \frac{n_s(t)}{A_c[1 + am_n(t)] + n_c(t)} \quad (6.40)$$

The output of an ideal envelope detector is independent of phase variations of the input. Thus the expression for $\phi(t)$ is of no interest, and we will

FIGURE 6.5 Phasor diagram of AM (drawn for $\theta = 0$)



concentrate on $r(t)$. The envelope detector is assumed to be ac-coupled so that

$$y_D(t) = r(t) - \overline{r(t)} \tag{6.41}$$

where $\overline{r(t)}$ is the average value of the envelope amplitude.

Equation (6.41) will be evaluated for two cases. First, we consider the case in which the $(\text{SNR})_T$ is large, and then we briefly consider the case in which the $(\text{SNR})_T$ is small. For the first case, the solution is simple. From Figure 6.5 and Equation (6.39), we see that if

$$|A_c[1 + am_n(t)] + n_c(t)| \gg |n_s(t)| \tag{6.42}$$

then *most of the time*

$$r(t) \cong A_c[1 + am_n(t)] + n_c(t) \tag{6.43}$$

yielding

$$y_D(t) \cong A_c am_n(t) + n_c(t) \tag{6.44}$$

This is the final result for the case in which the SNR is large.

Comparing (6.44) and (6.24) illustrates that the output of the envelope detector is equivalent to the output of the coherent detector if $(\text{SNR})_T$ is large. The detection gain for this case is given by (6.29).

For the case in which $(\text{SNR})_T$ is small, the analysis is somewhat more complex. In order to analyze this case, we recall from Chapter 5 that $n_c(t) \cos(\omega_c t + \theta) - n_s(t) \sin(\omega_c t + \theta)$ can be written in terms of envelope and phase, so that the envelope detector input can be written as

$$e(t) = A_c[1 + am_n(t)] \cos(\omega_c t + \theta) + r_n(t) \cos[\omega_c t + \theta + \phi_n(t)] \tag{6.45}$$

For $(\text{SNR})_T \ll 1$, the amplitude of $A_c[1 + am_n(t)]$ will usually be much smaller than $r_n(t)$. Thus the phasor diagram appears as illustrated in Figure 6.6. It can be seen that $r(t)$ is approximated by

$$r(t) \cong r_n(t) + A_c[1 + am_n(t)] \cos \phi_n(t) \tag{6.46}$$

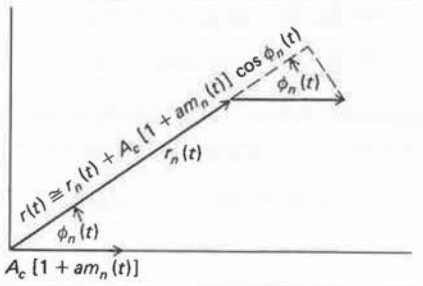
yielding

$$y_D(t) \cong r_n(t) + A_c[1 + am_n(t)] \cos \phi_n(t) - \overline{r(t)} \tag{6.47}$$

The principal component of $y_D(t)$ is the Rayleigh-distributed noise envelope, and *no* component of $y_D(t)$ is proportional to the signal. Note that since $n_c(t)$ and $n_s(t)$ are random, $\cos \phi_n(t)$ is also random. Thus the *signal* is *multiplied* by a random quantity. This multiplication of the signal by a function of the noise has a significantly worse degrading effect than does additive noise.

This severe loss of signal at low-input SNR is known as the *threshold effect*

FIGURE 6.6 Phasor diagram for AM with $(SNR)_T \ll 1$ (drawn for $\theta = 0$)



and results from the nonlinear action of the envelope detector. In coherent detectors, which are linear, the signal and noise are additive at the detector output *if* they are additive at the detector input. The result is that the signal retains its identity even when the input SNR is low. For this reason, coherent detection is often desirable when the noise is large.

Square-Law Detectors

The determination of the signal-to-noise ratio at the output of a nonlinear system is often a very difficult task. The *square-law detector,* however, is one system for which this is not the case. In this section, we conduct a simplified analysis to illustrate the phenomenon of thresholding, which is a characteristic of nonlinear systems. We will see that the detection gain is different for high and low signal-to-noise ratios. The intermediate region is known as the *threshold region.*

In the analysis that follows, the postdetection bandwidth will be assumed to be twice the message bandwidth W. As shown in the problem at the end of the chapter, this is not a necessary assumption, but it does result in a simplification of the analysis without having an impact on the threshold effect. We will also see that harmonic distortion is a problem with square-law detectors, an effect that often precludes their use.

Square-law detectors can be implemented as a squaring device followed by a lowpass filter. The response of a square-law detector to an AM signal is $r^2(t)$, where $r(t)$ is defined by (6.39). Thus the output can be written as

$$r^2(t) = \{A_c[1 + am_n(t)] + n_c(t)\}^2 + n_s^2(t) \tag{6.48}$$

We now determine the output SNR assuming sinusoidal modulation. Letting $m_n(t) = \cos \omega_m t$ and carrying out the indicated squaring operation, we obtain

$$r^2(t) = A_c^2 + 2A_c^2 a \cos \omega_m t + \tfrac{1}{2}A_c^2 a^2 + \tfrac{1}{2}A_c^2 a^2 \cos 2\omega_m t + 2A_c n_c(t) + 2A_c a n_c(t) \cos \omega_m t + n_c^2(t) + n_s^2(t) \tag{6.49}$$

Assuming that the detector output is ac-coupled, so that the dc terms are blocked, the signal and noise components of the output are written as

$$s_D(t) = 2A_c^2 a \cos \omega_m t \tag{6.50}$$

and

$$n_D(t) = 2A_c n_c(t) + 2A_c a n_c(t) \cos \omega_m t + n_c^2(t) + n_s^2(t) \tag{6.51}$$

respectively. The power in the signal component is

$$S_D = 2A_c^4 a^2 \tag{6.52}$$

and the noise power is

$$N_D = 4A_c^2 \overline{n_c^2(t)} + 2A_c^2 a^2 \overline{n_c^2(t)} + 2\{\overline{n_c^4(t)} - \overline{n_c^2(t)}^2\} \tag{6.53}$$

where the dc noise power must be subtracted because of the assumed ac coupling. The total predetection noise power is

$$\sigma_n^2 = N_0 B_T = 2N_0 W = \overline{n_c^2(t)} = \overline{n_s^2(t)} \tag{6.54}$$

where W is the bandwidth of the message signal. In order to evaluate the quantity $\overline{n_c^4(t)}$ in (6.53) we write

$$\overline{n_c^4(t)} = \int_{-\infty}^{\infty} n_c^4 \frac{1}{\sqrt{2\pi}\sigma_n} e^{-n_c^2/2\sigma_n^2} \, dn_c = 3\sigma_n^4 \tag{6.55}$$

Thus (6.53) becomes $N_D = 2A_c^2(2 + a^2)\sigma_n^2 + 4\sigma_n^4$, and the output SNR is

$$(\text{SNR})_D = \frac{A_c^2 a^2/\sigma_n^2}{(2 + a^2) + 2(\sigma_n^2/A_c^2)} \tag{6.56}$$

for the square-law detector.

It is interesting to compare (6.56) with baseband systems. The total transmitted power P_T of an AM system, with sinusoidal modulation, is

$$P_T = \tfrac{1}{2}A_c^2(1 + \tfrac{1}{2}a^2) \tag{6.57}$$

Substituting from (6.57) for A_c^2 and $2N_0W$ for σ_n^2 yields

$$(\text{SNR})_D = 2\left(\frac{a}{2 + a^2}\right)^2 \frac{P_T/N_0W}{1 + (N_0W/P_T)} \tag{6.58}$$

This expression clearly illustrates the threshold effect. For large values of P_T/N_0W,

$$(\text{SNR})_D \cong 2\left(\frac{a}{2+a^2}\right)^2 \frac{P_T}{N_0W} \tag{6.59}$$

For small values of P_T/N_0W,

$$(\text{SNR})_D \cong 2\left(\frac{a}{2+a^2}\right)^2 \left(\frac{P_T}{N_0W}\right)^2 \tag{6.60}$$

Figure 6.7 illustrates (6.58) for several values of the modulation index a.

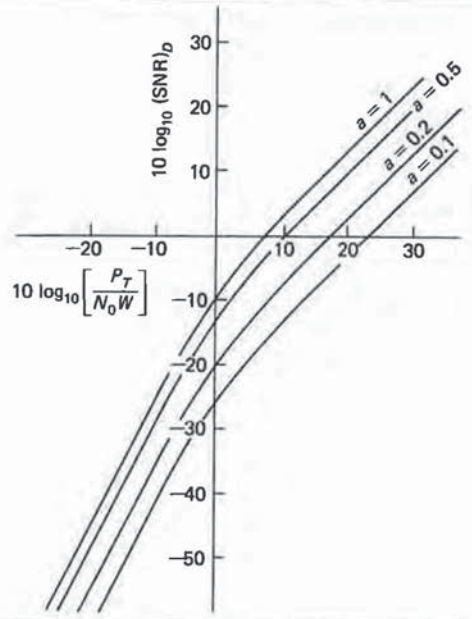
One term in the expression for $r^2(t)$ has been neglected. This term,

$$\tfrac{1}{2}A_c^2a^2\cos 2\omega_m t$$

represents *harmonic distortion* of the message due to the squaring operation. The power in this term is $\frac{1}{8}A_c^4a^4$, and the ratio of the harmonic distortion power D_D to the signal power S_D is

$$\frac{D_D}{S_D} = \frac{1}{16}a^2 \tag{6.61}$$

FIGURE 6.7 Performance of a square-law detector



Thus harmonic distortion can be reduced by decreasing the modulation index. However, as illustrated in Figure 6.7, a reduction of the modulation index also results in a decrease in the output signal-to-noise ratio.

The *linear envelope detector* defined by (6.39) is much more difficult to analyze over a wide range of signal-to-noise ratios because of the square root. However, to a first approximation, the performance of a linear envelope detector and a square-law envelope detector are the same. Harmonic distortion is also present in linear envelope detectors, but the amplitude of the distortion component is significantly less than that observed for square-law detectors. In addition, it can be shown that for high signal-to-noise ratios and a modulation index of unity, the performance of a linear envelope detector is better by approximately 1.8 dB than the performance of a square-law detector. (See Problem 6.10.)*

6.2 NOISE AND PHASE ERRORS IN COHERENT SYSTEMS

In the preceding section we investigated the performance of various types of demodulators. Our main interests were detection gain and calculation of the demodulated output SNR. Where coherent demodulation was used, the demodulation carrier was assumed to have *perfect* phase coherence. In a practical system, this is often not a valid assumption, since in systems where noise is present, perfect phase coherence is not possible. In this section we will allow the demodulation carrier to have a time-varying phase error, and we will determine the combined effect of noise at the demodulator input and phase error in the demodulation carrier.

A General Analysis

The demodulator model is illustrated in Figure 6.8. The signal portion of $e(t)$ will be assumed to be the quadrature double-sideband (QDSB) signal

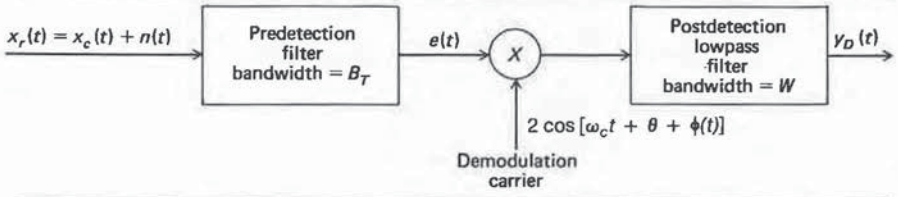
$$m_1(t) \cos(\omega_c t + \theta) + m_2(t) \sin(\omega_c t + \theta)$$

where any constant A_c is absorbed into $m_1(t)$ and $m_2(t)$ for notational simplicity.

Using this representation, we will obtain a general solution. After the analysis is complete, the DSB result is obtained by letting $m_1(t) = m(t)$ and $m_2(t) = 0$. The SSB result is obtained by letting $m_1(t) = m(t)$ and $m_2(t) = \pm \hat{m}(t)$, depending on the sideband of interest. For the QDSB system,

* For a detailed study of linear envelope detectors, see Bennett (1974).

FIGURE 6.8 Coherent demodulator with phase error



$y_D(t)$ is the demodulated output for the direct channel. The quadrature channel can be demodulated using a demodulation carrier of the form $2 \sin [\omega_c t + \theta + \phi(t)]$.

The noise portion of $e(t)$ is represented by

$$n_c(t) \cos (\omega_c t + \theta) - n_s(t) \sin (\omega_c t + \theta)$$

In the preceding expression,

$$\overline{n_c^2} = \overline{n_s^2} = N_0 B_T \tag{6.62}$$

where B_T is the bandwidth of the predetection filter and $\frac{1}{2}N_0$ is the double-sided power spectral density of the noise at the filter input. The phase error of the demodulation carrier is assumed to be a sample function of a zero-mean Gaussian process of known variance σ_ϕ^2. As before, the modulating signals are assumed to have zero mean.

With the preliminaries of defining the model and stating the assumptions disposed of, we now proceed with the analysis. The assumed performance criterion is mean-square error; that is, we will compute

$$\overline{\epsilon^2} = \overline{\{m_1(t) - y_D(t)\}^2} \tag{6.63}$$

for the three cases of interest: DSB, SSB, and QDSB. The multiplier input signal $e(t)$ is

$$\begin{aligned} e(t) = {} & m_1(t) \cos (\omega_c t + \theta) + m_2(t) \sin (\omega_c t + \theta) \\ & + n_c(t) \cos (\omega_c t + \theta) - n_s(t) \sin (\omega_c t + \theta) \end{aligned} \tag{6.64}$$

Multiplying by $2 \cos [\omega_c t + \theta + \phi(t)]$ and lowpass filtering gives us the generalized output

$$y_D(t) = [m_1(t) + n_c(t)] \cos \phi(t) - [m_2(t) - n_s(t)] \sin \phi(t) \tag{6.65}$$

The error $m_1(t) - y_D(t)$ can be written as

$$\epsilon = m_1 - (m_1 + n_c) \cos \phi + (m_2 - n_s) \sin \phi \tag{6.66}$$

where it is understood that ϵ, m_1, m_2, n_c, n_s, and ϕ are all functions of time. The mean-square error can be written as

$$\begin{aligned}\overline{\epsilon^2} = \overline{m_1^2} &- 2\overline{m_1(m_1 + n_c)\cos\phi} \\ &+ 2\overline{m_1(m_2 - n_s)\sin\phi} \\ &+ \overline{(m_1 + n_c)^2\cos^2\phi} \\ &- 2\overline{(m_1 + n_c)(m_2 - n_s)\sin\phi\cos\phi} \\ &+ \overline{(m_2 - n_s)^2\sin^2\phi}\end{aligned} \tag{6.67}$$

The variables m_1, m_2, n_c, n_s, and ϕ are all assumed to be uncorrelated. It should be pointed out that for the SSB case, the power spectra of $n_c(t)$ and $n_s(t)$ will not be symmetrical about ω_c. However, as pointed out in Section 5.5, $n_c(t)$ and $n_s(t)$ are still uncorrelated, since there is no time displacement. Thus the mean-square error can be written as

$$\overline{\epsilon^2} = \overline{m_1^2} - 2\overline{m_1^2\cos\phi} + \overline{m_1^2\cos^2\phi} + \overline{m_2^2\sin^2\phi} + \overline{n^2} \tag{6.68}$$

where

$$\overline{n_c^2} = \overline{n_s^2} = N_0B_T = \sigma_n^2 = \overline{n^2}$$

At this point in our development we must define the type of system being analyzed. First, let us assume the system of interest in *QDSB with equal power* in each modulating signal. Under this assumption, $\overline{m_1^2} = \overline{m_2^2} = \sigma_m^2$, and the mean-square error is

$$\overline{\epsilon_Q^2} = 2\sigma_m^2 - 2\sigma_m^2\,\overline{\cos\phi} + \sigma_n^2 \tag{6.69}$$

This expression can be easily evaluated for the case in which the maximum value of $|\phi(t)| \ll 1$. For this case,

$$\overline{\cos\phi} \cong \overline{(1 - \tfrac{1}{2}\phi^2)} = 1 - \tfrac{1}{2}\sigma_\phi^2$$

and

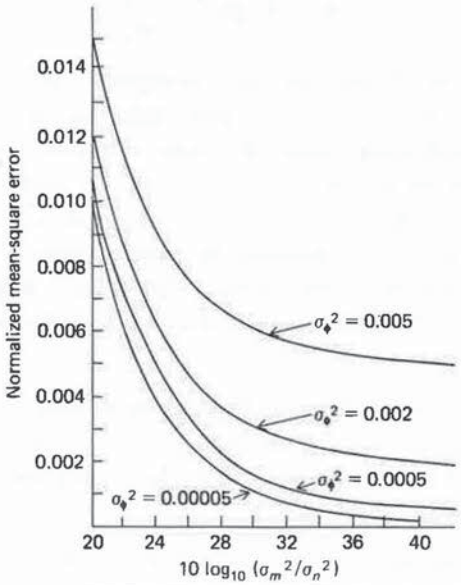
$$\overline{\epsilon_Q^2} = \sigma_m^2\sigma_\phi^2 + \sigma_n^2 \tag{6.70}$$

In order to have an easily interpreted measure of system performance, the mean-square error is usually normalized to σ_m^2. This yields

$$\overline{\epsilon_{NQ}^2} = \sigma_\phi^2 + \frac{\sigma_n^2}{\sigma_m^2} \tag{6.71}$$

The preceding expression is also valid for *the SSB case,* since an SSB signal is a QDSB signal with equal power in the direct and quadrature channels. The only possible point of confusion is that σ_n^2 may be different for the SSB

FIGURE 6.9 Mean-square error versus signal-to-noise ratio for QDSB system



and QDSB cases, since the SSB predetection filter bandwidth need only be half the bandwidth of the predetection filter for the QDSB case. Equation (6.71) is of such general interest that it is illustrated in Figure 6.9.

In order to compute the mean-square error *for a DSB system,* we simply let $m_2 = 0$ and $m_1 = m$ in (6.68). This yields

$$\overline{\epsilon_D^2} = \overline{m^2} - 2\overline{m^2}\,\overline{\cos\phi} + \overline{m^2}\,\overline{\cos^2\phi} + \overline{n^2} \tag{6.72}$$

or

$$\overline{\epsilon_D^2} = \sigma_m^2\overline{(1 - \cos\phi)^2} + \overline{n^2} \tag{6.73}$$

which, for small ϕ, can be approximated as

$$\overline{\epsilon_D^2} \cong \sigma_m^2(\tfrac{1}{4})\overline{\phi^4} + \overline{n^2} \tag{6.74}$$

In the preceding section [see (6.55)], we pointed out that if ϕ is zero-mean Gaussian with variance σ_ϕ^2,

$$\overline{\phi^4} = \overline{(\phi^2)^2} = 3\sigma_\phi^4 \tag{6.75}$$

Thus

$$\overline{\epsilon_D^2} \cong \tfrac{3}{4}\sigma_m^2\sigma_\phi^4 + \sigma_n^2 \tag{6.76}$$

and the normalized mean-square error becomes

$$\overline{\epsilon_{ND}^2} = \tfrac{3}{4}\sigma_\phi^4 + \frac{\sigma_n^2}{\sigma_m^2} \tag{6.77}$$

Several items are of interest when comparing (6.77) and (6.71). First, equal *output* SNR's imply equal *normalized* mean-square errors for $\sigma_\phi^2 = 0$. This is easy to understand, since the noise is additive at the output. The general expression for $y_D(t)$ is $y_D(t) = m(t) + n(t)$. The error is $n(t)$, and the normalized mean-square error is σ_n^2/σ_m^2. The analysis also illustrates that DSB systems are much less sensitive to phase errors of the demodulation carrier than are SSB or QDSB systems. This follows from the fact that if $\phi \ll 1$, the basic assumption made in the analysis, then $\sigma_\phi^4 \ll \sigma_\phi^2$.

▼ EXAMPLE 6.2 Assume that the demodulation phase error variance of a coherent demodulator is described by $\sigma_\phi^2 = 0.01$. The signal-to-noise ratio σ_m^2/σ_n^2 is 20 dB. If a DSB system is used, the normalized mean-square error is

$$\overline{\epsilon_{ND}^2} = \tfrac{3}{4}(0.01)^2 + 10^{-20/10} = 0.010075 \quad \text{(DSB)} \tag{6.78}$$

while for the SSB case the normalized mean-square error is

$$\overline{\epsilon_{ND}^2} = (0.01) + 10^{-20/10} = 0.02 \quad \text{(SSB)} \tag{6.79}$$

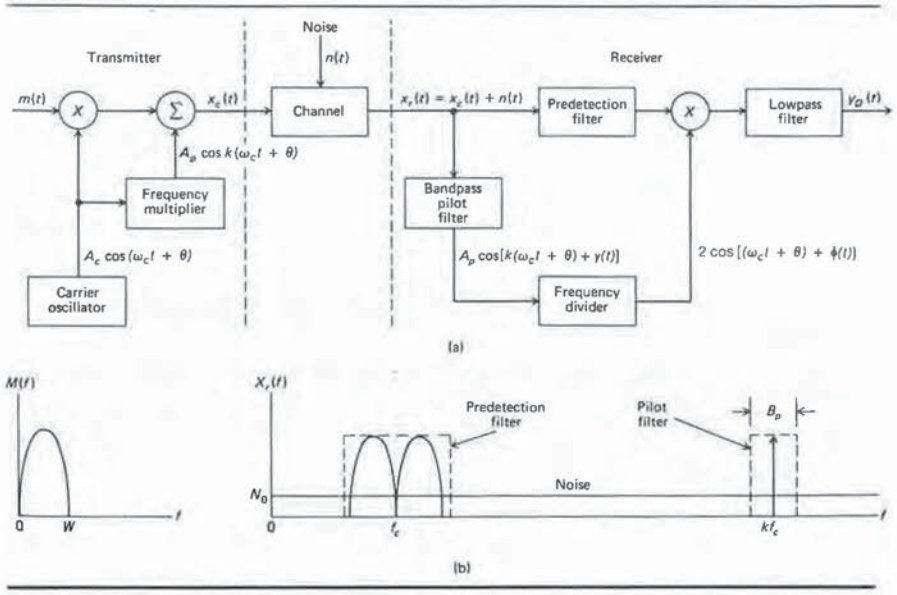
Note that for the DSB demodulator, the demodulation phase error can probably be neglected for most applications. For the SSB case the demodulation phase error variance must clearly be considered. Another way to look at the results of this example is to recognize that the use of DSB gives a factor-of-two improvement over the performance of a SSB system, at the cost of twice the bandwidth. ▲

Demodulation Phase Errors

At this point it is appropriate that we say a few words about the source of demodulation phase errors, the effects of which we have just analyzed. The demodulation carrier must be coherent with the original modulation carrier. Thus the modulation carrier must be the *source* of the demodulation carrier. This implies that the information necessary to generate the demodulation carrier passes through the channel and, consistent with the channel model, is perturbed by additive white noise.

One technique that can be used is illustrated in Figure 6.10(a). In this system, the demodulation carrier is derived from a *pilot signal,* which is har-

FIGURE 6.10 DSB system using pilot reconstruction. (a) System. (b) Single-sided spectra.



monically related to the carrier. The spectrum of the received signal is illustrated in Figure 6.10(b). As illustrated, the bandpass pilot filter extracts the pilot signal from the received signal. This pilot signal is then frequency-divided to yield the required demodulation carrier. Since the pilot is perturbed by channel noise, a pilot phase error $\gamma(t)$ is induced.

The statistics of $\gamma(t)$ are easily computed for the high SNR case. The pilot filter output can be expressed as

$$A_p \cos k(\omega_c t + \theta) + n_c(t) \cos k(\omega_c t + \theta) - n_s(t) \sin k(\omega_c t + \theta)$$

which can be written as

$$r(t) \cos [k(\omega_c t + \theta) + \gamma(t)]$$

where

$$\gamma(t) = \tan^{-1} \frac{n_s(t)}{A_p + n_c(t)} \tag{6.80}$$

In (6.80),

$$\overline{n_s^2} = \overline{n_c^2} = N_0 B_p \tag{6.81}$$

where B_p is the equivalent noise bandwidth of the pilot filter. For large signal-to-noise ratios, $A_p \gg n_c$ and the argument of the arctangent function is small. Thus

$$\gamma(t) \cong \frac{n_s(t)}{A_p} \tag{6.82}$$

and the phase error variance is

$$\sigma_\gamma^2 = \frac{\overline{n_s^2}}{A_p^2} = \frac{1}{2\rho} \tag{6.83}$$

where ρ is the pilot SNR measured in the equivalent noise bandwidth of the pilot filter.

A frequency divider can be thought of as a device that divides the total instantaneous phase by a constant. Thus, if a frequency divider has input $2 \cos [k(\omega_c t + \theta) + \gamma(t)]$, the output is $2 \cos [\omega_c t + \theta + (1/k)\gamma(t)]$. Therefore, in Figure 6.9,

$$\sigma_\phi^2 = \frac{1}{k^2} \sigma_\gamma^2 \tag{6.84}$$

As pointed out in Chapter 3, a phase-lock loop, adjusted so that a harmonic of the voltage-controlled oscillator output is phase-locked to the input signal, is an effective frequency divider. The assumption has been made that the frequency divider contains a limiter, so that its output is insensitive to variations in input amplitude. It also should be mentioned that a phase-lock loop can be used as the bandpass pilot filter. In this application, the VCO output is the pilot signal, and B_p is the equivalent noise bandwidth of the phase-lock loop.

The technique just described is not usually used with single-channel systems because of the added bandwidth and power required for the pilot. However, in multichannel systems, a number of channels can use the same pilot. It is even possible to use a common pilot signal for a number of completely separate communication systems.

6.3 NOISE IN ANGLE MODULATION SYSTEMS

Now that we have investigated the effect of noise on a linear modulation system, we turn our attention to angle modulation. We will find that there are great differences between linear and angle modulation when noise effects are considered. We will even find significant differences between PM and FM. Finally, we will see that FM can offer greatly improved performance over both linear modulation and PM systems in noisy environments.

FIGURE 6.11 Angle demodulation system



Demodulation in the Presence of Noise

Consider the system shown in Figure 6.11. The predetection filter bandwidth is B_T and can be determined by Carson's rule. In other words, B_T is approximately $2(D + 1)W$ Hz, where W is the bandwidth of the message signal and D is the deviation ratio, which is the peak frequency deviation divided by W. The input to the predetection filter is assumed to be the modulated carrier

$$x_c(t) = A_c \cos [\omega_c t + \theta + \phi(t)] \tag{6.85}$$

plus additive white noise that has the double-sided power spectral density $\frac{1}{2}N_0$ W/Hz.

At this point we recall that for PM,

$$\phi(t) = k_p m_n(t) \tag{6.86}$$

where k_p is the phase-deviation constant in radians per unit and $m_n(t)$ is the message signal normalized so that the peak value of $|m(t)|$ is unity. For the FM case,

$$\phi(t) = 2\pi f_d \int^t m_n(\alpha)\, d\alpha \tag{6.87}$$

where f_d is the deviation constant in hertz per unit. If the maximum value of $|m(t)|$ is not unity, as is usually the case, the scaling constant K, defined by $m(t) = Km_n(t)$, is contained in k_p or f_d. We shall analyze the PM and FM cases together and then specialize the results by replacing $\phi(t)$ by the proper function.

The output of the predetection filter can be written as

$$\begin{aligned} e_1(t) = {} & A_c \cos [\omega_c t + \theta + \phi(t)] \\ & + n_c t \cos (\omega_c t + \theta) - n_s(t) \sin (\omega_c t + \theta) \end{aligned} \tag{6.88}$$

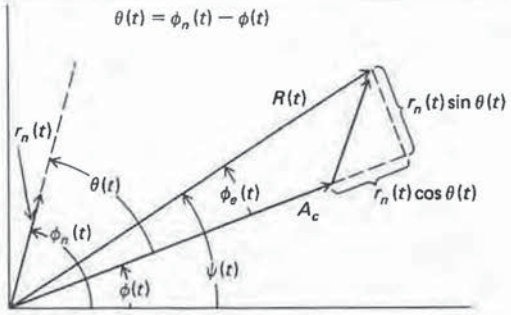
where

$$\overline{n_c^2} = \overline{n_s^2} = N_0 B_T \tag{6.89}$$

Equation (6.88) can be written as

$$e_1(t) = A_c \cos [\omega_c t + \theta + \phi(t)] + r_n(t) \cos [\omega_c t + \theta + \phi_n(t)] \tag{6.90}$$

FIGURE 6.12 Phasor diagram for angle demodulation, assuming $(SNR)_T > 1$ (drawn for $\theta = 0$)



where $r_n(t)$ is the Rayleigh-distributed noise envelope and $\phi_n(t)$ is the uniformally distributed phase. By replacing $\omega_c t + \phi_n(t)$ with $\omega_c t + \phi(t) + \phi_n(t) - \phi(t)$, we can write (6.90) as

$$\begin{aligned} e_1(t) = {} & A_c \cos [\omega_c t + \theta + \phi(t)] \\ & + r_n(t) \cos [\phi_n(t) - \phi(t)] \cos [\omega_c t + \theta + \phi(t)] \\ & - r_n(t) \sin [\phi_n(t) - \phi(t)] \sin [\omega_c t + \theta + \phi(t)] \end{aligned} \tag{6.91}$$

which is

$$\begin{aligned} e_1(t) = {} & \{A_c + r_n(t) \cos [\phi_n(t) - \phi(t)]\} \cos [\omega_c t + \theta + \phi(t)] \\ & - r_n(t) \sin [\phi_n(t) - \phi(t)] \sin [\omega_c t + \theta + \phi(t)] \end{aligned} \tag{6.92}$$

The preceding expression can be written as

$$e_1(t) = R(t) \cos [\omega_c t + \theta + \phi(t) + \phi_e(t)] \tag{6.93}$$

where $\phi_e(t)$ is the phase *deviation due to noise* and is given by

$$\phi_e(t) = \tan^{-1} \frac{r_n(t) \sin [\phi_n(t) - \phi(t)]}{A_c + r_n(t) \cos [\phi_n(t) - \phi(t)]} \tag{6.94}$$

Since $\phi_e(t)$ adds to $\phi(t)$, *which conveys the message signal,* it is the noise component of interest.

A phasor diagram of the predetection filter output is given in Figure 6.12 for the case where $A_c > r_n(t)$. If $e_1(t)$ is expressed as

$$e_1(t) = R(t) \cos [\omega_c t + \theta + \psi(t)] \tag{6.95}$$

the *phase deviation* of the discriminator input is

$$\psi(t) = \phi(t) + \phi_e(t) \tag{6.96}$$

If the predetection signal-to-noise ratio $(\text{SNR})_T$ is large, $A_c \gg r_n(t)$ *most of the time*. For this case, (6.94) becomes

$$\phi_e(t) \cong \frac{r_n(t)}{A_c} \sin [\phi_n(t) - \phi(t)] \tag{6.97}$$

so that $\psi(t)$ is

$$\psi(t) = \phi(t) + \frac{r_n(t)}{A_c} \sin [\phi_n(t) - \phi(t)] \tag{6.98}$$

It is important to note that the effect of the noise $r_n(t)$ is reduced if the transmitted signal amplitude A_c is increased. Thus the output noise is affected by the transmitted signal amplitude even for above-threshold operation.

If the predetection SNR is small, $A_c \ll r_n(t)$ most of the time, and the phasor diagram appears as shown in Figure 6.13. For this case, the expression for $\psi(t)$ can be found from

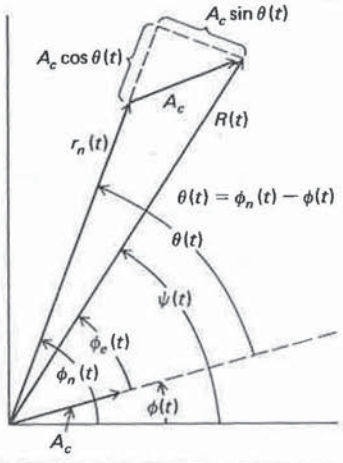
$$r_n(t)[\phi_n(t) - \psi(t)] \cong A_c \sin [\phi_n(t) - \phi(t)] \tag{6.99}$$

or

$$\psi(t) = \phi_n(t) - \frac{A_c}{r_n(t)} \sin [\phi_n(t) - \phi(t)] \tag{6.100}$$

Several comments concerning (6.98) and (6.100) are in order. The signal $m(t)$ is contained in $\phi(t)$, and $r_n(t)$ contains only noise. For the case of a large

FIGURE 6.13 Phasor diagram for angle demodulation, assuming $(\text{SNR})_T < 1$ (drawn for $\theta = 0$)



SNR, the phase of the predetection filter ouptut $\psi(t)$ is the sum of two terms. The first term is *signal alone,* and the second term contains components of both signal and noise. For the case of a small SNR, there is no term in $\psi(t)$ that can be identified as signal only. Actually, for small SNR, the only term containing the signal is *multiplied* by the noise. Thus the FM demodulator exhibits a threshold effect similar to that of the AM envelope detector. This effect will be examined in more detail in a later section.

The discriminator is assumed to contain a limiter and thus is insensitive to amplitude variations. Thus $R(t)$ can be assumed constant in (6.95). The discriminator output $y_D(t)$ for the PM case is given by

$$y_D(t) = K_D\psi(t) \tag{6.101}$$

and for the FM case is given by

$$y_D(t) = \frac{1}{2\pi} K_D \frac{d\psi}{dt} \tag{6.102}$$

where K_D is the discriminator constant. Substituting (6.96) into the preceding expression yields, *for the case of a large SNR,*

$$y_{DP}(t) = K_D\phi(t) + n_P(t) \tag{6.103}$$

for PM, and

$$y_{DF}(t) = \frac{1}{2\pi} K_D \frac{d\phi}{dt} + n_F(t) \tag{6.104}$$

for FM, where

$$n_P(t) = K_D\phi_e(t) = K_D \frac{r_n(t)}{A_c} \sin [\phi_n(t) - \phi(t)] \tag{6.105}$$

and

$$n_F(t) = \frac{K_D}{2\pi} \frac{d\phi_e}{dt} = \frac{K_D}{2\pi} \frac{d}{dt} \left\{ \frac{r_n(t)}{A_c} \sin [\phi_n(t) - \phi(t)] \right\} \tag{6.106}$$

are the output noise components for PM and FM, respectively. These outputs can be written in terms of the message signal as

$$y_{DP}(t) = K_D k_p m_n(t) + n_P(t) \tag{6.107}$$

and

$$y_{DF}(t) = K_D f_d m_n(t) + n_F(t) \tag{6.108}$$

The output signal power for PM is

$$S_{DP} = K_D{}^2 k_p{}^2 \overline{m_n^2} \tag{6.109}$$

and for FM is

$$S_{DF} = K_D^2 f_d^2 \overline{m_n^2} \tag{6.110}$$

Before the output signal-to-noise ratios can be calculated, the power spectral density of the output noise must be determined.

Output Noise Spectra

The power spectral density of the noise present at the discriminator output can, for the case of a large SNR, be computed from (6.105) and (6.106). As a first step in the analysis, we will set $\phi(t)$ equal to zero, so that $n_P(t)$ and $n_F(t)$ are functions of noise alone and contain no signal component. This assumption will greatly simplify our analysis, but it might seem that we are perhaps taking too many liberties. We could perform the analysis without assuming $\phi(t)$ equal to zero. The derivation would be much more complex and we would find that the effect of $\phi(t)$ is to produce frequency components in the demodulated output for $f > W$.* However, they can be removed by placing a lowpass filter in cascade with the discriminator output.

With $\phi(t) = 0$, the term $r_n(t) \sin [\phi_n(t) - \phi(t)]$ is equal to $n_s(t)$. Thus

$$n_P(t) = K_D \frac{n_s(t)}{A_c} \tag{6.111}$$

and

$$n_F(t) = \frac{K_D}{2\pi A_c} \frac{dn_s(t)}{dt} \tag{6.112}$$

From (6.111) it follows that the output noise power spectral density, for the PM case, is

$$S_{nP}(f) = \frac{K_D^2}{A_c^2} S_{ns}(f) = \frac{K_D^2}{A_c^2} N_0 \tag{6.113}$$

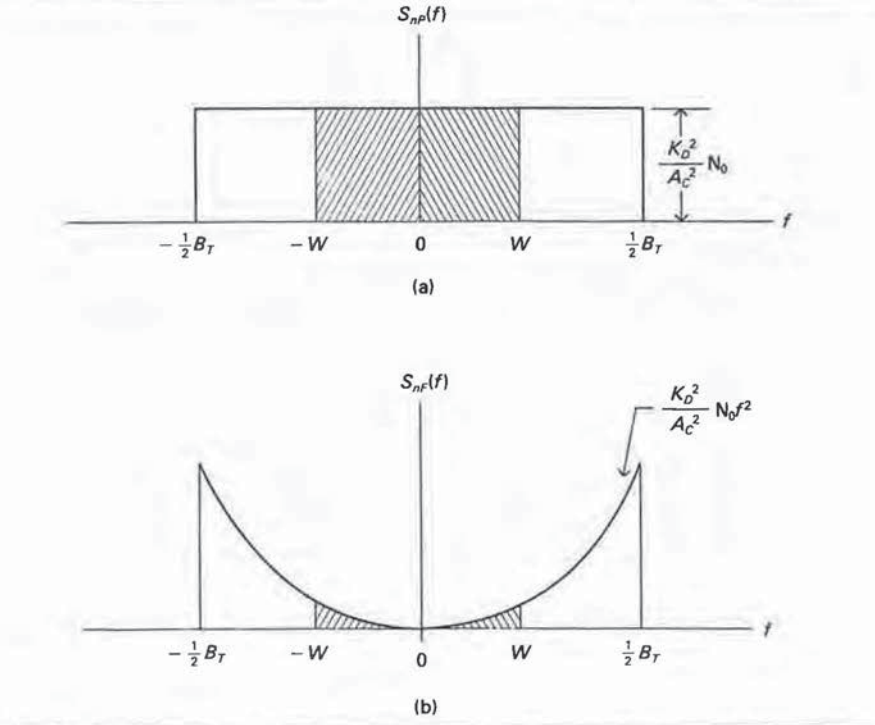
for $|f| < \frac{1}{2}B_T$ and zero otherwise, as illustrated in Figure 6.14(a). Thus the output noise power at the discriminator output is

$$N_{DP} = \frac{K_D^2}{A_c^2} \int_{-B_T/2}^{B_T/2} N_0 \, df = \frac{K_D^2}{A_c^2} N_0 B_T \tag{6.114}$$

The signal bandwidth is W. Since the predetection bandwidth B_T is greater than $2W$, the output SNR can be improved by following the discriminator

* An analysis that does not assume $\phi(t) = 0$ is contained in Downing (1964), pp. 96–98.

FIGURE 6.14 (a) Power spectral density for PM discriminator output, with portion for $|f| < W$ shaded. (b) Power spectral density for FM discriminator output, with portion for $|f| < W$ shaded.



with a lowpass filter of bandwidth W. This filter has no effect on the signal but reduces the output noise power to

$$N_{DP} = \frac{K_D^2}{A_c^2} \int_{-W}^{W} N_0 \, df = 2 \frac{K_D^2}{A_c^2} N_0 W \tag{6.115}$$

The output noise power is slightly more difficult to compute for the FM case because of the differentiation. It was shown in Chapter 5 that if $y(t) = dx/dt$, then $S_y(f) = (2\pi f)^2 S_x(f)$. Applying this result to (6.112) yields

$$S_{nF}(f) = \frac{K_D^2}{(2\pi)^2 A_c^2} (2\pi f)^2 N_0 = \frac{K_D^2}{A_c^2} N_0 f^2 \tag{6.116}$$

for $|f| < \frac{1}{2}B_T$ and zero otherwise. This spectrum is illustrated in Figure 6.14(b). The parabolic shape of the noise spectrum results from the differentiating action of the FM discriminator and has a profound effect on the performance of FM systems operating in the presence of noise. It is clear

from Figure 6.14(b) that low-frequency message-signal components are subjected to lower noise levels than are higher-frequency components. Once again, assuming that a lowpass filter having only sufficient bandwidth to pass the message follows the discriminator, the output noise power is

$$N_{DF} = \frac{K_D^2}{A_c^2} N_0 \int_{-W}^{W} f^2 \, df = \frac{2}{3} \frac{K_D^2}{A_c^2} N_0 W^3 \tag{6.117}$$

This quantity is indicated by the shaded area in Figure 6.14(b).

It is useful to write (6.115) for PM and (6.117) for FM in terms of P_T/N_0W. Since $P_T = A_c^2/2$ for both PM and FM, we have

$$\frac{P_T}{N_0 W} = \frac{A_c^2}{2N_0 W} \tag{6.118}$$

Thus, from (6.115) the discriminator output noise power for PM is

$$N_{DP} = K_D^2 \left(\frac{P_T}{N_0 W} \right)^{-1} \tag{6.119}$$

and from (6.117) the discriminator output noise power for FM is

$$N_{DF} = \tfrac{1}{3} K_D^2 W^2 \left(\frac{P_T}{N_0 W} \right)^{-1} \tag{6.120}$$

Note that for both PM and FM the noise power at the discriminator output is inversely proportional to P_T/N_0W.

Output Signal-to-Noise Ratios

The signal-to-noise ratio at the discriminator output is now easily determined. For the PM system, the discriminator output SNR is the signal power, defined by (6.109), divided by the noise power, defined by (6.119). This gives

$$(\text{SNR})_{DP} = \frac{K_D^2 k_p^2 \overline{m_n^2}}{(K_D^2 (P_T/N_0 W)^{-1}} \tag{6.121}$$

which gives

$$(\text{SNR})_{DP} = k_p^2 \overline{m_n^2} \frac{P_T}{N_0 W} \tag{6.122}$$

The above expression shows that the improvement of PM over linear modulation depends on the phase-deviation constant and the power in the modulating signal. It should be remembered that if the phase deviation of a PM

signal exceeds π radians, unique demodulation cannot be accomplished. Thus the peak value of $|k_p m_n(t)|$ is π, and the maximum value of $k_p^2 \overline{m_n^2}$ is π^2. This yields a *maximum* improvement of approximately 10 dB over baseband. In reality, the improvement is significantly less because $k_p^2 \overline{m_n^2}$ is typically much less than the maximum value of π^2. It should be pointed out that if the constraint that the output of the phase demodulator is continuous is imposed, then it is possible for $|k_p m_n(t)|$ to exceed π.

We now turn our attention to the FM case. For FM the signal-to-noise ratio at the discriminator output is the signal power, defined by (6.110) divided by the noise power, defined by (6.120). This gives

$$(\text{SNR})_{DF} = \frac{K_D^2 f_d^2 \overline{m_n^2}}{\frac{1}{3} K_D^2 \left(\frac{P_T}{N_0 W}\right)^{-1}} \tag{6.123}$$

which can be expressed as

$$(\text{SNR})_{DF} = 3 \left(\frac{f_d}{W}\right)^2 \overline{m_n^2} \frac{P_T}{N_0 W} \tag{6.124}$$

where P_T is the transmitted signal power $\frac{1}{2} A_c^2$. Since the ratio of peak deviation to W is the deviation ratio D, the output SNR is

$$(\text{SNR})_{DF} = 3D^2 \overline{m_n^2} \frac{P_T}{N_0 W} \tag{6.125}$$

where the maximum value of $|m_n(t)|$ is unity. Note that the maximum value of $m(t)$, together with f_d and W, determines D.

For the FM case, there is no restriction on the magnitude of the output SNR as there was for PM. As a matter of fact, it appears that we can increase D without bound, thereby increasing the output SNR to an arbitrarily large value. One price we pay for this enlarged SNR is excessive transmission bandwidth. For $D \gg 1$, the required bandwidth B_T is approximately $2DW$, which yields

$$(\text{SNR})_{DF} = \frac{3}{4} \left(\frac{B_T}{W}\right)^2 \overline{m_n^2} \left(\frac{P_T}{N_0 W}\right) \tag{6.126}$$

This expression illustrates the tradeoff that exists between bandwidth and output SNR. However, (6.126) is valid only if the discriminator input SNR is sufficiently large to result in operation *above threshold*. Thus the output SNR *cannot* be increased to any arbitrary value by simply increasing the deviation ratio and thus the transmission bandwidth. This effect will be studied in detail in a later section. First, however, we will study a simple technique for gaining additional improvement in the output SNR.

Performance Enhancement Through the Use of Deemphasis

In Chapter 3 we saw that preemphasis and deemphasis can be used to partially combat the effects of interference. These techniques can also be used to great advantage when noise is present in angle modulation systems.

As we saw in Chapter 3, the deemphasis filter is usually a first-order lowpass RC filter placed directly at the discriminator output. Prior to modulation, the signal is passed through a preemphasis filter with a transfer function so that the combination of the preemphasis and deemphasis filters has no net effect on the message signal. The deemphasis filter is followed by a lowpass filter, assumed to be ideal with bandwidth W, which eliminates the out-of-band noise. Assume the deemphasis filter to have the amplitude response

$$|H_{DE}(f)| = \frac{1}{\sqrt{1 + (f/f_3)^2}} \tag{6.127}$$

where f_3 is the 3-dB frequency $1/(2\pi RC)$ Hz. The total noise power output with deemphasis is

$$N_{DF} = \int_{-W}^{W} |H_{DE}(f)|^2 S_{nF}(f)\, df \tag{6.128}$$

Substituting $S_{nF}(f)$ from (6.116) and $|H_{DE}(f)|$ from (6.127) yields

$$N_{DF} = \frac{K_D^2}{A_c^2} N_0 f_3^2 \int_{-W}^{W} \frac{f^2}{f_3^2 + f^2}\, df \tag{6.129}$$

or

$$N_{DF} = 2\frac{K_D^2}{A_c^2} N_0 f_3^3 \left(\frac{W}{f_3} - \tan^{-1}\frac{W}{f_3}\right) \tag{6.130}$$

In a typical situation, $f_3 \ll W$, so $\tan^{-1}(W/f_3) \cong \frac{1}{2}\pi$, which is small compared to W/f_3. For this case,

$$N_{DF} = 2\left(\frac{K_D^2}{A_c^2}\right) N_0 f_3^2 W \tag{6.131}$$

and the output SNR becomes

$$(\text{SNR})_{DF} = \left(\frac{f_d}{f_3}\right)^2 \frac{\overline{m_n^2}}{} \frac{P_T}{N_0 W} \tag{6.132}$$

A comparison of (6.132) with (6.124) illustrates that for $f_3 \ll W$, the improvement gained through the use of preemphasis and deemphasis can be very significant in noisy environments.

EXAMPLE 6.3 Commerical FM operates with $f_d = 75$ kHz, $W = 15$ kHz, $D = 5$, and the standard value of 2.1 kHz for f_3. Assuming that $\overline{m_n^2} = 0.1$ we have, for FM without deemphasis,

$$(\text{SNR})_{DF} = 7.5 \frac{P_T}{N_0 W} \tag{6.133}$$

With deemphasis, the result is

$$(\text{SNR})_{DF,pe} = 128 \frac{P_T}{N_0 W} \tag{6.134}$$

With the chosen values, FM without deemphasis is 8.75 dB superior to baseband, and FM with deemphasis is 21 dB superior. This improvement more than justifies the use of deemphasis. ▲

As mentioned in Chapter 3, a price is paid for the SNR improvement gained by the use of preemphasis. The action of the preemphasis filter is to accentuate the high-frequency portion of the message signal. Thus preemphasis may increase the transmitter deviation and, consequently, the bandwidth required for signal transmission. Fortunately, many message signals of practical interest have relatively small energy in the high-frequency portion of their spectrum, so this effect is of little or no importance.

6.4 THRESHOLDS AND THRESHOLD EXTENSION IN FM

In Section 6.3 we observed that an FM discriminator exhibits a threshold effect. We now take a closer look at the threshold effect, and we examine techniques that can be used to improve the noise performance of FM systems by extending the threshold.

Threshold Effects in FM Discriminators

Significant insight into the mechanism by which threshold effects take place can be gained by performing a relatively simple laboratory experiment. On the input to an FM discriminator is placed an unmodulated sinusoid plus additive, bandlimited white noise having a power spectral density symmetrical about the frequency of the sinusoid. Starting out with a high SNR at the discriminator input, the noise power is gradually increased while continually observing the discriminator output on an oscilloscope. Initially, the discriminator output resembles bandlimited white noise. As the noise power spectral density is increased, thereby reducing the input SNR, a point is reached at which spikes or impulses appear in the discriminator output. The

initial appearance of these spikes denotes that the discriminator is operating in the region of threshold.

The statistics for these spikes were examined in Section 5.6. In this section we review the phenomenon of spike noise with specific application to FM demodulation. The system under consideration is that of Figure 6.11. For this case,

$$e_1(t) = A_c \cos(\omega_c t + \theta) + n_c(t) \cos(\omega_c t + \theta) - n_s(t) \sin(\omega_c t + \theta) \tag{6.135}$$

which is

$$e_1(t) = A_c \cos(\omega_c t + \theta) + r_n(t) \cos[\omega_c t + \theta + \phi_n(t)] \tag{6.136}$$

or

$$e_1(t) = R(t) \cos[\omega_c t + \theta + \psi(t)] \tag{6.137}$$

The phasor diagram of this signal is given in Figure 6.15. Like Figure 5.15, it illustrates the mechanism by which spikes occur. The signal amplitude is A_c and the angle is θ, since the carrier is assumed unmodulated. The noise amplitude is $r_n(t)$. The angle difference between signal and noise is $\phi_n(t)$. As threshold is approached, the noise amplitude grows until, at least part of the time, $|r_n(t)| > A_c$. Also, since $\phi_n(t)$ is uniformally distributed, the phase of the noise is sometimes in the region of $-\pi$. Thus the resultant phasor $R(t)$ can occasionally encircle the origin. When $R(t)$ is in the region of the origin, a relatively small change in the amplitude or phase of the noise results in a rapid change in $\psi(t)$. Since the discriminator output is proportional to the time rate of change $\psi(t)$, the discriminator output is very large as the origin is encircled. This is essentially the same effect that was observed in Chapter

FIGURE 6.15 Phasor diagram near threshold (spike output) (drawn for $\theta = 0$)

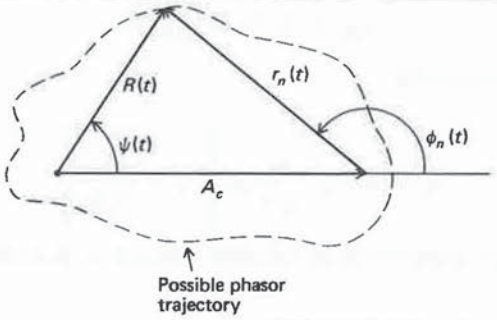
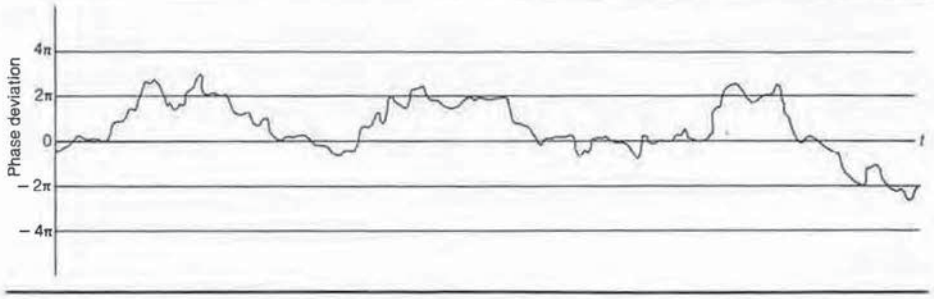


FIGURE 6.16 Phase deviation for a predetection signal-to-noise ratio of −4.0 dB



3 where the behavior of an FM discriminator operating in the presence of interference was studied.

The phase deviation $\psi(t)$ is illustrated in Figure 6.16 for the case in which the input signal-to-noise ratio is −4.0 dB. The origin encirclements can be observed by the 2π jumps in $\psi(t)$. The output of an FM discriminator for several predetection signal-to-noise ratios is shown in Figure 6.17. The decrease in spike noise as the signal-to-noise ratio is increased is clearly seen.

In Section 5.6 we showed that the power spectral density of spike noise is given by

$$S_{\dot{\psi}}(f) = (2\pi)^2(\nu + \overline{\delta\nu}) \tag{6.138}$$

where ν is the average number of impulses per second of an unmodulated carrier plus noise and $\delta\nu$ is the net increase of the spike rate due to modulation. Since the discriminator output is given by

$$y_D(t) = \frac{1}{2\pi} K_D \frac{d\psi}{dt} \tag{6.139}$$

The power spectral density due to spike noise at the discriminator output is

$$S_{D\delta} = K_D{}^2\nu + K_D{}^2\overline{\delta\nu} \tag{6.140}$$

Using (5.159) for ν yields

$$K_D{}^2\nu = K_D{}^2 \frac{B_T}{\sqrt{3}} Q\left[\sqrt{\frac{A_c{}^2}{N_0 B_T}}\right] \tag{6.141}$$

where $Q(x)$ is the probability Q-function defined in (4.186). Using (5.164) for $\overline{\delta\nu}$ yields

FIGURE 6.17 Output of FM discriminator due to input noise for various predetection signal-to-noise ratios. (a) Predetection SNR = −10 dB. (b) Predectection SNR = −4 dB. (c) Predetection SNR = 0 dB. (d) Predetection SNR = 6 dB. (e) Predetection SNR = 10 dB.



$$K_D^2 \delta\nu = K_D^2 |\overline{\delta f}| \exp\left[\frac{-A_c^2}{2N_0 B_T}\right] \tag{6.142}$$

Since the spike noise at the discriminator output is white, the spike noise power at the discriminator output is found by multiplying the power spectral density by the two-sided postdetection bandwidth, $2W$. Substituting (6.141)

and (6.142) into (6.140) and multiplying by $2W$ yields

$$N_{D\delta} = K_D^2 \frac{2B_T W}{\sqrt{3}} Q\left[\sqrt{\frac{A_c^2}{N_0 B_T}}\right] + K_D^2 (2W)|\overline{\delta f}| \exp\left[\frac{-A_c^2}{2N_0 B_T}\right] \quad (6.143)$$

for the spike noise power. Now that the spike noise power is known we can determine the total noise power at the discriminator output. After this is accomplished the output signal-to-noise ratio at the discriminator output is easily determined.

The total noise power at the discriminator output is the sum of the Gaussian noise power and spike noise power. The total noise power is therefore found by adding (6.143) to (6.120). This gives

$$N_D = \tfrac{1}{3} K_D^2 W^2 \left(\frac{P_T}{N_0 W}\right)^{-1} + K_D^2 \frac{2B_T W}{\sqrt{3}} Q\left[\sqrt{\frac{A_c^2}{N_0 B_T}}\right] + K_D^2 (2W)|\overline{\delta f}| \exp\left[-\frac{A_c^2}{2N_0 B_T}\right] \quad (6.144)$$

The signal power at the discriminator output is given by (6.110). Dividing the signal power by the noise power given above yields, after canceling the K_D^2 terms,

$$(\mathrm{SNR})_D = \frac{f_d^2 \overline{m_n^2}}{\tfrac{1}{3} W^2 \left(\frac{P_T}{N_0 W}\right)^{-1} + \frac{2B_T W}{\sqrt{3}} Q\left[\sqrt{\frac{A_n^2}{N_0 B_T}}\right] + 2W|\overline{\delta f}| \exp\left[\frac{-A_c^2}{2N_0 B_T}\right]} \quad (6.145)$$

This result can be placed in standard form by dividing numerator and denominator by the leading term in the denominator. This gives the final result

$$(\mathrm{SNR})_D = \frac{3\left(\frac{f_d}{W}\right)^2 \overline{m_n^2} \frac{P_T}{N_0 W}}{1 + 2\sqrt{3}\frac{B_T}{W}\left(\frac{P_T}{N_0 W}\right) Q\left[\sqrt{\frac{A_c^2}{N_0 B_T}}\right] + 6\frac{|\overline{\delta f}|}{W}\left(\frac{P_T}{N_0 W}\right) \exp\left[\frac{-A_c^2}{2N_0 B_T}\right]} \quad (6.146)$$

For operation above threshold, the region of input signal-to-noise ratios where spike noise is negligible, the last two terms in the denominator of the preceding expression are much less than one and may therefore be neglected. For this case the postdetection signal-to-noise ratio defined by the preceding expression is equivalent to the above threshold result given by (6.124). It is worth noting that the quantity $A_c^2/(2N_0B_T)$ appearing in each of the two spike noise terms is the predetection signal-to-noise ratio. Note that the message signal explicitly affects two terms in the expression for the postdetection signal-to-noise ratio through $\overline{m_n^2}$ and $|\overline{\delta f}|$. Thus, before the detection gain can be determined, a message signal must be assumed. This is the subject of the following example.

▼ EXAMPLE 6.4 In this example the detection gain of an FM discriminator is determined assuming the sinusoidal message signal

$$m_n(t) = \sin 2\pi Wt \tag{6.147}$$

The instantaneous frequency deviation is given by

$$f_d m_n(t) = f_d \sin 2\pi Wt \tag{6.148}$$

and the average of the absolute value of the frequency deviation is therefore given by

$$|\overline{\delta f}| = 2W \int_0^{1/2W} f_d \sin 2\pi Wt \, dt \tag{6.149}$$

Carrying out the integration yields

$$|\overline{\sigma f}| = \frac{2}{\pi}(f_d) \tag{6.150}$$

Note that f_d is the peak frequency deviation which, by definition of the modulation index, β, is βW. Thus

$$|\overline{\delta f}| = \frac{2}{\pi}\beta W \tag{6.151}$$

From Carson's rule we have

$$\frac{B_T}{W} = 2(\beta + 1) \tag{6.152}$$

Since the message signal is sinusoidal

$$\left(\frac{f_d}{W}\right)^2 \overline{m_n^2} = \tfrac{1}{2}\left(\frac{f_d}{W}\right)^2 = \tfrac{1}{2}\beta^2 \tag{6.153}$$

Finally, the predetection signal-to-noise ratio can be written

$$\frac{A_c^2}{2N_0B_T} = \frac{1}{2(\beta + 1)} \frac{P_T}{N_0W} \tag{6.154}$$

Substituting (6.153) and (6.154) into (6.146) yields

$$(\text{SNR})_D = \frac{1.5\beta^2 \dfrac{P_T}{N_0W}}{1 + 4\sqrt{3}(\beta + 1)\left(\dfrac{P_T}{N_0W}\right) Q\left[\sqrt{\dfrac{1}{(\beta + 1)} \dfrac{P_T}{N_0W}}\right] + \dfrac{12}{\pi}\beta\left(\dfrac{P_T}{N_0W}\right) \exp\left[\dfrac{-1}{2(\beta + 1)} \dfrac{P_T}{N_0W}\right]} \tag{6.155}$$

for the postdetection signal-to-noise ratio.

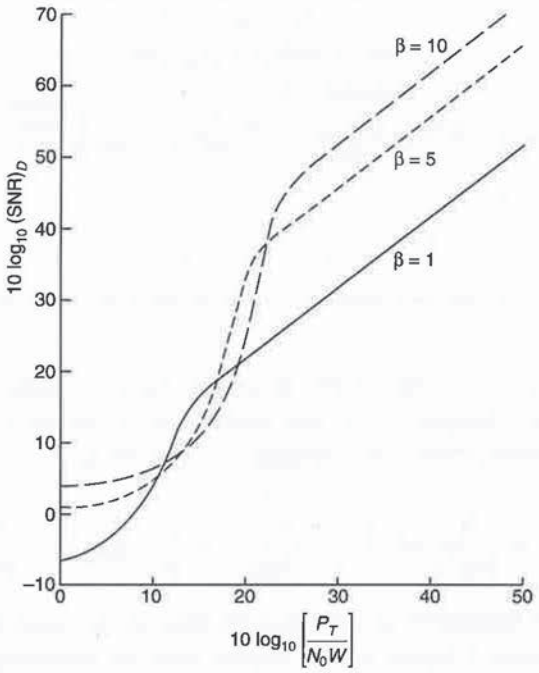
The postdetection signal-to-noise ratio is illustrated in Figure 6.18 as a function of P_T/N_0W. The threshold value of P_T/N_0W is the value of P_T/N_0W at which the postdetection SNR is 3 dB below the value of the postdetection SNR given by the above threshold analysis. In other words, the threshold value of P_T/N_0W is the value of P_T/N_0W for which the denominator of (6.155) is equal to 2. It should be noted from Figure 6.18 that the threshold value of P_T/N_0W increases as the modulation index β increases. The study of this effect is the subject of one of the computer exercises at the end of this chapter. ▲

In most communication systems it is important to maintain the operating point above threshold. There are two basic purposes served by Figure 6.18. First, we see the rapid convergence to the result of the above threshold analysis described by (6.124), with (6.153) used to allow (6.124) to be written in terms of the modulation index. The second important point emphasized by Figure 6.18 is the rapid deterioration of performance as the operating point moves into the below-threshold region.

Threshold Extension

There are several techniques that can be used for lowering the value of P_T/N_0W at which threshold occurs. Two systems, the frequency-compressive feedback loop and the phase-lock loop, have found widespread application in high-noise environments. Both these systems were presented briefly in Chapter 3.

FIGURE 6.18 FM system performance with sinusoidal modulation



The block diagram of a frequency-compressive feedback loop is repeated in Figure 6.19. The input to the loop is assumed to be

$$x_r(t) = A_c \cos [\omega_c t + \psi(t)] \tag{6.156}$$

where $\psi(t)$ represents the phase deviation of the input resulting from both modulation and noise. Amplitude perturbations of $x_r(t)$ due to noise are neglected because they do not affect the discriminator output. In Chapter 3 we showed that the input to the discriminator can be written as

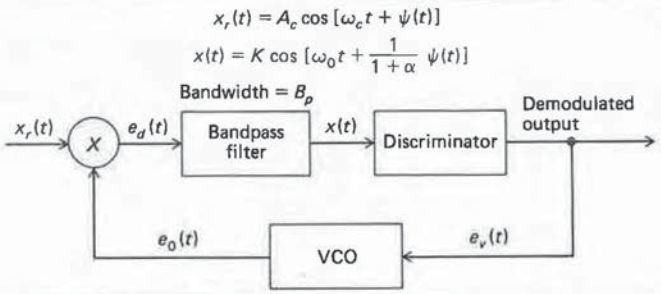
$$x(t) = K \cos \left[\omega_0 t + \frac{1}{1 + \alpha} \psi(t) \right] \tag{6.157}$$

where K is a constant related to the input amplitude and the VCO output amplitude and α is the loop gain. This is given, from (3.245), by

$$\alpha = \frac{1}{2\pi} K_D K_v \tag{6.158}$$

where $(1/2\pi)(K_D)$ is the discriminator gain and K_v is the gain of the VCO.

FIGURE 6.19 Frequency-compressive feedback loop



In the frequency-compressive feedback loop, a spike occurs when $[1/(1 + \alpha)]\psi(t)$ changes by 2π. The number of occurrences per second of this event, assuming sinusoidal modulation, is given by

$$\nu + \overline{\delta\nu} = \frac{B_p}{2\sqrt{3}} \operatorname{erfc}\left(\sqrt{\frac{A_c^2}{2N_0B_p}}\right) + \frac{2}{\pi} f_d \exp\left(-\frac{A_c^2}{2N_0B_p}\right) \quad (6.159)$$

where B_p is the bandwidth of the bandpass filter at the input to the discriminator. As shown in Chapter 3, B_p is smaller than the transmission bandwidth B_T due to bandwidth compression. Thus the rate of spikes is less with frequency-compressive feedback, and threshold is therefore extended.

Above threshold, (6.159) is negligible, and additive Gaussian noise is the only significant degradation of the output. For this case, both the signal and noise components are suppressed by $1 + \alpha$, as illustrated by (6.157). Thus, above threshold, the postdetection SNR is the same for the frequency-compressive feedback loop as for the ordinary discriminator.

The phase-lock loop also has been found to be useful for threshold extension. It is somewhat more difficult to analyze than the frequency-compressive feedback loop, and many developments have been published.* Thus we will not cover it here. We state, however, that the threshold extension obtained with the phase-lock loop is typically on the order of 2 to 3 dB. Even though this is a moderate extension, it can be important in high-noise environments.

6.5 NOISE IN PULSE MODULATION

We shall now investigate the performance of analog and digital pulse modulation systems in the presence of additive noise. We will see once again

* See Taub and Schilling (1986), pp. 419–422, for an introductory treatment.

that nonlinear systems allow a tradeoff between transmission bandwidth and postdetection signal-to-noise ratio.

Analog Pulse Modulation

For all analog pulse modulation techniques, demodulation involves the reconstruction of a continuous message signal from sample values that have been perturbed by noise. Assume that the received signal is of the form

$$x_r(t) = m_\delta(t) + n(t) \tag{6.160}$$

where $m_\delta(t)$ is the sampled message signal and $n(t)$ represents additive noise.

For pulse-amplitude modulation (PAM), the message signal is sampled, passed through the channel, and reconstructed using lowpass filtering. Thus, in effect, $m(t)$ is transmitted directly. Assuming that the output of the reconstruction filter is

$$y_D(t) = m(t) + n_0(t) \tag{6.161}$$

the maximum postdetection SNR is

$$(\text{SNR})_D = \frac{\overline{m^2(t)}}{N_0 W} \tag{6.162}$$

where W is the bandwidth of both $m(t)$ and the lowpass reconstruction filter. Thus PAM is essentially equivalent to continuous baseband transmission.

In both pulse-width modulation (PWM) and pulse-position modulation (PPM), the information resides in the position of pulse edges. Demodulation is accomplished by detecting pulse edges that have been perturbed by additive noise. If the channel bandwidth were infinite, perfectly rectangular pulses could be transmitted. For this case, additive noise would not cause a time perturbation of the pulse edge, and demodulation could be accomplished without error. Thus, for PWM and PPM, we can expect to see a tradeoff between postdetection SNR and transmission bandwidth.

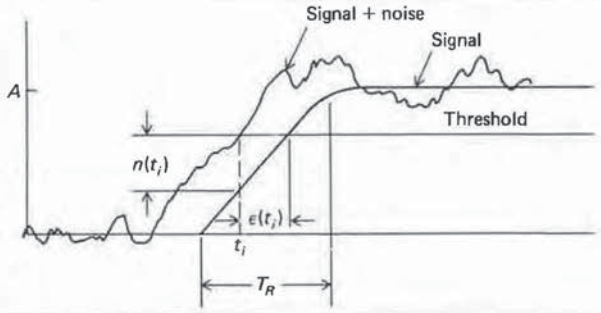
If the channel bandwidth is finite, the pulses will have nonzero risetime, and noise will perturb pulse detection. This is illustrated in Figure 6.20. Edge detection is accomplished by detecting the time at which the pulse crosses a threshold. The error $\epsilon(t_i)$ is related to the noise $n(t_i)$, the pulse amplitude A, and the risetime T_R by

$$\frac{\epsilon(t_i)}{T_R} = \frac{n(t_i)}{A} \tag{6.163}$$

Thus the mean-square error is given by

$$\overline{\epsilon^2(t_i)} = \left(\frac{T_R}{A}\right)^2 \overline{n^2(t_i)} \tag{6.164}$$

FIGURE 6.20 PWM and PPM demodulation in the presence of noise



which, assuming the noise process is stationary, is

$$\overline{\epsilon^2} = \left(\frac{T_R}{A}\right)^2 \overline{n^2} = \left(\frac{T_R}{A}\right)^2 N_0 B_T \tag{6.165}$$

where B_T is the transmission bandwidth.

The preceding expression can be placed in more useful form by recalling from Chapter 2 that risetime is related to bandwidth by

$$T_R \cong \frac{1}{2B_T} \tag{6.166}$$

so that

$$\overline{\epsilon^2} = \frac{N_0}{4A^2 B_T} \tag{6.167}$$

Assume that the pulse position, or pulse width, is defined by

$$\tau(t) = \tau_0 + \tau_1 m_n(t) \tag{6.168}$$

so that τ_1 represents the maximum pulse *displacement* for PPM and the maximum *change* in pulse width for PWM. Thus the signal power is

$$S_D = \tau_1^2 \overline{m_n^2} \tag{6.169}$$

If the sampling period is T_s, the average transmitted power is

$$P_T = \frac{\tau_0}{T_s} A^2 \tag{6.170}$$

since the average pulse duration is τ_0 if $m(t)$ is assumed to be zero mean. Substituting (6.170) into (6.167) yields

$$\overline{\epsilon^2} = \frac{N_0 \tau_0}{4 T_s P_T B_T} \tag{6.171}$$

which represents the postdetection noise power. Using (6.169) for signal power and (6.171) for noise power, the postdetection SNR becomes

$$(\text{SNR})_D = \frac{4\tau_1^2 \overline{m_n^2} T_s P_T B_T}{\tau_0 N_0} \tag{6.172}$$

which is, in terms of $P_T/N_0 W$,

$$(\text{SNR})_D = \frac{4\tau_1^2 \overline{m_n^2} T_s B_T W}{\tau_0} \left(\frac{P_T}{N_0 W} \right) \tag{6.173}$$

The tradeoff between bandwidth and postdetection SNR is evident from (6.173). As with FM, however, the transmission bandwidth cannot be increased without bound. Both PPM and PWM are nonlinear schemes and have definite thresholds. As B_T is increased, the predetection SNR is increased, and large instantaneous noise amplitudes are detected as pulse edges. Thus the preceding analysis is valid only for large signal-to-noise ratios. These systems will be revisited in Chapter 9, where a more complete analysis will be presented.

Pulse Code Modulation

Pulse code modulation (PCM) was briefly presented in Chapter 3. We now wish to analyze the performance of such systems.

There are two error sources in PCM, the error resulting from quantization and the error resulting from additive channel noise. As we saw in Chapter 3, quantization involves representing each input sample by one of q quantization levels. Each quantization level is then transmitted using a sequence of binary symbols to uniquely represent each quantization level.

The sampled and quantized message waveform can be represented as

$$m_{\delta q}(t) = \Sigma m(t)\, \delta(t - iT_s) + \Sigma \epsilon(t)\, \delta(t - iT_s) \tag{6.174}$$

where the first term represents the sampling operation and the second term represents quantization. The ith sample of $m_{\delta q}(t)$ is

$$m(t_i) + \epsilon(t_i)$$

where $t_i = iT_s$. Thus the signal-to-noise ratio resulting from quantization is

$$(\text{SNR})_Q = \frac{\overline{m^2(t_i)}}{\overline{\epsilon^2(t_i)}} = \frac{\overline{m^2(t)}}{\overline{\epsilon^2(t)}} \tag{6.175}$$

assuming stationarity. The quantization error term is easily evaluated for the case in which the quantization levels have uniform spacing S. The quantization error is bounded by $\pm\frac{1}{2}S$. Thus, assuming that $\epsilon(t)$ is uniformally distributed in the range

$$-\tfrac{1}{2}S \leq \epsilon(t_i) \leq \tfrac{1}{2}S$$

the mean-square error due to quantization is

$$\overline{\epsilon^2} = \frac{1}{S}\int_{-S/2}^{S/2} x^2\,dx = \tfrac{1}{12}S^2 \tag{6.176}$$

so that

$$(\text{SNR})_Q = 12\,\frac{\overline{m^2(t)}}{S^2} \tag{6.177}$$

The next step is to express $\overline{m^2(t)}$ in terms of q and S. If there are q quantization levels, each of width S, it follows that the peak-to-peak value of $m(t)$ is qS. Assuming that $m(t)$ is *uniformally* distributed in this range,

$$\overline{m^2(t)} = \tfrac{1}{12}q^2S^2 \tag{6.178}$$

Thus (6.177) becomes

$$(\text{SNR})_Q = q^2 = 2^{2n} \tag{6.179}$$

where n is the number of binary symbols used to represent each quantization level.

If the additive noise in the channel is so small that errors can be neglected, quantization is the only error source. Thus (6.179) becomes the postdetection SNR and is independent of P_T/N_0W. If quantization is not the only error source, the postdetection SNR depends not only on P_T/N_0W but also on the signaling scheme. An analysis of these various signaling schemes is the subject of Chapter 7.

An *approximate* analysis of PCM is easily carried out for the case in which the signal-to-noise ratio is assumed not to be large. Each sample value is transmitted as a group of n pulses, and any of these n pulses can be in error at the receiver due to noise effects in the channel. Chapter 7 deals with the calculation of the error probability as a function of both the received signal-to-noise ratio and the modulation technique used for transmission of the pulses. The group of n pulses defines the quantizing level and is referred to as a *digital word*. Each individual pulse is a digital symbol, or *bit*. We assume that the bit-error probability P_b is known (it will be after the next chapter). Each of the n bits in the digital word is received correctly with probability

$1 - P_b$. Assuming independence, the probability that all n bits are received correctly is $(1 - P_b)^n$. The word-error probability P_w is therefore given by

$$P_w = 1 - (1 - P_b)^n \tag{6.180}$$

The effect of a word error depends on which bit of the digital word is in error. If the least significant bit is in error, the resulting amplitude error is $\pm S$, or one quantizing level. If the next bit is in error, the resulting amplitude error is $\pm 2S$. The next bit results in an amplitude error of $\pm 4S$, and so forth. An error in the most significant bit results in an amplitude error of $\frac{1}{2}qS$. The effect of a word error is therefore an amplitude error in the range

$$-\tfrac{1}{2}qS \leq \epsilon_w \leq \tfrac{1}{2}qS$$

If we assume that ϵ_w is uniformally distributed in this range, then the mean-square word error is

$$\overline{\epsilon_w^2} = \tfrac{1}{12}q^2S^2 \tag{6.181}$$

which is equal to the signal power.

The total noise power at the output of a PCM system is given by

$$N_D = \overline{\epsilon^2}(1 - P_w) + \overline{\epsilon_w^2}P_w \tag{6.182}$$

The first term on the right-hand side of (6.182) is the contribution to the total noise due to quantizing error. It is equal to the quantizing error power weighted by the probability that all bits in the word are received correctly. The second term is the error power due to a word error weighted by the probability of word error. Using (6.182) for the signal power yields

$$(\text{SNR})_D = \frac{\frac{1}{12}q^2S^2}{\frac{1}{12}S^2(1 - P_w) + \frac{1}{12}S^2q^2P_w} \tag{6.183}$$

which can be written as

$$(\text{SNR})_D = \frac{1}{q^{-2}(1 - P_w) + P_w} \tag{6.184}$$

Equation (6.184) can be written in terms of the wordlength n using (6.179). The result is

$$(\text{SNR})_D = \frac{1}{2^{-2n} + P_w(1 - 2^{-2n})} \tag{6.185}$$

The term 2^{-2n} is completely determined by the number of quantizing levels, while the word-error probability P_w is largely determined by the signal-to-noise ratio P_T/N_0W.

▼ **EXAMPLE 6.5** The purpose of this example is to examine the postdetection signal-to-noise ratio for a PCM system. Before the postdetection signal-to-noise ratio, $(\text{SNR})_D$, can be numerically evaluated, the word-error probability, P_w, must be known. As shown by (6.180) the word-error probability depends upon the bit-error probability, the calculation of which is the subject of the following chapter. Borrowing a result from Chapter 7, however, will allow us to illustrate the threshold effect of PCM. If we assume frequency-shift keying (FSK), which, as was discussed in Chapter 3, is transmission using one frequency to represent a binary zero and a second frequency to represent a binary one, and a noncoherent receiver, the probability of bit error is

$$P_b = \tfrac{1}{2} \exp\left[-\frac{P_T}{2N_0 B_p}\right] \tag{6.186}$$

In the preceding expression B_p is the bit-rate bandwidth, which is the reciprocal of the time required for transmission of a single bit in the n-symbol PCM digital word. Substitution of (6.186) into (6.180) and substitution of the result into (6.185) yields the postdetection signal-to-noise ratio, $(\text{SNR})_D$. This result is shown in Figure 6.21. The threshold effect can be seen clearly.

Figure 6.21 shows that the postdetection signal-to-noise ratio approaches a value determined completely by the quantization noise. This can also be seen from (6.185) by letting the word-error probability tend to zero. The residual error, due to quantization, is defined by (6.179). Thus, above threshold,

$$(\text{SNR})_D \cong 2^{2n} \tag{6.187}$$

which, expressed in decibels, is

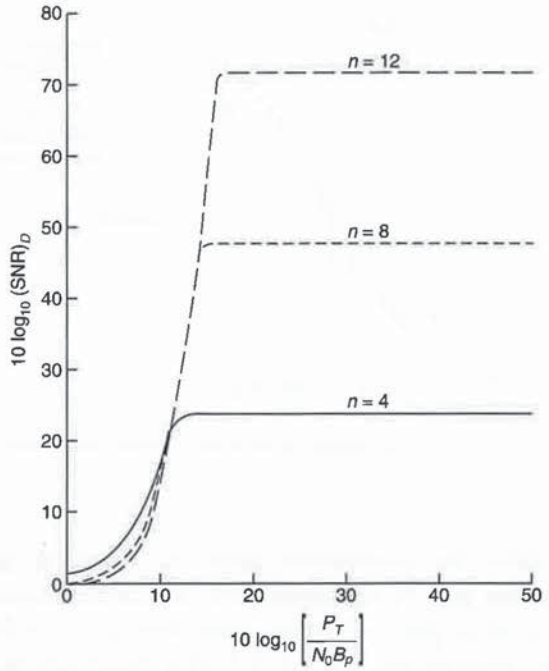
$$10 \log_{10} (\text{SNR})_D = 6.02n \tag{6.188}$$

We therefore gain 6 dB in signal-to-noise ratio for every bit added to the quantizer wordlength. However, the longer digital word means that more bits must be transmitted for each sample taken of the original time-domain signal, $m(t)$. This increases the bandwidth requirements of the system. Thus, the improved signal-to-noise ratio comes at the expense of a higher bit rate or system bandwidth. We therefore see again the tradeoff between signal-to-noise ratio and bandwidth. It is worth noting that, above threshold, the postdetection signal-to-noise ratio is independent of the modulation method. ▲

Companding

As we saw in Chapter 3, a PCM signal is formed by modulating a digital signal. The digital signal is generated from an analog signal through the

FIGURE 6.21 Signal-to-noise ratio at output of PCM system (FSK modulation used with noncoherent receiver)



operations of sampling the analog waveform, quantizing the resulting samples, and encoding the quantized samples into digital words of finite length. These three operations are collectively referred to as analog-to-digital (A/D) conversion. The inverse process of forming an analog signal from a digital signal is known as digital-to-analog (D/A) conversion.

In the preceding section we saw that significant errors can result from the quantizing process if the wordlength n is chosen too small for a particular application. The result of these errors is expressed by the signal-to-quantization-noise ratio expressed by (6.179). Keep in mind that (6.179) was developed for the case of a uniformly distributed signal.

The level of quantizing noise added to a given sample, Equation (6.176), is independent of the signal amplitude, and small amplitude signals will therefore suffer more from quantization effects than large amplitude signals. This can be seen from (6.177). There are essentially two ways to combat this problem. First, the quantizing steps can be made small for small amplitudes and large for large amplitude portions of the signal. This scheme is known as nonuniform quantizing. An example of a nonuniform quantizer is the Max