Seventh Edition

Principles of Communications

Systems, Modulation, and Noise

Rodger E. Ziemer
William H. Tranter

CHAPTER 9

PRINCIPLES OF DIGITAL DATA TRANSMISSION IN NOISE

In Chapter 8 we studied the effects of noise in analog communication systems. We now consider digital data modulation system performance in noise. Instead of being concerned with continuous-time, continuous-level message signals, we are concerned with the transmission of information from sources that produce discrete-valued symbols. That is, the input signal to the transmitter block of Figure 1.1 would be a signal that assumes only discrete values. Recall that we started the discussion of digital data transmission systems in Chapter 5, but without consideration of the effects of noise.

The purpose of this chapter is to consider various systems for the transmission of digital data and their relative performances. Before beginning, however, let us consider the block diagram of a digital data transmission system, shown in Figure 9.1, which is somewhat more detailed than Figure 1.1. The focus of our attention will be on the portion of the system between the optional blocks labeled *Encoder* and *Decoder*. In order to gain a better perspective of the overall problem of digital data transmission, we will briefly discuss the operations performed by the blocks shown as dashed lines.

As discussed previously in Chapters 4 and 5, while many sources result in message signals that are inherently digital, such as from computers, it is often advantageous to represent analog signals in digital form (referred to as *analog-to-digital conversion*) for transmission and then convert them back to analog form upon reception (referred to as *digital-to-analog conversion*), as discussed in the preceding chapter. Pulse-code modulation (PCM), introduced in Chapter 4, is an example of a modulation technique that can be employed to transmit analog messages in digital form. The signal-to-noise ratio performance characteristics of a PCM system, which were presented in Chapter 8, show one advantage of this system to be the option of exchanging bandwidth for signal-to-noise ratio improvement.[1]

Throughout most of this chapter we will make the assumption that source symbols occur with equal probability. Many discrete-time sources naturally produce symbols with equal probability. As an example, a binary computer file, which may be transmitted through a channel, frequently contains a nearly equal number of 1s and 0s. If source symbols do not occur with nearly equal probably, we will see in Chapter 12 that a process called source coding

[1] A device for converting voice signals from analog-to-digital and from digital-to-analog form is known as a *vocoder*.

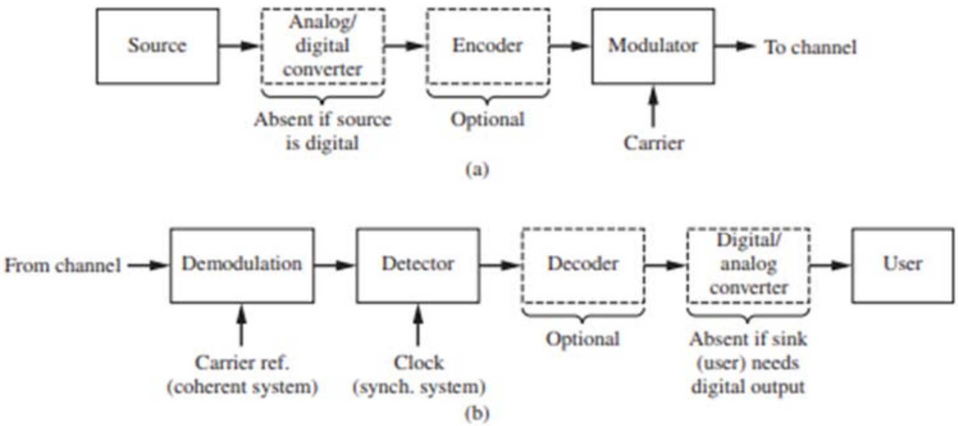

Figure 9.1
Block diagram of a digital data transmission system. (a) Transmitter. (b) Receiver.

(compression) can be used to create a new set of source symbols in which the binary states, 1 and 0, are equally likely, or nearly so. The mapping from the original set to the new set of source symbols is deterministic so that the original set of source symbols can be recovered from the data output at the receiver. The use of source coding is not restricted to binary sources. We will see in Chapter 12 that the transmission of equally likely symbols ensures that the information transmitted with each source symbol is maximized and, therefore, the channel is used efficiently. In order to understand the process of source coding, we need a rigorous definition of information, which will be accomplished in Chapter 12.

Regardless of whether a source is purely digital or an analog source that has been converted to digital, it may be advantageous to add or remove redundant digits to the digital signal. Such procedures, referred to as forward error-correction coding, are performed by the encoder-decoder blocks of Figure 9.1 and also will be considered in Chapter 12.

We now consider the basic system in Figure 9.1, shown as the blocks with solid lines. If the digital signals at the modulator input take on one of only two possible values, the communication system is referred to as *binary*. If one of $M > 2$ possible values is available, the system is referred to as M-ary. For long-distance transmission, these digital baseband signals from the source may modulate a carrier before transmission, as briefly mentioned in Chapter 5. The result is referred to as *amplitude-shift keying* (ASK), *phase-shift keying* (PSK), or *frequency-shift keying* (FSK) if it is amplitude, phase, or frequency, respectively, that is varied in accordance with the baseband signal. An important M-ary modulation scheme, *quadriphase-shift keying* (QPSK), is often employed in situations in which bandwidth efficiency is a consideration. Other schemes related to QPSK include offset QPSK and minimum-shift keying (MSK). These schemes will be discussed in Chapter 10.

A digital communication system is referred to as *coherent* if a local reference is available for demodulation that is in phase with the transmitted carrier (with fixed-phase shifts due to transmission delays accounted for). Otherwise, it is referred to as *noncoherent*. Likewise, if a periodic signal is available at the receiver that is in synchronism with the transmitted sequence of digital signals (referred to as a clock), the system is referred to as synchronous (i.e., the data streams at transmitter and receiver are in lockstep); if a signaling technique is employed in which such a clock is unnecessary (e.g., timing markers might be built into the data blocks, an examole being split phase as discussed in Chapter 5), the system is called *asynchronous*.

The primary measure of system performance for digital data communication systems is the probability of error, P_E. In this chapter we will obtain expressions for P_E for various types of digital communication systems. We are, of course, interested in receiver structures that give minimum P_E for given conditions. Synchronous detection in a white Gaussian-noise background requires a *correlation* or a *matched-filter* detector to give minimum P_E for fixed-signal and noise conditions.

We begin our consideration of digital data transmission systems in Section 9.1 with the analysis of a simple, synchronous baseband system that employs a special case of the matched-filter detector known as an *integrate-and-dump detector*. This analysis is then generalized in Section 9.2 to the matched-filter receiver, and these results are specialized to consideration of several coherent signaling schemes. Section 9.3 considers two schemes not requiring a coherent reference for demodulation. In Section 9.4, digital pulse-amplitude modulation is considered, which is an example of an M-ary modulation scheme. We will see that it allows the trade-off of bandwidth for P_E performance. Section 9.5 provides a comparison of the digital modulation schemes on the basis of power and bandwidth. After analyzing these modulation schemes, which operate in an ideal environment in the sense that infinite bandwidth is available, we look at zero-ISI signaling through bandlimited baseband channels in Section 9.6. In Sections 9.7 and 9.8, the effect of multipath interference and signal fading on data transmission is analyzed, and in Section 9.9, the use of equalizing filters to mitigate the effects of channel distortion is examined.

9.1 BASEBAND DATA TRANSMISSION IN WHITE GAUSSIAN NOISE

Consider the binary digital data communication system illustrated in Figure 9.2(a), in which the transmitted signal consists of a sequence of constant-amplitude pulses of either A or $-A$ units in amplitude and T seconds in duration. A typical transmitted sequence is shown in Figure 9.2(b). We may think of a positive pulse as representing a logic 1 and a negative pulse as representing a logic 0 from the data source. Each T-second pulse is called a *binit* for binary digit or, more simply, a *bit*. (In Chapter 12, the term *bit* will take on a more precise meaning.)

As in Chapter 8, the channel is idealized as simply adding white Gaussian noise with double-sided power spectral density $\frac{1}{2}N_0$ W/Hz to the signal. A typical sample function of the received signal plus noise is shown in Figure 9.2(c). For sketching purposes, it is assumed that the noise is bandlimited, although it is modeled as white noise later when the performance of the receiver is analyzed. It is assumed that the starting and ending times of each pulse are known by the receiver. The problem of acquiring this information, referred to as *synchronization*, will not be considered at this time.

The function of the receiver is to decide whether the transmitted signal was A or $-A$ during each bit period. A straightforward way of accomplishing this is to pass the signal-pulse noise through a lowpass predetection filter, sample its output at some time within each T-second interval, and determine the sign of the sample. If the sample is greater than zero, the decision is made that $+A$ was transmitted. If the sample is less than zero, the decision is that $-A$ was transmitted. With such a receiver structure, however, we do not take advantage of everything known about the signal. Since the starting and ending times of the pulses are known, a better procedure is to compare the area of the received signal-plus-noise waveform

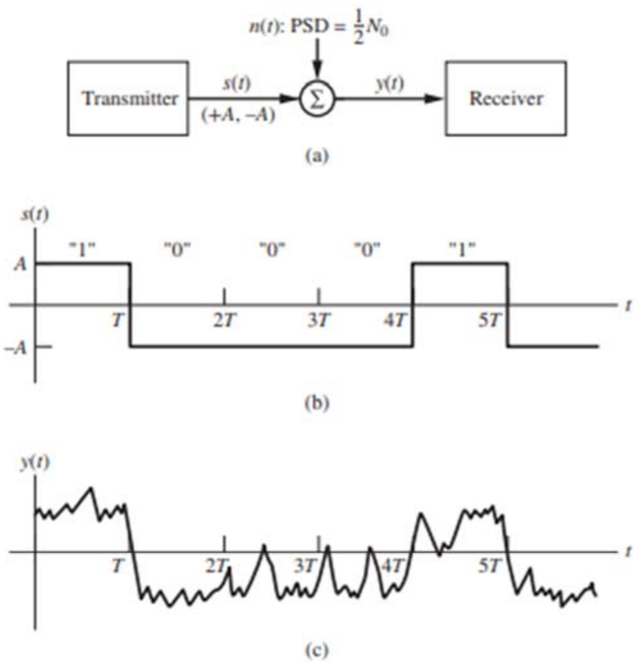

Figure 9.2
System model and waveforms for synchronous baseband digital data transmission. (a) Baseband digital data communication system. (b) Typical transmitted sequence. (c) Received sequence plus noise.

(data) with zero at the end of each signaling interval by integrating the received data over the T-second signaling interval. Of course, a noise component is present at the output of the integrator, but since the input noise has zero mean, it takes on positive and negative values with equal probability. Thus, the output noise component has zero mean. The proposed receiver structure and a typical waveform at the output of the integrator are shown in Figure 9.3 where t_0 is the start of an arbitrary signaling interval. For obvious reasons, this receiver is referred to as an integrate-and-dump detector because charge is dumped after each integration.

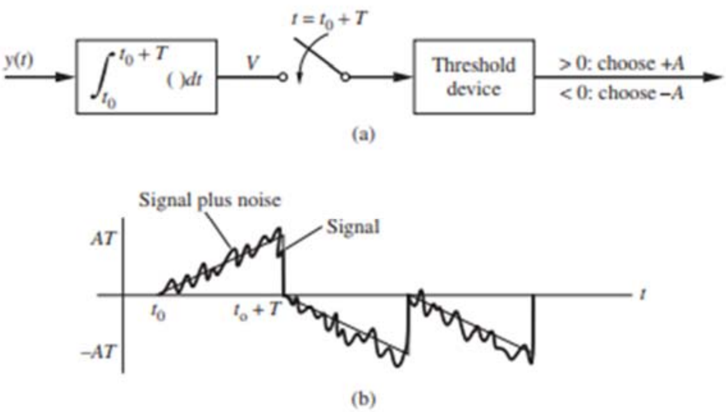

Figure 9.3
Receiver structure and integrator output. (a) Integrate-and-dump receiver. (b) Output from the integrator.

The question to be answered is: How well does this receiver perform, and on what parameters does its performance depend? As mentioned previously, a useful criterion of performance is probability of error, and it is this we now compute. The output of the integrator at the end of a signaling interval is

$$
\begin{aligned}
V &= \int_{t_0}^{t_0+T} [s(t) + n(t)]\, dt \\
&= \begin{cases} +AT + N & \text{if } +A \text{ is sent} \\ -AT + N & \text{if } -A \text{ is sent} \end{cases}
\end{aligned} \tag{9.1}
$$

where N is a random variable defined as

$$
N = \int_{t_0}^{t_0+T} n(t)\, dt \tag{9.2}
$$

Since N results from a linear operation on a sample function from a Gaussian process, it is a Gaussian random variable. It has mean

$$
E\{N\} = E\left\{\int_{t_0}^{t_0+T} n(t)\, dt\right\} = \int_{t_0}^{t_0+T} E\{n(t)\}\, dt = 0 \tag{9.3}
$$

since $n(t)$ has zero mean. Its variance is therefore

$$
\begin{aligned}
\operatorname{var}\{N\} &= E\left\{N^2\right\} = E\left\{\left[\int_{t_0}^{t_0+T} n(t)\, dt\right]^2\right\} \\
&= \int_{t_0}^{t_0+T}\int_{t_0}^{t_0+T} E\{n(t)n(\sigma)\}\, dt\, d\sigma \\
&= \int_{t_0}^{t_0+T}\int_{t_0}^{t_0+T} \frac{1}{2}N_0\delta(t-\sigma)\, dt\, d\sigma
\end{aligned} \tag{9.4}
$$

where we have made the substitution $E\{n(t)n(\sigma)\} = \frac{1}{2}N_0\delta(t-\sigma)$. Using the sifting property of the delta function, we obtain

$$
\begin{aligned}
\operatorname{var}\{N\} &= \int_{t_0}^{t_0+T} \frac{1}{2}N_0\, d\sigma \\
&= \frac{1}{2}N_0 T
\end{aligned} \tag{9.5}
$$

Thus, the pdf of N is

$$
f_N(\eta) = \frac{e^{-\eta^2/N_0T}}{\sqrt{\pi N_0 T}} \tag{9.6}
$$

where η is used as the dummy variable for N to avoid confusion with $n(t)$.

There are two ways in which errors occur. If $+A$ is transmitted, an error occurs if $AT + N < 0$, that is, if $N < -AT$. From (9.6), the probability of this event is

$$
P(\text{error}|A \text{ sent}) = P(E|A) = \int_{-\infty}^{-AT} \frac{e^{-\eta^2/N_0T}}{\sqrt{\pi N_0 T}}\, d\eta \tag{9.7}
$$

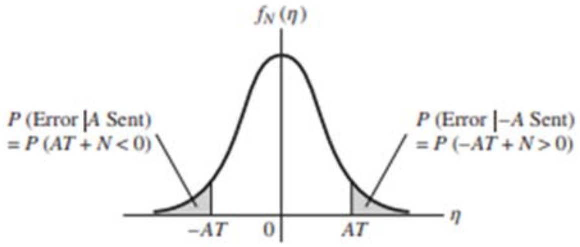

Figure 9.4
Illustration of error probabilities for binary signaling.

which is the area to the left of $\eta = -AT$ in Figure 9.4. Letting $u = \sqrt{2/N_0T}\eta$ and using the evenness of the integrand, we can write this as

$$P(E|A) = \int_{\sqrt{2A^2T/N_0}}^{\infty} \frac{e^{-u^2/2}}{\sqrt{2\pi}}\, du \triangleq Q\left(\sqrt{\frac{2A^2T}{N_0}}\right) \tag{9.8}$$

where $Q(\cdot)$ is the Q-function.[2] The other way in which an error can occur is if $-A$ is transmitted and $-AT + N > 0$. The probability of this event is the same as the probability that $N > AT$, which can be written as

$$P(E|-A) = \int_{AT}^{\infty} \frac{e^{-\eta^2/N_0T}}{\sqrt{\pi N_0 T}}\, d\eta \triangleq Q\left(\sqrt{\frac{2A^2T}{N_0}}\right) \tag{9.9}$$

which is the area to the right of $\eta = AT$ in Figure 9.4. The average probability of error is

$$P_E = P(E|+A)P(+A) + P(E|-A)P(-A) \tag{9.10}$$

Substituting (9.8) and (9.9) into (9.10) and noting that $P(+A) + P(-A) = 1$, where $P(A)$ is the probability that $+A$ is transmitted, we obtain

$$P_E = Q\left(\sqrt{\frac{2A^2T}{N_0}}\right) \tag{9.11}$$

Thus, the important parameter is A^2T/N_0. We can interpret this ratio in two ways. First, since the energy in each signal pulse is

$$E_b = \int_{t_0}^{t_0+T} A^2\, dt = A^2T \tag{9.12}$$

we see that the ratio of signal energy per pulse to noise power spectral density is

$$z = \frac{A^2T}{N_0} = \frac{E_b}{N_0} \tag{9.13}$$

where E_b is called the *energy per bit* because each signal pulse ($+A$ or $-A$) carries one *bit* of information. Second, we recall that a rectangular pulse of duration T seconds has amplitude

[2]See Appendix F.1 for a discussion and tabulation of the Q-function.

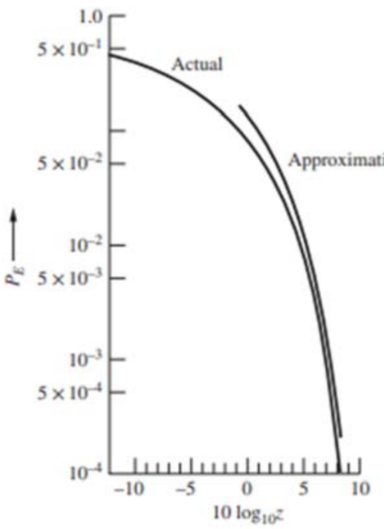

Figure 9.5
P_E for antipodal baseband digital signaling.

spectrum $AT\,\text{sinc}Tf$ and that $B_p = 1/T$ is a rough measure of its bandwidth. Thus,

$$\frac{E_b}{N_0} = \frac{A^2}{N_0(1/T)} = \frac{A^2}{N_0 B_p} \tag{9.14}$$

can be interpreted as the ratio of signal power to noise power in the signal bandwidth. The bandwidth B_p is sometimes referred to as the bit-rate bandwidth. We will refer to z as the signal-to-noise ratio (SNR). An often-used reference to this signal-to-noise ratio in the digital communications industry is ''e-b-over-n-naught.''[3]

A plot of P_E versus z is shown in Figure 9.5, where z is given in decibels. Also shown is an approximation for P_E using the asymptotic expansion for the Q-function:

$$Q(u) \cong \frac{e^{-u^2/2}}{u\sqrt{2\pi}}, \quad u \gg 1 \tag{9.15}$$

Using this approximation,

$$P_E \cong \frac{e^{-z}}{2\sqrt{\pi z}}, \quad z \gg 1 \tag{9.16}$$

which shows that P_E essentially decreases exponentially with increasing z. Figure 9.5 shows that the approximation of (9.16) is close to the true result of (9.11) for $z \gtrapprox 3$ dB.

[3]A somewhat curious term in use by some is ''ebno.''

EXAMPLE 9.1

Digital data is to be transmitted through a baseband system with $N_0 = 10^{-7}$ W/Hz and the received-signal amplitude $A = 20$ mV. (a) If 10^3 bits per second (bps) are transmitted, what is P_E? (b) If 10^4 bps are transmitted, to what value must A be adjusted in order to attain the same P_E as in part (a)?

Solution

To solve part (a), note that

$$z = \frac{A^2T}{N_0} = \frac{(0.02)^2(10^{-3})}{10^{-7}} = 4 \tag{9.17}$$

Using (9.16), $P_E \cong e^{-4}/2\sqrt{4\pi} = 2.58 \times 10^{-3}$. Part (b) is solved by finding A such that $A^2(10^{-4})/(10^{-7}) = 4$, which gives $A = 63.2$ mV.

■

EXAMPLE 9.2

The noise power spectral density is the same as in the preceding example, but a bandwidth of 5000 Hz is available. (a) What is the maximum data rate that can be supported by the channel? (b) Find the transmitter power required to give a probability of error of 10^{-6} at the data rate found in part (a).

Solution

(a) Since a rectangular pulse has Fourier transform

$$\Pi(t/T) \leftrightarrow T\,\text{sinc}(fT)$$

we take the signal bandwidth to be that of the first null of the sinc function. Therefore, $1/T = 5000$ Hz, which implies a maximum data rate of $R = 5000$ bps. (b) To find the transmitter power to give $P_E = 10^{-6}$, we solve

$$10^{-6} = Q\left[\sqrt{2A^2T/N_0}\right] = Q\left[\sqrt{2z}\right] \tag{9.18}$$

Using the approximation (9.15) for the error function, we need to solve

$$10^{-6} = \frac{e^{-z}}{2\sqrt{\pi z}}$$

iteratively. This gives the result

$$z \cong 10.53 \text{ dB} = 11.31 \text{ (ratio)}$$

Thus, $A^2T/N_0 = 11.31$, or

$$A^2 = (11.31)N_0/T = (11.31)\left(10^{-7}\right)(5000) = 5.655 \text{ mW} \quad (\text{actually V}^2 \times 10^{-3})$$

This corresponds to a signal amplitude of approximately 75.2 mV.

■

9.2 BINARY SYNCHRONOUS DATA TRANSMISSION WITH ARBITRARY SIGNAL SHAPES

In Section 9.1 we analyzed a simple baseband digital communication system. As in the case of analog transmission, it is often necessary to utilize modulation to condition a digital message signal so that it is suitable for transmission through a channel. Thus, instead of the constant-level signals considered in Section 9.1, we will let a logic 1 be represented by $s_1(t)$ and a logic 0 by $s_2(t)$. The only restriction on $s_1(t)$ and $s_2(t)$ is that they must have finite energy in a T-second interval. The energies of $s_1(t)$ and $s_2(t)$ are denoted by

$$E_1 \triangleq \int_{-\infty}^{\infty} s_1^2(t)\, dt \tag{9.19}$$

and

$$E_2 \triangleq \int_{-\infty}^{\infty} s_2^2(t)\, dt \tag{9.20}$$

respectively. In Table 9.1, several commonly used choices for $s_1(t)$ and $s_2(t)$ are given.

9.2.1 Receiver Structure and Error Probability

A possible receiver structure for detecting $s_1(t)$ or $s_2(t)$ in additive white Gaussian noise is shown in Figure 9.6. Since the signals chosen may have zero-average value over a T-second interval (see the examples in Table 9.1), we can no longer employ an integrator followed by a threshold device as in the case of constant-amplitude signals. Instead of the integrator, we employ a filter with, as yet, unspecified impulse response $h(t)$ and corresponding frequency response function $H(f)$. The received signal plus noise is either

$$y(t) = s_1(t) + n(t) \tag{9.21}$$

or

$$y(t) = s_2(t) + n(t) \tag{9.22}$$

where the noise, as before, is assumed to be white with double-sided power spectral density $\frac{1}{2}N_0$.

We can assume, without loss of generality; that the signaling interval under consideration is $0 \le t \le T$. (The filter initial conditions are set to zero at $t = 0$.)

To find P_E, we again note that an error can occur in either one of two ways. Assume that $s_1(t)$ and $s_2(t)$ were chosen such that $s_{01}(T) < s_{02}(T)$, where $s_{01}(t)$ and $s_{02}(t)$ are the outputs

Table 9.1 Possible Signal Choices for Binary Digital Signaling

Case	$s_1(t)$	$s_2(t)$	Type of signaling
1	0	$A\cos(\omega_c t)\,\Pi\left(\frac{t-T/2}{T}\right)$	Amplitude-shift keying
2	$A\sin(\omega_c t + \cos^{-1} m)\Pi\left(\frac{t-T/2}{T}\right)$	$A\sin(\omega_c t - \cos^{-1} m)\Pi\left(\frac{t-T/2}{T}\right)$	Phase-shift keying with carrier ($\cos^{-1} m \triangleq$ modulation index)
3	$A\cos(\omega_c t)\,\Pi\left(\frac{t-T/2}{T}\right)$	$A\cos\left[(\omega_c + \Delta\omega)\, t\right]\Pi\left(\frac{t-T/2}{T}\right)$	Frequency-shift keying

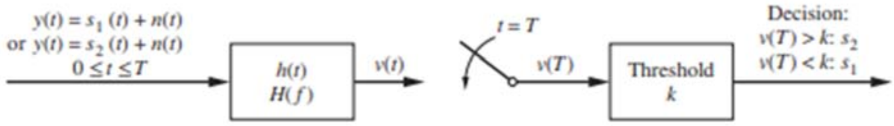

Figure 9.6
A possible receiver structure for detecting binary signals in white Gaussian noise.

of the filter due to $s_1(t)$ and $s_2(t)$, respectively, at the input. If not, the roles of $s_1(t)$ and $s_2(t)$ at the input can be reversed to ensure this. Referring to Figure 9.6, if $v(T) > k$, where k is the threshold, we decide that $s_2(t)$ was sent; if $v(T) < k$, we decide that $s_1(t)$ was sent. Letting $n_0(t)$ be the noise component at the filter output, an error is made if $s_1(t)$ is sent and $v(T) = s_{01}(T) + n_0(T) > k$; if $s_2(t)$ is sent, an error occurs if $v(T) = s_{02}(T) + n_0(T) < k$. Since $n_0(t)$ is the result of passing white Gaussian noise through a fixed linear filter, it is a Gaussian process. Its power spectral density is

$$S_{n_0}(f) = \frac{1}{2} N_0 |H(f)|^2 \tag{9.23}$$

Because the filter is fixed, $n_0(t)$ is a stationary Gaussian random process with mean zero and variance

$$\sigma_0^2 = \int_{-\infty}^{\infty} \frac{1}{2} N_0 |H(f)|^2 \, df \tag{9.24}$$

Since $n_0(t)$ is stationary, $N = n_0(T)$ is a random variable with mean zero and variance σ_0^2. Its pdf is

$$f_N(\eta) = \frac{e^{-\eta^2/2\sigma_0^2}}{\sqrt{2\pi\sigma_0^2}} \tag{9.25}$$

Given that $s_1(t)$ is transmitted, the sampler output is

$$V \triangleq v(T) = s_{01}(T) + N \tag{9.26}$$

and if $s_2(t)$ is transmitted, the sampler output is

$$V \triangleq v(T) = s_{02}(T) + N \tag{9.27}$$

These are also Gaussian random variables, since they result from linear operations on Gaussian random variables. They have means $s_{01}(T)$ and $s_{02}(T)$, respectively, and the same variance as N, that is, σ_0^2. Thus, the conditional pdfs of V given $s_1(t)$ is transmitted, $f_V(v \mid s_1(t))$, and given $s_2(t)$ is transmitted, $f_V(v \mid s_2(t))$, are as shown in Figure 9.7. Also illustrated is a decision threshold k.

From Figure 9.7, we see that the probability of error, given $s_1(t)$ is transmitted, is

$$\begin{aligned} P(E \mid s_1(t)) &= \int_k^{\infty} f_V(v|s_1(t))\, dv \\ &= \int_k^{\infty} \frac{e^{-[v - s_{01}(T)]^2/2\sigma_0^2}}{\sqrt{2\pi\sigma_0^2}}\, dv \end{aligned} \tag{9.28}$$

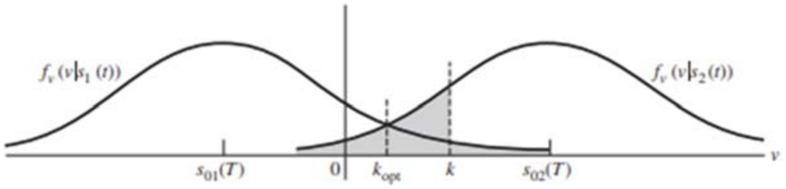

Figure 9.7
Conditional probability density functions of the filter output at time $t = T$.

which is the area under $f_V(v|s_1(t))$ to the right of $v = k$. Similarly, the probability of error, given $s_2(t)$ is transmitted, which is the area under $f_V(v \mid s_2(t))$ to the left of $v = k$, is given by

$$P\left(E|s_2(t)\right) = \int_{-\infty}^{k} \frac{e^{-[v-s_{02}(T)]^2/2\sigma_0^2}}{\sqrt{2\pi\sigma_0^2}}\, dv \tag{9.29}$$

Assuming that $s_1(t)$ and $s_2(t)$ are *a priori* equally probable,[4] the average probability of error is

$$P_E = \frac{1}{2}P[E|s_1(t)] + \frac{1}{2}P[E|s_2(t)] \tag{9.30}$$

The task now is to minimize this error probability by adjusting the threshold k and the impulse response $h(t)$.

Because of the equal *a priori* probabilities for $s_1(t)$ and $s_2(t)$ and the symmetrical shapes of $f_V(v|s_1(t))$ and $f_V(v|s_2(t))$, it is reasonable that the optimum choice for k is the intersection of the conditional pdfs, which is

$$k_{\text{opt}} = \frac{1}{2}[s_{01}(T) + s_{02}(T)] \tag{9.31}$$

The optimum threshold is illustrated in Figure 9.7 and can be derived by differentiating (9.30) with respect to k after substitution of (9.28) and (9.29). Because of the symmetry of the pdfs, the probabilities of either type of error, (9.28) or (9.29), are equal for this choice of k.

With this choice of k, the probability of error given by (9.30) reduces to

$$P_E = Q\left[\frac{s_{02}(T) - s_{01}(T)}{2\sigma_0}\right] \tag{9.32}$$

Thus, we see that P_E is a function of the difference between the two output signals at $t = T$. Remembering that the Q-function decreases monotonically with increasing argument, we see that P_E decreases with increasing distance between the two output signals, a reasonable result. We will encounter this interpretation again in Chapters 10 and 11, where we discuss concepts of signal space.

We now consider the minimization of P_E by proper choice of $h(t)$. This will lead us to the matched filter.

[4]See Problem 9.10 for the case of unequal *a priori* probabilities.

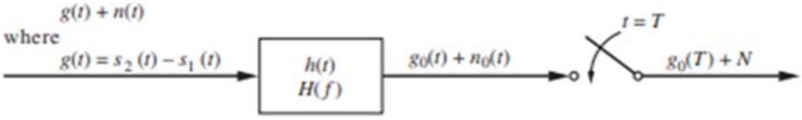

Figure 9.8
Choosing $H(f)$ to minimize P_E.

9.2.2 The Matched Filter

For a given choice of $s_1(t)$ and $s_2(t)$, we wish to determine an $H(f)$, or equivalently, an $h(t)$ in (9.32), that maximizes

$$\zeta = \frac{s_{02}(T) - s_{01}(T)}{\sigma_0} \tag{9.33}$$

which follows because the Q-function is monotonically decreasing as its argument increases. Letting $g(t) = s_2(t) - s_1(t)$, the problem is to find the $H(f)$ that maximizes $\zeta = g_0(T)/\sigma_0$, where $g_0(t)$ is the signal portion of the output due to the input, $g(t)$.[5] This situation is illustrated in Figure 9.8.

We can equally well consider the maximization of

$$\zeta^2 = \frac{g_0^2(T)}{\sigma_0^2} = \left.\frac{g_0^2(t)}{E\left\{n_0^2(t)\right\}}\right|_{t=T} \tag{9.34}$$

Since the input noise is stationary,

$$E\left\{n_0^2(t)\right\} = E\left\{n_0^2(T)\right\} = \frac{N_0}{2}\int_{-\infty}^{\infty} |H(f)|^2\, df \tag{9.35}$$

We can write $g_0(t)$ in terms of $H(f)$ and the Fourier transform of $g(t)$, $G(f)$, as

$$g_0(t) = \mathfrak{F}^{-1}[G(f)H(f)] = \int_{-\infty}^{\infty} H(f)G(f)e^{j2\pi ft}\, df \tag{9.36}$$

Setting $t = T$ in (9.36) and using this result along with (9.35) in (9.34), we obtain

$$\zeta^2 = \frac{\left|\int_{-\infty}^{\infty} H(f)G(f)e^{j2\pi fT}\, df\right|^2}{\frac{1}{2}N_0\int_{-\infty}^{\infty}|H(f)|^2\, df} \tag{9.37}$$

To maximize this equation with respect to $H(f)$, we employ *Schwarz's inequality*. Schwarz's inequality is a generalization of the inequality

$$|\mathbf{A}\cdot\mathbf{B}| = |AB\cos\theta| \le |\mathbf{A}|\,|\mathbf{B}| \tag{9.38}$$

where $\mathbf{A}$ and $\mathbf{B}$ are ordinary vectors, with θ the angle between them, and $\mathbf{A}\cdot\mathbf{B}$ denotes their inner, or dot, product. Since $|\cos\theta|$ equals unity if and only if θ equals zero or an integer multiple of π, equality holds if and only if $\mathbf{A}$ equals $\alpha\mathbf{B}$, where α is a constant ($\alpha > 0$ corresponds to $\theta = 0$ while $\alpha < 0$ corresponds to $\theta = \pi$). Considering the case of two complex

[5]Note that $g(t)$ is a fictitious signal in that the difference $s_{02}(T) - s_{01}(T)$ does not appear specifically in the receiver of Figure 9.8. How it relates to the detection of digital signals will be apparent later.

functions $X(f)$ and $Y(f)$, and defining the inner product as

$$\int_{-\infty}^{\infty} X(f)Y^*(f)\,df$$

Schwarz's inequality assumes the form[6]

$$\left|\int_{-\infty}^{\infty} X(f)Y^*(f)\,df\right| \le \sqrt{\int_{-\infty}^{\infty} |X(f)|^2\,df}\sqrt{\int_{-\infty}^{\infty} |Y(f)|^2\,df} \tag{9.39}$$

Equality holds if and only if $X(f) = \alpha Y(f)$ where α is, in general, a complex constant. We will prove Schwarz's inequality in Chapter 11 with the aid of signal-space notation.

We now return to our original problem, that of finding the $H(f)$ that maximizes (9.37). We replace $X(f)$ in (9.39) squared with $H(f)$ and $Y^*(f)$ with $G(f)e^{j2\pi Tf}$. Thus,

$$\zeta^2 = \frac{2}{N_0}\frac{\left|\int_{-\infty}^{\infty} X(f)Y^*(f)\,df\right|^2}{\int_{-\infty}^{\infty} |H(f)|^2\,df} \le \frac{2}{N_0}\frac{\int_{-\infty}^{\infty} |H(f)|^2\,df \int_{-\infty}^{\infty} |G(f)|^2\,df}{\int_{-\infty}^{\infty} |H(f)|^2\,df} \tag{9.40}$$

Canceling the integral over $|H(f)|^2$ in the numerator and denominator, we find the maximum value of ζ^2 to be

$$\zeta^2_{\max} = \frac{2}{N_0}\int_{-\infty}^{\infty} |G(f)|^2\,df = \frac{2E_g}{N_0} \tag{9.41}$$

where $E_g = \int_{-\infty}^{\infty} |G(f)|^2\,df$ is the energy contained in $g(t)$, which follows by Rayleigh's energy theorem. Equality holds in (9.40) if and only if

$$H(f) = \alpha G^*(f)e^{-j2\pi Tf} \tag{9.42}$$

where α is an arbitrary constant. Since α just fixes the gain of the filter (signal and noise are amplified the same), we can set it to unity. Thus, the optimum choice for $H(f)$, $H_0(f)$, is

$$H_0(f) = G^*(f)e^{-j2\pi Tf} \tag{9.43}$$

The impulse response corresponding to this choice of $H_0(f)$ is

$$\begin{aligned} h_0(t) &= \mathfrak{F}^{-1}[H_0(f)] \\ &= \int_{-\infty}^{\infty} G^*(f)\,e^{-j2\pi Tf}e^{j2\pi ft}\,df \\ &= \int_{-\infty}^{\infty} G(-f)\,e^{-j2\pi f(T-t)}\,df \\ &= \int_{-\infty}^{\infty} G(f')\,e^{j2\pi f'(T-t)}\,df' \end{aligned} \tag{9.44}$$

where the substitution $f' = -f$ in the integrand of the third integral to get the fourth integral. Recognizing this as the inverse Fourier transform of $g(t)$ with t replaced by $T - t$, we obtain

$$h_0(t) = g(T-t) = s_2(T-t) - s_1(T-t) \tag{9.45}$$

[6]If more convenient for a given application, one could equally well work with the square of Schwarz's inequality.

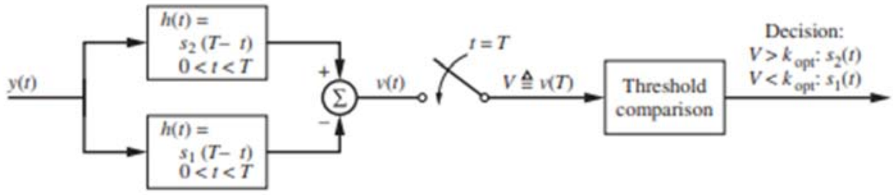

Figure 9.9
Matched-filter receiver for binary signaling in white Gaussian noise.

Thus, in terms of the original signals, the optimum receiver corresponds to passing the received signal plus noise through two parallel filters whose impulse responses are the time reverses of $s_1(t)$ and $s_2(t)$, respectively, and comparing the difference of their outputs at time T with the threshold given by (9.31). This operation is illustrated in Figure 9.9.

EXAMPLE 9.3

Consider the pulse signal

$$s(t) = \begin{cases} A, & 0 \le t \le T \\ 0, & \text{otherwise} \end{cases} \tag{9.46}$$

A filter matched to this signal has the impulse response

$$h_0(t) = s(t_0 - t) = \begin{cases} A, & 0 \le t_0 - t \le T \text{ or } t_0 - T \le t \le t_0 \\ 0, & \text{otherwise} \end{cases} \tag{9.47}$$

where the parameter t_0 will be fixed later. We note that if $t_0 < T$, the filter will be noncausal, since it will have nonzero impulse response for $t < 0$. The response of the filter to $s(t)$ is

$$y(t) = h_0(t) * s(t) = \int_{-\infty}^{\infty} h_0(\tau) s(t - \tau)\, d\tau \tag{9.48}$$

The factors in the integrand are shown in Figure 9.10(a). The resulting integrations are familiar from our previous considerations of linear systems, and the filter output is easily found to be as shown in Figure 9.10(b). Note that the peak output signal occurs at $t = t_0$. This is also the time of peak signal-to-rms-noise ratio, since the noise is stationary.

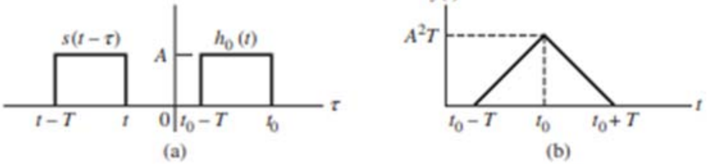

Figure 9.10
Signals pertinent to finding the matched-filter response of Example 9.3.

■

EXAMPLE 9.4

For a given value of N_0, consider the peak signal–squared-to-mean-square-noise ratio at the output of a matched filter for the two pulses

$$g_1(t) = A\Pi\left(\frac{t-t_0}{T}\right) \tag{9.49}$$

and

$$g_2(t) = B\cos\left[\frac{2\pi(t-t_0)}{T}\right]\Pi\left(\frac{t-t_0}{T}\right) \tag{9.50}$$

Relate A and B such that both pulses provide the same signal-to-noise ratio at the matched-filter output.

Solution

Since the peak signal–squared-to-mean-square-noise ratio at the matched-filter output is $2E_g/N_0$ and N_0 is the same for both cases, we can obtain equal signal-to-noise ratios for both cases by computing the energy of each pulse and setting the two energies equal. The results are

$$E_{g_1} = \int_{t_0-T/2}^{t_0+T/2} A^2\,dt = A^2T \tag{9.51}$$

and

$$E_{g_2} = \int_{t_0-T/2}^{t_0+T/2} B^2\cos^2\left[\frac{2\pi(t-t_0)}{T}\right]dt = \frac{B^2T}{2} \tag{9.52}$$

Setting these equal, we have that $A = B/\sqrt{2}$ to give equal signal-to-noise ratios. The peak signal–squared-to-mean-square-noise ratio, from (9.41), is

$$\zeta_{\max}^2 = \frac{2E_g}{N_0} = \frac{2A^2T}{N_0} = \frac{B^2T}{N_0} \tag{9.53}$$

■

9.2.3 Error Probability for the Matched-Filter Receiver

From (9.33) substituted into (9.32), the error probability for the matched-filter receiver of Figure 9.9 is

$$P_E = Q\left(\frac{\zeta}{2}\right) \tag{9.54}$$

where ζ has the maximum value

$$\zeta_{\max} = \left[\frac{2}{N_0}\int_{-\infty}^{\infty}|G(f)|^2\,df\right]^{1/2} = \left[\frac{2}{N_0}\int_{-\infty}^{\infty}|S_2(f) - S_1(f)|^2\,df\right]^{1/2} \tag{9.55}$$

given by (9.41). Using Parseval's theorem, we can write $\zeta_{\max}^2$ in terms of $g(t) = s_2(t) - s_1(t)$ as

$$\begin{aligned}\zeta_{\max}^2 &= \frac{2}{N_0}\int_{-\infty}^{\infty}\left[s_2(t) - s_1(t)\right]^2 dt \\ &= \frac{2}{N_0}\left\{\int_{-\infty}^{\infty} s_2^2(t)\,dt + \int_{-\infty}^{\infty} s_1^2(t)\,dt - 2\int_{-\infty}^{\infty} s_1(t)s_2(t)\,dt\right\}\end{aligned} \tag{9.56}$$

where $s_1(t)$ and $s_2(t)$ are assumed real. From (9.19) and (9.20), we see that the first two terms inside the braces are E_1 and E_2, the energies of $s_1(t)$ and $s_2(t)$, respectively. We define

$$\rho_{12} = \frac{1}{\sqrt{E_1 E_2}} \int_{-\infty}^{\infty} s_1(t)\, s_2(t)\, dt \tag{9.57}$$

as the correlation coefficient of $s_1(t)$ and $s_2(t)$. Just as for random variables, ρ_{12} is a measure of the similarity between $s_1(t)$ and $s_2(t)$ and is normalized such that $-1 \le \rho_{12} \le 1$ (ρ_{12} achieves its end points for $s_1(t) = \pm k s_2(t)$, where k is a positive constant). Thus,

$$\zeta_{\max}^2 = \frac{2}{N_0}\left(E_1 + E_2 - 2\sqrt{E_1 E_2}\rho_{12}\right) \tag{9.58}$$

and the error probability is

$$\begin{aligned} P_E &= Q\left[\left(\frac{E_1 + E_2 - 2\sqrt{E_1 E_2}\rho_{12}}{2N_0}\right)^{1/2}\right] \\ &= Q\left[\left(\frac{\frac{1}{2}(E_1 + E_2) - \sqrt{E_1 E_2}\rho_{12}}{N_0}\right)^{1/2}\right] \\ &= Q\left[\left(\frac{E_b}{N_0}\left(1 - \frac{\sqrt{E_1 E_2}}{E}\rho_{12}\right)\right)^{1/2}\right] \end{aligned} \tag{9.59}$$

where $E_b = \frac{1}{2}(E_1 + E_2)$ is the average received-signal energy, since $s_1(t)$ and $s_2(t)$ are transmitted with equal *a priori* probability. It is apparent from (9.59) that in addition to depending on the signal energies, as in the constant-signal case, P_E also depends on the similarity between the signals through ρ_{12}. We note that (9.58) takes on its maximum value of $(2/N_0)(\sqrt{E_1} + \sqrt{E_2})^2$ for $\rho_{12} = -1$, which gives the minimum value of P_E possible through choice of $s_1(t)$ and $s_2(t)$. This is reasonable, for then the transmitted signals are as dissimilar as possible. Finally, we can write (9.59) as

$$P_E = Q\left[\sqrt{(1 - R_{12})\frac{E_b}{N_0}}\right] \tag{9.60}$$

where $z = E_b/N_0$ is the average energy per bit divided by noise power spectral density as it was for the baseband system. The parameter R_{12} is defined as

$$R_{12} = \frac{2\sqrt{E_1 E_2}}{E_1 + E_2}\rho_{12} = \frac{\sqrt{E_1 E_2}}{E_b}\rho_{12} \tag{9.61}$$

and is a convenient parameter related to the correlation coefficient, but which should *not* be confused with a correlation function. The minimum value of R_{12} is -1, which is attained for $E_1 = E_2$ and $\rho_{12} = -1$. For this value of R_{12},

$$P_E = Q\left(\sqrt{\frac{2E_b}{N_0}}\right) \tag{9.62}$$

which is identical to (9.11), the result for the baseband antipodal system.

The probability of error versus the signal-to-noise ratio is compared in Figure 9.11 for $R_{12} = 0$ (orthogonal signals) and $R_{12} = -1$ (antipodal signals).

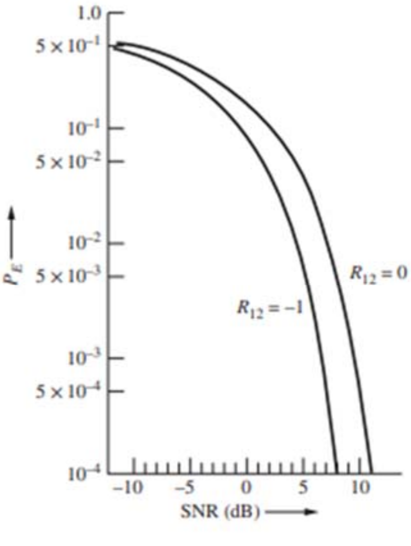

Figure 9.11
Probability of error for arbitrary waveshape case with $R_{12} = 0$ and $R_{12} = -1$.

COMPUTER EXAMPLE 9.1

A MATLAB program for computing the error probability for several values of correlation coefficient, R_{12}, is given below. Entering the vector [−1 0] in response to the first query reproduces the curves of Figure 9.11. Note that the user-defined function qfn(·) is used because MATLAB includes a function for erfc(u), but not $Q(u) = \frac{1}{2}\text{erfc}\left(u/\sqrt{2}\right)$.

```
% file: c9ce1
% Bit error probability for binary signaling;
% vector of correlation coefficients allowed
%
clf
R12 = input('Enter vector of desired R_1_2 values; <= 3 values ');
A = char('-','-.',':','--');
LR = length(R12);
z_dB = 0:.3:15;              % Vector of desired values of Eb/N0 in dB
z = 10.^(z_dB/10);           % Convert dB to ratios
for k = 1:LR                 % Loop for various desired values of R12
   P_E=qfn(sqrt(z*(1-R12(k)))); % Probability of error for vector of
                                  z-values
                             % Plot probability of error versus Eb/N0 in dB
   semilogy(z_dB,P_E,A(k,:)),axis([0   15   10^(-6)   1]),xlabel('E_b/N_0,
      dB'),ylabel('P_E'),...
   if k==1
      hold on; grid      % Hold plot for plots for other values of R12
   end
end
if LR == 1                   % Plot legends for R12 values
   legend(['R_1_2 = ',num2str(R12(1))],1)
elseif LR == 2
   legend(['R_1_2 = ',num2str(R12(1))],['R_1_2 = ',num2str(R12(2))],1)
elseif LR == 3
   legend(['R_1_2 = ',num2str(R12(1))],['R_1_2 = ',num2str(R12(2))],
```

```
        ['R,1,2 = ',num2str(R12(3))],1)
% This function computes the Gaussian Q-function
%
function Q=qfn(x)
Q = 0.5*erfc(x/sqrt(2));
% End of script file
```

■

9.2.4 Correlator Implementation of the Matched-Filter Receiver

In Figure 9.9, the optimum receiver involves two filters with impulse responses equal to the time reverse of the respective signals being detected. An alternative receiver structure can be obtained by noting that the matched filter in Figure 9.12(a) can be replaced by a multiplier-integrator cascade as shown in Figure 9.12(b). Such a series of operations is referred to as *correlation detection*.

To show that the operations given in Figure 9.12 are equivalent, we will show that $v(T)$ in Figure 9.12(a) is equal to $v'(T)$ in Figure 9.12(b). The output of the matched filter in Figure 9.12(a) is

$$v(t) = h(t) * y(t) = \int_0^T s(T-\tau)y(t-\tau)\,d\tau \tag{9.63}$$

which follows because $h(t) = s(T-t), 0 \le t < T$, and zero otherwise. Letting $t = T$ and changing variables in the integrand to $\alpha = T - \tau$, we obtain

$$v(T) = \int_0^T s(\alpha)\,y(\alpha)\,d\alpha \tag{9.64}$$

Considering next the output of the correlator configuration in Figure 9.12(b), we obtain

$$v'(T) = \int_0^T y(t)\,s(t)\,dt \tag{9.65}$$

which is identical to (9.64). Thus, the matched filters for $s_1(t)$ and $s_2(t)$ in Figure 9.9 can be replaced by correlation operations with $s_1(t)$ and $s_2(t)$, respectively, and the receiver operation will not be changed. We note that the integrate-and-dump receiver for the constant-signal case of Section 9.1 is actually a correlation or, equivalently, a matched-filter receiver.

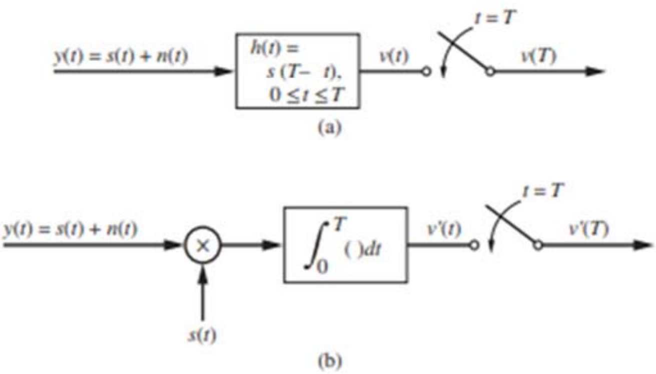

Figure 9.12
Equivalence of the matched-filter and correlator receivers.
(a) Matched-filter sampler.
(b) Correlator sampler.

9.2.5 Optimum Threshold

The optimum threshold for binary signal detection is given by (9.31), where $s_{01}(T)$ and $s_{02}(T)$ are the outputs of the detection filter in Figure 9.6 at time T due to the input signals $s_1(t)$ and $s_2(t)$, respectively. We now know that the optimum detection filter is a matched filter, matched to the difference of the input signals, and has the impulse response given by (9.45). From the superposition integral, we have

$$\begin{aligned} s_{01}(T) &= \int_{-\infty}^{\infty} h(\lambda) s_1(T-\lambda)\, d\lambda \\ &= \int_{-\infty}^{\infty} \left[s_2(T-\lambda) - s_1(T-\lambda)\right] s_1(T-\lambda)\, d\lambda \\ &= \int_{-\infty}^{\infty} s_2(u)\, s_1(u)\, du - \int_{-\infty}^{\infty} \left[s_1(u)\right]^2 du \\ &= \sqrt{E_1 E_2}\,\rho_{12} - E_1 \end{aligned} \tag{9.66}$$

where the substitution $u = T - \tau$ has been used to go from the second equation to the third, and the definition of the correlation coefficient (9.57) has been used to get the last equation along with the definition of energy of a signal. Similarly, it follows that

$$\begin{aligned} s_{02}(T) &= \int_{-\infty}^{\infty} \left[s_2(T-\lambda) - s_1(T-\lambda)\right] s_2(T-\lambda)\, d\lambda \\ &= \int_{-\infty}^{\infty} \left[s_2(u)\right]^2 du - \int_{-\infty}^{\infty} s_2(u)\, s_1(u)\, du \\ &= E_2 - \sqrt{E_1 E_2}\,\rho_{12} \end{aligned} \tag{9.67}$$

Substituting (9.66) and (9.67) into (9.31), we find the optimum threshold to be

$$k_{\text{opt}} = \frac{1}{2}\left(E_2 - E_1\right) \tag{9.68}$$

Note that equal energy signals will always result in an optimum threshold of zero. Also note that the waveshape of the signals, as manifested through the correlation coefficient, has no effect on the optimum threshold. Only the signal energies do.

9.2.6 Nonwhite (Colored) Noise Backgrounds

The question naturally arises about the optimum receiver for nonwhite noise backgrounds. Usually, the noise in a receiver system is generated primarily in the front-end stages and is due to thermal agitation of electrons in the electronic components (see Appendix A). This type of noise is well approximated as white. If a bandlimited channel precedes the introduction of the white noise, then we need only work with modified transmitted signals. If, for some reason, a bandlimiting filter follows the introduction of the white noise (for example, an intermediate-frequency amplifier following the radio-frequency amplifier and mixers where most of the noise is generated in a heterodyne receiver), we can use a simple artifice to approximate the matched-filter receiver. The colored noise plus signal is passed through a "whitening filter" with a frequency response function that is the inverse square root of the noise spectral density. Thus, the output of this whitening filter is white noise plus a signal component that has been transformed by the whitening filter. We then build a matched-filter receiver with impulse response that is the difference of the time-reverse of the "whitened" signals. The cascade of a whitening filter and matched filter (matched to the whitened signals) is called a *whitened matched filter*. This combination provides only an approximately optimum receiver for two reasons. Since the whitening filters will spread the received signals beyond the T-second signaling interval, two types of degradation will result: (1) the signal energy spread beyond the interval under consideration is not used by the matched filter in making a decision; (2) previous signals spread out by the whitening filter will interfere with the matched-filtering operation on the signal on which a decision is being made. The latter is referred to as *intersymbol interference*, as first discussed in Chapter 5, and is explored further in Sections 9.7 and 9.9. It is apparent that degradation due to these effects is minimized if the signal *duration* is short compared with T, such as in a pulsed radar system. Finally, signal intervals adjacent to the interval being used in the decision process contain information that is relevant to making a decision on the basis of the correlation of the noise. In short, the whitened matched-filter receiver is nearly optimum if the signaling interval is large compared with the inverse bandwidth of the whitening filter. The question of bandlimited channels, and nonwhite background noise, is explored further in Section 9.6.

9.2.7 Receiver Implementation Imperfections

In the theory developed in this section, it is assumed that the signals are known *exactly* at the receiver. This is, of course, an idealized situation. Two possible deviations from this assumption are: (1) the phase of the receiver's replica of the transmitted signal may be in error, and (2) the exact arrival time of the received signal may be in error. These are called *synchronization* errors. The first case is explored in this section, and the latter is explored in the problems. Methods of synchronization are discussed in Chapter 10.

9.2.8 Error Probabilities for Coherent Binary Signaling

We now compare the performance of several commonly used coherent binary signaling schemes. Then we will examine noncoherent systems. To obtain the error probability for coherent systems, the results of Section 9.2 will be applied directly. The three types of coherent systems to be considered in this section are amplitude-shift keyed (ASK), phase-shift keyed (PSK), and frequency-shift keyed (FSK). Typical transmitted waveforms for these three types of digital modulation are shown in Figure 9.13. We also will consider the effect of an imperfect phase reference on the performance of a coherent PSK system. Such systems are often referred to as partially coherent.

Amplitude-Shift Keying (ASK)

In Table 9.1, $s_1(t)$ and $s_2(t)$ for ASK are given as 0 and $A\cos(\omega_c t)\,\Pi\left[(t - T/2)/T\right]$, where $f_c = \omega_c/2\pi$ is the carrier frequency. We note that the transmitter for such a system simply consists of an oscillator that is gated on and off; accordingly, binary ASK with one amplitude set to 0 is often referred to as *on-off keying*. It is important to note that the oscillator runs continuously as the on-off gating is carried out.