9

SIGNAL DESIGN FOR BAND-LIMITED CHANNELS

In previous chapters, we considered the transmission of digital information through an additive gaussian noise channel. In effect, no bandwidth constraint was imposed on the signal design and the communication system design.

In this chapter, we consider the problem of signal design when the channel is band-limited to some specified bandwidth W Hz. Under this condition, the channel may be modeled as a linear filter having an equivalent lowpass frequency response $C(f)$ that is zero for $|f| > W$.

The first topic that is treated is the design of the signal pulse $g(t)$ in a linearly modulated signal, represented as

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

that efficiently utilizes the total available channel bandwidth W. We shall see that when the channel is ideal for $|f| \leq W$, a signal pulse can be designed that allows us to transmit at symbol rates comparable to or exceeding the channel bandwidth W. On the other hand, when the channel is not ideal, signal transmission at a symbol rate equal to or exceeding W results in intersymbol interference (ISI) among a number of adjacent symbols.

The second topic that is treated in this chapter is the use of coding to shape the spectrum of the transmitted signal and, thus, to avoid the problem of ISI.

We begin our discussion with a general characterization of band-limited, linear filter channels.

9-1 CHARACTERIZATION OF BAND-LIMITED CHANNELS

Of the various channels available for digital communications, telephone channels are by far the most widely used. Such channels are characterized as

band-limited linear filters. This is certainly the proper characterization when frequency-division multiplexing (FDM) is used as a means for establishing channels in the telephone network. Recent additions to the telephone network employ pulse-code modulation (PCM) for digitizing and encoding the analog signal and time-division multiplexing (TDM) for establishing multiple channels. Nevertheless, filtering is still used on the analog signal prior to sampling and encoding. Consequently, even though the present telephone network employs a mixture of FDM and TDM for transmission, the linear filter model for telephone channels is still appropriate.

For our purposes, a band-limited channel such as a telephone channel will be characterized as a linear filter having an equivalent lowpass frequency response characteristic $C(f)$. Its equivalent lowpass impulse response is denoted by $c(t)$. Then, if a signal of the form

$$s(t) = \text{Re}\left[v(t)e^{j2\pi f_c t}\right] \tag{9-1-1}$$

is transmitted over a bandpass telephone channel, the equivalent lowpass received signal is

$$r_l(t) = \int_{-\infty}^{\infty} v(\tau)c(t-\tau)\,d\tau + z(t) \tag{9-1-2}$$

where the integral represents the convolution of $c(t)$ with $v(t)$, and $z(t)$ denotes the additive noise. Alternatively, the signal term can be represented in the frequency domain as $V(f)C(f)$, where $V(f)$ is the Fourier transform of $v(t)$.

If the channel is band-limited to W Hz then $C(f) = 0$ for $|f| > W$. As a consequence, any frequency components in $V(f)$ above $|f| = W$ will not be passed by the channel. For this reason, we limit the bandwidth of the transmitted signal to W Hz also.

Within the bandwidth of the channel, we may express the frequency response $C(f)$ as

$$C(f) = |C(f)|\,e^{j\theta(f)} \tag{9-1-3}$$

where $|C(f)|$ is the amplitude response characteristic and $\theta(f)$ is the phase response characteristic. Furthermore, the envelope delay characteristic is defined as

$$\tau(f) = -\frac{1}{2\pi}\frac{d\theta(f)}{df} \tag{9-1-4}$$

A channel is said to be *nondistorting* or *ideal* if the amplitude response $|C(f)|$ is constant for all $|f| \leq W$ and $\theta(f)$ is a linear function of frequency, i.e., $\tau(f)$ is a constant for all $|f| \leq W$. On the other hand, if $|C(f)|$ is not constant for all $|f| \leq W$, we say that the channel *distorts the transmitted signal* $V(f)$ *in amplitude*, and, if $\tau(f)$ is not constant for all $|f| \leq W$, we say that the channel *distorts the signal* $V(f)$ *in delay.*

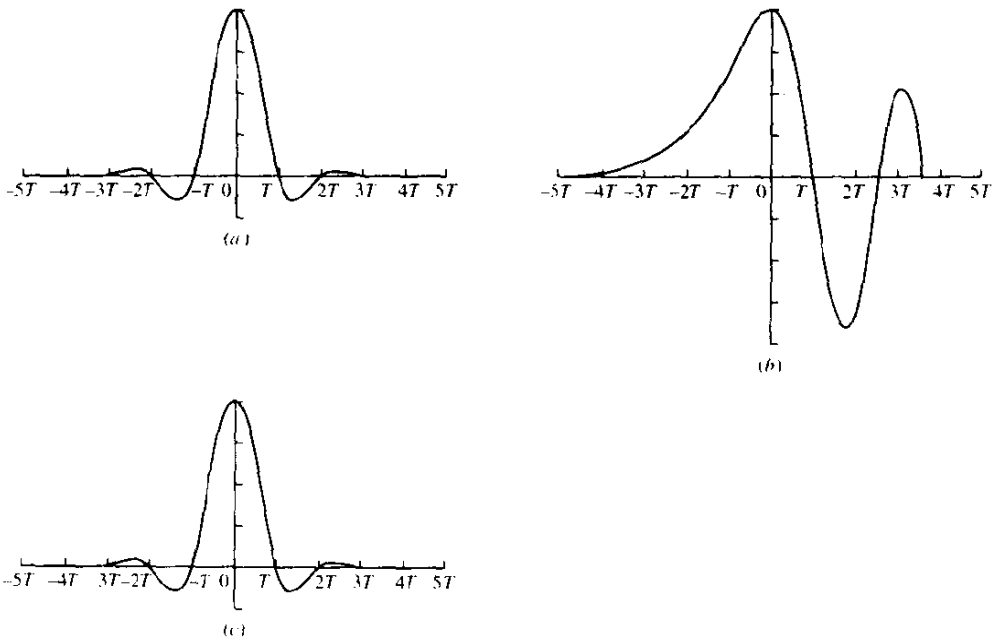


FIGURE 9-1-1 Effect of channel distortion: (*a*) channel input; (*b*) channel output; (*c*) equalizer output.

As a result of the amplitude and delay distortion caused by the nonideal channel frequency response characteristic $C(f)$, a succession of pulses transmitted through the channel at rates comparable to the bandwidth W are smeared to the point that they are no longer distinguishable as well-defined pulses at the receiving terminal. Instead, they overlap and, thus, we have intersymbol interference. As an example of the effect of delay distortion on a transmitted pulse, Fig. 9-1-1(*a*) illustrates a band-limited pulse having zeros periodically spaced in time at points labeled $\pm T$, $\pm 2T$, etc. If information is conveyed by the pulse amplitude, as in PAM, for example, then one can transmit a sequence of pulses, each of which has a peak at the periodic zeros of the other pulses. However, transmission of the pulse through a channel modeled as having a linear envelope delay characteristic $\tau(f)$ [quadratic phase $\theta(f)$] results in the received pulse shown in Fig. 9-1-1(*b*) having zero-crossings that are no longer periodically spaced. Consequently, a sequence of successive pulses would be smeared into one another and the peaks of the pulses would no longer be distinguishable. Thus, the channel delay distortion results in intersymbol interference. As will be discussed in Chapter 10, it is possible to compensate for the nonideal frequency response characteristic of the channel by use of a filter or equalizer at the demodulator. Figure 9-1-1(*c*) illustrates the output of a linear equalizer that compensates for the linear distortion in the channel.

The extent of the intersymbol interference on a telephone channel can be

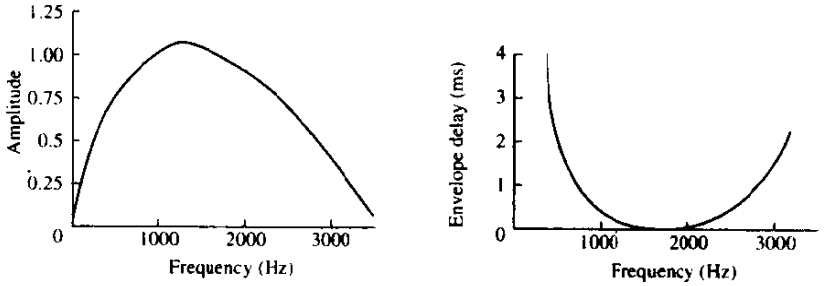


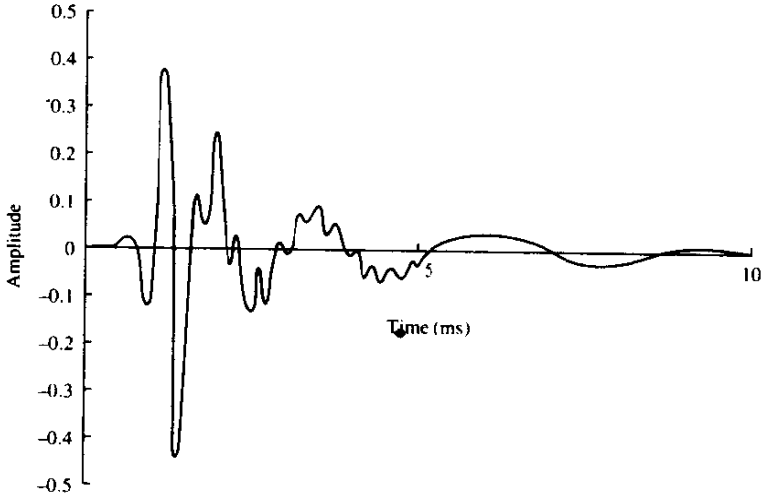
FIGURE 9-1-2 Average amplitude and delay characteristics of medium-range telephone channel.

appreciated by observing a frequency response characteristic of the channel. Figure 9-1-2 illustrates the measured average amplitude and delay as functions of frequency for a medium-range (180–725 mi) telephone channel of the switched telecommunications network as given by Duffy and Tratcher (1971). We observe that the usable band of the channel extends from about 300 Hz to about 3000 Hz. The corresponding impulse response of this average channel is shown in Fig. 9-1-3. Its duration is about 10 ms. In comparison, the transmitted symbol rates on such a channel may be of the order of 2500 pulses or symbols per second. Hence, intersymbol interference might extend over 20–30 symbols.

In addition to linear distortion, signals transmitted through telephone channels are subject to other impairments, specifically nonlinear distortion, frequency offset, phase jitter, impulse noise and thermal noise.

Nonlinear distortion in telephone channels arises from nonlinearities in

FIGURE 9-1-3 Impulse response of average channel with amplitude and delay shown in Fig. 9-1-2.



amplifiers and compandors used in the telephone system. This type of distortion is usually small and it is very difficult to correct.

A small *frequency offset*, usually less than 5 Hz, results from the use of carrier equipment in the telephone channel. Such an offset cannot be tolerated in high-speed digital transmission systems that use synchronous phase-coherent demodulation. The offset is usually compensated for by the carrier recovery loop in the demodulator.

Phase jitter is basically a low-index frequency modulation of the transmitted signal with the low frequency harmonics of the power line frequency (50–60 Hz). Phase jitter poses a serious problem in digital transmission of high rates. However, it can be tracked and compensated for, to some extent, at the demodulator.

Impulse noise is an additive disturbance. It arises primarily from the switching equipment in the telephone system. *Thermal* (gaussian) *noise* is also present at levels of 20–30 dB below the signal.

The degree to which one must be concerned with these channel impairments depends on the transmission rate over the channel and the modulation technique. For rates below 1800 bits/s ($R/W < 1$), one can choose a modulation technique, e.g., FSK, that is relatively insensitive to the amount of distortion encountered on typical telephone channels from all the sources listed above. For rates between 1800 and 2400 bits/s ($R/W \approx 1$), a more bandwidth-efficient modulation technique such as four-phase PSK is usually employed. At these rates, some form of compromise equalization is often employed to compensate for the average amplitude and delay distortion in the channel. In addition, the carrier recovery method is designed to compensate for the frequency offset. The other channel impairments are not that serious in their effects on the error rate performance at these rates. At transmission rates above 2400 bits/s ($R/W > 1$), bandwidth-efficient coded modulation techniques such as trellis-coded QAM, PAM, and PSK are employed. For such rates, special attention must be paid to linear distortion, frequency offset, and phase jitter. Linear distortion is usually compensated for by means of an adaptive equalizer. Phase jitter is handled by a combination of signal design and some type of phase compensation at the demodulator. At rates above 9600 bits/s, special attention must be paid not only to linear distortion, phase jitter, and frequency offset, but also to the other channel impairments mentioned above.

Unfortunately, a channel model that encompasses all the impairments listed above becomes difficult to analyze. For mathematical tractability the channel model that is adopted in this and the next two chapters is a linear filter that introduces amplitude and delay distortion and adds gaussian noise.

Besides the telephone channels, there are other physical channels that exhibit some form of time dispersion, and thus, introduce intersymbol interference. Radio channels such as shortwave ionospheric propagation (HF) and tropospheric scatter are two examples of time-dispersive channels. In these channels, time dispersion and, hence, intersymbol interference is the result of multiple propagation paths with different path delays. The number of paths

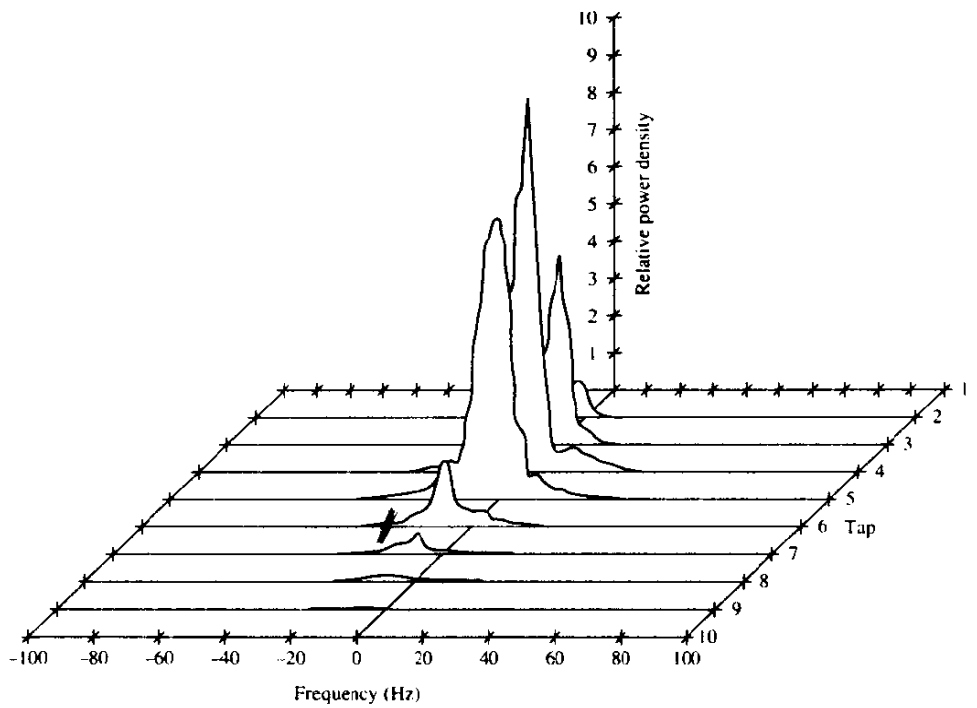


FIGURE 9-1-4 Scattering function of a medium-range tropospheric scatter channel.

and the relative time delays among the paths vary with time, and, for this reason, these radio channels are usually called *time-variant multipath channels*. The time-variant multipath conditions give rise to a wide variety of frequency response characteristics. Consequently the frequency response characterization that is used for telephone channels is inappropriate for time-variant multipath channels. Instead, these radio channels are characterized statistically, as explained in more detail in Chapter 14, in terms of the scattering function, which, in brief, is a two-dimensional representation of the average received signal power as a function of relative time delay and Doppler frequency.

For illustrative purposes, a scattering function measured on a medium-range (150 mi) trophospheric scatter channel is shown in Fig. 9-1-4. The total time duration (multipath spread) of the channel response is approximately 0.7 μs on the average, and the spread between "half-power points" in Doppler frequency is a little less than 1 Hz on the strongest path and somewhat larger on the other paths. Typically, if one is transmitting at a rate of 10^7 symbols/s over such a channel, the multipath spread of 0.7 μs will result in intersymbol interference that spans about seven symbols.

In this chapter, we deal exclusively with the linear time-invariant filter model for a band-limited channel. The adaptive equalization techniques presented in Chapters 10 and 11 for combating intersymbol interference are also applicable to time-invariant multipath channels, under the condition that

the time variations in the channel are relatively slow in comparison to the total channel bandwidth or, equivalently, to the symbol transmission rate over the channel.

9-2 SIGNAL DESIGN FOR BAND-LIMITED CHANNELS

It was shown in Chapter 4 that the equivalent lowpass transmitted signal for several different types of digital modulation techniques has the common form

$$v(t) = \sum_{n=0}^{\infty} I_n g(t - nT) \tag{9-2-1}$$

where $\{I_n\}$ represents the discrete information-bearing sequence of symbols and $g(t)$ is a pulse that, for the purposes of this discussion, is assumed to have a band-limited frequency response characteristic $G(f)$, i.e., $G(f) = 0$ for $|f| > W$. This signal is transmitted over a channel having a frequency response $C(f)$, also limited to $|f| \leq W$. Consequently, the received signal can be represented as

$$r_l(t) = \sum_{n=0}^{\infty} I_n h(t - nT) + z(t) \tag{9-2-2}$$

where

$$h(t) = \int_{-\infty}^{\infty} g(\tau)c(t - \tau)\, d\tau \tag{9-2-3}$$

and $z(t)$ represents the additive white Gaussian noise.

Let us suppose that the received signal is passed first through a filter and then sampled at a rate $1/T$ samples/s. We shall show in a subsequent section that the optimum filter from the point of view of signal detection is one matched to the received pulse. That is, the frequency response of the receiving filter is $H^*(f)$. We denote the output of the receiving filter as

$$y(t) = \sum_{n=0}^{\infty} I_n x(t - nT) + v(t) \tag{9-2-4}$$

where $x(t)$ is the pulse representing the response of the receiving filter to the input pulse $h(t)$ and $v(t)$ is the response of the receiving filter to the noise $z(t)$.

Now, if $y(t)$ is sampled at times $t = kT + \tau_0$, $k = 0, 1, \ldots$, we have

$$y(kT + \tau_0) \equiv y_k = \sum_{n=0}^{\infty} I_n x(kT - nT + \tau_0) + v(kT + \tau_0) \tag{9-2-5}$$

or, equivalently,

$$y_k = \sum_{n=0}^{\infty} I_n x_{k-n} + v_k, \qquad k = 0, 1, \ldots \tag{9-2-6}$$

where τ_0 is the transmission delay through the channel. The sample values can be expressed as

$$y_k = x_0\left(I_k + \frac{1}{x_0}\sum_{\substack{n=0\\ n\neq k}}^{\infty} I_n x_{k-n}\right) + \nu_k, \qquad k = 0, 1, \ldots \tag{9-2-7}$$

We regard x_0 as an arbitrary scale factor, which we arbitrarily set equal to unity for convenience. Then

$$y_k = I_k + \sum_{\substack{n=0\\ n\neq k}}^{\infty} I_n x_{k-n} + \nu_k \tag{9-2-8}$$

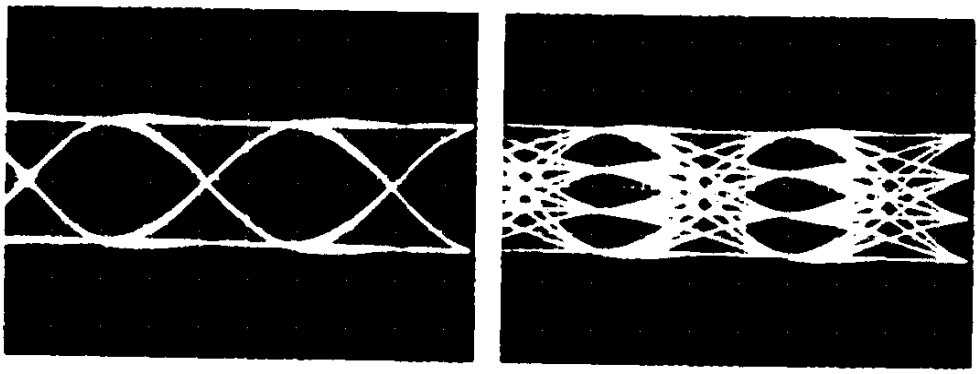
The term I_k represents the desired information symbol at the k th sampling instant, the term

$$\sum_{\substack{n=0\\ n\neq k}}^{\infty} I_n x_{k-n}$$

represents the intersymbol interference (ISI), and ν_k is the additive gaussian noise variable at the kth sampling instant.

The amount of intersymbol interference and noise in a digital communications system can be viewed on an oscilloscope. For PAM signals, we can display the received signal $y(t)$ on the vertical input with the horizontal sweep rate set at $1/T$. The resulting oscilloscope display is called an *eye pattern* because of its resemblance to the human eye. For example, Fig. 9-2-1 illustrates the eye patterns for binary and four-level PAM modulation. The effect of ISI is to cause the eye to close, thereby reducing the margin for additive noise to cause errors. Figure 9-2-2 graphically illustrates the effect of intersymbol interference in reducing the opening of a binary eye. Note that intersymbol interference distorts the position of the zero-crossings and causes

FIGURE 9-2-1 Examples of eye patterns for binary and quaternary amplitude shift keying (or PAM).



BINARY QUATERNARY

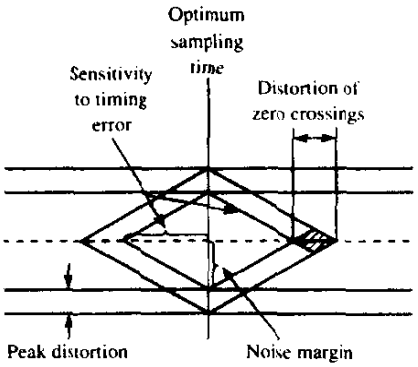


FIGURE 9-2-2 Effect of intersymbol interference on eye opening.

a reduction in the eye opening. Thus, it causes the system to be more sensitive to a synchronization error.

For PSK and QAM it is customary to display the "eye pattern" as a two-dimensional scatter diagram illustrating the sampled values $\{y_k\}$ that represent the decision variables at the sampling instants. Figure 9-2-3 illustrates such an eye pattern for an 8-PSK signal. In the absence of intersymbol interference and noise, the superimposed signals at the sampling instants would result in eight distinct points corresponding to the eight transmitted signal phases. Intersymbol interference and noise result in a deviation of the received samples $\{y_k\}$ from the desired 8-PSK signal. The larger the intersymbol interference and noise, the larger the scattering of the received signal samples relative to the transmitted signal points.

Below, we consider the problem of signal design under the condition that there is no intersymbol interference at the sampling instants.

9-2-1 DESIGN OF BAND-LIMITED SIGNALS FOR NO INTERSYMBOL INTERFERENCE—THE NYQUIST CRITERION

For the discussion in this section and in Section 9-2-2, we assume that the band-limited channel has ideal frequency response characteristics, i.e., $C(f) = 1$

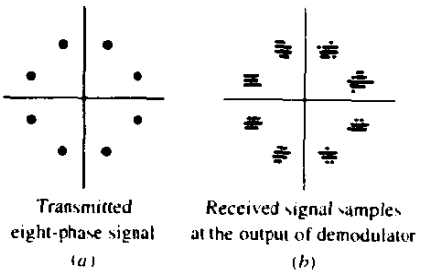


FIGURE 9-2-3 Two-dimensional digital "eye patterns."

for $|f| \leq W$. Then the pulse $x(t)$ has a spectral characteristic $X(f) = |G(f)|^2$, where

$$x(t) = \int_{-W}^{W} X(f) e^{j2\pi ft}\, df \tag{9-2-9}$$

We are interested in determining the spectral properties of the pulse $x(t)$ and, hence, the transmitted pulse $g(t)$, that results in no intersymbol interference. Since

$$y_k = I_k + \sum_{\substack{n=0 \\ n \neq k}}^{\infty} I_n x_{k-n} + v_k \tag{9-2-10}$$

the condition for no intersymbol interference is

$$x(t = kT) \equiv x_k = \begin{cases} 1 & (k = 0) \\ 0 & (k \neq 0) \end{cases} \tag{9-2-11}$$

Below, we derive the necessary and sufficient condition on $X(f)$ in order for $x(t)$ to satisfy the above relation. This condition is known as the *Nyquist pulse-shaping criterion* or *Nyquist condition for zero ISI* and is stated in the following theorem.

Theorem (Nyquist)

The necessary and sufficient condition for $x(t)$ to satisfy

$$x(nT) = \begin{cases} 1 & (n = 0) \\ 0 & (n \neq 0) \end{cases} \tag{9-2-12}$$

is that its Fourier transform $X(f)$ satisfy

$$\sum_{m=-\infty}^{\infty} X(f + m/T) = T \tag{9-2-13}$$

Proof

In general, $x(t)$ is the inverse Fourier transform of $X(f)$. Hence,

$$x(t) = \int_{-\infty}^{\infty} X(f) e^{j2\pi ft}\, df \tag{9-2-14}$$

At the sampling instants $t = nT$, this relation becomes

$$x(nT) = \int_{-\infty}^{\infty} X(f) e^{j2\pi fnT}\, df \tag{9-2-15}$$

Let us break up the integral in (9-2-15) into integrals covering the finite range of $1/T$. Thus, we obtain

$$\begin{aligned} x(nT) &= \sum_{m=-\infty}^{\infty} \int_{(2m-1)/2T}^{(2m+1)/2T} X(f)e^{j2\pi fnT}\,df \\ &= \sum_{m=-\infty}^{\infty} \int_{-1/2T}^{1/2T} X(f+m/T)e^{j2\pi fnT}\,dt \\ &= \int_{-1/2T}^{1/2T} \left[\sum_{m=-\infty}^{\infty} X(f+m/T)\right] e^{j2\pi fnT}\,df \\ &= \int_{-1/2T}^{1/2T} B(f)e^{j2\pi fnT}\,df \end{aligned} \tag{9-2-16}$$

where we have defined $B(f)$ as

$$B(f) = \sum_{m=-\infty}^{\infty} X(f+m/T) \tag{9-2-17}$$

Obviously $B(f)$ is a periodic function with period $1/T$, and, therefore, it can be expanded in terms of its Fourier series coefficients $\{b_n\}$ as

$$B(f) = \sum_{n=-\infty}^{\infty} b_n e^{j2\pi nfT} \tag{9-2-18}$$

where

$$b_n = T\int_{-1/2T}^{1/2T} B(f)e^{-j2\pi nfT}\,df \tag{9-2-19}$$

Comparing (9-2-19) and (9-2-16), we obtain

$$b_n = Tx(-nT) \tag{9-2-20}$$

Therefore, the necessary and sufficient condition for (9-2-10) to be satisfied is that

$$b_n = \begin{cases} T & (n=0) \\ 0 & (n\neq 0) \end{cases} \tag{9-2-21}$$

which, when substituted into (9-2-18), yields

$$B(f) = T \tag{9-2-22}$$

or, equivalently,

$$\sum_{m=-\infty}^{\infty} X(f+m/T) = T \tag{9-2-23}$$

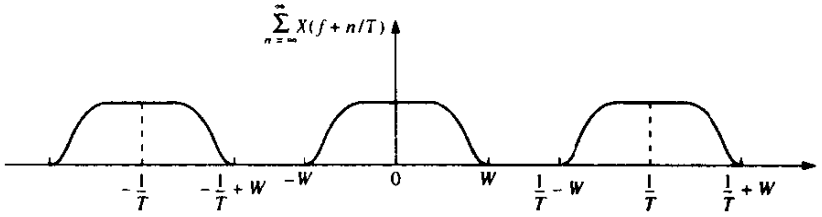


FIGURE 9-2-4 Plot of $B(f)$ for the case $T < 1/2W$.

This concludes the proof of the theorem.

Now suppose that the channel has a bandwidth of W. Then $C(f) \equiv 0$ for $|f| > W$ and, consequently, $X(f) = 0$ for $|f| > W$. We distinguish three cases.

1 When $T < 1/2W$, or, equivalently, $1/T > 2W$, since $B(f) = \sum_{n=-\infty}^{+\infty} X(f + n/T)$ consists of nonoverlapping replicas of $X(f)$, separated by $1/T$ as shown in Fig. 9-2-4, there is no choice for $X(f)$ to ensure $B(f) \equiv T$ in this case and there is no way that we can design a system with no ISI.

2 When $T = 1/2W$, or, equivalently, $1/T = 2W$ (the Nyquist rate), the replications of $X(f)$, separated by $1/T$, are as shown in Fig. 9-2-5. It is clear that in this case there exists only one $X(f)$ that results in $B(f) = T$, namely,

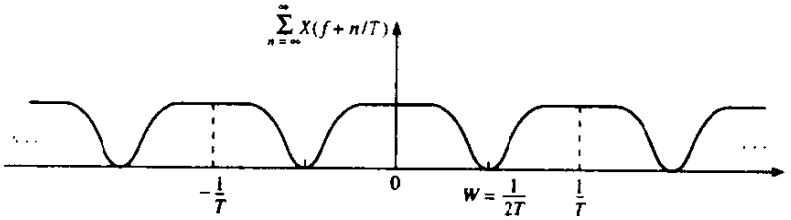
$$X(f) = \begin{cases} T & (|f| < W) \\ 0 & (\text{otherwise}) \end{cases} \tag{9-2-24}$$

which corresponds to the pulse

$$x(t) = \frac{\sin(\pi t/T)}{\pi t/T} \equiv \text{sinc}\left(\frac{\pi t}{T}\right) \tag{9-2-25}$$

This means that the smallest value of T for which transmission with zero ISI is possible is $T = 1/2W$, and for this value, $x(t)$ has to be a sinc function. The difficulty with this choice of $x(t)$ is that it is noncausal and therefore nonrealizable. To make it realizable, usually a delayed version of it, i.e., $\text{sinc}[\pi(t - t_0)/T]$ is used and t_0 is chosen such that for $t < 0$, we have $\text{sinc}[\pi(t - t_0)/T] \approx 0$. Of course, with this choice of $x(t)$, the sampling time

FIGURE 9-2-5 Plot of $B(f)$ for the case $T = 1/2W$.



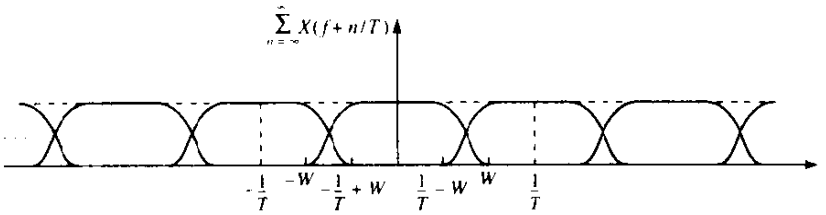


FIGURE 9-2-6 Plot of $B(f)$ for the case $T > 1/2W$.

must also be shifted to $mT + t_0$. A second difficulty with this pulse shape is that its rate of convergence to zero is slow. The tails of $x(t)$ decay as $1/t$; consequently, a small mistiming error in sampling the output of the matched filter at the demodulator results in an infinite series of ISI components. Such a series is not absolutely summable because of the $1/t$ rate of decay of the pulse, and, hence, the sum of the resulting ISI does not converge.

3 When $T > 1/2W$, $B(f)$ consists of overlapping replications of $X(f)$ separated by $1/T$, as shown in Fig. 9-2-6. In this case, there exist numerous choices for $X(f)$ such that $B(f) \equiv T$.

A particular pulse spectrum, for the $T > 1/2W$ case, that has desirable spectral properties and has been widely used in practice is the raised cosine spectrum. The raised cosine frequency characteristic is given as (see Problem 9-11)

$$X_{rc}(f) = \begin{cases} T & \left(0 \leq |f| \leq \dfrac{1-\beta}{2T}\right) \\ \dfrac{T}{2}\left\{1 + \cos\left[\dfrac{\pi T}{\beta}\left(|f| - \dfrac{1-\beta}{2T}\right)\right]\right\} & \left(\dfrac{1-\beta}{2T} \leq |f| \leq \dfrac{1+\beta}{2T}\right) \\ 0 & \left(|f| > \dfrac{1+\beta}{2T}\right) \end{cases} \tag{9-2-26}$$

where β is called the *rolloff factor*, and takes values in the range $0 \leq \beta \leq 1$. The bandwidth occupied by the signal beyond the Nyquist frequency $1/2T$ is called the *excess bandwidth* and is usually expressed as a percentage of the Nyquist frequency. For example, when $\beta = \frac{1}{2}$, the excess bandwidth is 50%, and when $\beta = 1$, the excess bandwidth is 100%. The pulse $x(t)$, having the raised cosine spectrum, is

$$\begin{aligned} x(t) &= \frac{\sin(\pi t/T)}{\pi t/T} \frac{\cos(\pi\beta t/T)}{1 - 4\beta^2 t^2/T^2} \\ &= \operatorname{sinc}(\pi t/T) \frac{\cos(\pi\beta t/T)}{1 - 4\beta^2 t^2/T^2} \end{aligned} \tag{9-2-27}$$

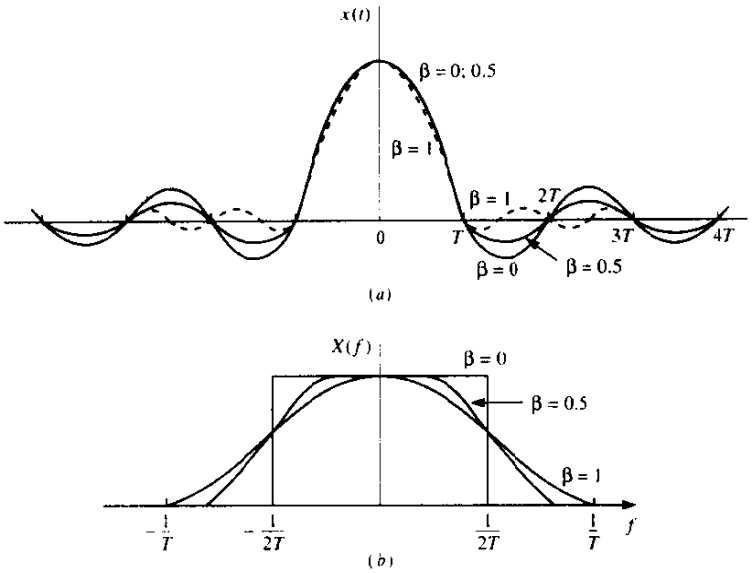


FIGURE 9-2-7 Pulses having a raised cosine spectrum.

Note that $x(t)$ is normalized so that $x(0) = 1$. Figure 9-2-7 illustrates the raised cosine spectral characteristics and the corresponding pulses for $\beta = 0$, $\frac{1}{2}$ and 1. Note that for $\beta = 0$, the pulse reduces to $x(t) = \text{sinc}(\pi t/T)$, and the symbol rate $1/T = 2W$. When $\beta = 1$, the symbol rate is $1/T = W$. In general, the tails of $x(t)$ decay as $1/t^3$ for $\beta > 0$. Consequently, a mistiming error in sampling leads to a series of ISI components that converges to a finite value.

Due to the smooth characteristics of the raised cosine spectrum, it is possible to design practical filters for the transmitter and the receiver that approximate the overall desired frequency response. In the special case where the channel is ideal, i.e., $C(f) = 1$, $|f| \leq W$, we have

$$X_{rc}(f) = G_T(f)G_R(f), \tag{9-2-28}$$

where $G_T(f)$ and $G_R(f)$ are the frequency responses of the two filters. In this case, if the receiver filter is matched to the transmitter filter, we have $X_{rc}(f) = G_T(f)G_R(f) = |G_T(f)|^2$. Ideally,

$$G_T(f) = \sqrt{|X_{rc}(f)|}\, e^{-j2\pi f t_0} \tag{9-2-29}$$

and $G_R(f) = G_T^*(f)$, where t_0 is some nominal delay that is required to ensure physical realizability of the filter. Thus, the overall raised cosine spectral characteristic is split evenly between the transmitting filter and the receiving filter. Note also that an additional delay is necessary to ensure the physical realizability of the receiving filter.

9-2-2 Design of Band-Limited Signals with Controlled ISI—Partial-Response Signals

As we have observed from our discussion of signal design for zero ISI, it is necessary to reduce the symbol rate $1/T$ below the Nyquist rate of $2W$ symbols/s to realize practical transmitting and receiving filters. On the other hand, suppose we choose to relax the condition of zero ISI and, thus, achieve a symbol transmission rate of $2W$ symbols/s. By allowing for a controlled amount of ISI, we can achieve this symbol rate.

We have already seen that the condition for zero ISI is $x(nT) = 0$ for $n \neq 0$. However, suppose that we design the band-limited signal to have controlled ISI at one time instant. This means that we allow one additional nonzero value in the samples $\{x(nT)\}$. The ISI that we introduce is deterministic or "controlled" and, hence, it can be taken into account at the receiver, as discussed below.

One special case that leads to (approximately), physically realizable transmitting and receiving filters is specified by the samples†

$$x(nT) = \begin{cases} 1 & (n = 0, 1) \\ 0 & (\text{otherwise}) \end{cases} \tag{9-2-30}$$

Now, using (9-2-20), we obtain

$$b_n = \begin{cases} T & (n = 0, -1) \\ 0 & (\text{otherwise}) \end{cases} \tag{9-2-31}$$

which, when substituted into (9-2-18), yields

$$B(f) = T + Te^{-j2\pi fT} \tag{9-2-32}$$

As in the preceding section, it is impossible to satisfy the above equation for $T < 1/2W$. However, for $T = 1/2W$, we obtain

$$\begin{aligned} X(f) &= \begin{cases} \dfrac{1}{2W}(1 + e^{-j\pi f/W}) & (|f| < W) \\ 0 & (\text{otherwise}) \end{cases} \\ &= \begin{cases} \dfrac{1}{W} e^{-j\pi f/2W} \cos\dfrac{\pi f}{2W} & (|f| < W) \\ 0 & (\text{otherwise}) \end{cases} \end{aligned} \tag{9-2-33}$$

Therefore, $x(t)$ is given by

$$x(t) = \text{sinc}\,(2\pi Wt) + \text{sinc}\,[2\pi(Wt - \tfrac{1}{2})] \tag{9-2-34}$$

This pulse is called a *duobinary signal pulse*. It is illustrated along with its

†It is convenient to deal with samples of $x(t)$ that are normalized to unity for $n = 0, 1$.

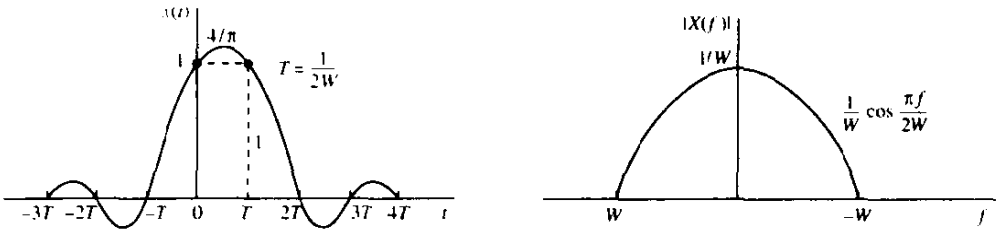


FIGURE 9-2-8 Time domain and frequency domain characteristics of a duobinary signal.

magnitude spectrum in Fig. 9-2-8. Note that the spectrum decays to zero smoothly, which means that physically realizable filters can be designed that approximate this spectrum very closely. Thus, a symbol rate of $2W$ is achieved.

Another special case that leads to (approximately) physically realizable transmitting and receiving filters is specified by the samples

$$x\left(\frac{n}{2W}\right) = x(nT) = \begin{cases} 1 & (n = -1) \\ -1 & (n = 1) \\ 0 & (\text{otherwise}) \end{cases} \tag{9-2-35}$$

The corresponding pulse $x(t)$ is given as

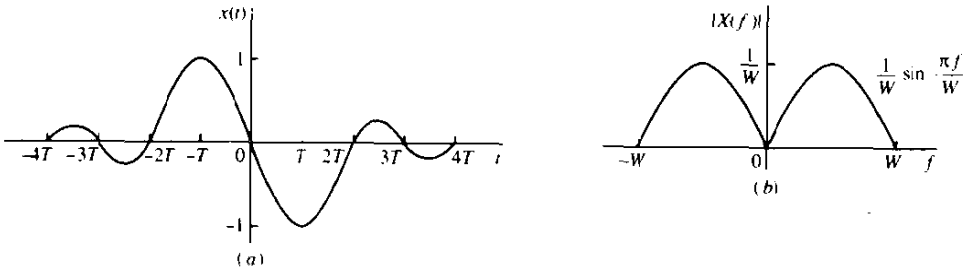
$$x(t) = \text{sinc}\left[\frac{\pi(t+T)}{T}\right] - \text{sinc}\left[\frac{\pi(t-T)}{T}\right] \tag{9-2-36}$$

and its spectrum is

$$X(f) = \begin{cases} \dfrac{1}{2W}(e^{j\pi f/W} - e^{-j\pi f/W}) = \dfrac{j}{W}\sin\dfrac{\pi f}{W} & |f| \leq W \\ 0 & |f| > W \end{cases} \tag{9-2-37}$$

This pulse and its magnitude spectrum are illustrated in Fig. 9-2-9. It is called a *modified duobinary signal pulse*. It is interesting to note that the spectrum of

FIGURE 9-2-9 Time domain and frequency domain characteristics of a modified duobinary signal.



this signal has a zero at $f = 0$, making it suitable for transmission over a channel that does not pass d.c.

One can obtain other interesting and physically realizable filter characteristics, as shown by Kretzmer (1966) and Lucky *et al.* (1968), by selecting different values for the samples $\{x(n/2W)\}$ and more than two nonzero samples. However, as we select more nonzero samples, the problem of unraveling the controlled ISI becomes more cumbersome and impractical.

In general, the class of bandlimited signals pulses that have the form

$$x(t) = \sum_{n=-\infty}^{\infty} x\left(\frac{n}{2W}\right) \operatorname{sinc}\left[2\pi W\left(t - \frac{n}{2W}\right)\right] \tag{9-2-38}$$

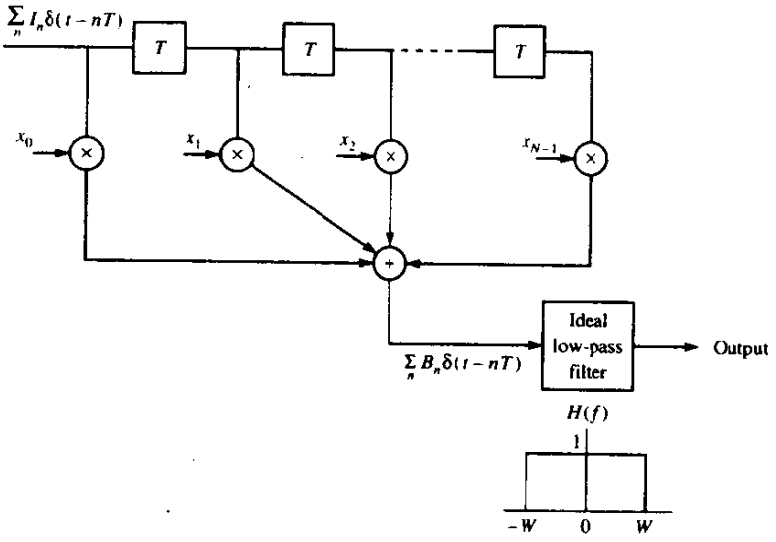
and their corresponding spectra

$$X(f) = \begin{cases} \dfrac{1}{2W} \displaystyle\sum_{n=-\infty}^{\infty} x\left(\frac{n}{2W}\right) e^{-jn\pi f/W} & (|f| \leq W) \\ 0 & (|f| \leq W) \end{cases} \tag{9-2-39}$$

are called *partial-response signals* when controlled ISI is purposely introduced by selecting two or more nonzero samples from the set $\{x(n/2W)\}$. The resulting signal pulses allow us to transmit information symbols at the Nyquist rate of $2W$ symbols/s. The detection of the received symbols in the presence of controlled ISI is described below.

Alternative Characterization of Partial-Response Signals We conclude this subsection by presenting another interpretation of a partial-response signal. Suppose that the partial-response signal is generated, as shown in Fig. 9-2-10, by passing the discrete-time sequence $\{I_n\}$ through a discrete-time filter

FIGURE 9-2-10 An alternative method for generating a partial-response signal.



with coefficients $x_n \equiv x(n/2W)$, $n = 0, 1, \ldots, N-1$, and using the output sequence $\{B_n\}$ from this filter to excite periodically with an input $B_n\delta(t - nT)$ an analog filter having an impulse response $\text{sinc}(2\pi Wt)$. The resulting output signal is identical to the partial-response signal given by (9-2-38).

Since

$$B_n = \sum_{k=0}^{N-1} x_k I_{n-k} \tag{9-2-40}$$

the sequence of symbols $\{B_n\}$ is correlated as a consequence of the filtering performed on the sequence $\{I_n\}$. In fact, the autocorrelation function of the sequence $\{B_n\}$ is

$$\begin{aligned}\phi(m) &= E(B_n B_{n+m}) \\ &= \sum_{k=0}^{N-1}\sum_{l=0}^{N-1} x_k x_l E(I_{n-k} I_{n+m-l})\end{aligned} \tag{9-2-41}$$

When the input sequence is zero-mean and white,

$$E(I_{n-k} I_{n+m-l}) = \delta_{m+k-l} \tag{9-2-42}$$

where we have used the normalization $E(I_n^2) = 1$. Substitution of (9-2-42), into (9-2-41) yields the desired autocorrelation function for $\{B_n\}$ in the form

$$\phi(m) = \sum_{k=0}^{N-1-|m|} x_k x_{k+|m|}, \qquad m = 0, \pm 1, \ldots, \pm(N-1) \tag{9-2-43}$$

The corresponding power spectral density is

$$\begin{aligned}\Phi(f) &= \sum_{m=-(N-1)}^{N-1} \phi(m) e^{-j2\pi fmT} \\ &= \left| \sum_{m=0}^{N-1} x_m e^{-j2\pi fmT} \right|^2\end{aligned} \tag{9-2-44}$$

where $T = 1/2W$ and $|f| \leq 1/2T = W$.

9-2-3 Data Detection for Controlled ISI

In this section, we describe two methods for detecting the information symbols at the receiver when the received signal contains controlled ISI. One is a symbol-by-symbol detection method that is relatively easy to implement. The second method is based on the maximum-likelihood criterion for detecting a sequence of symbols. The latter method minimizes the probability of error but is a little more complex to implement. In particular, we consider the detection of the duobinary and the modified duobinary partial response signals. In both

cases, we assume that the desired spectral characteristic $X(f)$ for the partial response signal is split evenly between the transmitting and receiving filters, i.e., $|G_T(f)| = |G_R(f)| = |X(f)|^{1/2}$. This treatment is based on PAM signals, but it is easily generalized to QAM and PSK.

Symbol-by-Symbol Suboptimum Detection For the duobinary signal pulse, $x(nT) = 1$, for $n = 0, 1$, and zero otherwise. Hence, the samples at the output of the receiving filter (demodulator) have the form

$$y_m = B_m + \nu_m = I_m + I_{m-1} + \nu_m \tag{9-2-45}$$

where $\{I_m\}$ is the transmitted sequence of amplitudes and $\{\nu_m\}$ is a sequence of additive gaussian noise samples. Let us ignore the noise for the moment and consider the binary case where $I_m = \pm 1$ with equal probability. Then B_m takes on one of three possible values, namely, $B_m = -2, 0, 2$ with corresponding probabilities 1/4, 1/2, 1/4. If I_{m-1} is the detected symbol from the $(m-1)$ th signaling interval, its effect on B_m, the received signal in the m th signaling interval, can be eliminated by subtraction, thus allowing I_m to be detected. This process can be repeated sequentially for every received symbol.

The major problem with this procedure is that errors arising from the additive noise tend to propagate. For example, if I_{m-1} is in error, its effect on B_m is not eliminated but, in fact, it is reinforced by the incorrect subtraction. Consequently, the detection of B_m is also likely to be in error.

Error propagation can be avoided by *precoding* the data at the transmitter instead of eliminating the controlled ISI by subtraction at the receiver. The precoding is performed on the binary data sequence prior to modulation. From the data sequence $\{D_n\}$ of 1s and 0s that is to be transmitted, a new sequence $\{P_n\}$, called the *precoded sequence*, is generated. For the duobinary signal, the precoded sequence is defined as

$$P_m = D_m \ominus P_{m-1}, \qquad m = 1, 2, \ldots \tag{9-2-46}$$

where $\ominus$ denotes modulo-2 subtraction.† Then we set $I_m = -1$ if $P_m = 0$ and $I_m = 1$ if $P_m = 1$, i.e., $I_m = 2P_m - 1$. Note that this precoding operation is identical to that described in Section 4-3-2 in the context of our discussion of an NRZI signal.

The noise-free samples at the output of the receiving filter are given by

$$\begin{aligned} B_m &= I_m + I_{m-1} \\ &= (2P_m - 1) + (2P_{m-1} - 1) \\ &= 2(P_m + P_{m-1} - 1) \end{aligned} \tag{9-2-47}$$

Consequently,

$$P_m + P_{m-1} = \tfrac{1}{2}B_m + 1 \tag{9-2-48}$$

†Although this is identical to modulo-2 addition, it is convenient to view the precoding operation for duobinary in terms of modulo-2 subtraction.

TABLE 9-2-1 BINARY SIGNALING WITH DUOBINARY PULSES

Data sequence D_n		1	1	1	0	1	0	0	1	0	0	0	1	1	0	1
Precoded sequence P_n	0	1	0	1	1	0	0	0	1	1	1	1	0	1	1	0
Transmitted sequence I_m	−1	1	−1	1	1	−1	−1	−1	1	1	1	1	−1	1	1	−1
Received sequence B_n		0	0	0	2	0	−2	−2	0	2	2	2	0	0	2	0
Decoded sequence D_n		1	1	1	0	1	0	0	1	0	0	0	1	1	0	1

Since $D_m = P_m \oplus P_{m-1}$, it follows that the data sequence D_m is obtained from B_m using the relation

$$D_m = \tfrac{1}{2}B_m + 1 \pmod 2 \tag{9-2-49}$$

Consequently, if $B_m = \pm 2$ then $D_m = 0$, and if $B_m = 0$ then $D_m = 1$. An example that illustrates the precoding and decoding operations is given in Table 9-2-1. In the presence of additive noise, the sampled outputs from the receiving filter are given by (9-2-45). In this case $y_m = B_m + v_m$ is compared with the two thresholds set at $+1$ and -1. The data sequence $\{D_n\}$ is obtained according to the detection rule

$$D_m = \begin{cases} 1 & (|y_m| < 1) \\ 0 & (|y_m| \geq 1) \end{cases} \tag{9-2-50}$$

The extension from binary PAM to multilevel PAM signaling using the duobinary pulses is straightforward. In this case the M-level amplitude sequence $\{I_m\}$ results in a (noise-free) sequence

$$B_m = I_m + I_{m-1}, \qquad m = 1, 2, \ldots \tag{9-2-51}$$

which has $2M - 1$ possible equally spaced levels. The amplitude levels are determined from the relation

$$I_m = 2P_m - (M - 1) \tag{9-2-52}$$

where $\{P_m\}$ is the precoded sequence that is obtained from an M-level data sequence $\{D_m\}$ according to the relation

$$P_m = D_m \ominus P_{m-1} \pmod M \tag{9-2-53}$$

where the possible values of the sequence $\{D_m\}$ are $0, 1, 2, \ldots, M - 1$.

In the absence of noise, the samples at the output of the receiving filter may be expressed as

$$B_m = I_m + I_{m-1} = 2[P_m + P_{m-1} - (M - 1)] \tag{9-2-54}$$

TABLE 9-2-2 FOUR-LEVEL SIGNAL TRANSMISSION WITH DUOBINARY PULSES

Data sequence D_m		0	0	1	3	1	2	0	3	3	2	0	1	0
Precoded sequence P_m	0	0	0	1	2	3	3	1	2	1	1	3	2	2
Transmitted sequence I_m	−3	−3	−3	−1	1	3	3	−1	1	−1	−1	3	1	1
Received sequence B_m		−6	−6	−4	0	4	6	2	0	0	−2	2	4	2
Decoded sequence D_m		0	0	1	3	1	2	0	3	3	2	0	1	0

Hence,

$$P_m + P_{m-1} = \tfrac{1}{2}B_m + (M-1) \tag{9-2-55}$$

Since $D_m = P_m + P_{m-1} \pmod M$, it follows that

$$D_m = \tfrac{1}{2}B_m + (M-1) \pmod M \tag{9-2-56}$$

An example illustrating multilevel precoding and decoding is given in Table 9-2-2.

In the presence of noise, the received signal-plus-noise is quantized to the nearest of the possible signal levels and the rule given above is used on the quantized values to recover the data sequence.

In the case of the modified duobinary pulse, the controlled ISI is specified by the values $x(n/2W) = -1$, for $n = 1$, $x(n/2W) = 1$ for $n = -1$, and zero otherwise. Consequently, the noise-free sampled output from the receiving filter is given as

$$B_m = I_m - I_{m-2} \tag{9-2-57}$$

where the M-level sequence $\{I_m\}$ is obtained by mapping a precoded sequence according to the relation (9-2-52) and

$$P_m = D_m \ominus P_{m-2} \pmod M \tag{9-2-58}$$

From these relations, it is easy to show that the detection rule for recovering the data sequence $\{D_m\}$ from $\{B_m\}$ in the absence of noise is

$$D_m = \tfrac{1}{2}B_m \pmod M \tag{9-2-59}$$

As demonstrated above, the precoding of the data at the transmitter makes it possible to detect the received data on a symbol-by-symbol basis without having to look back at previously detected symbols. Thus, error propagation is avoided.

The symbol-by-symbol detection rule described above is not the optimum detection scheme for partial response signals due to the memory inherent in

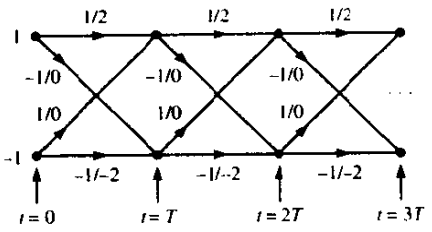


FIGURE 9-2-11 Trellis for duobinary partial response signal.

the received signal. Nevertheless, symbol-by-symbol detection is relatively simple to implement and is used in many practical applications involving duobinary and modified duobinary pulse signals. Its performance is evaluated in the following section.

Maximum-Likelihood Sequence Detection It is clear from the above discussion that partial-response waveforms are signal waveforms with memory. This memory is conveniently represented by a trellis. For example, the trellis for the duobinary partial-response signal for binary data transmission is illustrated in Fig. 9-2-11. For binary modulation, this trellis contains two states, corresponding to the two possible input values of I_m, i.e., $I_m = \pm 1$. Each branch in the trellis is labeled by two numbers. The first number on the left is the new data bit, i.e., $I_{m+1} = \pm 1$. This number determines the transition to the new state. The number on the right is the received signal level.

The duobinary signal has a memory of length $L = 1$. Hence, for binary modulation the trellis has $S_t = 2$ states. In general, for M-ary modulation, the number of trellis states is M^L.

The optimum maximum-likelihood (ML) sequence detector selects the most probable path through the trellis upon observing the received data sequence $\{y_m\}$ at the sampling instants $t = mT$, $m = 1, 2, \ldots$. In general, each node in the trellis will have M incoming paths and M corresponding metrics. One out of the M incoming paths is selected as the most probable, based on the values of the metrics and the other $M - 1$ paths and their metrics are discarded. The surviving path at each node is then extended to M new paths, one for each of the M possible input symbols, and the search process continues. This is basically the Viterbi algorithm for performing the trellis search.

For the class of partial response signals, the received sequence $\{y_m, 1 \leq m \leq N\}$ is generally described statistically by the joint pdf $f(\mathbf{y}_N \mid \mathbf{I}_N)$, where $\mathbf{y}_N = [y_1 \; y_2 \; \cdots \; y_N]'$ and $\mathbf{I}_N = [I_1 \; I_2 \; \cdots \; I_N]'$ and $N > L$. When the additive noise is zero-mean gaussian, $f(\mathbf{y}_N \; \mathbf{I}_N)$ is a multivariate gaussian pdf, i.e.,

$$f(\mathbf{y}_N \mid \mathbf{I}_N) = \frac{1}{(2\pi \det \mathbf{C})^{N/2}} \exp\left[-\tfrac{1}{2}(\mathbf{y}_N - \mathbf{B}_N)'\mathbf{C}^{-1}(\mathbf{y}_N - \mathbf{B}_N)\right] \quad (9\text{-}2\text{-}60)$$

where $\mathbf{B}_N = [B_1 \; B_2 \; \cdots \; B_N]'$, is the mean of the vector $\mathbf{y}_N$ and $\mathbf{C}$ is the $N \times N$ covariance matrix of $\mathbf{y}_N$. Then, the ML sequence detector selects the sequence through the trellis that maximizes the pdf $f(\mathbf{y}_N \mid \mathbf{I}_N)$.

The computation for finding the most probable sequence through the trellis is simplified by taking the natural logarithms of $f(\mathbf{y}_N \mid \mathbf{I}_N)$. Thus,

$$\ln f(\mathbf{y}_N \mid \mathbf{I}_N) = -\tfrac{1}{2}N \ln (2\pi \det \mathbf{C}) - \tfrac{1}{2}(\mathbf{y}_N - \mathbf{B}_N)'\mathbf{C}^{-1}(\mathbf{y}_N - \mathbf{B}_N) \quad (9\text{-}2\text{-}61)$$

Given the received sequence $\{y_m\}$, the data sequence $\{I_m\}$ that maximizes $\ln f(\mathbf{y}_N \mid \mathbf{I}_N)$ is identical to the sequence $\{I_N\}$ that minimizes $(\mathbf{y}_N - \mathbf{B}_N)'\mathbf{C}^{-1}(\mathbf{y}_N - \mathbf{B}_N)$, i.e.

$$\hat{\mathbf{I}}_N = \arg\min_{\mathbf{I}_N} [(\mathbf{y}_N - \mathbf{B}_N)'\mathbf{C}^{-1}(\mathbf{y}_N - \mathbf{B}_N)] \quad (9\text{-}2\text{-}62)$$

The metric computations in the trellis search are complicated by the correlation of the noise samples at the output of the matched filter for the partial response signal. For example, in the case of the duobinary signal waveform, the correlation of the noise sequence $\{v_m\}$ is over two successive signal samples. Hence, v_m and v_{m+k} are correlated for $k = 1$ and uncorrelated for $k > 1$. In general, a partial response signal waveform with memory L will result in a correlated noise sequence at the output of the matched filter, which satisfies the condition $E[v_m v_{m+k}] = 0$ for $k > L$. In such a case, the Viterbi algorithm for performing the trellis search may be modified as described in Chapter 10.

Some simplification in the metric computations result if we ignore the noise correlation by assuming that $E(v_m v_{m+k}) = 0$ for $k > 0$. Then, by assumption, the covariance matrix $\mathbf{C} = \sigma_v^2 \mathbf{1}_N$, where $\sigma_v^2 = E[v_m^2]$ and $\mathbf{1}_N$ is the $N \times N$ identity matrix.† In this case, (9-2-62) simplifies to

$$\begin{aligned}\hat{\mathbf{I}}_N &= \arg\min_{\mathbf{I}_N} [(\mathbf{y}_N - \mathbf{B}_N)'(\mathbf{y}_N - \mathbf{B}_N)] \\ &= \arg\min_{\mathbf{I}_N} \left[\sum_{m=1}^{N} \left(y_m - \sum_{k=0}^{L} x_k I_{m-k} \right)^2 \right] \end{aligned} \quad (9\text{-}2\text{-}63)$$

where

$$B_m = \sum_{k=0}^{L} x_k I_{m-k}$$

and $x_k = x(kT)$ are the sampled values of the partial response signal waveform. In this case, the metric computations at each node of the trellis have the form

$$DM_m(\mathbf{I}_m) = DM_{m-1}(\mathbf{I}_{m-1}) + \left(y_m - \sum_{k=0}^{L} x_k I_{m-k} \right)^2 \quad (9\text{-}2\text{-}64)$$

where $DM_m(\mathbf{I}_m)$ are the distance metrics at time $t = mT$, $DM_{m-1}(\mathbf{I}_{m-1})$ are the distance metrics at time $t = (m-1)T$ and the second term on the right-hand side of (9-2-64) represents the new increments to the metrics based on the new received sample y_m.

†We are using $\mathbf{1}_N$ here to avoid confusion with $\mathbf{I}_N$.

As indicated in Section 5-1-4, ML sequence detection introduces a variable delay in detecting each transmitted information symbol. In practice, the variable delay is avoided by truncating the surviving sequences to N_t most recent symbols, where $N_t \gg 5L$, thus achieving a fixed delay. In the case that the M^L surviving sequences at time $t = mT$ disagree on the symbol I_{m-N_t}, the symbol in the most probable surviving sequence may be chosen. The loss in performance resulting from this truncation is negligible if $N_t > 5L$.

9-2-4 Signal Design for Channels with Distortion

In Sections 9-2-1 and 9-2-2, we described signal design criteria for the modulation filter at the transmitter and the demodulation filter at the receiver when the channel is ideal. In this section, we perform the signal design under the condition that the channel distorts the transmitted signal. We assume that the channel frequency response $C(f)$ is known for $|f| \leq W$ and that $C(f) = 0$ for $|f| > W$. The criterion for the optimization of the filter responses $G_T(f)$ and $G_R(f)$ is the maximization of the SNR at the output of the demodulation filter or equivalently, at the input to the detector. The additive channel noise is assumed to be gaussian with power spectral density $\Phi_{nn}(f)$. Figure 9-2-12 illustrates the overall system under consideration.

For the signal component at the output of the demodulator, we must satisfy the condition

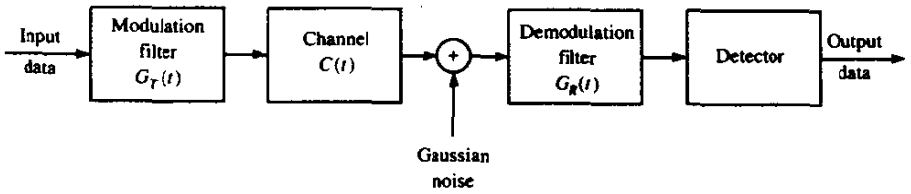
$$G_T(f)C(f)G_R(f) = X_d(f)e^{-j2\pi f t_0}, \quad |f| \leq W \tag{9-2-65}$$

where $X_d(f)$ is the desired frequency response of the cascade of the modulator, channel, and demodulator, and t_0 is a time delay that is necessary to ensure the physical realizability of the modulation and demodulation filters. The desired frequency response $X_d(f)$ may be selected to yield either zero ISI or controlled ISI at the sampling instants. We shall carry out the optimization for zero ISI by selecting $X_d(f) = X_{rc}(f)$, where $X_{rc}(f)$ is the raised cosine spectrum with an arbitrary rolloff factor.

The noise at the output of the demodulation filter may be expressed as

$$v(t) = \int_{-\infty}^{\infty} n(t-\tau)g_R(\tau)\,d\tau \tag{9-2-66}$$

FIGURE 9-2-12 System model for the design of the modulation and demodulation filters.



where $n(t)$ is the input to the filter. Since $n(t)$ is zero-mean gaussian, $v(t)$ is zero-mean gaussian, with a power spectral density

$$\Phi_{vv}(f) = \Phi_{nn}(f)\,|G_R(f)|^2 \tag{9-2-67}$$

For simplicity, we consider binary PAM transmission. Then, the sampled output of the matched filter is

$$y_m = x_0 I_m + v_m = I_m + v_m \tag{9-2-68}$$

where x_0 is normalized† to unity, $I_m = \pm d$, and v_m represents the noise term, which is zero-mean gaussian with variance

$$\sigma_v^2 = \int_{-\infty}^{\infty} \Phi_{nn}(f)\,|G_R(f)|^2\,df \tag{9-2-69}$$

Consequently, the probability of error is

$$P_2 = \frac{1}{\sqrt{2\pi}} \int_{d/\sigma_v}^{\infty} e^{-y^2/2}\,dy = Q(\sqrt{d^2/\sigma_v^2}) \tag{9-2-70}$$

The probability of error is minimized by maximizing the SNR $= d^2/\sigma_v^2$, or, equivalently, by minimizing the noise-to-signal ratio σ_v^2/d^2. But d^2 is related to the transmitted signal power as follows:

$$\begin{aligned} P_{av} &= \frac{E(I_m^2)}{T} \int_{-\infty}^{\infty} g_T^2(t)\,dt = \frac{d^2}{T} \int_{-\infty}^{\infty} g_T^2(t)\,dt \\ \frac{1}{d^2} &= \frac{1}{P_{av}T} \int_{-\infty}^{\infty} |G_T(f)|^2\,df \end{aligned} \tag{9-2-71}$$

However, $G_T(f)$ must be chosen to satisfy the zero ISI condition. Consequently,

$$|G_T(f)| = \frac{|X_{rc}(f)|}{|C(f)|\,|G_R(f)|}, \qquad |f| \leq W \tag{9-2-72}$$

and $G_T(f) = 0$ for $|f| \geq W$. Hence

$$\frac{1}{d^2} = \frac{1}{P_{av}T} \int_{-W}^{W} \frac{|X_{rc}(f)|^2}{|C(f)|^2\,|G_R(f)|^2}\,df \tag{9-2-73}$$

Therefore, the noise-to-signal ratio that must be minimized with respect to $|G_R(f)|$ for $|f| \leq W$ is

$$\frac{\sigma_v^2}{d^2} = \frac{1}{P_{av}T} \int_{-W}^{W} \Phi_{nn}(f)\,|G_R(f)|^2\,df \int_{-W}^{W} \frac{|X_{rc}(f)|^2}{|C(f)|^2\,|G_R(f)|^2}\,df \tag{9-2-74}$$

†By setting $x_0 = 1$ and $I_m = \pm d$, the scaling by x_0 is incorporated into the parameter d.

The optimum $|G_R(f)|$ can be found by applying the Cauchy–Schwartz inequality,

$$\int_{-\infty}^{\infty} |U_1(f)|^2\, df \int_{-\infty}^{\infty} |U_2(f)|^2\, df \geq \left[\int_{-\infty}^{\infty} |U_1(f)|\,|U_2(f)|\, df\right]^2 \tag{9-2-75}$$

where $|U_1(f)|$ and $|U_2(f)|$ are defined as

$$\begin{aligned} |U_1(f)| &= |\sqrt{\Phi_{nn}(f)}|\,|G_R(f)| \\ |U_2(f)| &= \frac{|X_{rc}(f)|}{|C(f)|\,|G_R(f)|} \end{aligned} \tag{9-2-76}$$

The minimum value of (9-2-74) is obtained when $|U_1(f)|$ is proportional to $|U_2(f)|$, or, equivalently, when

$$|G_R(f)| = K\frac{|X_{rc}(f)|^{1/2}}{[\Phi_{nn}(f)]^{1/4}\,|C(f)|^{1/2}}, \qquad |f| \leq W \tag{9-2-77}$$

where K is an arbitrary constant. The corresponding modulation filter has a magnitude characteristic

$$|G_T(f)| = \frac{1}{K}\frac{|X_{rc}(f)|^{1/2}\,[\Phi_{nn}(f)]^{1/4}}{|C(f)|^{1/2}}, \qquad |f| \leq W \tag{9-2-78}$$

Finally, the maximum SNR achieved by these optimum transmitting and receiving filters is

$$\frac{d^2}{\sigma_\nu^2} = \frac{P_{\text{av}}T}{\{\int_{-W}^{W} |X_{rc}(f)|\,[\Phi_{nn}(f)]^{1/2}\,|C(f)|^{-1}\, df\}^2} \tag{9-2-79}$$

We note that the optimum modulation and demodulation filters are specified in magnitude only. The phase characteristics for $G_T(f)$ and $G_R(f)$ may be selected so as to satisfy the condition in (9-2-65), i.e.,

$$\Theta_T(f) + \Theta_c(f) + \Theta_R(f) = 2\pi f t_0 \tag{9-2-80}$$

where $\Theta_T(f)$, $\Theta_c(f)$, and $\Theta_R(f)$ are the phase characteristics of the modulation filter, the channel, and the demodulation filter, respectively.

In the special case where the additive noise at the input to the demodulator is white gaussian with spectral density $\frac{1}{2}N_0$, the optimum filter characteristics specified by (9-2-77) and (9-2-78) reduce to

$$\begin{aligned} |G_R(f)| &= K_1\frac{|X_{rc}(f)|^{1/2}}{|C(f)|^{1/2}}, \qquad |f| \leq W \\ |G_T(f)| &= K_2\frac{|X_{rc}(f)|^{1/2}}{|C(f)|^{1/2}}, \qquad |f| \leq W \end{aligned} \tag{9-2-81}$$

where K_1 and K_2 are arbitrary scale factors. Note that, in this case, $|G_R(f)|$ is the matched filter to $|G_T(f)|$. The corresponding SNR at the detector, given by (9-2-79) reduces to

$$\frac{d^2}{\sigma_v^2} = \frac{2P_{av}T}{N_0}\left[\int_{-W}^{W} \frac{|X_{rc}(f)|}{|C(f)|}\,df\right]^{-2} \tag{9-2-82}$$

Example 9-2-1

Let us determine the optimum transmitting and receiving filters for a binary communication system that transmits data at a rate of 4800 bits/s over a channel with frequency (magnitude) response

$$|C(f)| = \frac{1}{\sqrt{1 + (f/W)^2}}, \qquad |f| \leq W \tag{9-2-83}$$

where $W = 4800$ Hz. The additive noise is zero-mean, white, gaussian with spectral density $\frac{1}{2}N_0 = 10^{-15}\,W/Hz$.

Since $W = 1/T = 4800$, we use a signal pulse with a raised cosine spectrum and $\beta = 1$. Thus,

$$\begin{aligned} X_{rc}(f) &= \tfrac{1}{2}T[1 + \cos(\pi T\,|f|)] \\ &= T\cos^2\left(\frac{\pi\,|f|}{9600}\right) \end{aligned} \tag{9-2-84}$$

Then,

$$|G_T(f) = |G_R(f)| = \left[1 + \left(\frac{f}{4800}\right)^2\right]^{1/4} \cos\left(\frac{\pi\,|f|}{9600}\right), \qquad |f| \leq 4800 \tag{9-2-85}$$

and $|G_T(f)| = |G_R(f)| = 0$, otherwise. Figure 9-2-13 illustrates the filter characteristic $G_T(f)$.

One can now use these optimum filters to determine the amount of transmitted energy $\mathscr{E}$ required to achieve a specified error probability. This problem is left as an exercise for the reader.

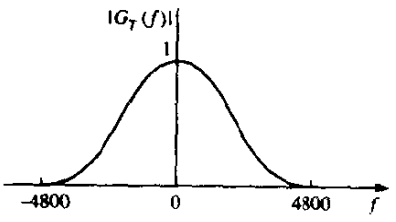


FIGURE 9-2-13 Frequency response of optimum transmitter filter.

9-3 PROBABILITY OF ERROR IN DETECTION OF PAM

In this section, we evaluate the performance of the receiver for demodulating and detecting an M-ary PAM signal in the presence of additive, white, gaussian noise at its input. First, we consider the case in which the transmitter and receiver filters $G_T(f)$ and $G_R(f)$ are designed for zero ISI. Then, we consider the case in which $G_T(f)$ and $G_R(f)$ are designed such that $x(t) = g_T(t) \star g_R(t)$ is either a duobinary signal or a modified duobinary signal.

9-3-1 Probability of Error for Detection of PAM with Zero ISI

In the absence of ISI, the received signal sample at the output of the receiving matched filter has the form

$$y_m = x_0 I_m + v_m \tag{9-3-1}$$

where

$$x_0 = \int_{-W}^{W} |G_T(f)|^2 \, df = \mathscr{E}_g \tag{9-3-2}$$

and v_m is the additive gaussian noise that has zero mean and variance

$$\sigma_v^2 = \tfrac{1}{2}\mathscr{E}_g N_0 \tag{9-3-3}$$

In general, I_m takes one of M possible equally spaced amplitude values with equal probability. Given a particular amplitude level, the problem is to determine the probability of error.

The problem of evaluating the probability of error for digital PAM in a band-limited, additive white gaussian noise channel, in the absence of ISI, is identical to the evaluation of the error probability for M-ary PAM as given in Section 5-2. The final result that is obtained from the derivation is

$$P_M = \frac{2(M-1)}{M} Q\left(\sqrt{\frac{2\mathscr{E}_g}{N_0}}\right) \tag{9-3-4}$$

But $\mathscr{E}_g = 3\mathscr{E}_{\text{av}}/(M^2-1)$, $\mathscr{E}_{\text{av}} = k\mathscr{E}_{b\text{av}}$ is the average energy per symbol and $\mathscr{E}_{b\text{av}}$ is the average energy per bit. Hence,

$$P_M = \frac{2(M-1)}{M} Q\left(\sqrt{\frac{6(\log_2 M)\mathscr{E}_{b\text{av}}}{(M^2-1)N_0}}\right) \tag{9-3-5}$$

This is exactly the form for the probability of error of M-ary PAM derived in Section 5-2 (see (5-2-46)). In the treatment of PAM given in this chapter, we imposed the additional constraint that the transmitted signal is band-limited to the bandwidth allocated for the channel. Consequently, the transmitted signal pulses were designed to be band-limited and to have zero ISI.

In contrast, no bandwidth constraint was imposed on the PAM signals considered in Section 5-2. Nevertheless, the receivers (demodulators and detectors) in both cases are optimum (matched filters) for the corresponding

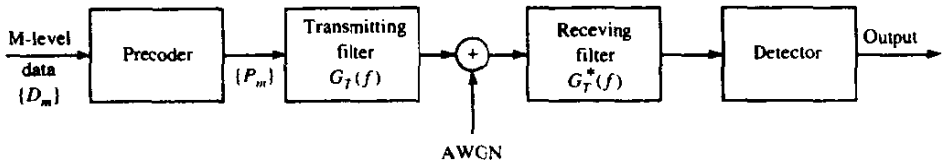


FIGURE 9-3-1 *Block diagram of modulator and demodulator for partial-response signals.*

transmitted signals. Consequently, no loss in error rate performance results from the bandwidth constraint when the signal pulse is designed for zero ISI and the channel does not distort the transmitted signal.

9-3-2 Probability of Error for Detection of Partial-Response Signals

In this section we determine the probability of error for detection of digital M-ary PAM signaling using duobinary and modified duobinary pulses. The channel is assumed to be an ideal bandlimited channel with additive white gaussian noise. The model for the communication system is shown in Fig. 9-3-1.

We consider two types of detectors. The first is the symbol-by-symbol detector and the second is the optimum ML sequence detector described in the previous section.

Symbol-by-Symbol Detector At the transmitter, the M-level data sequence $\{D_m\}$ is precoded as described previously. The precoder output is mapped into one of M possible amplitude levels. Then the transmitting filter with frequency response $G_T(f)$ has an output

$$v(t) = \sum_{n=-\infty}^{\infty} I_n g_T(t - nT) \tag{9-3-6}$$

The partial-response function $X(f)$ is divided equally between the transmitting and receiving filters. Hence, the receiving filter is matched to the transmitted pulse, and the cascade of the two filters results in the frequency characteristic

$$|G_T(f)G_R(f)| = |X(f)| \tag{9-3-7}$$

The matched filter output is sampled at $t = nT = n/2W$ and the samples are fed to the decoder. For the duobinary signal, the output of the matched filter at the sampling instant may be expressed as

$$y_m = I_m + I_{m-1} + \nu_m = B_m + \nu_m \tag{9-3-8}$$

where ν_m is the additive noise component. Similarly, the output of the matched filter for the modified duobinary signal is

$$y_m = I_m - I_{m-2} + \nu_m = B_m + \nu_m \tag{9-3-9}$$

For binary transmission, let $I_m = \pm d$, where $2d$ is the distance between signal levels. Then, the corresponding values of B_m are $(2d, 0, -2d)$. For M-ary PAM signal transmission, where $I_m = \pm d, \pm 3d, \ldots, \pm(M-1)d$, the received signal levels are $B_m = 0, \pm 2d, \pm 4d, \ldots, \pm 2(M-1)d$. Hence, the number of received levels is $2M-1$, and the scale factor d is equivalent to $x_0 = \mathscr{E}_g$.

The input transmitted symbols $\{I_m\}$ are assumed to be equally probable. Then, for duobinary and modified duobinary signals, it is easily demonstrated that, in the absence of noise, the received output levels have a (triangular) probability distribution of the form

$$P(B = 2md) = \frac{M - |m|}{M^2}, \qquad m = 0, \pm 1, \pm 2, \ldots, \pm(M-1) \tag{9-3-10}$$

where B denotes the noise-free received level and $2d$ is the distance between any two adjacent received signal levels.

The channel corrupts the signal transmitted through it by the addition of white gaussian noise with zero mean and power spectral density $\frac{1}{2}N_0$.

We assume that a symbol error occurs whenever the magnitude of the additive noise exceeds the distance d. This assumption neglects the rare event that a large noise component with magnitude exceeding d may result in a received signal level that yields a correct symbol decision. The noise component v_m is zero-mean gaussian with variance

$$\begin{aligned}\sigma_v^2 &= \tfrac{1}{2}N_0 \int_{-W}^{W} |G_R(f)|^2\, df \\ &= \tfrac{1}{2}N_0 \int_{-W}^{W} |X(f)|\, df = 2N_0/\pi\end{aligned} \tag{9-3-11}$$

for both the duobinary and the modified duobinary signals. Hence, an upper bound on the symbol probability of error is

$$\begin{aligned}P_M < & \sum_{m=-(M-2)}^{M-2} P(|y - 2md| > d \mid B = 2md) P(B = 2md) \\ & + 2P(y + 2(M-1)d > d \mid B = -2(M-1)d) P(B = -2(M-1)d) \\ = & P(|y| > d \mid b = 0)\left[2 \sum_{m=0}^{M-1} P(B = 2md) - P(B = 0) - P(B = -2(M-1)d)\right] \\ = & (1 - M^{-2}) P(|y| > d \mid B = 0)\end{aligned} \tag{9-3-12}$$

But

$$\begin{aligned}P(|y| > d \mid B = 0) &= \frac{2}{\sqrt{2\pi}\sigma_v} \int_d^{\infty} e^{-x^2/2\sigma_v^2}\, dx \\ &= 2Q(\sqrt{\pi d^2/2N_0})\end{aligned} \tag{9-3-13}$$

Therefore, the average probability of a symbol error is upper-bounded as

$$P_M < 2(1 - M^{-2})Q(\sqrt{\pi d^2/2N_0}) \tag{9-3-14}$$

The scale factor d in (9-3-14) can be eliminated by expressing it in terms of the average power transmitted into the channel. For the M-ary PAM signal in which the transmitted levels are equally probable, the average power at the output of the transmitting filter is

$$\begin{aligned} P_{av} &= \frac{E(I_m^2)}{T}\int_{-W}^{W} |G_T(f)|^2\, df \\ &= \frac{E(I_m^2)}{T}\int_{-W}^{W} |X(f)|\, df = \frac{4}{\pi T}E(I_m^2) \end{aligned} \tag{9-3-15}$$

where $E(I_m^2)$ is the mean square value of the M signal levels, which is

$$E(I_m^2) = \tfrac{1}{3}d^2(M^2 - 1) \tag{9-3-16}$$

Therefore,

$$d^2 = \frac{3\pi P_{av} T}{4(M^2 - 1)} \tag{9-3-17}$$

By substituting the value of d^2 from (9-3-17) into (9-3-14), we obtain the upper bound on the symbol error probability as

$$P_M < 2\left(1 - \frac{1}{M^2}\right)Q\left(\sqrt{\left(\frac{\pi}{4}\right)^2 \frac{6}{M^2 - 1}\frac{\mathscr{E}_{av}}{N_0}}\right) \tag{9-3-18}$$

where $\mathscr{E}_{av}$ is the average energy per transmitted symbol, which can be also expressed in terms of the average bit energy as $\mathscr{E}_{av} = k\mathscr{E}_{bav} = (\log_2 M)\mathscr{E}_{bav}$.

The expression in (9-3-18) for the probability of error of M-ary PAM holds for both duobinary and modified duobinary partial-response signals. If we compare this result with the error probability of M-ary PAM with zero ISI, which can be obtained by using a signal pulse with a raised cosine spectrum, we note that the performance of partial response duobinary or modified duobinary has a loss of $(\frac{1}{4}\pi)^2$, or 2.1 dB. This loss in SNR is due to the fact that the detector for the partial response signals makes decisions on a symbol-by symbol basis, thus ignoring the inherent memory contained in the received signal at the input to the detector.

Maximum-Likelihood Sequence Detector The ML sequence detector searches through the trellis for the most probable transmitted sequence $\{I_m\}$ as previously described in Section 9-2-3. At each stage of the search process the detector compares the metrics of paths that merge at each of the nodes and selects the path that is most probable at each node. The performance of the detector may be evaluated by determining the probability of error events, based on a euclidean distance metric, as was done for soft-decision decoding of convolutional codes. The general derivation is given in Section 10-1-4. In the

case of the duobinary and modified duobinary signals, it is demonstrated that the 2.1 dB loss inherent in the suboptimum symbol-by-symbol detector is completely recovered by the ML sequence detector.

9-3-3 Probability of Error for Optimum Signals in a Channel with Distortion

In Section 9-2-4, we derived the filter responses for the modulation and demodulation filters that maximize the SNR at the input to the detector when there is channel distortion. When the filters are designed for zero ISI at the sampling instants, the probability of error for M-ary PAM is

$$P_M = \frac{2(M-1)}{M} Q(\sqrt{d^2/\sigma_v^2}) \tag{9-3-19}$$

The parameter d is related to the average transmitted power as

$$\begin{aligned} P_{\text{av}} &= \frac{E[I_m^2]}{T} \int_{-W}^{W} |G_T(f)|^2 \, df \\ &= \frac{(M^2-1)d^2}{3T} \int_{-W}^{W} |G_T(f)|^2 \, df \end{aligned} \tag{9-3-20}$$

and the noise variance is given by (9-2-69). For AWGN, (9-3-19) may be expressed as

$$P_M = \frac{2(M-1)}{M} Q\left(\sqrt{\frac{6\mathscr{E}_{\text{av}}}{(M^2-1)N_0} \left[\int_{-W}^{W} \frac{|X_{rc}(f)|}{|C(f)|} df \right]^{-2}} \right) \tag{9-3-21}$$

Finally, we observe that the loss due to channel distortion is

$$20 \log_{10} \left[\int_{-W}^{W} \frac{|X_{rc}(f)|}{|C(f)|} df \right] \tag{9-3-22}$$

Note that when $C(f) = 1$ for $|f| \leq W$, the channel is ideal and

$$\int_{-W}^{W} X_{rc}(f) \, df = 1 \tag{9-3-23}$$

so that no loss is incurred. On the other hand, when there is amplitude distortion, $|C(f)| < 1$ for some range of frequencies in the band $|f| \leq W$ and, hence, there is a loss in SNR incurred, as given by (9-3-22). This loss is independent of channel phase distortion, because phase distortion has been perfectly compensated, as implied by (9-2-80). The loss given by (9-3-22) is due entirely to amplitude distortion and is a measure of the noise enhancement

resulting from the receiving filter, which compensates for the channel distortion.

9-4 MODULATION CODES FOR SPECTRUM SHAPING

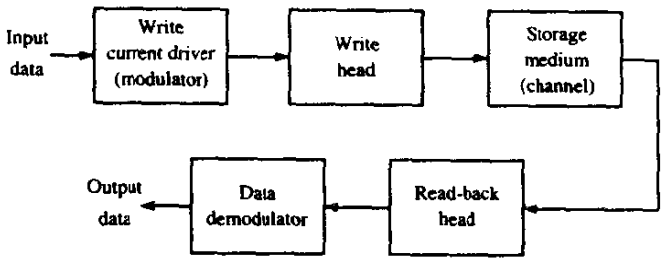
We have observed that the power spectral density of a digital communication signal can be controlled and shaped by selecting the transmitted signal pulse $g(t)$ and by introducing correlation through coding, which is used to combat channel distortion and noise in transmission. Coding for spectrum shaping is introduced following the channel encoding so that the spectrum of the transmitted signal matches the spectral characteristics of a baseband or equivalent lowpass channel.

Codes that are used for spectrum shaping are generally called either *modulation codes,* or *line codes,* or *data translation codes.* Such codes generally place restrictions on the sequence of bits into the modulator and, thus, introduce correlation and, hence, memory into the transmitted signal. It is this type of coding that is treated in this section.

Modulation codes are usually employed in magnetic recording, in optical recording, and in digital communications over cable systems to achieve spectral shaping and to eliminate or minimize the d.c. content in the transmitted (or stored) baseband signal. In magnetic recording channels, the modulation code is designed to increase the distance between transitions in the recorded waveform and, thus, intersymbol interference effects are also reduced.

As an example of the use of a modulation code, let us consider a magnetic recording system, which consists of the elements shown in the block diagram of Fig. 9-4-1. The binary data sequence to be stored is used to generate a write current. This current may be viewed as the output for the "modulator." The most commonly used method to map the information sequence into the write current waveform is NRZI, which was described in Section 4-3-2. Recall that in NRZI, a transition from one amplitude to another (A to $-$A or $-$A to A) occurs only when the information bit is a 1. No transition occurs when the information bit is a 0, i.e., the amplitude level remains the same as in the previous signal interval. The positive amplitude pulse results in magnetizing

FIGURE 9-4-1 Block diagram of magnetic storage read/write system.



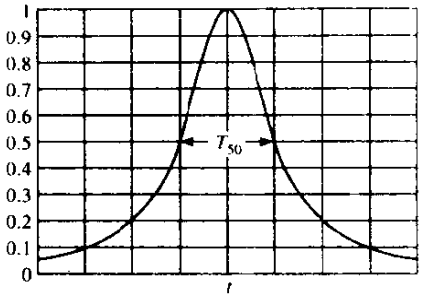


FIGURE 9-4-2 Read-back pulse in magnetic recording system.

the medium in one (direction) polarity and the negative pulse magnetizes the medium in the opposite (direction) polarity.

Since the input data sequence is basically random with equally probable 1 s and 0 s, we shall encounter level transitions from A to $-A$ or $-A$ to A with probability 1/2 for every data bit. The readback signal for a positive transition ($-A$ to A) is a pulse that is well-modeled mathematically as

$$g(t) = \frac{1}{1 + (2t/T_{50})^2} \tag{9-4-1}$$

where T_{50} is defined as the width of the pulse at its 50% amplitude level, as shown in Fig. 9-4-2. Similarly, the readback signal for a negative transition (A to $-A$) is the pulse $-g(t)$. The value of T_{50} is determined by the characteristics of the medium, the read/write heads, and the distance of the head to the medium.

Now, suppose we write a positive transition followed by a negative transition. Let's vary the time interval between the two transitions, which we denote as T_b (the bit time interval). Figure 9-4-3 illustrates the readback signal pulses, which are obtained by a superposition of $p(t)$ with $-p(t - T_b)$. The parameter $\Delta = T_{50}/T_b$ is defined as the *normalized density. The* closer the bit transitions (T_b small), the larger will be the value of the normalized density and, hence, the larger will be the packing density. We notice that as Δ is

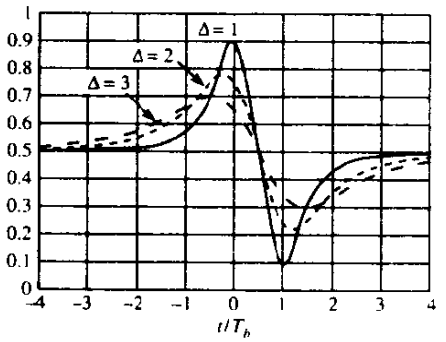


FIGURE 9-4-3 Read-back signal response to a pulse.

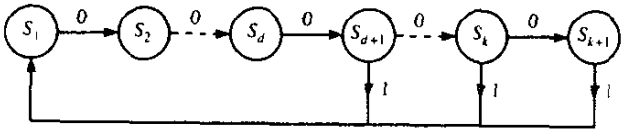
increased, the peak amplitudes of the readback signal are reduced and are also shifted in time from the desired time instants. In other words, the pulses interfere with one another, thus limiting the density with which we can write. This problem serves as a motivation to design modulation codes that take the original data sequence and transform (encode) it into another sequence that results in a write waveform in which amplitude transitions are spaced farther apart. For example, if we use NRZI, the encoded sequence into the modulator must contain one or more 0s between 1s.

The second problem encountered in magnetic recording is the need to avoid (or minimize) having a d.c. content in the modulated signal (the write current) due to the frequency response characteristics of the readback system and associated electronics. This requirement also arises in digital communication over cable channels. This problem can be overcome by altering (encoding) the data sequence into the modulator. A class of codes that satisfy these objectives are the modulation codes described below.

Runlength-Limited Codes Codes that have a restriction on the number of consecutive 1s or 0s in a sequence are generally called *runlength-limited codes*. These codes are generally described by two parameters, say d and κ, where d denotes the minimum number of 0s between two 1s in a sequence, and κ denotes the maximum number of 0s between two 1s in a sequence. When used with NRZI modulation, the effect of placing d zeros between successive 1s is to spread the transitions farther apart, thus reducing the overlap in the channel response due to successive transitions and hence reducing the intersymbol interference. Setting an upper limit κ on the runlength of 0s ensures that transitions occur frequently enough so that symbol timing information can be recovered from the received modulated signal. Runlength-limited codes are usually called (d, κ) codes.†

The (d, κ) code sequence constraints may be represented by a finite-state sequential machine with $\kappa + 1$ states, denoted as S_i, $1 \leq i \leq \kappa + 1$, as shown in Fig. 9-4-4. We observe that an output data bit 0 takes the sequence from state S_i to S_{i+1}, $i \leq \kappa$. The output data bit 1 takes the sequence to state S_1. The output bit from the encoder may be a 1 only when the sequence is in state S_i, $d + 1 \leq i \leq \kappa + 1$. When the sequence is in state $S_{\kappa+1}$, the output bit is always 1.

FIGURE 9-4-4 Finite-state sequential machine for a (d, κ)-coded sequence.



†In fact, they are usually called (d, k) codes, where k is the maximum runlength of zeros. We have substituted the Greek letter kappa κ for k, to avoid confusion with our previous use of k.

The finite-state sequential machine may also be represented by a *state transition matrix*, denoted as **D**, which is a square $(\kappa+1)\times(\kappa+1)$ with elements d_{ij}, where

$$d_{i1} = 1 \quad (i \geq d+1)$$
$$d_{ij} = \begin{cases} 1 & (j = i+1) \\ 0 & (\text{otherwise}) \end{cases} \tag{9-4-2}$$

Example 9-4-1

Let us determine the state transition matrix for a $(d, \kappa) = (1, 3)$ code. The $(1, 3)$ code has four states. From Fig. 9-4-4, we obtain its state transition matrix, which is

$$\mathbf{D} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \tag{9-4-3}$$

An important parameter of any (d, κ) code is the number of sequences of a certain length, say n, that satisfy the (d, κ) constraints. As n is allowed to increase, the number of sequences $N(n)$ that satisfy the (d, κ) constraint also increases. The number of information bits that can be uniquely represented with $N(n)$ code sequences is

$$k = \lfloor \log_2 N(n) \rfloor$$

where $\lfloor x \rfloor$ denotes the largest integer contained in x. The maximum code rate is then $R_c = k/n$.

The capacity of a (d, κ) code is defined as

$$C(d, \kappa) = \lim_{n \to \infty} \frac{1}{n} \log_2 N(n) \tag{9-4-4}$$

Clearly, $C(d, \kappa)$ is the maximum possible rate that can be achieved with the (d, κ) constraints. Shannon (1948) showed that the capacity is given as

$$C(d, \kappa) = \log_2 \lambda_{\max} \tag{9-4-5}$$

where $\lambda_{\max}$ is the largest real eigenvalue of the state transition matrix **D**.

Example 9-4-2

Let us determine the capacity of a $(d, \kappa) = (1, 3)$ code. Using the state-transition matrix given in Example 9-4-1 for the $(1, 3)$ code, we have

$$\det(\mathbf{D} - \lambda \mathbf{I}) = \det \begin{bmatrix} -\lambda & 1 & 0 & 0 \\ 1 & -\lambda & 1 & 0 \\ 1 & 0 & -\lambda & 1 \\ 1 & 0 & 0 & -\lambda \end{bmatrix}$$
$$= \lambda^4 - \lambda^2 - \lambda - 1 = 0 \tag{9-4-6}$$

TABLE 9-4-1 CAPACITY $C(d, \kappa)$ VERSUS RUNLENGTH PARAMETERS d AND κ

κ	$d=0$	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$	$d=6$
2	.8791	.4057					
3	.9468	.5515	.2878				
4	.9752	.6174	.4057	.2232			
5	.9881	.6509	.4650	.3218	.1823		
6	.9942	.6690	.4979	.3746	.2269	.1542	
7	.9971	.6793	.5174	.4057	.3142	2281	.1335
8	.9986	.6853	.5293	.4251	.3432	.2709	.1993
9	.9993	.6888	.5369	.4376	.3620	.2979	.2382
10	.9996	.6909	.5418	.4460	.3746	.3158	.2633
11	.9998	.6922	.5450	.4516	.3833	.3285	.2804
12	.9999	.6930	.5471	.4555	.3894	.3369	.2924
13	.9999	.6935	.5485	.4583	.3937	.3432	.3011
14	.9999	.6938	.5495	.4602	.3968	.3478	.3074
15	.9999	.6939	.5501	.4615	.3991	.3513	.3122
∞	1.000	.6942	.5515	.4650	.4057	.3620	.3282

The maximum real root of this polynomial is found to be $\lambda_{max} = 1.4656$. Therefore, the capacity $C(1, 3) = \log_2 \lambda_{max} = 0.5515$.

The capacities of (d, κ) codes for $0 \leq d \leq 6$ and $2 \leq \kappa \leq 15$ are given in Table 9-4-1. We observe that $C(d, \kappa) < \frac{1}{2}$ for $d \geq 3$ and any value of κ. The most commonly used codes for magnetic recording employ $d \leq 2$; hence, their rate R_c is at least $\frac{1}{2}$.

Now let us turn our attention to the construction of some runlength-limited codes. In general, (d, κ) codes can be constructed either as fixed-length codes or as variable-length codes. In a fixed-length code, each bit or block of k bits is encoded into a block of $n > k$ bits.

In principle, the construction of a fixed-length code is straightforward. For a given block length n, we may select the subset of the 2^n code words that satisfy the specified runlength constraints. From this subset, we eliminate code words that do not satisfy the runlength constraints when concatenated. Thus, we obtain a set of code words that satisfy the constraints and can be used in the mapping of the input data bits to the encoder. The encoding and decoding operations can be performed by use of a look-up table.

Example 9-4-3

Let us construct a $d = 0$, $\kappa = 2$ code of length $n = 3$, and determine its efficiency. By listing all the code words, we find that the following five code words satisfy the (0, 2) constraint: (0 1 0), (0 1 1), (1 0 1), (1 1 0), (1 1 1). We may select any four of these code words and use then to encode the pairs of

data bits (00, 01, 10, 11). Thus, we have a rate $k/n = 2/3$ code that satisfies the (0, 2) constraint.

The fixed-length code in this example is not very efficient. The capacity is $C(0, 2) = 0.8791$, so that this code has an *efficiency* of

$$\text{efficiency} = \frac{R_c}{C(d, \kappa)} = \frac{2/3}{0.8791} = 0.76$$

Surely, better (0, 2) codes can be constructed by increasing the block length n.

In the following example, we place no restriction on the maximum runlength of zeros.

Example 9-4-4

Let us construct a $d = 1$, $\kappa = \infty$ code of length $n = 5$. In this case, we are placing no constraint on the number of consecutive zeros. To construct the code, we select from the set of 32 possible code words those that satisfy the $d = 1$ constraint. There are eight such code words, which implies that we can encode three information bits with each code word. The code is given in Table 9-4-2. Note that the first bit of each code word is a 0, whereas the last bit may be either 0 or 1. Consequently, the $d = 1$ constraint is satisfied when these code words are concatenated. This code has a rate $R_c = 3/5$. When compared with the capacity $C(1, \infty) = 0.6942$ obtained from Table 9-4-1, the code efficiency is 0.864, which is quite acceptable.

The code construction method described in the two examples above produces fixed-length (d, κ) codes that are *state-independent*. By state-independent, we mean that fixed-length code words can be concatenated without violating the (d, κ) constraints. In general, fixed-length state-independent (d, κ) codes require large block lengths, except in cases such as those in the examples above where d is small. Simpler (shorter-length) codes

TABLE 9-4-2 FIXED LENGTH $d = 1$, $\kappa = \infty$ CODE

Input data bits	Output coded sequence
0 0 0	0 0 0 0 0
0 0 1	0 0 0 0 1
0 1 0	0 0 0 1 0
0 1 1	0 0 1 0 0
1 0 0	0 0 1 0 1
1 0 1	0 1 0 0 0
1 1 0	0 1 0 0 1
1 1 1	0 1 0 1 0

are generally possible by allowing for state-dependence and for variable length code words. Below, we consider codes for which both the input blocks to the encoder and the output blocks may have variable length. For the code words to be uniquely decodable at the receiver, the variable-length code should satisfy the prefix condition, described in Chapter 3.

Example 9-4-5

A very simple uniquely decodable variable-length $d = 0$, $\kappa = 2$ code is

$$0 \rightarrow 01$$
$$10 \rightarrow 10$$
$$11 \rightarrow 11$$

The code in the above example has a fixed output block size but a variable input block size. In general, both the input and output blocks may be variable. The following example illustrates the latter case.

Example 9-4-6

Let us construct a (2, 7) variable block size code. The solution to this code construction is certainly not unique, nor is it trivial. We picked this example because the (2, 7) code has been widely used by IBM in many of its disk storage systems. The code is listed in Table 9-4-3. We observe that the input data blocks of 2, 3, and 4 bits are mapped into output data blocks of 4, 6, and 8 bits, respectively. Hence, the code rate is $R_c = 1/2$. Since this is the code rate for all code words, the code is called a *fixed-rate* code. This code has an efficiency of $0.5/0.5174 = 0.966$. Note that this code satisfies the prefix condition.

TABLE 9-4-3 CODE BOOK FOR VARIABLE-LENGTH (2, 7) CODE

Input data bits	Output coded sequence
1 0	1 0 0 0
1 1	0 1 0 0
0 1 1	0 0 0 1 0 0
0 1 0	0 0 1 0 0 0
0 0 0	1 0 0 1 0 0
0 0 1 1	0 0 1 0 0 1 0 0
0 0 1 0	0 0 0 0 1 0 0 0

TABLE 9-4-4 ENCODER FOR (1,3) MILLER CODE

Input data bits	Output coded sequence
0	x 0
1	0 1

$x = 0$, if preceding input bit is 1
$x = 1$, if preceding input bit is 0

Another code that has been widely used in magnetic recording is the rate 1/2, $(d, \kappa) = (1, 3)$ code in Table 9-4-4. We observe that when the information bit is a 0, the first output bit is 1 if the previous input bit was 0, or a 0 if the previous input bit was a 1. When the information bit is a 1, the encoder output is 01. Decoding of this code is simple. The first bit of the two-bit block is redundant and may be discarded. The second bit is the information bit. This code is usually called the *Miller code*. We observe that this is a state-dependent code, which is described by the state diagram shown in Fig. 9-4-5. There are two states labeled S_1 and S_2 with transitions as shown in the figure. When the encoder is a state S_1, an input bit 1 results in the encoder staying in state S_1 and outputs 01. This is denoted as 1/01. If the input bit is a 0, the encoder enters state S_2 and outputs 00. This is denoted as 0/00. Similarly, if the encoder is in state S_2, an input bit 0 causes no transition and the encoder output is 10. On the other hand, if the input bit is a 1, the encoder enters state S_1 and outputs 01. Figure 9-4-6 shows the trellis for the Miller code.

The Mapping of Coded Bits into Signal Waveforms The output sequence from a (d, κ) encoder is mapped by the modulator into signal waveforms for transmission over the channel. If the binary digit 1 is mapped into a rectangular pulse of amplitude A and the binary digit 0 is mapped into a

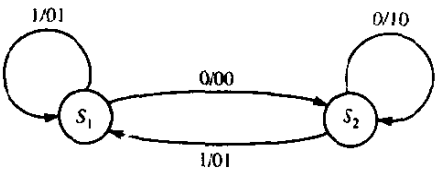


FIGURE 9-4-5 State diagrams for $d = 1$, $\kappa = 3$ (Miller) code.

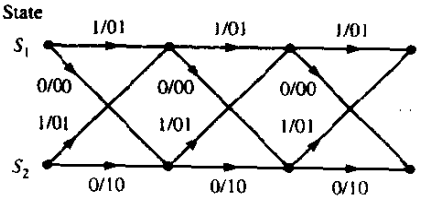


FIGURE 9-4-6 Trellis for $d = 1$, $\kappa = 3$ (Miller) code.

rectangular pulse of amplitude $-A$, the result is a (d, κ) coded NRZ modulated signal. Note that the duration of the rectangular pulses is $T_c = R_c/R_b = R_c T_b$, where R_b is the information (bit) rate into the encoder, T_b is the corresponding (uncoded) bit interval, and R_c is the code rate for the (d, κ) code.

When the (d, κ) code is a state-independent fixed-length code with code rate $R_c = k/n$, we may consider each n-bit block as generating one signal waveform of duration nT_c. Thus, we have $M = 2^k$ signal waveforms, one for each of the 2^k possible k-bit data blocks. These coded waveforms have the general form given by (4-3-6) and (4-3-38). In this case, there is no dependence between the transmission of successive waveforms.

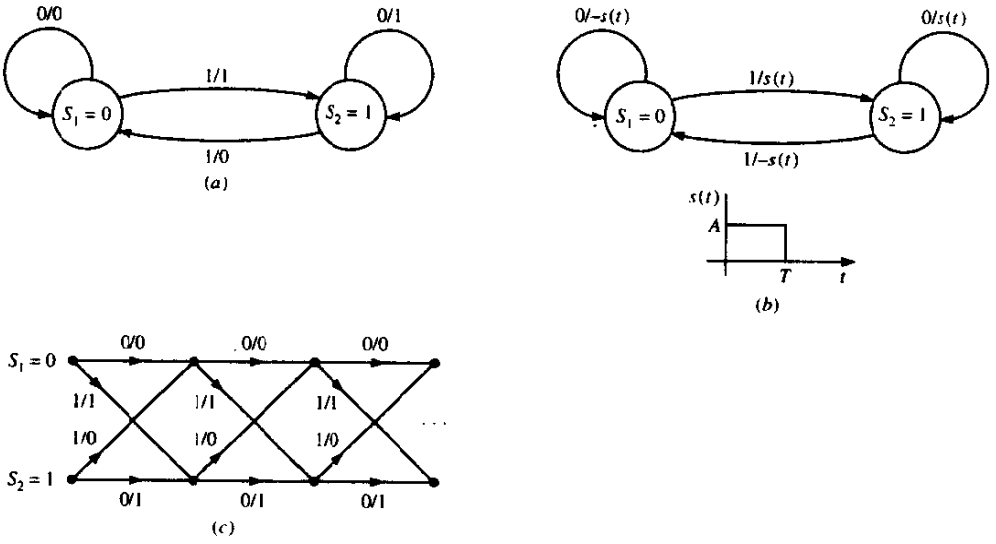
In contrast to the situation considered above, the modulation signal is no longer memoryless when NRZI is used and/or the (d, κ) code is state-dependent. Let us consider the effect of mapping the coded bits into an NRZI signal waveform.

Recall that the state dependence in the NRZI signal is due to the differential encoding of the information sequence. The differential encoding is a form of precoding, which is described mathematically as

$$p_k = d_k \oplus p_{k-1}$$

where $\{d_k\}$ is the binary sequence into the precoder, $\{p_k\}$ is the output binary sequence from the precoder, and $\oplus$ denotes modulo-2 addition. This encoding is characterized by the state diagram shown in Fig. 9-4-7(a). Then, the sequence $\{p_k\}$ is transmitted by NRZ. Thus, when $p_k = 1$, the modulator output is a rectangular pulse of amplitude A, and when $p_k = 0$, the modulator

FIGURE 9-4-7 State and trellis diagrams for NRZI signal.



output is a rectangular pulse of amplitude $-A$. When the signal waveforms are superimposed on the state diagram of Fig. 9-4-7(a), we obtain the corresponding state diagram shown in Fig. 9-4-7(b). The corresponding trellis is shown in Fig. 9-4-7(c).

When the output of a state-dependent (d, κ) encoder is followed by an NRZI modulator, we may simply combine the two-state diagrams into a single-state diagram for the (d, κ) code with precoding. A similar combination can be performed with the corresponding trellises. The following example illustrates the approach for the (1,3) Miller code followed by NRZI modulation.

Example 9-4-7

Let us determine the state diagram of the combined (1,3) Miller code followed by the precoding inherent in NRZI modulation. Since the (1,3) Miller code has two states and the precoder has two states, the state diagram for the combined encoder has four states, which we denote as $(S_M, S_N) = (\sigma_1, s_1), (\sigma_1, s_2), (\sigma_2, s_1), (\sigma_2, s_2)$, where $S_M = \{\sigma_1, \sigma_2\}$ represents the two states of the Miller code and $S_N = \{s_1, s_2\}$ represents the two states of the precoder for NRZI. For each data input bit into the Miller encoder, we obtain two output bits which are then precoded to yield two precoded output bits. The resulting state diagram is shown in Fig. 9-4-8, where the first bit denotes the information bit into the Miller encoder and the next two bits represent the corresponding output of the precoder.

The trellis diagram for the Miller precoded sequence may be obtained directly from the combined state diagram or from a combination of the trellises of the two codes. The result of this combination is the four-state trellis, one stage of which is shown in Fig. 9-4-9.

It is left as an exercise for the reader to show that the four signal waveforms obtained by mapping each pair of bits of the Miller-precoded sequence into an

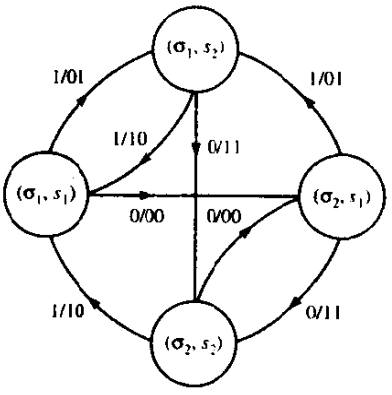


FIGURE 9-4-8 State diagram of the Miller code followed by the precoder.

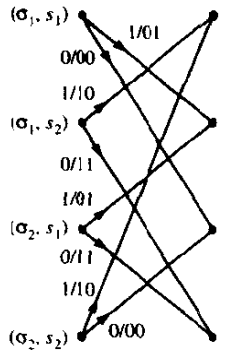


FIGURE 9-4-9 One stage of trellis diagram for the Miller code followed by the precoder.

NRZ signal are biorthogonal and that the resulting modulated signal waveform is identical to the delay modulation that was described in Section 4-3-2.

From the state diagram of a state-dependent runlength-limited code, one can obtain the transition probability matrix, as described in Section 4-3-2. Then, the power spectral density of the code may be determined, as shown in Section 4-4-3.

9-5 BIBLIOGRAPHICAL NOTES AND REFERENCES

The pioneering work on signal design for bandwidth-constrained channels was done by Nyquist (1928). The use of binary partial response signals was originally proposed by Lender (1963), and was later generalized by Kretzmer (1966). Other early work on problems dealing with intersymbol interference (ISI) and transmitter and receiver optimization with constraints on ISI was done by Gerst and Diamond (1961), Tufts (1965), Smith (1965), and Berger and Tufts (1967). "Faster than Nyquist" transmission has been studied by Mazo (1975) and Foschini (1984).

Modulation codes were also first introduced by Shannon (1948). Some of the early work on the construction of runlength-limited codes is found in the papers by Freiman and Wyner (1964), Gabor (1967), Franaszek (1968, 1969, 1970), Tang and Bahl (1970), and Jacoby (1977). More recent work is found in papers by Adler Coppersmith and Hassner (1983), and Karabed and Siegel (1991). The motivation for most of the work on runlength-limited codes was provided by applications to magnetic and optical recording. A well-written tutorial paper on runlength-limited codes has been published by Immink (1990).

PROBLEMS

9-1 A channel is said to be *distortionless* if the response $y(t)$ to an input $x(t)$ is $Kx(t - t_0)$, where K and t_0 are constants. Show that if the frequency response of the channel is $A(f)e^{j\theta(f)}$, where $A(f)$ and $\theta(f)$ are real, the necessary and

sufficient conditions for distortionless transmission are $A(f) = K$ and $\theta(f) = 2\pi f t_0 \pm n\pi$, $n = 0, 1, 2, \ldots$.

9-2 The raised-cosine spectral characteristic is given by (9-2-26).

a Show that the corresponding impulse response is

$$x(t) = \frac{\sin(\pi t/T)}{\pi t/T} \frac{\cos(\beta \pi t/T)}{1 - 4\beta^2 t^2/T^2}$$

b Determine the Hilbert transform of $x(t)$ when $\beta = 1$.

c Does $\hat{x}(t)$ possess the desirable properties of $x(t)$ that make it appropriate for data transmission? Explain.

d Determine the envelope of the SSB suppressed-carrier signal generated from $x(t)$.

9-3 **a** Show that (Poisson sum formula)

$$x(t) = \sum_{k=-\infty}^{\infty} g(t)h(t - kT) \Rightarrow X(f) = \frac{1}{T} \sum_{n=-\infty}^{\infty} H\left(\frac{n}{T}\right) G\left(f - \frac{n}{T}\right)$$

Hint: Make a Fourier-series expansion of the periodic factor

$$\sum_{k=-\infty}^{\infty} h(t - kT)$$

b Using the result in (a), verify the following versions of the Poisson sum:

$$\sum_{k=-\infty}^{\infty} h(kT) = \frac{1}{T} \sum_{n=-\infty}^{\infty} H\left(\frac{n}{T}\right) \tag{i}$$

$$\sum_{k=-\infty}^{\infty} h(t - kT) = \frac{1}{T} \sum_{n=-\infty}^{\infty} H\left(\frac{n}{T}\right) \exp\left(\frac{j2\pi nt}{T}\right) \tag{ii}$$

$$\sum_{k=-\infty}^{\infty} h(kT) \exp(-j2\pi kTf) = \frac{1}{T} \sum_{n=-\infty}^{\infty} H\left(f - \frac{n}{T}\right) \tag{iii}$$

c Derive the condition for no intersymbol interference (Nyquist criterion) by using the Poisson sum formula.

9-4 Suppose a digital communications system employs gaussian-shaped pulses of the form

$$x(t) = \exp(-\pi a^2 t^2)$$

To reduce the level of intersymbol interference to a relatively small amount, we impose the condition that $x(T) = 0.01$, where T is the symbol interval. The bandwidth W of the pulse $x(t)$ is defined as that value of W for which $X(W)/X(0) = 0.01$, where $X(f)$ is the Fourier transform of $x(t)$. Determine the value of W and compare this value to that of raised-cosine spectrum with 100% rolloff.

9-5 A band-limited signal having bandwidth W can be represented as

$$x(t) = \sum_{n=-\infty}^{\infty} x_n \frac{\sin[2\pi W(t - n/2W)]}{2\pi W(t - n/2W)}$$

a Determine the spectrum $X(f)$ and plot $|X(f)|$ for the following cases:

$$x_0 = 2, \quad x_1 = 1, \quad x_2 = -1, \quad x_n = 0, \quad n \neq 0, 1, 2 \tag{i}$$

$$x_{-1} = -1, \quad x_0 = 2, \quad x_1 = -1, \quad x_n = 0, \quad n \neq -1, 0, 1 \tag{ii}$$

b Plot $x(t)$ for these two cases.

c If these signals are used for binary signal transmission, determine the number of received levels possible at the sampling instants $t = nT = n/2W$, and the probabilities of occurrence of the received levels. Assume that the binary digits at the transmitter are equally probable.

9-6 A 4 kHz bandpass channel is to be used for transmission of data at a rate of 9600 bits/s. If $\frac{1}{2}N_0 = 10^{-10}$ W/Hz is the spectral density of the additive, zero-mean gaussian noise in the channel, design a QAM modulation and determine the average power that achieves a bit error probability of 10^{-6}. Use a signal pulse with a raised-cosine spectrum having a roll-off factor of at least 50%.

9-7 Determine the bit rate that can be transmitted through a 4 kHz voice-band telephone (bandpass) channel if the following modulation methods are used: (a) binary PAM; (b) four-phase PSK; (c) 8-point QAM; (d) binary orthogonal FSK, with noncoherent detection; (e) orthogonal four-FSK with noncoherent detection; (f) orthogonal 8-FSK with noncoherent detection. For (a)–(c), assume that the transmitter pulse shape has a raised-cosine spectrum with a 50% roll-off.

9-8 An ideal voice-band telephone line channel has a bandpass frequency response characteristic spanning the frequency range 600–3000 Hz.

a Design an $M = 4$ PSK (quadrature PSK or QPSK) system for transmitting data at a rate of 2400 bits/s and a carrier frequency $f_c = 1800$ Hz. For spectral shaping, use a raised-cosine frequency-response characteristic. Sketch a block diagram of the system and describe the functional operation of each block.

b Repeat (a) for a bit rate $R = 4800$ bits/s.

9-9 A voice-band telephone channel passes the frequencies in the band from 300 to 3300 Hz. It is desired to design a modem that transmits at a symbol rate of 2400 symbols/s, with the objective of achieving 9600 bits/s. Select an appropriate QAM signal constellation, carrier frequency, and the roll-off factor of a pulse with a raised cosine spectrum that utilizes the entire frequency band. Sketch the spectrum of the transmitted signal pulse and indicate the important frequencies.

9-10 A communication system for a voice-band (3 kHz) channel is designed for a received SNR at the detector of 30 dB when the transmitter power is $P_S = -3$ dBW. Determine the value of P_S if it is desired to expand the bandwidth of the system to 10 kHz, while maintaining the same SNR at the detector.

9-11 Show that a pulse having the raised cosine spectrum given by (9-2-26) satisfies the Nyquist criterion given by (9-2-13) for any value of the roll-off factor β.

9-12 Show that, for any value of β, the raised cosine spectrum given by (9-2-26) satisfies

$$\int_{-\infty}^{\infty} X_{rc}(f)\, df = 1$$

[*Hint:* Use the fact that $X_{rc}(f)$ satisfies the Nyquist criterion given by (9-2-13).]

9-13 The Nyquist criterion gives the necessary and sufficient condition for the spectrum $X(f)$ of the pulse $x(t)$ that yields zero ISI. Prove that for any pulse that is band-limited to $|f| < 1/T$, the zero-ISI condition is satisfied if Re $[X(f)]$, for $f > 0$, consists of a rectangular function plus an arbitrary odd function around $f = 1/2T$, and Im $[X(f)]$ is any arbitrary even function around $f = 1/2T$.

9-14 A voice-band telephone channel has a passband characteristic in the frequency range 300 Hz $< f <$ 3000 Hz.

a Select a symbol rate and a power efficient constellation size to achieve 9600 bits/s signal transmission.

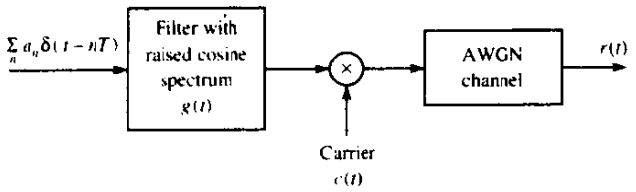


FIGURE P9-16

b If a square-root raised cosine pulse is used for the transmitter pulse $g(t)$, select the roll-off factor. Assume that the channel has an ideal frequency response characteristic.

9-15 Design an M-ary PAM system that transmits digital information over an ideal channel with bandwidth $W = 2400$ Hz. The bit rate is 14 400 bit/s. Specify the number of transmitted points, the number of received signal points using a duobinary signal pulse, and the required $\mathscr{E}_b$ to achieve an error probability of 10^{-6}. The additive noise is zero-mean gaussian with a power spectral density 10^{-4} W/Hz.

9-16 A binary PAM signal is generated by exciting a raised cosine roll-off filter with a 50% roll-off factor and is then DSB-SC amplitude-modulated on a sinusoidal carrier as illustrated in Fig. P9-16. The bit rate is 2400 bit/s.

a Determine the spectrum of the modulated binary PAM signal and sketch it.

b Draw the block diagram illustrating the optimum demodulator/detector for the received signal, which is equal to the transmitted signal plus additive white gaussian noise.

9-17 The elements of the sequence $\{a_n\}_{n=-\infty}^{\infty}$ are independent binary random variables taking values of ± 1 with equal probability. This data sequence is used to modulate the basic pulse $g(t)$ shown in Fig. P9-17(a). The modulated signal is

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

a Find the power spectral density of $X(t)$.

b If $g_1(t)$ (shown in Fig. 9-17b) is used instead of $g(t)$, how would the power spectrum in (a) change?

c In (b) assume we want to have a null in the spectrum at $f = 1/3T$. This is done by a precoding of the form $b_n = a_n + \alpha a_{n-3}$. Find the α that provides the desired null.

d Is it possible to employ a precoding of the form $b_n = a_n + \sum_{i=1}^{N} \alpha_i a_{n-i}$ for some finite N such that the final power spectrum will be identical to zero for $1/3T \leq |f| \leq 1/2T$? If yes, how? If no, why? [*Hint:* Use properties of analytic functions.]

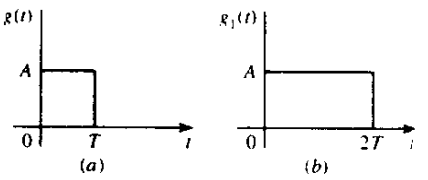


FIGURE P9-17

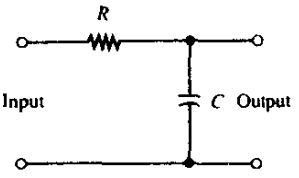


FIGURE P9-22

9-18 Consider the transmission of data via PAM over a voice-band telephone channel that has a bandwidth of 3000 Hz. Show how the symbol rate varies as a function of the excess bandwidth. In particular, determine the symbol rate for an excess bandwidth of 25%, 33%, 50%, 67%, 75%, and 100%.

9-19 The binary sequence 10010110010 is the input to a precoder whose output is used to modulate a duobinary transmitting filter. Construct a table as in Table 9-2-1 showing the precoded sequence, the transmitted amplitude levels, the received signal levels and the decoded sequence.

9-20 Repeat Problem 9-19 for a modified duobinary signal pulse.

9-21 A precoder for a partial response signal fails to work if the desired partial response at $n = 0$ is zero modulo M. For example, consider the desired response for $M = 2$:

$$x(nT) = \begin{cases} 2 & (n = 0) \\ 1 & (n = 1) \\ -1 & (n = 2) \\ 0 & \text{(otherwise)} \end{cases}$$

Show why this response cannot be precoded.

9-22 Consider the RC lowpass filter shown in Fig. P9-22, where $\tau = RC = 10^{-6}$.

a Determine and sketch the envelope (group) delay of the filter as a function of frequency.

b Suppose that the input to the filter is a lowpass signal of bandwidth $\Delta f = 1$ kHz. Determine the effect of the RC filter on this signal.

9-23 A microwave radio channel has a frequency response

$$C(f) = 1 + 0.3 \cos 2\pi fT$$

Determine the frequency response characteristic of the optimum transmitting and receiving filters that yield zero ISI at a rate of $1/T$ symbols/s and have a 50% excess bandwidth. Assume that the additive noise spectrum is flat.

9-24 $M = 4$ PAM modulation is used for transmitting at a bit rate of 9600 bit/s on a channel having a frequency response

$$C(f) = \frac{1}{1 + j(f/2400)}$$

for $|f| \leq 2400$, and $C(f) = 0$ otherwise. The additive noise is zero-mean, white Gaussian with power spectral density $\frac{1}{2}N_0$ W/Hz. Determine the (magnitude) frequency response characteristic of the optimum transmitting and receiving filters.

9-25 Determine the capacity of a $(0, 1)$ runlength-limited code. Compare its capacity with that of a $(1, \infty)$ code and explain the relationship.

9-26 A ternary signal format is designed for a channel that does not pass d.c. The

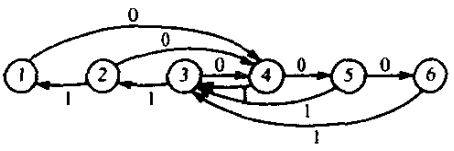


FIGURE P9-31

binary input information sequence is transmitted by mapping a 1 into either a positive pulse or a negative pulse, and a zero is transmitted by the absence of a pulse. Hence, for the transmission of 1s, the polarity of the pulses alternate. This is called an AMI (alternate mark inversion) code. Determine the capacity of the code.

9-27 Give an alternative description of the AMI code described in Problem 9-26 using the running digit sum (RDS) with the constraint that the RDS can take only the values 0 and +1.

9-28 (kBnT codes) From Problem 9-26, note that the AMI code is a "pseudo-ternary" code in that it transmits one bit per symbol using a ternary alphabet, which has the capacity of $\log_2 3 = 1.58$ bits. Such a code does not provide sufficient spectral shaping. Better spectral shaping is achieved by the class of block codes designated as kBnT, where k denotes the number of information bits and n denotes the number of ternary symbols per block. By selecting the largest k possible for each n, we obtain the following table:

k	n	Code
1	1	1B1T
3	2	3B2T
4	3	4B3T
6	4	6B4T

Determine the efficiency of these codes by computing the ratio of the code in bits/symbol divided by $\log_2 3$. Note that 1B1T is the AMI code.

9-29 This problem deals with the capacity of two (d, κ) codes.

a Determine the capacity of a (d, κ) code that has the following state transition matrix:

$$\mathbf{D} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

b Repeat (a) for

$$\mathbf{D} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

c Comment on the differences between (a) and (b).

9-30 A simplified model of the telegraph code consists of two symbols (Blahut, 1990). A dot consists of one time unit of line closure followed by one time unit of line

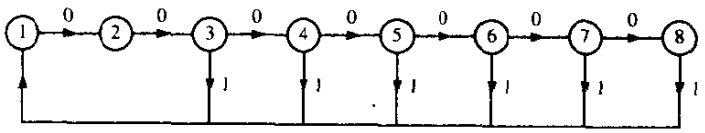


FIGURE P9-32

open. A dash consists of three units of line closure followed by one time unit of line open.

a Viewing this code as a constrained code with symbols of equal duration, give the constraints.

b Determine the state-transition matrix.

c Determine the capacity.

9-31 Determine the state-transition matrix for the runlength-constrained code described by the state diagram shown in Fig. P9-31. Sketch the corresponding trellis.

9-32 Determine the state-transition matrix for the (2, 7) runlength-limited code specified by the state diagram shown in Fig. P9-32.