18BEBME406 ANALOG AND DIGITAL COMMNICATION 3 0 0 3 100

OBJECTIVES:

The student should be made to:

- Understand analog and digital communication techniques.
- Learn data and pulse communication techniques.
- Be familiarized with source and Error control coding.
- Gain knowledge on multi-user radio communication.

UNIT I ANALOG COMMUNICATION

Introduction to Communication Systems - Modulation – Types - Need for Modulation. Theory of Amplitude Modulation - Evolution and Description of SSB Techniques - Theory of Frequency and Phase Modulation – Comparison of Analog Communication Systems (AM – FM – PM).

UNIT II PULSE AND DATA COMMUNICATION

Pulse Communication: Pulse Amplitude Modulation (PAM) – Pulse Time Modulation (PTM) –Pulse code Modulation (PCM) - Comparison of various Pulse Communication System (PAM –PTM – PCM). Data Communication: History of Data Communication - Standards Organizations for Data Communication- Data Communication Circuits - Data Communication Codes - Data communication Hardware - serial and parallel interfaces.

UNIT III DIGITAL COMMUNICATION

Amplitude Shift Keying (ASK) – Frequency Shift Keying (FSK)–Phase Shift Keying (PSK) – BPSK– QPSK – Quadrature Amplitude Modulation (QAM) – 8 QAM – 16 QAM – Bandwidth Efficiency–Comparison of various Digital Communication System (ASK – FSK – PSK – QAM).

UNIT IV SOURCE AND ERROR CONTROL CODING

Entropy, Source encoding theorem, Shannon fano coding, Huffman coding, mutual information, channel capacity, Error Control Coding, linear block codes, cyclic codes - ARQ Techniques.

UNIT V MULTI-USER RADIO COMMUNICATION

Global System for Mobile Communications (GSM) - Code division multiple access (CDMA) – Cellular Concept and Frequency Reuse - Channel Assignment and Handover Techniques - Overview of Multiple Access Schemes - Satellite Communication - Bluetooth.

TEXT BOOK:

1. Wayne Tomasi, ―Advanced Electronic Communication Systems‖, 6th Edition, Pearson Education, 2009

References ;

Simon Haykin, ―Communication Systems‖, 4th Edition, John Wiley & Sons, 2004
2. Rappaport T.S, "Wireless Communications: Principles and Practice", 2nd Edition, Pearson
Education, 2007

3. H.Taub, D L Schilling and G Saha, ―Principles of Communication‖, 3rd Edition, Pearson
Education, 2007.
4. B. P.Lathi, ―Modern Analog and Digital Communication Systems‖, 3rd Edition, Oxford
University Press, 2007.
5. Blake, ―Electronic Communication Systems‖, Thomson Delmar Publications, 2002.
6. Martin S.Roden, ―Analog and Digital Communication System‖, 3rd Edition, Prentice Hall of
India, 2002.

KARPAGAM ACADEMY OF HIGHER EDUCATION

(Deemed to be University Established Under Section 3 of UGC Act 1956)

FACULTY OF ENGINEERING

DEPARTMENT OF BIOMEDICAL ENGINEERING

LECTURE PLAN

STAFF NAME : G.ARAVINDH **COURSE : BE - BME**
SUBJECT CODE : 18BEEC406 **SEMESTER:IV**
SUBJECT NAME : ANALOG AND DIGITAL COMMUNICATION

S.No	Description of Portion to be Covered	No. of Hours	Referred Books / Websites
UNIT I ANALOG COMMUNICATION			
1	Introduction to Communication Systems	1	T1 Page No 62 to 72
2	Modulation – Types - Need for Modulation.	2	T1 Page No 185 to 186
3	Theory of Amplitude Modulation	2	T1 Page No 193 to 195
4	Evolution and Description of SSB Techniques	2	T1 Page No 193 to 195
5	Theory of Frequency and Phase Modulation	1	T1 Page No 201 to 204
6	Comparison of Analog Communication Systems (AM – FM – PM).	1	T1 Page No 211 to 212
UNIT II PULSE AND DATA COMMUNICATION			
7	Pulse Communication: Pulse Amplitude Modulation (PAM)	1	R2 Page No 242 to 244
8	Pulse Time Modulation (PTM)	1	R2 Page No 251 to 253
9	Pulse code Modulation (PCM)	2	R2 Page No 265 to 268
10	Comparison of various Pulse Communication System (PAM –PTM – PCM)	2	R2 Page No 269 to 271
11	History of Data Communication - Standards Organizations for Data Communication	1	R2 Page No 281 to 283
12	Data Communication Circuits - Data Communication Codes	1	R2 Page No 291 to 294
13	Data communication Hardware - serial and parallel interfaces.	1	R2 Page No 296 to 299
UNIT III DIGITAL COMMUNICATION			
14	Amplitude Shift Keying (ASK)	1	T1 Page No 201 to 209
15	Frequency Shift Keying (FSK)	2	T1 Page No 209 to 210
16	Phase Shift Keying (PSK) – BPSK– QPSK	2	T1 Page No 217 to 218
17	Quadrature Amplitude Modulation (QAM) – 8 QAM – 16 QAM	2	T1 Page No 227 to 229
18	Bandwidth Efficiency–Comparison of various Digital Communication System (ASK – FSK – PSK – QAM)	2	T1 Page No 218 to 223
UNIT IV SOURCE AND ERROR CONTROL CODING			
19	Entropy-Source encoding theorem	1	R4 Page No 119 to 122
20	Shannon fano coding	2	T1 Page No 248 to 253

21	Huffman coding	2	T1 Page No 253 to 258
22	Mutual information, channel capacity	2	T1 Page No 259 to 261
23	Error Control Coding, linear block codes	1	T1 Page No 261 to 267
24	Cyclic codes - ARQ Techniques	1	T1 Page No 271 to 279
UNIT V MULTI-USER RADIO COMMUNICATION			
25	Global System for Mobile Communications (GSM)	2	R4 Page No 220 to 229
26	Code division multiple access (CDMA)	1	R4 Page No 229 to 236
27	Cellular Concept and Frequency Reuse	2	T1 Page No 490 to 493
28	Channel Assignment and Handover Techniques	2	R4 Page No 119 to 122
29	Overview of Multiple Access Schemes - Satellite Communication - Bluetooth.	2	T1 Page No 462 to 467
		Total	45 Hours

TEXT BOOKS

S.No.	Author(s) Name	Title of the Book	Publisher	Year of Publication
1	Wayne Tomasi	Advanced Electronic Communication Systems	Pearson Education,6th Edition	2009

REFERENCE BOOKS

S.No.	Author(s) Name	Title of the Book	Publisher	Year of Publication
1	Simon Haykins	Communication Systems	John Wiley PHI, New Delhi	2004
2	John G Proakis	Digital Communication	Mc Graw Hill, Inc, 5th Edition	2008
3	Taub & Schiling	Principle of Digital Communication	Tata Mc Graw-Hill, New Delhi	2003
4	Bernard Sklar	Digital Communication - Fundamentals and Applications	Prentice Hall P T R, Upper Saddle River, New Jersey 07458	2001

UNIT I SAMPLING AND QUANTIZATION

Sampling Process – Aliasing – Instantaneous sampling – Natural Sampling – Flat Sampling – Quantization of signals – sampling and quantizing effects – channel effects – SNR for quantization pulses – data formatting techniques – Time division multiplexing.

SAMPLING PROCESS

A message signal may originate from a digital or analog source. If the message signal is analog in nature, then it has to be converted into digital form before it can transmitted by digital means. The process by which the continuous-time signal is converted into a discrete– time signal is called Sampling. Sampling operation is performed in accordance with the sampling theorem.

The Sampling Theorem

The link between an analog waveform and its sampled version is provided by what is known as the *sampling process*. This process can be implemented in several ways, the most popular being the *sample-and-hold* operation. In this operation, a switch and storage mechanism (such as a transistor and a capacitor, or a shutter and a filmstrip) form a sequence of samples of the continuous input waveform. The output of the sampling process is called *pulse amplitude modulation* (PAM) because the successive output intervals can be described as a sequence of pulses with amplitudes derived from the input waveform samples. The analog waveform can be approximately retrieved from a PAM waveform by simple low-pass filtering. An important question: how closely can a filtered PAM waveform approximate the original input waveform? This question can be answered by reviewing the *sampling theorem*, which states the following [1]: A bandlimited signal having no spectral components above f_m hertz can be determined uniquely by values sampled at uniform intervals of

$$T_s \leq \frac{1}{2f_m} \text{ sec} \tag{2.1}$$

This particular statement is also known as the *uniform sampling theorem*. Stated another way, the upper limit on T_s can be expressed in terms of the sampling rate, denoted $f_s = 1/T_s$. The restriction, stated in terms of the sampling rate, is known as the *Nyquist criterion*. The statement is

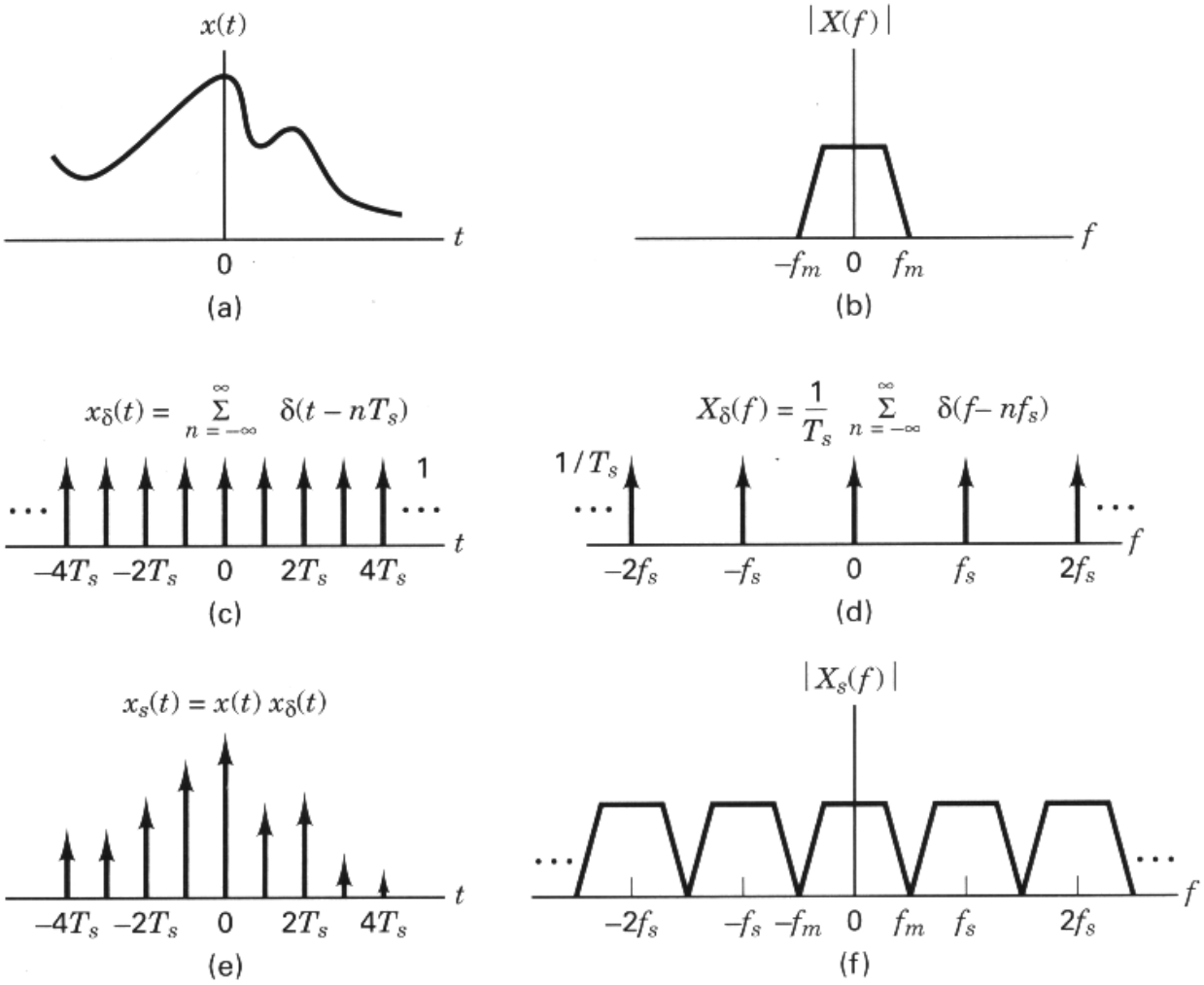
$$f_s \geq 2f_m \tag{2.2}$$

The sampling rate $f_s = 2f_m$ is also called the *Nyquist rate*. The Nyquist criterion is a theoretically sufficient condition to allow an analog signal to be *reconstructed completely* from a set of uniformly spaced discrete-time samples. In the sections that follow, the validity of the sampling theorem is demonstrated using different sampling approaches.

Impulse Sampling

Here we demonstrate the validity of the sampling theorem using the frequency convolution property of the Fourier transform. Let us first examine the case of *ideal sampling* with a sequence of unit impulse functions. Assume an analog waveform, $x(t)$, as shown in Figure 2.6a, with a Fourier transform, $X(f)$, which is zero outside the interval $(-f_m < f < f_m)$, as shown in Figure 2.6b. The sampling of $x(t)$ can be viewed as the product of $x(t)$ with a periodic train of unit impulse functions $x_\delta(t)$, shown in Figure 2.6c and defined as

$$x_\delta(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \tag{2.3}$$



Sampling theorem using the frequency convolution property of the Fourier transform.

where T_s is the sampling period and $\delta(t)$ is the unit impulse or Dirac delta function defined in Section 1.2.5. Let us choose $T_s = 1/2f_m$, so that the Nyquist criterion is just satisfied.

The *sifting property* of the impulse function (see Section A.4.1) states that

$$x(t)\delta(t - t_0) = x(t_0)\delta(t - t_0) \tag{2.4}$$

Using this property, we can see that $x_s(t)$, the sampled version of $x(t)$ shown in Figure 2.6e, is given by

$$\begin{aligned} x_s(t) = x(t)x_\delta(t) &= \sum_{n=-\infty}^{\infty} x(t)\delta(t - nT_s) \\ &= \sum_{n=-\infty}^{\infty} x(nT_s)\delta(t - nT_s) \end{aligned} \tag{2.5}$$

Using the *frequency convolution property* of the Fourier transform (see Section A.5.3), we can transform the time-domain product $x(t)x_\delta(t)$ of Equation (2.5) to the frequency-domain convolution $X(f) * X_\delta(f)$, where

$$X_\delta(f) = \frac{1}{T_s} \sum_{n=-\infty}^{\infty} \delta(f - nf_s) \tag{2.6}$$

is the Fourier transform of the impulse train $x_\delta(t)$ and where $f_s = 1/T_s$ is the sampling frequency. Notice that the Fourier transform of an impulse train is another impulse train; the values of the periods of the two trains are reciprocally related to one another. Figures 2.6c and d illustrate the impulse train $x_\delta(t)$ and its Fourier transform $X_\delta(f)$, respectively.

Convolution with an impulse function simply shifts the original function as follows:

$$X(f) * \delta(f - nf_s) = X(f - nf_s) \tag{2.7}$$

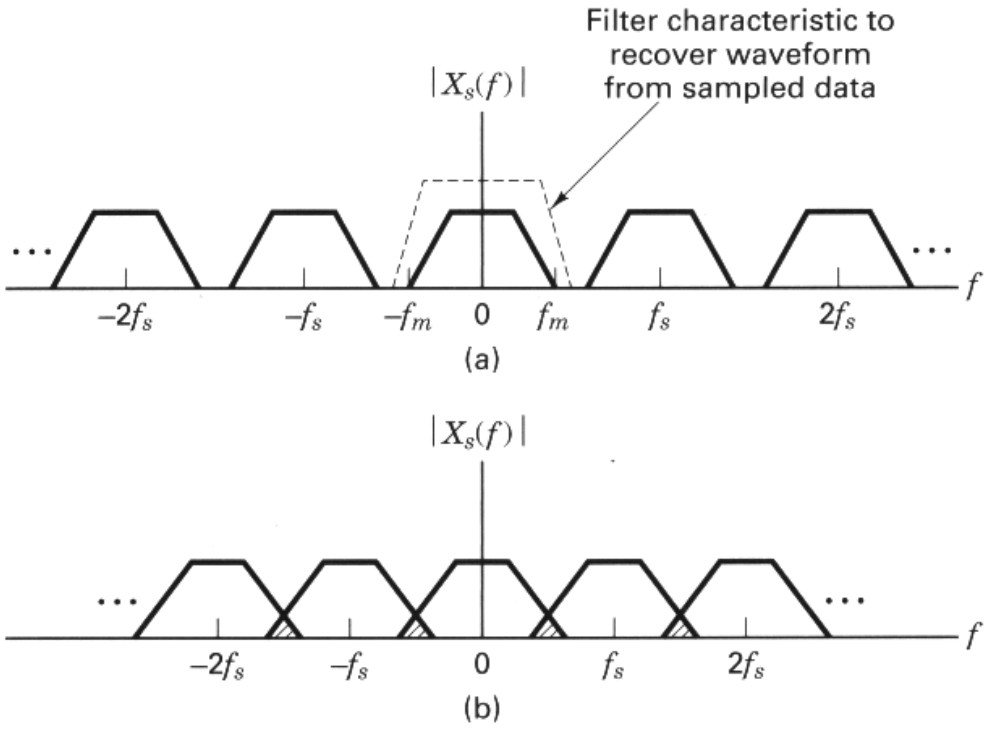
We can now solve for the transform $X_s(f)$ of the sampled waveform:

$$\begin{aligned} X_s(f) = X(f) * X_\delta(f) &= X(f) * \left[\frac{1}{T_s} \sum_{n=-\infty}^{\infty} \delta(f - nf_s) \right] \\ &= \frac{1}{T_s} \sum_{n=-\infty}^{\infty} X(f - nf_s) \end{aligned} \tag{2.8}$$

We therefore conclude that within the original bandwidth, the spectrum $X_s(f)$ of the sampled signal $x_s(t)$ is, to within a constant factor $(1/T_s)$, exactly the same as that of $x(t)$. In addition, the spectrum repeats itself periodically in frequency every f_s hertz. The sifting property of an impulse function makes the convolving of an impulse train with another function easy to visualize. The impulses act as sampling functions. Hence, convolution can be performed graphically by sweeping the impulse train $X_\delta(f)$ in Figure 2.6d past the transform $|X(f)|$ in Figure 2.6b. This sampling of $|X(f)|$ at each step in the sweep replicates $|X(f)|$ at each of the frequency positions of the impulse train, resulting in $|X_s(f)|$, shown in Figure 2.6f.

When the sampling rate is chosen, as it has been here, such that $f_s = 2f_m$, each spectral replicate is separated from each of its neighbors by a frequency band exactly equal to f_s hertz, and the analog waveform can theoretically be completely recovered from the samples, by the use of filtering. However, a filter with infinitely steep sides would be required. It should be clear that if $f_s > 2f_m$, the replications will move farther apart in frequency, as shown in Figure 2.7a, making it easier to perform the filtering operation. A typical low-pass filter characteristic that might be used to separate the baseband spectrum from those at higher frequencies is shown in the figure. When the sampling rate is reduced, such that $f_s < 2f_m$, the replications will overlap, as shown in Figure 2.7b, and some information will be lost. The phenomenon, the result of undersampling (sampling at too low a rate), is called *aliasing*. The Nyquist rate, $f_s = 2f_m$, is the sampling rate below which aliasing occurs; to avoid aliasing, the Nyquist criterion, $f_s \geq 2f_m$, must be satisfied.

As a matter of practical consideration, neither waveforms of engineering interest nor realizable bandlimiting filters are strictly bandlimited. Perfectly bandlimited signals do not occur in nature (see Section 1.7.2); thus, realizable signals, even though we may think of them as bandlimited, always contain some aliasing. These signals and filters can, however, be considered to be "essentially" bandlimited. By



Spectra for various sampling rates. (a) Sampled spectrum ($f_s > 2f_m$). (b) Sampled spectrum ($f_s < 2f_m$).

this we mean that a bandwidth can be determined beyond which the spectral components are attenuated to a level that is considered negligible.

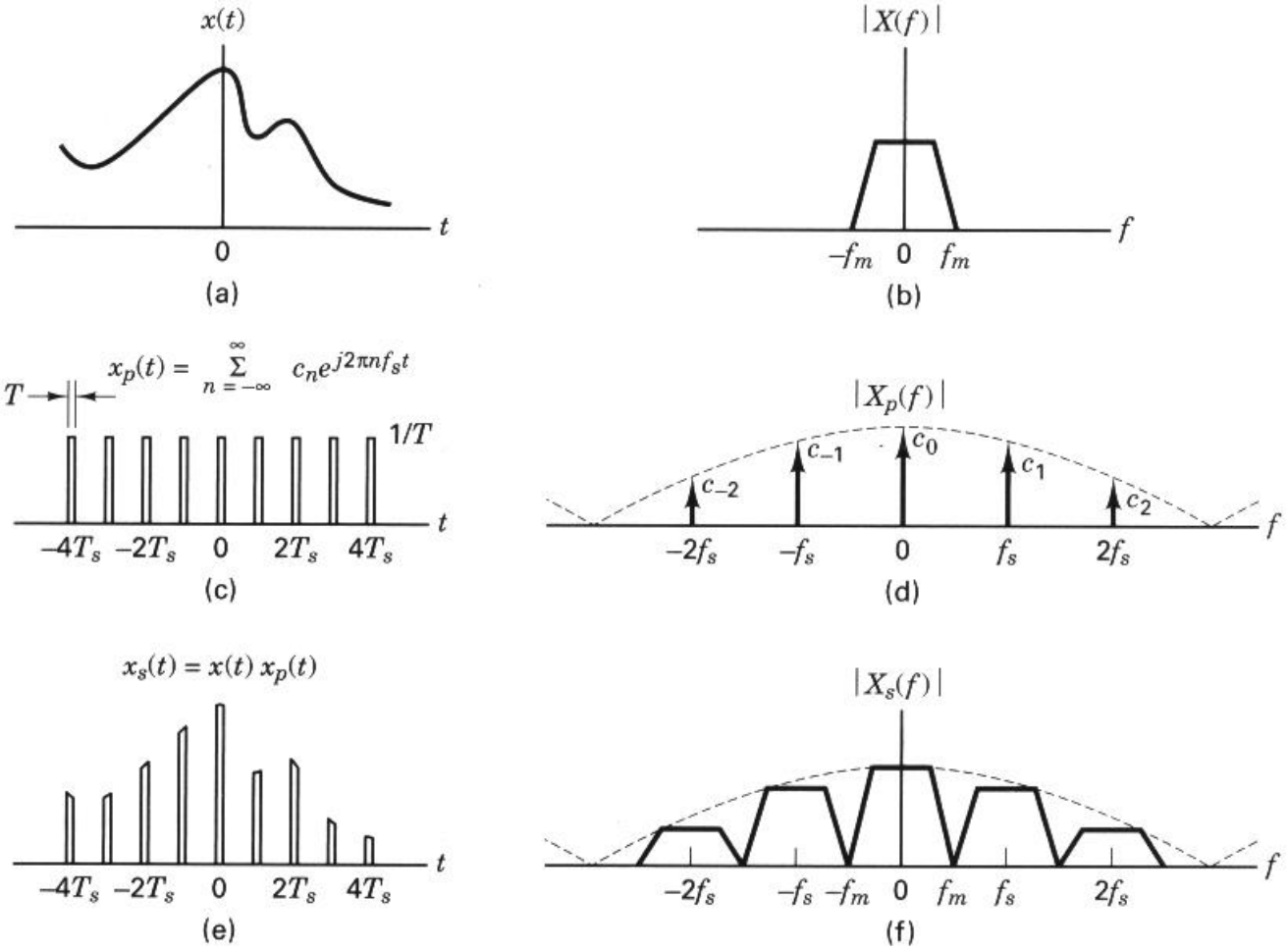
Natural Sampling

Here we demonstrate the validity of the sampling theorem using the frequency shifting property of the Fourier transform. Although instantaneous sampling is a convenient model, a more practical way of accomplishing the sampling of a bandlimited analog signal $x(t)$ is to multiply $x(t)$, shown in Figure 2.8a, by the pulse train or switching waveform $x_p(t)$, shown in Figure 2.8c. Each pulse in $x_p(t)$ has width T and amplitude $1/T$. Multiplication by $x_p(t)$ can be viewed as the opening and closing of a switch. As before, the sampling frequency is designated f_s, and its reciprocal, the time period between samples, is designated T_s. The resulting sampled-data sequence, $x_s(t)$, is illustrated in Figure 2.8e and is expressed as

$$x_s(t) = x(t)x_p(t) \tag{2.9}$$

The sampling here is termed *natural sampling,* since the top of each pulse in the $x_s(t)$ sequence retains the shape of its corresponding analog segment during the pulse interval. Using Equation (A.13), we can express the periodic pulse train as a Fourier series in the form

$$x_p(t) = \sum_{n=-\infty}^{\infty} c_n e^{j\,2\pi n f_s t} \tag{2.10}$$



Sampling theorem using the frequency shifting property of the Fourier transform.

where the sampling rate, $f_s = 1/T_s$, is chosen equal to $2f_m$, so that the Nyquist criterion is just satisfied. From Equation (A.24), $c_n = (1/T_s)\ \text{sinc}\ (nT/T_s)$, where T is the pulse width, $1/T$ is the pulse amplitude, and

$$\text{sinc}\ y = \frac{\sin \pi y}{\pi y}$$

The envelope of the magnitude spectrum of the pulse train, seen as a dashed line in Figure 2.8d, has the characteristic sinc shape. Combining Equations (2.9) and (2.10) yields

$$x_s(t) = x(t) \sum_{n=-\infty}^{\infty} c_n e^{j\,2\pi n f_s t} \tag{2.11}$$

The transform $X_s(f)$ of the sampled waveform is found as follows:

$$X_s(f) = \mathcal{F}\left\{x(t) \sum_{n=-\infty}^{\infty} c_n e^{j\,2\pi n f_s t}\right\} \tag{2.12}$$

For linear systems, we can interchange the operations of summation and Fourier transformation. Therefore, we can write

$$X_s(f) = \sum_{n=-\infty}^{\infty} c_n \mathcal{F}\{x(t)e^{j2\pi n f_s t}\} \tag{2.13}$$

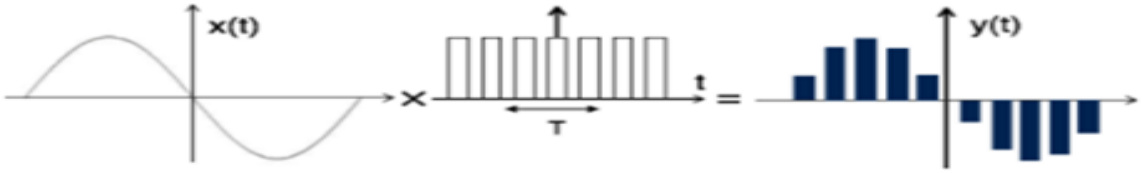
Using the *frequency translation* property of the Fourier transform (see Section A.3.2), we solve for $X_s(f)$ as follows:

$$X_s(f) = \sum_{n=-\infty}^{\infty} c_n X(f - nf_s) \tag{2.14}$$

Similar to the unit impulse sampling case, Equation (2.14) and Figure 2.8f illustrate that $X_s(f)$ is a replication of $X(f)$, periodically repeated in frequency every f_s hertz. In this natural-sampled case, however, we see that $X_s(f)$ is weighted by the Fourier series coefficients of the pulse train, compared with a constant value in the impulse-sampled case. It is satisfying to note that *in the limit,* as the pulse width, T, approaches zero, c_n approaches $1/T_s$ for all n (see the example that follows), and

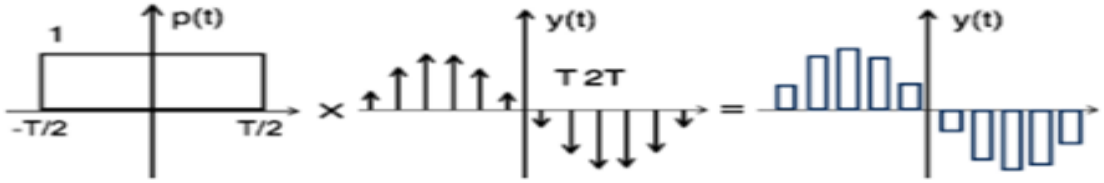
Flat Top Sampling

During transmission, noise is introduced at top of the transmission pulse which can be easily removed if the pulse is in the form of flat top. Here, the top of the samples are flat i.e. they have constant amplitude. Hence, it is called as flat top sampling or practical sampling. Flat top sampling makes use of sample and hold circuit.



Theoretically, the sampled signal can be obtained by convolution of rectangular pulse p(t) with ideally sampled signal say $y_s(t)$ as shown in the diagram:

i.e. $y(t) = p(t) \times y_s(t) \ldots\ldots(1)$



To get the sampled spectrum, consider Fourier transform on both sides for equation 1

$Y[\omega] = F.T[P(t) \times y_s(t)]$

By the knowledge of convolution property,

$Y[\omega] = P(\omega)\, Y_s(\omega)$

Here $P(\omega) = TSa(\frac{\omega T}{2}) = 2\sin\omega T/\omega$

Sample-and-Hold Operation

The simplest and thus most popular sampling method, *sample and hold,* can be described by the convolution of the sampled pulse train, $[x(t)x_\delta(t)]$, shown in Figure 2.6e, with a unity amplitude rectangular pulse $p(t)$ of pulse width T_s. This time, convolution results in the *flattop* sampled sequence

$$\begin{aligned} x_s(t) &= p(t) * [x(t)x_\delta(t)] \\ &= p(t) * \left[x(t) \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \right] \end{aligned} \tag{2.15}$$

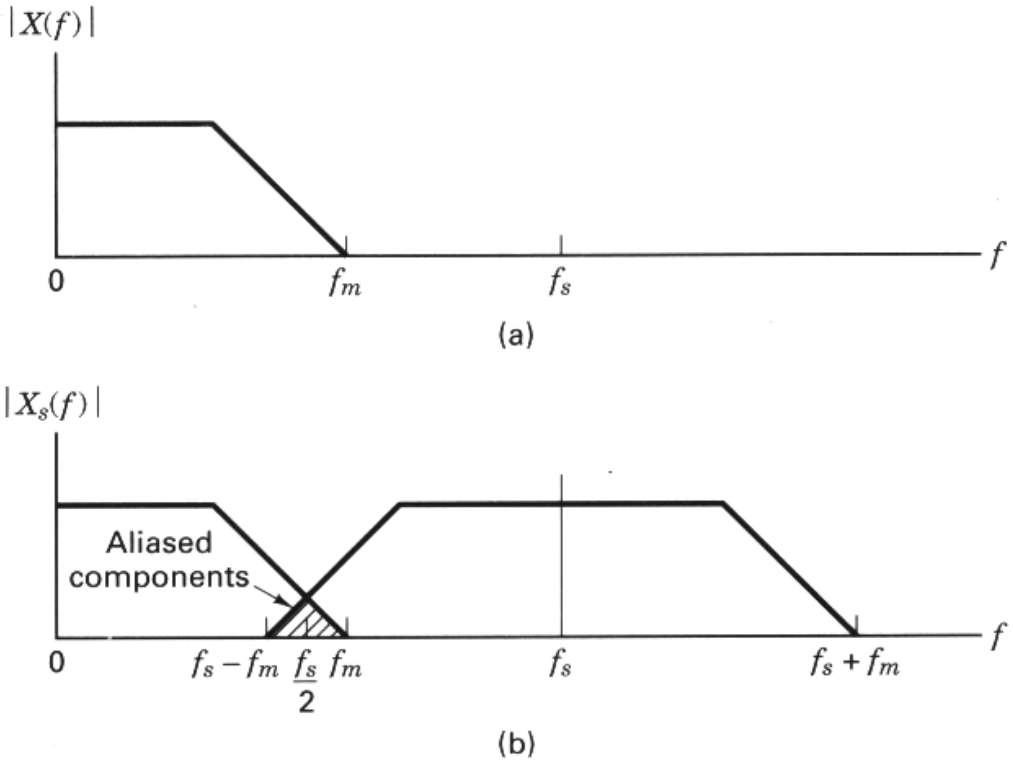
The Fourier transform, $X_s(f)$, of the time convolution in Equation (2.15) is the frequency-domain product of the transform $P(f)$ of the rectangular pulse and the periodic spectrum, shown in Figure 2.6f, of the impulse-sampled data:

$$\begin{aligned} X_s(f) &= P(f)\mathcal{F}\left\{ x(t) \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \right\} \\ &= P(f)\left\{ X(f) * \left[\frac{1}{T_s} \sum_{n=-\infty}^{\infty} \delta(f - nf_s) \right] \right\} \\ &= P(f) \frac{1}{T_s} \sum_{n=-\infty}^{\infty} X(f - nf_s) \end{aligned} \tag{2.16}$$

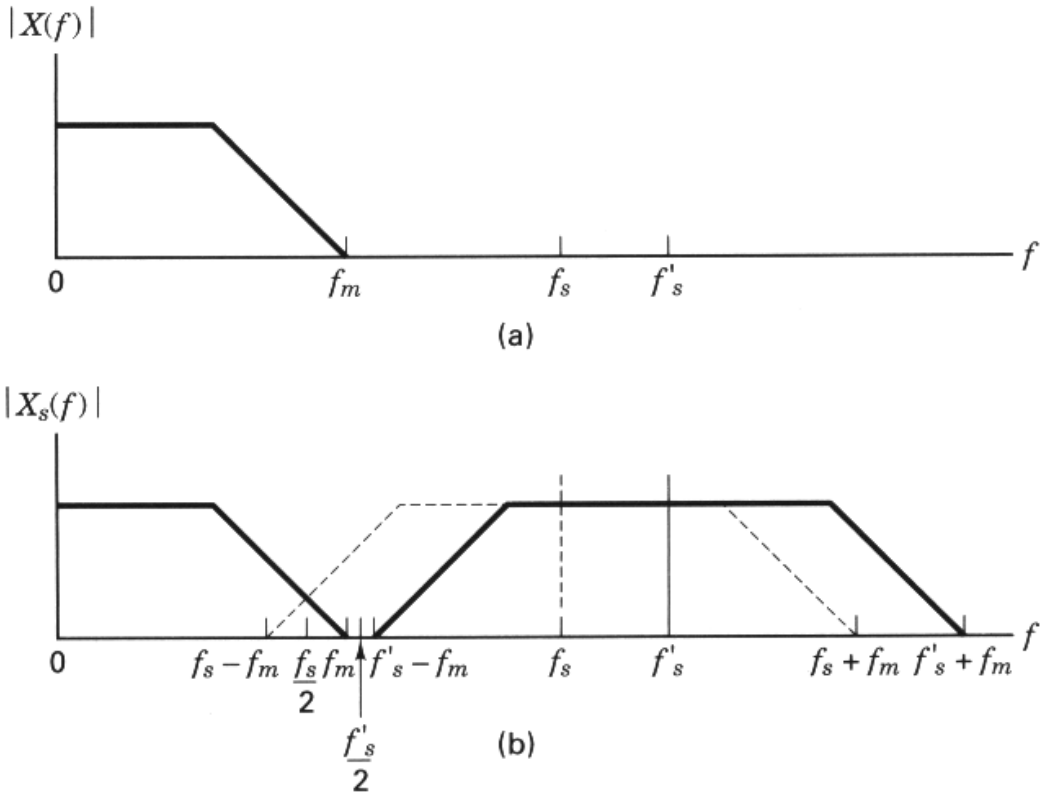
Here, $P(f)$ is of the form T_s sinc fT_s. The effect of this product operation results in a spectrum similar in appearance to the natural-sampled example presented in Figure 2.8f. The most obvious effect of the hold operation is the significant attenuation of the higher-frequency spectral replicates (compare Figure 2.8f to Figure 2.6f), which is a desired effect. Additional analog postfiltering is usually required to finish the filtering process by further attenuating the residual spectral components located at the multiples of the sample rate. A secondary effect of the hold operation is the nonuniform spectral gain $P(f)$ applied to the desired baseband spectrum shown in Equation (2.16). The postfiltering operation can compensate for this attenuation by incorporating the inverse of $P(f)$ over the signal passband.

Aliasing

A detailed view of the positive half of the baseband spectrum and one of the replicates from Figure 2.7b. It illustrates aliasing in the frequency domain. The overlapped region, shown in Figure 2.9b, contains that part of the spectrum which is aliased due to *undersampling.* The aliased spectral components represent ambiguous data that appear in the frequency band between $(f_s - f_m)$ and f_m. Figure 2.10 illustrates that a higher sampling rate f'_s, can eliminate the aliasing by separat-



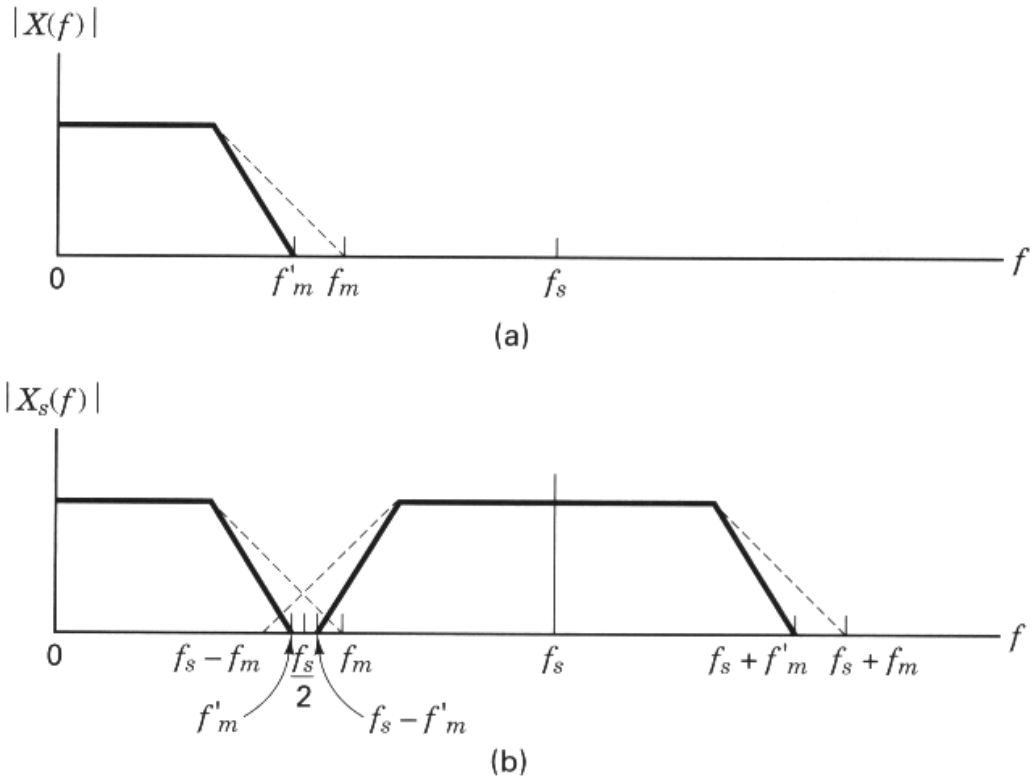
Aliasing in the frequency domain. (a) Continuous signal spectrum. (b) Sampled signal spectrum.



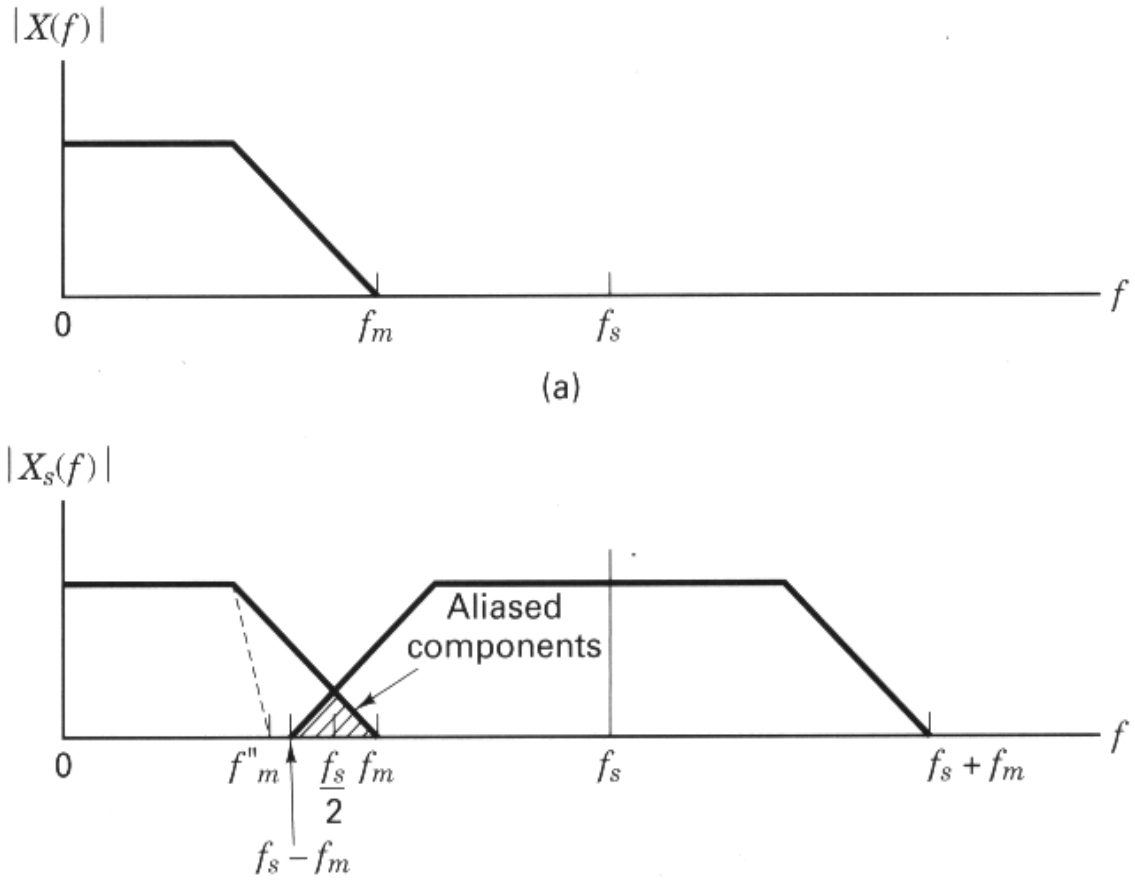
Higher sampling rate eliminates aliasing. (a) Continuous signal spectrum. (b) Sampled signal spectrum.

ing the spectral replicates; the resulting spectrum in Figure 2.10b corresponds to the case in Figure 2.7a. Figures 2.11 and 2.12 illustrate two ways of eliminating aliasing using *antialiasing filters.* In Figure 2.11 the analog signal is *prefiltered* so that the new maximum frequency, f'_m, is reduced to $f_s/2$ or less. Thus there are no aliased components seen in Figure 2.11b, since $f_s > 2f'_m$. Eliminating the aliasing terms prior to sampling is good engineering practice. When the signal structure is well known, the aliased terms can be eliminated after sampling, with a low-pass filter operating on the sampled data [2]. In Figure 2.12 the aliased components are removed by *postfiltering* after sampling; the filter cutoff frequency, f''_m, removes the aliased components; f''_m needs to be less than $(f_s - f_m)$. Notice that the filtering techniques for eliminating the aliased portion of the spectrum in Figures 2.11 and 2.12 *will result in a loss* of some of the signal information. For this reason, the sample rate, cutoff bandwidth, and filter type selected for a particular signal bandwidth are all interrelated.

Realizable filters require a nonzero bandwidth for the transition between the passband and the required out-of-band attenuation. This is called the *transition bandwidth.* To minimize the system sample rate, we desire that the antialiasing filter have a small transition bandwidth. Filter complexity and cost rise sharply with narrower transition bandwidth, so a trade-off is required between the cost of a small transition bandwidth and the costs of the higher sampling rate, which are those of more storage and higher transmission rates. In many systems the answer has been to make the transition bandwidth between 10 and 20% of the signal band-



Sharper-cutoff filters eliminate aliasing. (a) Continuous signal spectrum. (b) Sampled signal spectrum.



Postfilter eliminates aliased portion of spectrum. (a) Continuous signal spectrum. (b) Sampled signal spectrum.

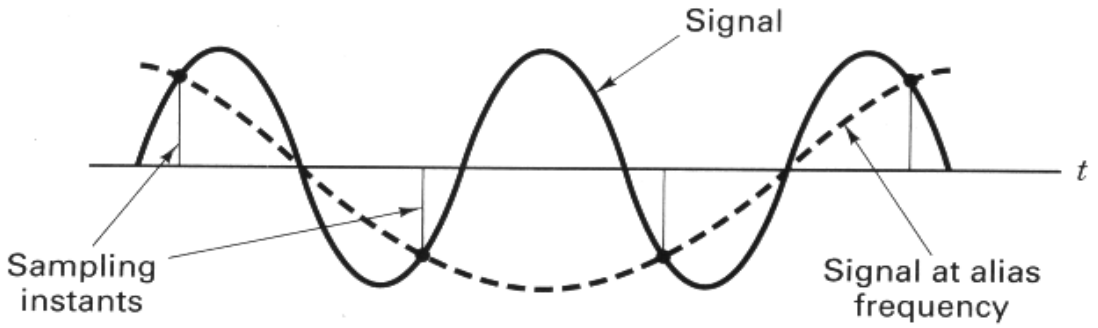
width. If we account for the 20% transition bandwidth of the antialiasing filter, we have an *engineer's version* of the Nyquist sampling rate:

$$f_s \geq 2.2f_m \tag{2.17}$$

Figure 2.13 provides some insight into aliasing as seen in the time domain. The sampling instants of the solid-line sinusoid have been chosen so that the sinusoidal signal is undersampled. Notice that the resulting ambiguity allows one to draw a totally different (dashed-line) sinusoid, following the undersampled points.

Why Oversample?

Oversampling is the most economic solution for the task of transforming an analog signal to a digital signal, or the reverse, transforming a digital signal to an analog signal. This is so because signal processing performed with high performance ana-



Alias frequency generated by sub-Nyquist sampling rate.

log equipment is typically much more costly than using digital signal processing equipment to perform the same task. Consider the task of transforming analog signals to digital signals. When this task is performed without the benefit of over-sampling, the process is characterized by three simple steps, performed in the order that follows.

Without Oversampling

1. The signal passes through a high performance analog lowpass filter to limit its bandwidth.
2. The filtered signal is sampled at the Nyquist rate for the (approximated) bandlimited signal. As described in Section 1.7.2, a strictly bandlimited signal is not realizable.
3. The samples are processed by an analog-to-digital converter that maps the continuous-valued samples to a finite list of discrete output levels.

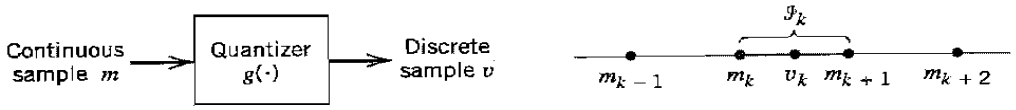
When this task is performed with the benefit of over-sampling, the process is best described as five simple steps, performed in the order that follows.

With Oversampling

1. The signal is passed through a low performance (less costly) analog low-pass filter (prefilter) to limit its bandwidth.
2. The pre-filtered signal is sampled at the (now higher) Nyquist rate for the (approximated) bandlimited signal.
3. The samples are processed by an analog-to-digital converter that maps the continuous-valued samples to a finite list of discrete output levels.
4. The digital samples are then processed by a high performance digital filter to reduce the bandwidth of the digital samples.
5. The sample rate at the output of the digital filter is reduced in proportion to the bandwidth reduction obtained by this digital filter.

Quantization of Signals

A continuous signal, such as voice, has a continuous range of amplitudes and therefore its samples have a continuous amplitude range. In other words, within the finite amplitude



Description of a memoryless quantizer.

range of the signal, we find an infinite number of amplitude levels. It is not necessary in fact to transmit the exact amplitudes of the samples. Any human sense (the ear or the eye), as ultimate receiver, can detect only finite intensity differences. This means that the original continuous signal may be *approximated* by a signal constructed of discrete amplitudes selected on a minimum error basis from an available set. The existence of a finite number of discrete amplitude levels is a basic condition of pulse-code modulation. Clearly, if we assign the discrete amplitude levels with sufficiently close spacing, we may make the approximated signal practically indistinguishable from the original continuous signal.

Amplitude *quantization* is defined as *the process of transforming the sample amplitude* $m(nT_s)$ *of a message signal* $m(t)$ *at time* $t = nT_s$ *into a discrete amplitude* $v(nT_s)$ *taken from a finite set of possible amplitudes*. We assume that the quantization process is *memoryless* and *instantaneous*, which means that the transformation at time $t = nT_s$ is not affected by earlier or later samples of the message signal. This simple form of scalar quantization, though not optimum, is commonly used in practice.

When dealing with a memoryless quantizer, we may simplify the notation by dropping the time index. We may thus use the symbol m in place of $m(nT_s)$, as indicated in the block diagram of a quantizer shown in Figure 3.9*a*. Then, as shown in Figure. 3.9*b*, the signal amplitude m is specified by the index k if it lies inside the *partition cell*

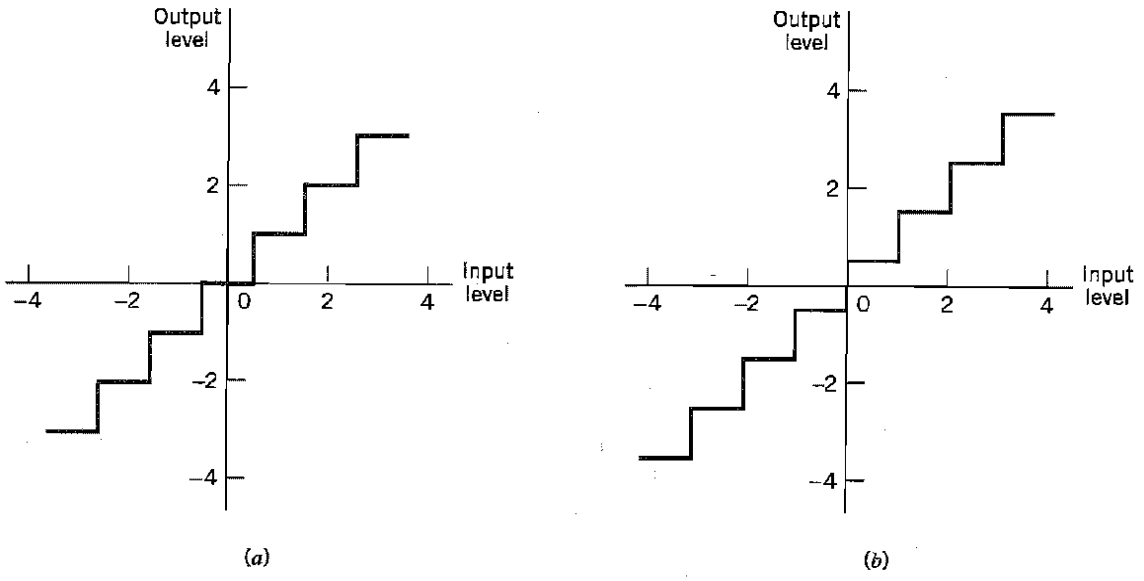
$$\mathcal{J}_k: \{m_k < m \leq m_{k+1}\}, \qquad k = 1, 2, \ldots, L \tag{3.21}$$

where L is the total number of amplitude levels used in the quantizer. The discrete amplitudes m_k, $k = 1, 2, \ldots, L$, at the quantizer input are called *decision levels* or *decision thresholds*. At the quantizer output, the index k is transformed into an amplitude v_k that represents all amplitudes of the cell $\mathcal{J}_k$; the discrete amplitudes v_k, $k = 1, 2, \ldots, L$, are called *representation levels* or *reconstruction levels*, and the spacing between two adjacent representation levels is called a *quantum* or *step-size*. Thus, the quantizer output v equals v_k if the input signal sample m belongs to the interval $\mathcal{J}_k$. The mapping (see Figure 3.9*a*)

$$v = g(m) \tag{3.22}$$

is the *quantizer characteristic*, which is a staircase function by definition.

Quantizers can be of a *uniform* or *nonuniform* type. In a uniform quantizer, the representation levels are uniformly spaced; otherwise, the quantizer is nonuniform. In this section, we consider only uniform quantizers; nonuniform quantizers are considered in Section 3.7. The quantizer characteristic can also be of *midtread* or *midrise type*. Figure 3.10*a* shows the input–output characteristic of a uniform quantizer of the midtread type, which is so called because the origin lies in the middle of a tread of the staircaselike graph. Figure 3.10*b* shows the corresponding input–output characteristic of a uniform quantizer of the midrise type, in which the origin lies in the middle of a rising part of the staircaselike graph.



Two types of quantization: (*a*) midtread and (*b*) midrise.

QUANTIZATION NOISE

The use of quantization introduces an error defined as the difference between the input signal m and the output signal v. The error is called *quantization noise*. Figure 3.11 illustrates a typical variation of the quantization noise as a function of time, assuming the use of a uniform quantizer of the midtread type.

Let the quantizer input m be the sample value of a zero-mean random variable M. (If the input has a nonzero mean, we can always remove it by subtracting the mean from the input and then adding it back after quantization.) A quantizer $g(\cdot)$ maps the input

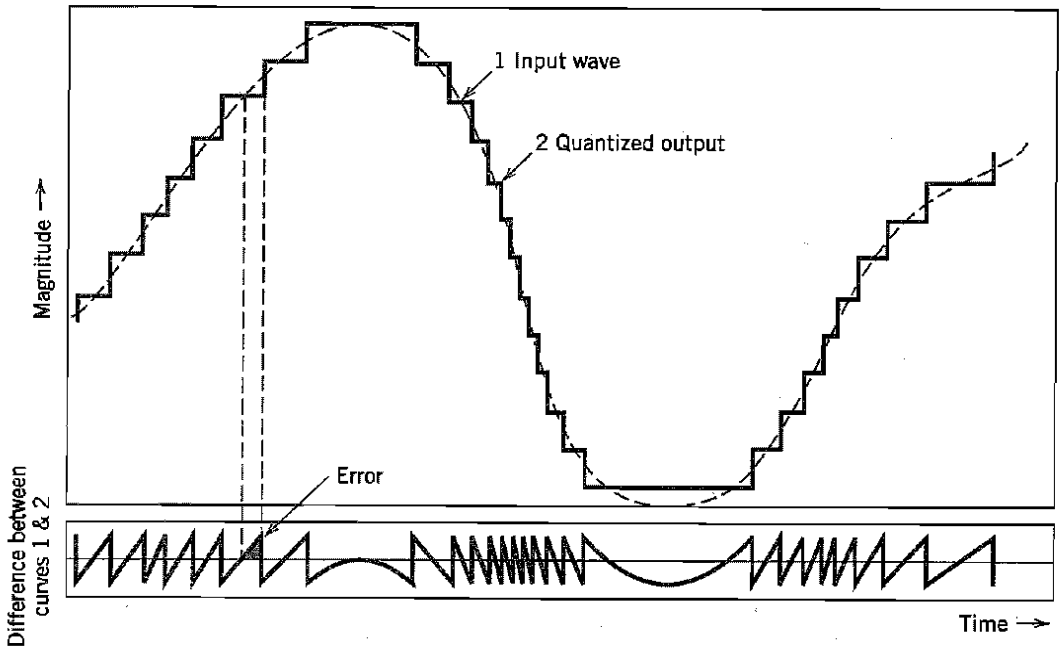


Illustration of the quantization process.

random variable M of continuous amplitude into a discrete random variable V; their respective sample values m and v are related by Equation (3.22). Let the quantization error be denoted by the random variable Q of sample value q. We may thus write

$$q = m - v \tag{3.23}$$

or, correspondingly,

$$Q = M - V \tag{3.24}$$

With the input M having zero mean, and the quantizer assumed to be symmetric as in Figure 3.10, it follows that the quantizer output V and therefore the quantization error Q, will also have zero mean. Thus for a partial statistical characterization of the quantizer in terms of output signal-to-(quantization) noise ratio, we need only find the mean-square value of the quantization error Q.

Consider then an input m of continuous amplitude in the range $(-m_{max}, m_{max})$. Assuming a uniform quantizer of the midrise type illustrated in Figure 3.10*b*, we find that the step-size of the quantizer is given by

$$\Delta = \frac{2m_{max}}{L} \tag{3.25}$$

where L is the total number of representation levels. For a uniform quantizer, the quantization error Q will have its sample values bounded by $-\Delta/2 \leq q \leq \Delta/2$. If the step-size Δ is sufficiently small (i.e., the number of representation levels L is sufficiently large), it is reasonable to assume that the quantization error Q is a *uniformly distributed* random variable, and the interfering effect of the quantization noise on the quantizer input is similar to that of thermal noise. We may thus express the probability density function of the quantization error Q as follows:

$$f_Q(q) = \begin{cases} \frac{1}{\Delta}, & -\frac{\Delta}{2} < q \leq \frac{\Delta}{2} \\ 0, & \text{otherwise} \end{cases} \tag{3.26}$$

For this to be true, however, we must ensure that the incoming signal does *not* overload the quantizer. Then, with the mean of the quantization error being zero, its variance σ_Q^2 is the same as the mean-square value:

$$\begin{aligned} \sigma_Q^2 &= E[Q^2] \\ &= \int_{-\Delta/2}^{\Delta/2} q^2 f_Q(q)\, dq \end{aligned} \tag{3.27}$$

Substituting Equation (3.26) into (3.27), we get

$$\begin{aligned} \sigma_Q^2 &= \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} q^2\, dq \\ &= \frac{\Delta^2}{12} \end{aligned} \tag{3.28}$$

Typically, the L-ary number k, denoting the kth representation level of the quantizer, is transmitted to the receiver in binary form. Let R denote the number of *bits per sample* used in the construction of the binary code. We may then write

$$L = 2^R \tag{3.29}$$

or, equivalently,

$$R = \log_2 L \tag{3.30}$$

Hence, substituting Equation (3.29) into (3.25), we get the step size

$$\Delta = \frac{2m_{\max}}{2^R} \tag{3.31}$$

Thus the use of Equation (3.31) in (3.28) yields

$$\sigma_Q^2 = \tfrac{1}{3}m_{\max}^2 2^{-2R} \tag{3.32}$$

Let P denote the average power of the message signal $m(t)$. We may then express the *output signal-to-noise ratio* of a uniform quantizer as

$$\begin{aligned}(\text{SNR})_O &= \frac{P}{\sigma_Q^2} \\ &= \left(\frac{3P}{m_{\max}^2}\right)2^{2R}\end{aligned} \tag{3.33}$$

Equation (3.33) shows that the output signal-to-noise ratio of the quantizer increases *exponentially* with increasing number of bits per sample, R. Recognizing that an increase in R requires a proportionate increase in the channel (transmission) bandwidth B_T, we thus see that the use of a binary code for the representation of a message signal (as in pulse-code modulation) provides a more efficient method than either frequency modulation (FM) or pulse-position modulation (PPM) for the trade-off of increased channel bandwidth for improved noise performance. In making this statement, we presume that the FM and PPM systems are limited by receiver noise, whereas the binary-coded modulation system is limited by quantization noise.

SOURCES OF CORRUPTION

The analog signal recovered from the sampled, quantized, and transmitted pulses will contain corruption from several sources. The sources of corruption are related to (1) sampling and quantizing effects, and (2) channel effects.

Sampling and Quantizing Effects

Quantization Noise

The distortion inherent in quantization is a round-off or truncation error. The process of encoding the PAM signal into a quantized PAM signal involves discarding some of the original analog information. This distortion, introduced by the need to approximate the analog waveform with quantized samples, is referred to as *quantization noise;* the amount of such noise is inversely proportional to the number of levels employed in the quantization process. (The signal-to-noise ratio of quantized pulses is treated in Sections 2.5.3 and 13.2.)

Quantizer Saturation

The quantizer (or analog-to-digital converter) allocates L levels to the task of approximating the continuous range of inputs with a finite set of outputs. The range of inputs for which the difference between the input and output is small is

called the *operating range* of the converter. If the input exceeds this range, the difference between the input and the output becomes large, and we say that the converter is operating in *saturation.* Saturation errors, being large, are more objectionable than quantizing noise. Generally, saturation is avoided by the use of automatic gain control (AGC), which effectively extends the operating range of the converter. (Chapter 13 covers quantizer saturation in greater detail.)

Timing Jitter

Our analysis of the sampling theorem predicted precise reconstruction of the signal based on uniformly spaced samples of the signal. If there is a slight jitter in the position of the sample, the sampling is no longer uniform. Although exact reconstruction is still possible if the sample positions are accurately known, the jitter is usually a random process and thus the sample positions are not accurately known. The effect of the jitter is equivalent to frequency modulation (FM) of the baseband signal. If the jitter is random, a low-level wideband spectral contribution is induced whose properties are very close to those of the quantizing noise. If the jitter exhibits periodic components, as might be found in data extracted from a tape recorder, the periodic FM will induce low-level spectral lines in the data. Timing jitter can be controlled with very good power supply isolation and stable clock references.

Channel Effects

Channel Noise

Thermal noise, interference from other users, and interference from circuit switching transients can cause errors in detecting the pulses carrying the digitized samples. Channel-induced errors can degrade the reconstructed signal quality quite quickly. This rapid degradation of output signal quality with channel-induced errors is called a *threshold effect.* If the channel noise is small, there will be no problem detecting the presence of the waveforms. Thus, small noise does not corrupt the reconstruct signals. In this case, the only noise present in the reconstruction is the quantization noise. On the other hand, if the channel noise is large enough to affect our ability to detect the waveforms, the resulting detection error causes reconstruction errors. A large difference in behavior can occur for very small changes in channel noise level.

Intersymbol Interference

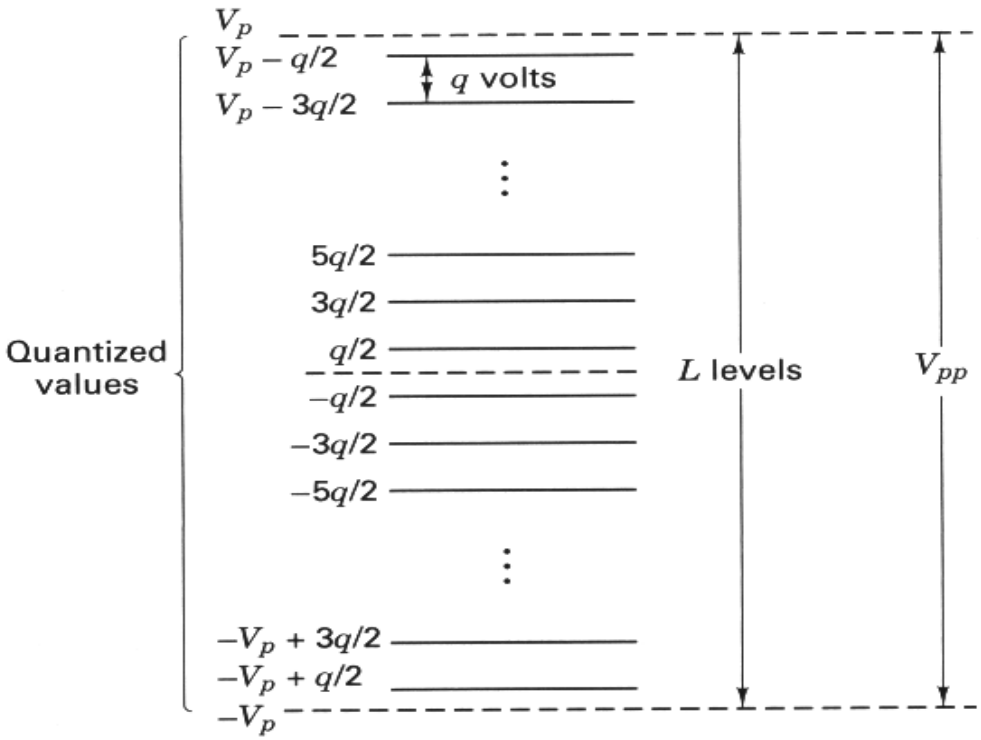
The channel is always bandlimited. A bandlimited channel disperses or spreads a pulse waveform passing through it (see Section 1.6.4). When the channel bandwidth is much greater than the pulse bandwidth, the spreading of the pulse will be slight. When the channel bandwidth is close to the signal bandwidth, the spreading will exceed a symbol duration and cause signal pulses to overlap. This overlapping is called *intersymbol interference* (ISI). Like any other source of interference, ISI causes system degradation (higher error rates); it is a particularly

insidious form of interference because raising the signal power to overcome the interference will not always improve the error performance. (Details of how ISI is handled are presented in the next chapter, in Sections 3.3 and 3.4.)

Signal-to-Noise Ratio for Quantized Pulses

Figure illustrates an L-level linear quantizer for an analog signal with a peak-to-peak voltage range of $V_{pp} = V_p - (-V_p) = 2V_p$ volts. The quantized pulses assume positive and negative values, as shown in the figure. The step size between quantization levels, called the *quantile interval,* is denoted q volts. When the quantization levels are uniformly distributed over the full range, the quantizer is called a *uniform or linear quantizer.* Each sample value of the analog waveform is approximated with a quantized pulse; the approximation will result in an error no larger than $q/2$ in the positive direction or $-q/2$ in the negative direction. The degradation of the signal due to quantization is therefore limited to half a quantile interval, $\pm q/2$ volts.

A useful figure of merit for the uniform quantizer is the quantizer variance (mean-square error assuming zero mean). If we assume that the quantization error, e, is uniformly distributed over a single quantile interval q-wide (i.e., the analog input takes on all values with equal probability), the quantizer error variance is found to be



Quantization levels.

$$\sigma^2 = \int_{-q/2}^{+q/2} e^2 p(e)\, de \tag{2.18a}$$

$$= \int_{-q/2}^{+q/2} e^2 \frac{1}{q}\, de = \frac{q^2}{12} \tag{2.18b}$$

where $p(e) = 1/q$ is the (uniform) probability density function of the quantization error. The variance, σ^2, corresponds to the *average quantization noise power.* The peak power of the analog signal (normalized to 1 Ω) can be expressed as

$$V_p^2 = \left(\frac{V_{pp}}{2}\right)^2 = \left(\frac{Lq}{2}\right)^2 = \frac{L^2q^2}{4} \tag{2.19}$$

where L is the number of quantization levels. Equations (2.18) and (2.19) combined yield the ratio of *peak* signal power to *average* quantization noise power $(S/N)_q$, assuming that there are no errors due to ISI or channel noise:

$$\left(\frac{S}{N}\right)_q = \frac{L^2q^2/4}{q^2/12} = 3L^2 \tag{2.20}$$

It is intuitively satisfying to see that $(S/N)_q$ improves as a function of the number of quantization levels squared. In the limit (as $L \to \infty$), the signal approaches the PAM format (with no quantization), and the signal-to-quantization noise ratio is infinite; in other words, with an infinite number of quantization levels, there is zero quantization noise.

Data Formatting Techniques

Encoding is the process of converting the data or a given sequence of characters, symbols, alphabets etc., into a specified format, for the secured transmission of data. **Decoding** is the reverse process of encoding which is to extract the information from the converted format.

Data Encoding

Encoding is the process of using various patterns of voltage or current levels to represent **1s** and **0s** of the digital signals on the transmission link. The common types of line encoding are Unipolar, Polar, Bipolar, and Manchester.

Encoding Techniques

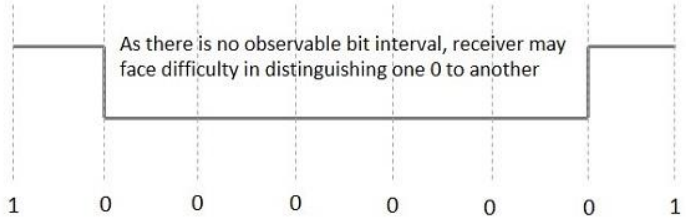
The data encoding technique is divided into the following types, depending upon the type of data conversion.

- **Analog data to Analog signals** − The modulation techniques such as Amplitude Modulation, Frequency Modulation and Phase Modulation of analog signals, fall under this category.
- **Analog data to Digital signals** − This process can be termed as digitization, which is done by Pulse Code Modulation (PCM). Hence, it is nothing but digital modulation. As we have already discussed, sampling and quantization are the important factors in this. Delta Modulation gives a better output than PCM.
- **Digital data to Analog signals** − The modulation techniques such as Amplitude Shift Keying (ASK), Frequency Shift Keying (FSK), Phase Shift Keying (PSK), etc., fall under this category. These will be discussed in subsequent chapters.
- **Digital data to Digital signals** − These are in this section. There are several ways to map digital data to digital signals. Some of them are −

Non Return to Zero (NRZ)

NRZ Codes has **1** for High voltage level and **0** for Low voltage level. The main behavior of NRZ codes is that the voltage level remains constant during bit interval. The end or start of a bit will not be indicated and it will maintain the same voltage state, if the value of the previous bit and the value of the present bit are same.

The following figure explains the concept of NRZ coding.



NRZ Coding

If the above example is considered, as there is a long sequence of constant voltage level and the clock synchronization may be lost due to the absence of bit interval, it becomes difficult for the receiver to differentiate between 0 and 1.

There are two variations in NRZ namely −

NRZ - L (NRZ – LEVEL)

There is a change in the polarity of the signal, only when the incoming signal changes from 1 to 0 or from 0 to 1. It is the same as NRZ, however, the first bit of the input signal should have a change of polarity.

NRZ - I (NRZ – INVERTED)

If a **1** occurs at the incoming signal, then there occurs a transition at the beginning of the bit interval. For a **0** at the incoming signal, there is no transition at the beginning of the bit interval.

NRZ codes has a **disadvantage** that the synchronization of the transmitter clock with the receiver clock gets completely disturbed, when there is a string of **1s** and **0s**. Hence, a separate clock line needs to be provided.

Bi-phase Encoding

The signal level is checked twice for every bit time, both initially and in the middle. Hence, the clock rate is double the data transfer rate and thus the modulation rate is also doubled. The clock is taken from the signal itself. The bandwidth required for this coding is greater.

There are two types of Bi-phase Encoding.

- Bi-phase Manchester
- Differential Manchester

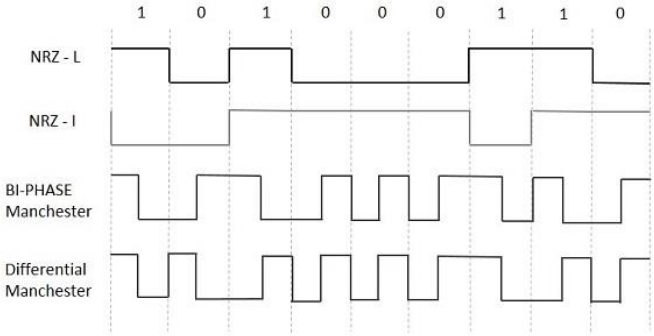
Bi-phase Manchester

In this type of coding, the transition is done at the middle of the bit-interval. The transition for the resultant pulse is from High to Low in the middle of the interval, for the input bit 1. While the transition is from Low to High for the input bit **0**.

Differential Manchester

In this type of coding, there always occurs a transition in the middle of the bit interval. If there occurs a transition at the beginning of the bit interval, then the input bit is **0**. If no transition occurs at the beginning of the bit interval, then the input bit is **1**.

The following figure illustrates the waveforms of NRZ-L, NRZ-I, Bi-phase Manchester and Differential Manchester coding for different digital inputs.



Block Coding

Among the types of block coding, the famous ones are 4B/5B encoding and 8B/6T encoding. The number of bits are processed in different manners, in both of these processes.

4B/5B Encoding

In Manchester encoding, to send the data, the clocks with double speed is required rather than NRZ coding. Here, as the name implies, 4 bits of code is mapped with 5 bits, with a minimum number of **1** bits in the group.

The clock synchronization problem in NRZ-I encoding is avoided by assigning an equivalent word of 5 bits in the place of each block of 4 consecutive bits. These 5-bit words are predetermined in a dictionary.

The basic idea of selecting a 5-bit code is that, it should have **one leading 0** and it should have **no more than two trailing 0s**. Hence, these words are chosen such that two transactions take place per block of bits.

8B/6T Encoding

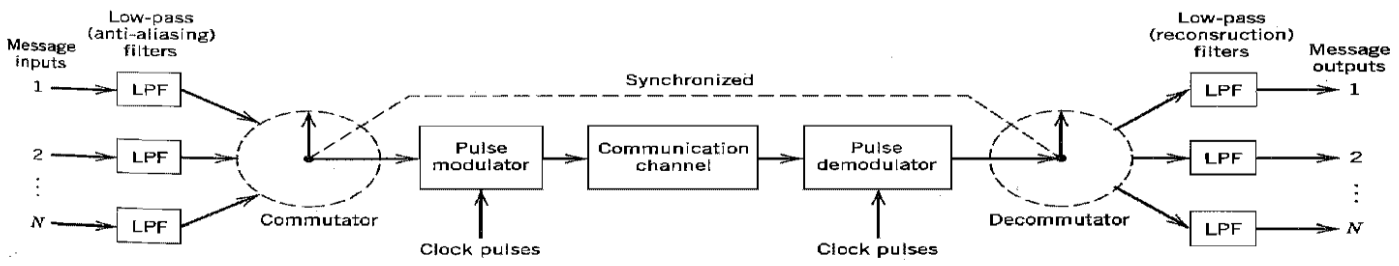
We have used two voltage levels to send a single bit over a single signal. But if we use more than 3 voltage levels, we can send more bits per signal.

Time-Division Multiplexing

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

The concept of TDM is illustrated by the block diagram shown in Figure 3.19. Each input message signal is first restricted in bandwidth by a low-pass anti-aliasing filter to remove the frequencies that are nonessential to an adequate signal representation. The low-pass filter outputs are then applied to a *commutator*, which is usually implemented using electronic switching circuitry. The function of the commutator is twofold: (1) to take a narrow sample of each of the N input messages at a rate f_s that is slightly higher than $2W$, where W is the cutoff frequency of the anti-aliasing filter, and (2) to sequentially interleave these N samples inside the sampling interval T_s. Indeed, this latter function is the essence of the time-division multiplexing operation. Following the commutation process, the multiplexed signal is applied to a *pulse modulator*, the purpose of which is to transform the multiplexed signal into a form suitable for transmission over the common channel. It is clear that the use of time-division multiplexing introduces a bandwidth expansion factor N, because the scheme must squeeze N samples derived from N independent message sources into a time slot equal to one sampling interval. At the receiving end of the system, the received signal is applied to a *pulse demodulator*, which performs the reverse operation of the pulse modulator. The narrow samples produced at the pulse demodulator output are distributed to the appropriate low-pass reconstruction filters by means of a *decommutator*, which operates in *synchronism* with the commutator in the transmitter. This synchronization is essential for a satisfactory operation of the system. The way this synchronization is implemented depends naturally on the method of pulse modulation used to transmit the multiplexed sequence of samples.

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



Block diagram of TDM system.

However, unlike FDM, to a first-order approximation TDM is immune to nonlinearities in the channel as a source of cross-talk. The reason for this behavior is that different message signals are not simultaneously applied to the channel.

SYNCHRONIZATION

In applications using PCM, it is natural to multiplex different messages sources by time division, whereby each source keeps its individuality throughout the journey from the transmitter to the receiver. This individuality accounts for the comparative ease with which message sources may be dropped or reinserted in a time-division multiplex system. As the number of independent message sources is increased, the time interval that may be allotted to each source has to be reduced, since all of them must be accommodated into a time interval equal to the reciprocal of the sampling rate. This, in turn, means that the allowable duration of a code word representing a single sample is reduced. However, pulses tend to become more difficult to generate and to transmit as their duration is reduced. Furthermore, if the pulses become too short, impairments in the transmission medium begin to interfere with the proper operation of the system. Accordingly, in practice, it is necessary to restrict the number of independent message sources that can be included within a time-division group.

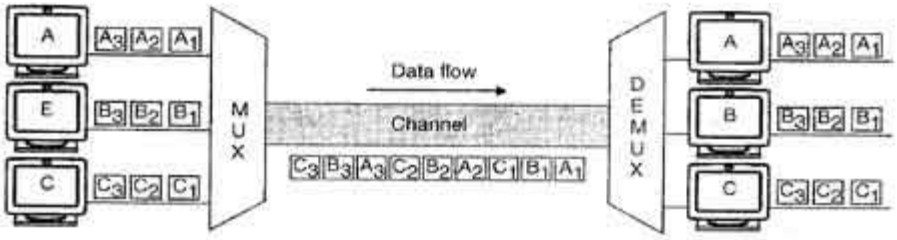
In any event, for a PCM system with time-division multiplexing to operate satisfactorily, it is necessary that the timing operations at the receiver, except for the time lost in transmission and regenerative repeating, follow closely the corresponding operations at the transmitter. In a general way, this amounts to requiring a local clock at the receiver to keep the same time as a distant standard clock at the transmitter, except that the local clock is somewhat slower by an amount corresponding to the time required to transport the message signals from the transmitter to the receiver. One possible procedure to synchronize the transmitter and receiver clocks is to set aside a code element or pulse at the end of a *frame* (consisting of a code word derived from each of the independent message sources in succession) and to transmit this pulse every other frame only. In such a case, the receiver includes a circuit that would search for the pattern of 1s and 0s alternating at half the frame rate, and thereby establish synchronization between the transmitter and receiver.

When the transmission path is interrupted, it is highly unlikely that transmitter and receiver clocks will continue to indicate the same time for long. Accordingly, in carrying out a synchronization process, we must set up an orderly procedure for detecting the synchronizing pulse. The procedure consists of observing the code elements one by one until the synchronizing pulse is detected. That is, after observing a particular code element long enough to establish the absence of the synchronizing pulse, the receiver clock is set back by one code element and the next code element is observed. This *searching process* is repeated until the synchronizing pulse is detected. Clearly, the time required for synchronization depends on the epoch at which proper transmission is re-established.

Time Division Multiplexing

1. TDM is the digital multiplexing technique.
2. In TDM, the channel/link is not divided on the basis of frequency but on the basis of time.
3. Total time available in the channel is divided between several users.
4. Each user is allotted a particular a time interval called time slot or time slice during which the data is transmitted by that user.
5. Thus each sending device takes control of entire bandwidth of the channel for fixed amount of time.

6. In TDM the data rate capacity of the transmission medium should be greater than the data rate required by sending or receiving devices.
7. In TDM all the signals to be transmitted are not transmitted simultaneously. Instead, they are transmitted one-by-one.
8. Thus each signal will be transmitted for a very short time. One cycle or frame is said to be complete when all the signals are transmitted once on the transmission channel.
9. The TDM system can be used to multiplex analog or digital signals, however it is more suitable for the digital signal multiplexing.
10. The TDM signal in the form of frames is transmitted on the common communication medium.



Time Division Multiplexing (TDM)

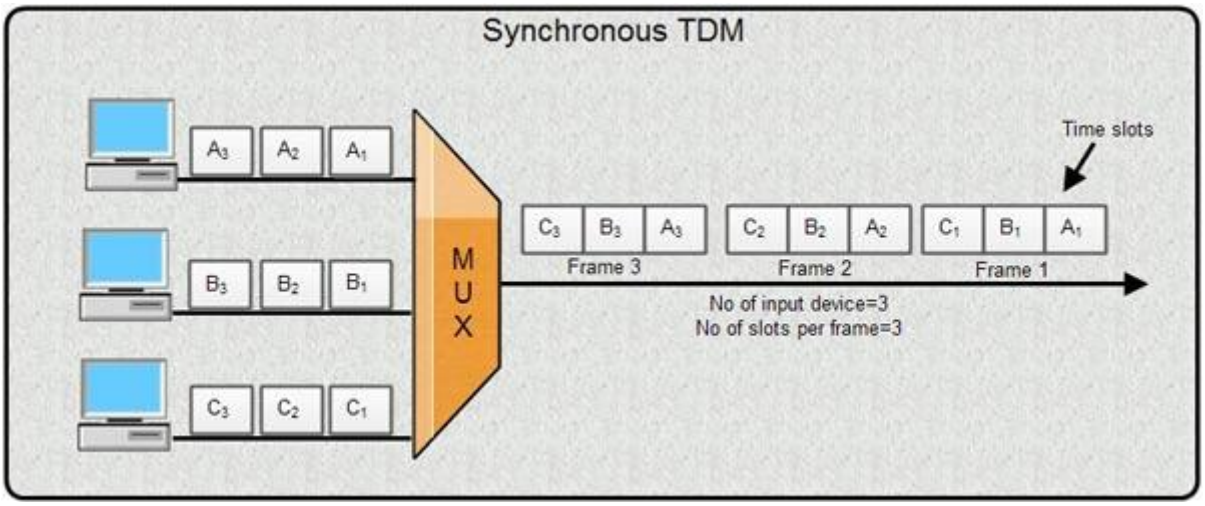
Types of TDM

1. Synchronous TDM
2. Asynchronous TDM

Synchronous TDM (STDM)

1. In synchronous TDM, each device is given same **time slot** to transmit the data over the link, irrespective of the fact that the device has any data to transmit or not. Hence the name Synchronous TDM. Synchronous TDM requires that the total speed of various input lines should not exceed the capacity of path.

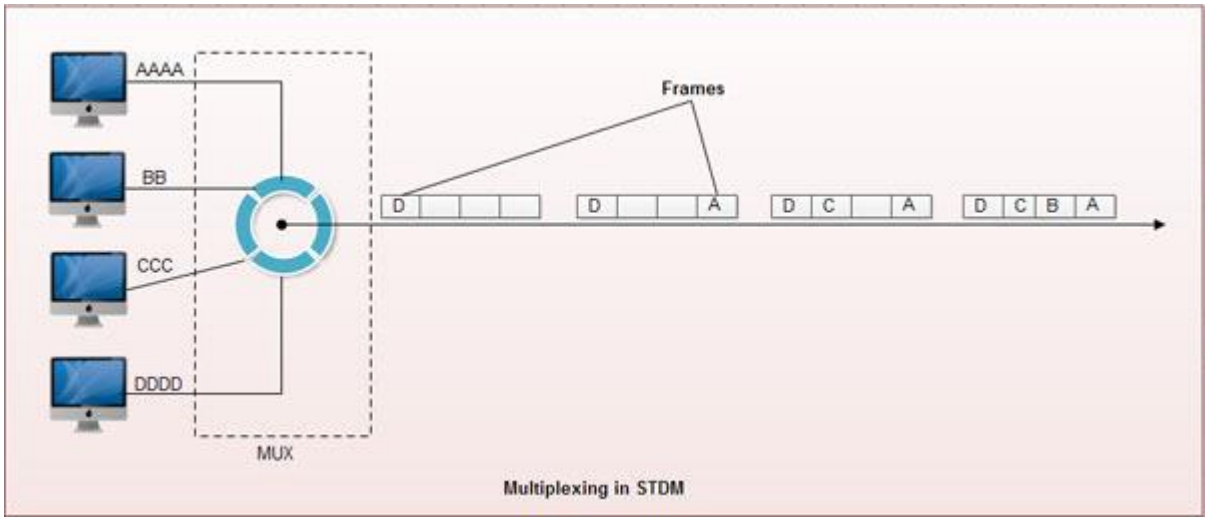
2. Each device places its data onto the link when its **time slot** arrives *i.e.* each device is given the possession of line turn by turn.
3. If any device does not have data to send then its time slot remains empty.
4. The various time slots are organized into **frames** and each frame consists of one or more time slots dedicated to each sending device.
5. If there are *n* sending devices, there will be *n* slots in frame *i.e.* one slot for each device.



6. As show in fig, there are 3 input devices, so there are 3 slots in each frame.

Multiplexing Process in STDM

1. In STDM every device is given the opportunity to transmit a specific amount of data onto the link.
2. Each device gets its turn in fixed order and for fixed amount of time. This process is known as interleaving.
3. We can say that the operation of STDM is similar to that of a fast interleaved switch. The switch opens in front of a device; the device gets a chance to place the data onto the link.
4. Such an interleaving may be done on the basis of a hit, a byte or by any other data unit.
5. In STDM, the interleaved units are of same size *i.e.* if one device sends a byte, other will also send a byte and so on.
6. As shown in the fig. interleaving is done by a character (one byte). Each frame consists of four slots as there are four input devices. The slots of some devices go empty if they do not have any data to send.
7. At the receiver, demultiplexer decomposes each frame by extracting each character in turn. As a character is removed from frame, it is passed to the appropriate receiving device.



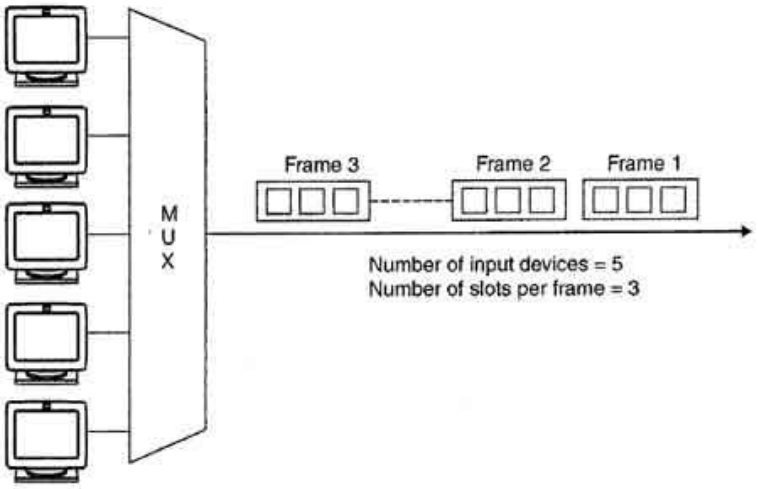
Disadvantages of Synchronous TDM

1. The channel capacity cannot be fully utilized. Some of the slots go empty in certain frames. As shown in fig only first two frames are completely filled. The last three frames have 6 empty slot. It means out of 20 slots in all, 6 slots are empty. This wastes the l/4th capacity of links.
2. The capacity of single communication line that is used to carry the various transmission should be greater than the total speed of input lines.

Asynchronous TDM

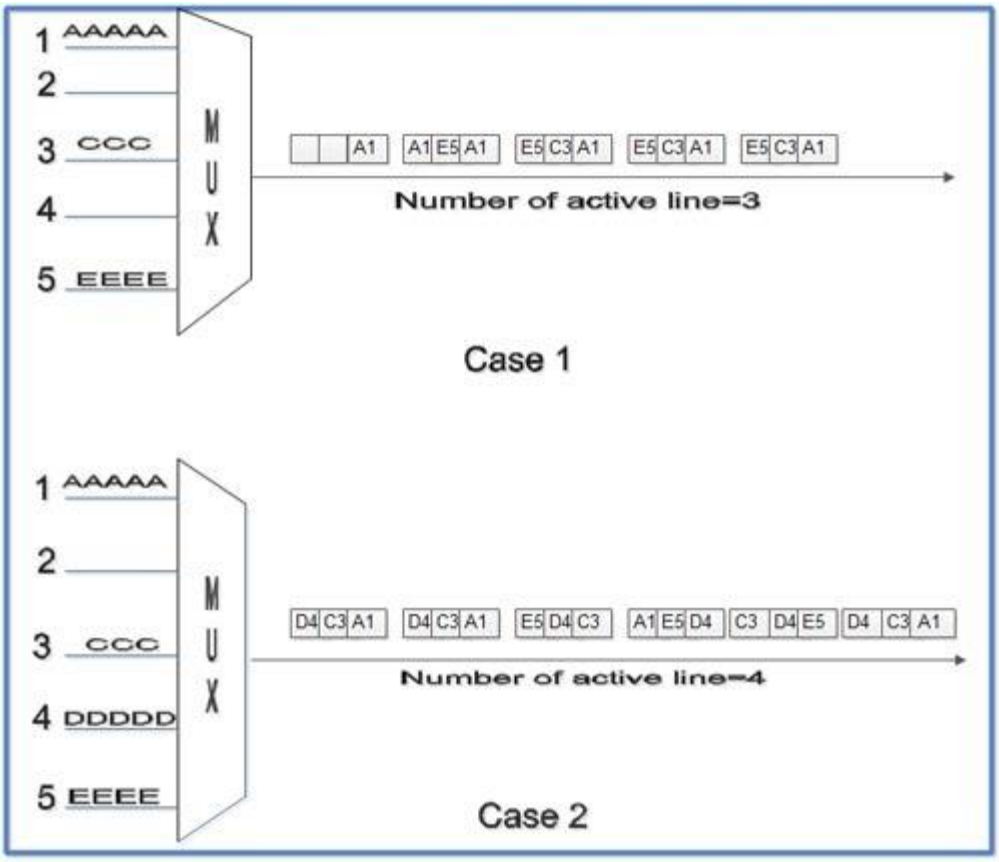
1. It is also known as statistical time division multiplexing.
2. Asynchronous TDM is called so because is this type of multiplexing, time slots are not fixed *i.e.* the slots are flexible.
3. Here, the total speed of input lines can be greater than the capacity of the path.
4. In synchronous TDM, if we have *n* input lines then there are *n* slots in one frame. But in asynchronous it is not so.
5. In asynchronous TDM, if we have *n* input lines then the frame contains not more than *m* slots, with *m* less than *n (m < n).*
6. In asynchronous TDM, the number of time slots in a frame is based on a statistical analysis of number of input lines.
7. In this system slots are not predefined, the slots are allocated to any of the device that has data to send.

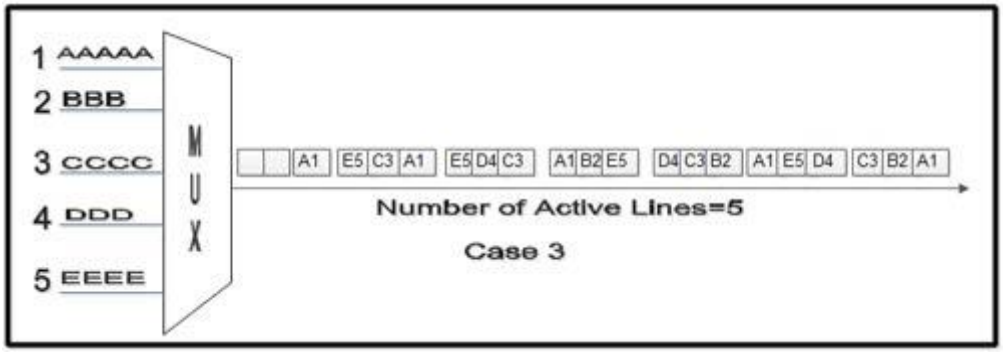
8. The multiplexer scans the various input lines, accepts the data from the lines that have data to send, fills the frame and then sends the frame across the link.
9. If there are not enough data to fill all the slots in a frame, then the frames are transmitted partially filled.
10. Asynchronous Time Division Multiplexing is depicted in fig. Here we have five input lines and three slots per frame.



Asynchronous TDM

11. In Case 1, only three out of five input lines place data onto the link *i.e.* number of input lines and number of slots per frame are same.
12. In Case 2, four out of five input lines are active. Here number of input line is one more than the number of slots per frame.
13. In Case 3, all five input lines are active.
In all these cases, multiplexer scans the various lines in order and fills the frames and transmits them across the channel.
The distribution of various slots in the frames is not symmetrical. In case 2, device 1 occupies first slot in first frame, second slot in second frame and third slot in third frame.





Advantages of TDM :

1. Full available channel bandwidth can be utilized for each channel.
2. lntermodulation distortion is absent.
3. TDM circuitry is not very complex.
4. The problem of crosstalk is not severe.

Disadvantages of TDM :

1. Synchronization is essential for proper operation.
2. Due to slow narrowband fading, all the TDM channels may get wiped out.

UNIT II-DIGITAL MODULATION SYSTEMS

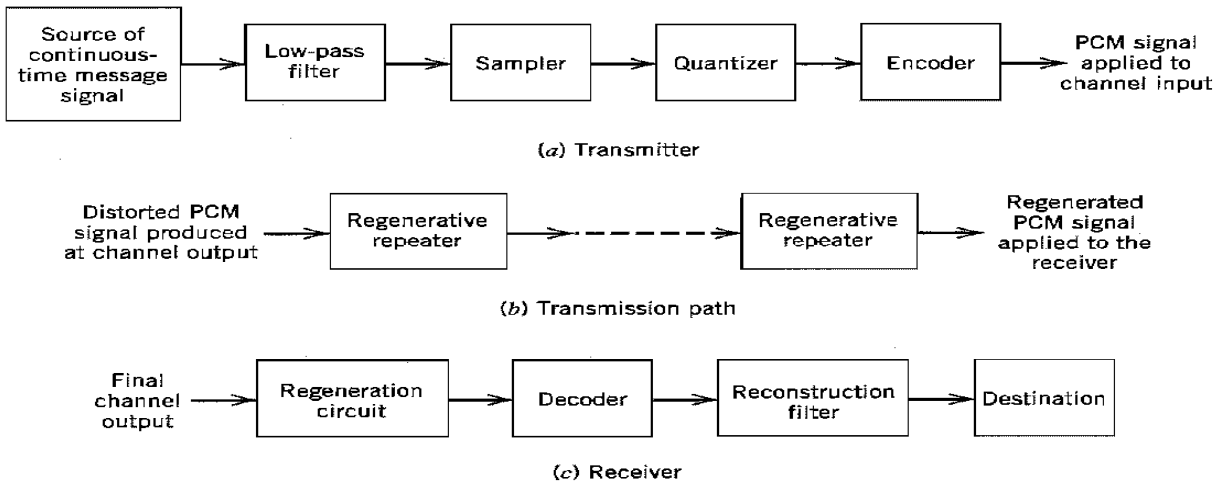
PCM Systems – Noise Considerations in PCM system – Overall Signal-to- noise ratio for PCM system – Threshold effect – Channel Capacity – Virtues, Limitations & Modification of PCM system – PCM Signal Multiplexing – Differential PCM – Delta Modulation – Noise Considerations in Delta Modulation – SNR Calculations – Comparison of PCM, DPCM & DM.

Pulse-Code Modulation

With the sampling and quantization processes at our disposal, we are now ready to describe pulse-code modulation, which, as mentioned previously, is the most basic form of digital pulse modulation. In *pulse-code modulation (PCM), a message signal is represented by a sequence of coded pulses, which is accomplished by representing the signal in discrete form in both time and amplitude.* The basic operations performed in the transmitter of a PCM system are *sampling*, *quantizing*, and *encoding*, as shown in Figure 3.13*a*; the low-pass filter prior to sampling is included to prevent aliasing of the message signal. The quantizing and encoding operations are usually performed in the same circuit, which is called an *analog-to-digital converter*. The basic operations in the receiver are *regeneration* of impaired signals, *decoding*, and *reconstruction* of the train of quantized samples, as shown in Figure 3.13*c*. Regeneration also occurs at intermediate points along the transmission path as necessary, as indicated in Figure 3.13*b*. When time-division multiplexing is used, it becomes necessary to synchronize the receiver to the transmitter for the overall system to operate satisfactorily, as discussed in Section 3.9. In what follows, we describe the various operations that constitute a basic PCM system.

SAMPLING

The incoming message signal is sampled with a train of narrow rectangular pulses so as to closely approximate the instantaneous sampling process. To ensure perfect reconstruction of the message signal at the receiver, the sampling rate must be greater than twice the highest frequency component W of the message signal in accordance with the sampling theorem. In practice, a low-pass anti-aliasing filter is used at the front end of the sampler to exclude frequencies greater than W before sampling. Thus the application of sampling



The basic elements of a PCM system.

permits the reduction of the continuously varying message signal (of some finite duration) to a limited number of discrete values per second.

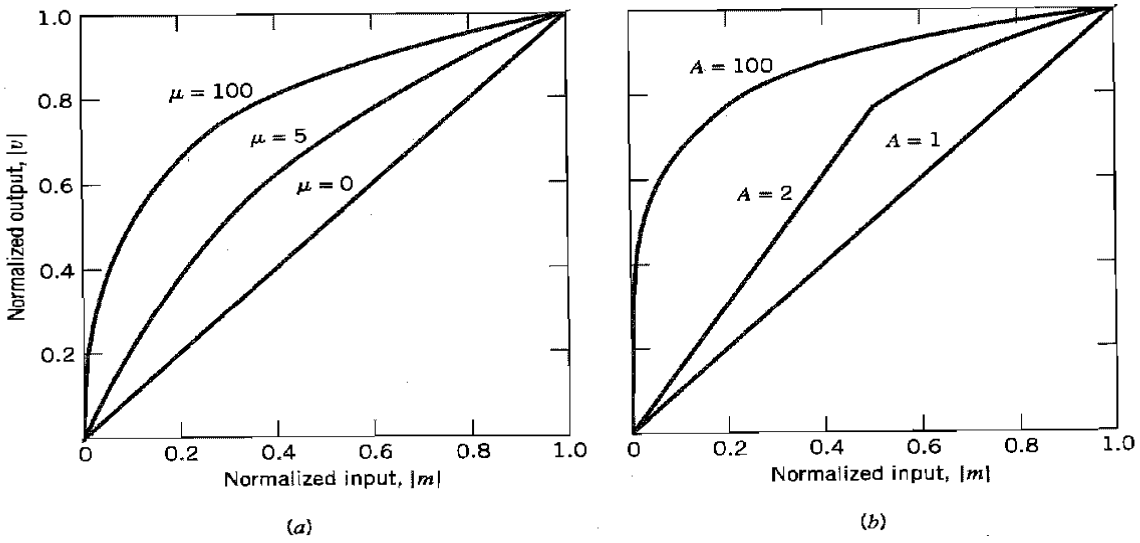
QUANTIZATION

The sampled version of the message signal is then quantized, thereby providing a new representation of the signal that is discrete in both time and amplitude. The quantization process may follow a uniform law as described in Section 3.6. In telephonic communication, however, it is preferable to use a variable separation between the representation levels. For example, the range of voltages covered by voice signals, from the peaks of loud talk to the weak passages of weak talk, is on the order of 1000 to 1. By using a *nonuniform quantizer* with the feature that the step-size increases as the separation from the origin of the input–output amplitude characteristic is increased, the large end steps of the quantizer can take care of possible excursions of the voice signal into the large amplitude ranges that occur relatively infrequently. In other words, the weak passages, which need more protection, are favored at the expense of the loud passages. In this way, a nearly uniform percentage precision is achieved throughout the greater part of the amplitude range of the input signal, with the result that fewer steps are needed than would be the case if a uniform quantizer were used.

The use of a nonuniform quantizer is equivalent to passing the baseband signal through a *compressor* and then applying the compressed signal to a uniform quantizer. A particular form of compression law that is used in practice is the so-called μ-*law*,[5] which is defined by

$$|v| = \frac{\log(1 + \mu|m|)}{\log(1 + \mu)} \tag{3.48}$$

where m and v are the normalized input and output voltages, and μ is a positive constant. In Figure 3.14*a*, we have plotted the μ-law for three different values of μ. The case of uniform quantization corresponds to $\mu = 0$. For a given value of μ, the reciprocal slope



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

of the compression curve, which defines the quantum steps, is given by the derivative of $|m|$ with respect to $|v|$; that is,

$$\frac{d|m|}{d|v|} = \frac{\log(1+\mu)}{\mu}(1+\mu|m|) \tag{3.49}$$

We see therefore that the μ-law is neither strictly linear not strictly logarithmic, but it is approximately linear at low input levels corresponding to $\mu|m| \ll 1$, and approximately logarithmic at high input levels corresponding to $\mu|m| \gg 1$.

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

$$|v| = \begin{cases} \dfrac{A|m|}{1+\log A}, & 0 \le |m| \le \dfrac{1}{A} \\ \dfrac{1+\log(A|m|)}{1+\log A}, & \dfrac{1}{A} \le |m| \le 1 \end{cases} \tag{3.50}$$

which is plotted in Figure 3.14*b* for varying A. The case of uniform quantization corresponds to $A = 1$. The reciprocal slope of this second compression curve is given by the derivative of $|m|$ with respect to $|v|$, as shown by (depending on the value assigned to the normalized input $|m|$)

$$\frac{d|m|}{d|v|} = \begin{cases} \dfrac{1+\log A}{A}, & 0 \le |m| \le \dfrac{1}{A} \\ (1+A)|m|, & \dfrac{1}{A} \le |m| \le 1 \end{cases} \tag{3.51}$$

To restore the signal samples to their correct relative level, we must, of course, use a device in the receiver with a characteristic complementary to the compressor. Such a device is called an *expander*. Ideally, the compression and expansion laws are exactly inverse so that, except for the effect of quantization, the expander output is equal to the compressor input. The combination of a *com*pressor and an exp*ander* is called a *compander*.

For both the μ-law and A-law, the dynamic range capability of the compander improves with increasing μ and A, respectively. The SNR for low-level signals increases at the expense of the SNR for high-level signals. To accommodate these two conflicting requirements (i.e., a reasonable SNR for both low- and high-level signals), a compromise is usually made in choosing the value of paramenter μ for the μ-law and parameter A for the A-law. The typical values used in practice are: $\mu = 255$ and $A = 87.6$.

It is also of interest to note that in actual PCM systems, the companding circuitry does not produce an exact replica of the nonlinear compression curves shown in Figure 3.14. Rather, it provides a *piecewise linear* approximation to the desired curve. By using a large enough number of linear segments, the approximation can approach the true compression curve very closely. This form of approximation is illustrated in Example 3.2.

ENCODING

In combining the processes of sampling and quantization, the specification of a continuous message (baseband) signal becomes limited to a discrete set of values, but not in the form best suited to transmission over a telephone line or radio path. To exploit the advantages of sampling and quantizing for the purpose of making the transmitted signal more robust to noise, interference and other channel impairments, we require the use of an *encoding*

Binary number system for R = 4 bits/sample

Ordinal Number of Representation Level	*Level Number Expressed as Sum of Powers of 2*	*Binary Number*
0		0000
1	2^0	0001
2	2^1	0010
3	$2^1 + 2^0$	0011
4	2^2	0100
5	$2^2 + 2^0$	0101
6	$2^2 + 2^1$	0110
7	$2^2 + 2^1 + 2^0$	0111
8	2^3	1000
9	$2^3 + 2^0$	1001
10	$2^3 + 2^1$	1010
11	$2^3 + 2^1 + 2^0$	1011
12	$2^3 + 2^2$	1100
13	$2^3 + 2^2 + 2^0$	1101
14	$2^3 + 2^2 + 2^1$	1110
15	$2^3 + 2^2 + 2^1 + 2^0$	1111

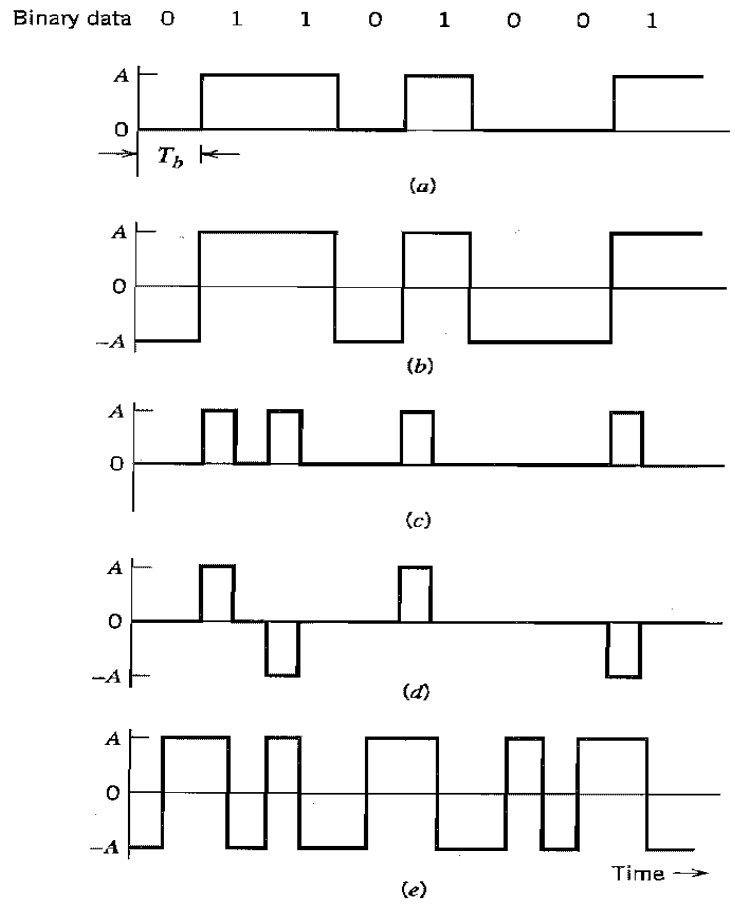
process to translate the discrete set of sample values to a more appropriate form of signal. Any plan for representing each of this discrete set of values as a particular arrangement of discrete events is called a *code*. One of the discrete events in a code is called a *code element* or *symbol*. For example, the presence or absence of a pulse is a symbol. A particular arrangement of symbols used in a code to represent a single value of the discrete set is called a *code word* or *character*.

In a *binary code*, each symbol may be either of two distinct values or kinds, such as the presence or absence of a pulse. The two symbols of a binary code are customarily denoted as 0 and 1. In a *ternary code*, each symbol may be one of three distinct values or kinds, and so on for other codes. However, *the maximum advantage over the effects of noise in a transmission medium is obtained by using a binary code, because a binary symbol withstands a relatively high level of noise and is easy to regenerate*. Suppose that, in a binary code, each code word consists of R bits: *bit* is an acronym for *binary digit*; thus R denotes the number of *bits per sample*. Then, using such a code, we may represent a total of 2^R distinct numbers. For example, a sample quantized into one of 256 levels may be represented by an 8-bit code word.

There are several ways of establishing a one-to-one correspondence between representation levels and code words. A convenient method is to express the ordinal number of the representation level as a binary number. In the binary number system, each digit has a place-value that is a power of 2, as illustrated in Table 3.2 for the case of four bits per sample (i.e., $R = 4$).

Line Codes

Any of several line codes can be used for the electrical representation of a binary data stream. Figure 3.15 displays the waveforms of five important line codes for the example data stream 01101001. Figure 3.16 displays their individual power spectra (for



Line codes for the electrical representations of binary data. (*a*) Unipolar NRZ signaling. (*b*) Polar NRZ signaling. (*c*) Unipolar RZ signaling. (*d*) Bipolar RZ signaling. (*e*) Split-phase or Manchester code.

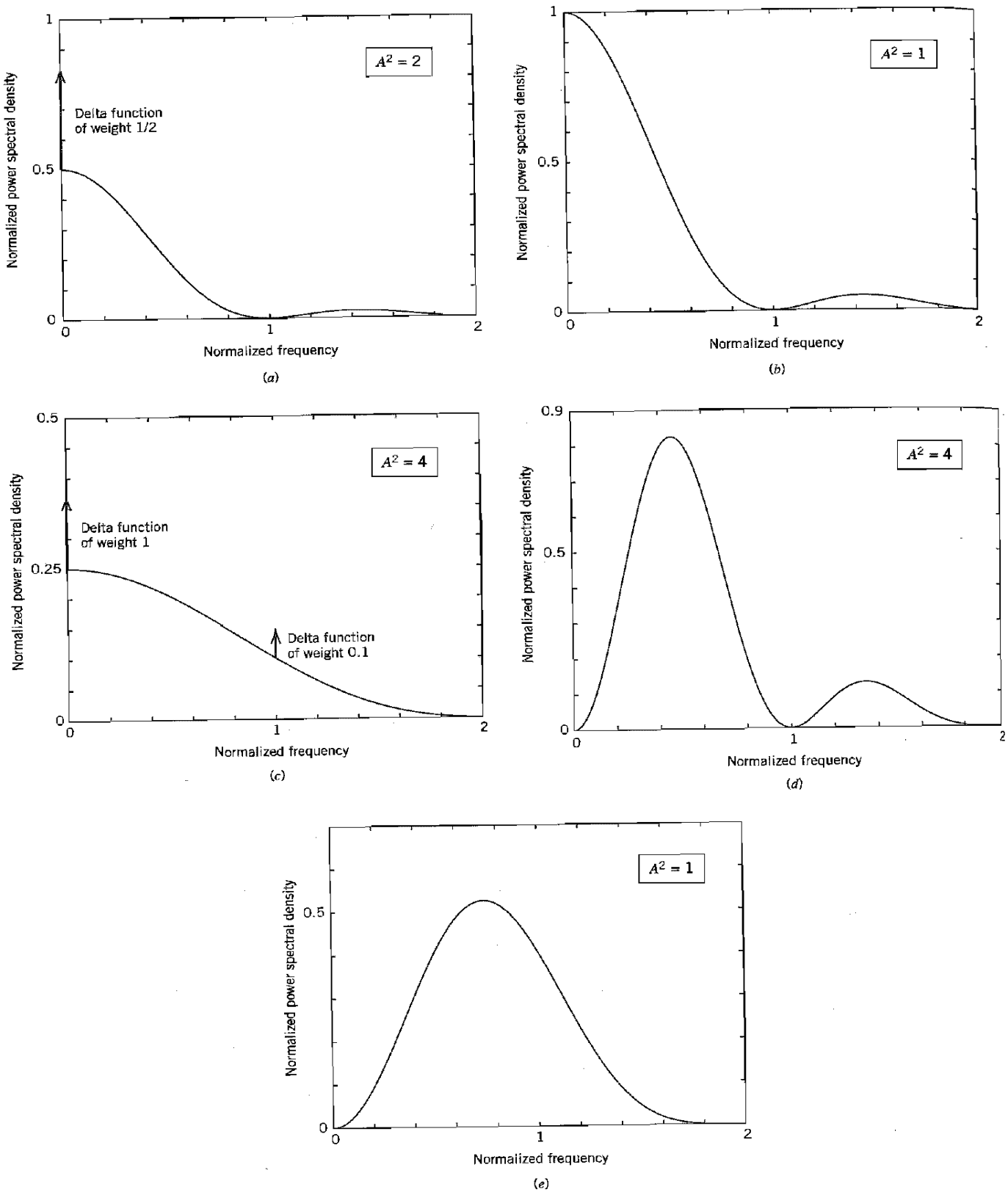
positive frequencies) for randomly generated binary data, assuming that (1) symbols 0 and 1 are equiprobable, (2) the average power is normalized to unity, and (3) the frequency f is normalized with respect to the bit rate $1/T_b$. (For the formulas used to plot the power spectra of Figure 3.16, the reader is referred to Problem 3.11.) The five line codes illustrated in Figure 3.15 are described here:

1. *Unipolar nonreturn-to-zero (NRZ) signaling*

In this line code, symbol 1 is represented by transmitting a pulse of amplitude A for the duration of the symbol, and symbol 0 is represented by switching off the pulse, as in Figure 3.15*a*. This line code is also referred to as *on-off signaling*. Disadvantages of on-off signaling are the waste of power due to the transmitted DC level and the fact that the power spectrum of the transmitted signal does not approach zero at zero frequency.

2. *Polar nonreturn-to-zero (NRZ) signaling*

In this second line code, symbols 1 and 0 are represented by transmitting pulses of amplitudes $+A$ and $-A$, respectively, as illustrated in Figure 3.15*b*. This line code is relatively easy to generate but its disadvantage is that the power spectrum of the signal is large near zero frequency.



Power spectra of line codes: (*a*) Unipolar NRZ signal. (*b*) Polar NRZ signal. (*c*) Unipolar RZ signal. (*d*) Bipolar RZ signal. (*e*) Manchester-encoded signal. The frequency is normalized with respect to the bit rate $1/T_b$, and the average power is normalized to unity.

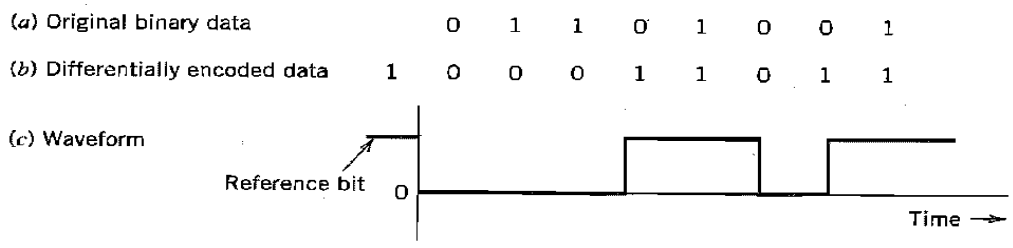
3. *Unipolar return-to-zero (RZ) signaling*
In this other line code, symbol 1 is represented by a rectangular pulse of amplitude A and half-symbol width, and symbol 0 is represented by transmitting *no* pulse, as illustrated in Figure 3.15*c*. An attractive feature of this line code is the presence of delta functions at $f = 0, \pm 1/T_b$ in the power spectrum of the transmitted signal, which can be used for bit-timing recovery at the receiver. However, its disadvantage is that it requires 3 dB more power than polar return-to-zero signaling for the same probability of symbol error; this issue is addressed in Chapter 4 under Problem 4.10.

4. *Bipolar return-to-zero (BRZ) signaling*
This line code uses three amplitude levels as indicated in Figure 3.15*d*. Specifically, positive and negative pulses of equal amplitude (i.e., $+A$ and $-A$) are used alternately for symbol 1, with each pulse having a half-symbol width; no pulse is always used for symbol 0. A useful property of the BRZ signaling is that the power spectrum of the transmitted signal has no DC component and relatively insignificant low-frequency components for the case when symbols 1 and 0 occur with equal probability. This line code is also called *alternate mark inversion* (AMI) signaling.

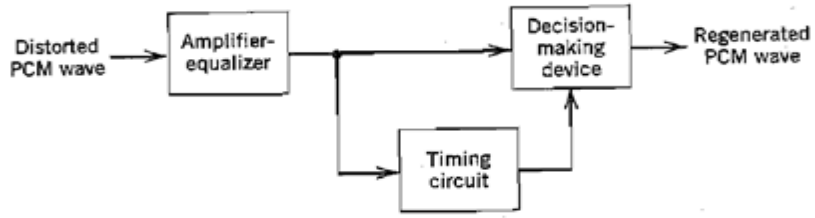
5. *Split-phase (Manchester code)*
In this method of signaling, illustrated in Figure 3.15*e*, symbol 1 is represented by a positive pulse of amplitude A followed by a negative pulse of amplitude $-A$, with both pulses being half-symbol wide. For symbol 0, the polarities of these two pulses are reversed. The Manchester code suppresses the DC component and has relatively insignificant low-frequency components, regardless of the signal statistics. This property is essential in some applications.

Differential Encoding

This method is used to encode information in terms of *signal transitions*. In particular, a transition is used to designate symbol 0 in the incoming binary data stream, while no transition is used to designate symbol 1, as illustrated in Figure 3.17. In Figure 3.17*b* we show the differentially encoded data stream for the example data specified in Figure 3.17*a*. The original binary data stream used here is the same as that used in Figure 3.15. The waveform of the differentially encoded data is shown in Figure 3.17*c*, assuming the use of unipolar nonreturn-to-zero signaling. From Figure 3.17 it is apparent that a differentially encoded signal may be inverted without affecting its interpretation. The original binary information is recovered simply by comparing the polarity of adjacent binary symbols to establish whether or not a transition has occurred. Note that differential encoding requires the use of a *reference bit* before initiating the encoding process. In Figure 3.17, symbol 1 is used as the reference bit.



(*a*) Original binary data. (*b*) Differentially encoded data, assuming reference bit 1. (*c*) Waveform of differentially encoded data using unipolar NRZ signaling.



Block diagram of regenerative repeater.

REGENERATION

The most important feature of PCM systems lies in the ability to control the effects of distortion and noise produced by transmitting a PCM signal through a channel. This capability is accomplished by reconstructing the PCM signal by means of a chain of *regenerative repeaters* located at sufficiently close spacing along the transmission route. As illustrated in Figure 3.18, three basic functions are performed by a regenerative repeater: *equalization*, *timing*, and *decision making*. The equalizer shapes the received pulses so as to compensate for the effects of amplitude and phase distortions produced by the nonideal transmission characteristics of the channel. The timing circuitry provides a periodic pulse train, derived from the received pulses, for sampling the equalized pulses at the instants of time where the signal-to-noise ratio is a maximum. Each sample so extracted is compared to a predetermined *threshold* in the decision-making device. In each bit interval, a decision is then made whether the received symbol is a 1 or a 0 on the basis of whether the threshold is exceeded or not. If the threshold is exceeded, a clean new pulse representing symbol 1 is transmitted to the next repeater. Otherwise, another clean new pulse representing symbol 0 is transmitted. In this way, the accumulation of distortion and noise in a repeater span is completely removed, provided that the disturbance is not too large to cause an error in the decision-making process. Ideally, except for delay, the regenerated signal is exactly the same as the signal originally transmitted. In practice, however, the regenerated signal departs from the original signal for two main reasons:

1. The unavoidable presence of channel noise and interference causes the repeater to make wrong decisions occasionally, thereby introducing *bit errors* into the regenerated signal.
2. If the spacing between received pulses deviates from its assigned value, a *jitter* is introduced into the regenerated pulse position, thereby causing distortion.

DECODING

The first operation in the receiver is to regenerate (i.e., reshape and clean up) the received pulses one last time. These clean pulses are then regrouped into code words and decoded (i.e., mapped back) into a quantized PAM signal. The *decoding* process involves generating a pulse the amplitude of which is the linear sum of all the pulses in the code word, with each pulse being weighted by its place value $(2^0, 2^1, 2^2, \ldots, 2^{R-1})$ in the code, where R is the number of bits per sample.

FILTERING

The final operation in the receiver is to recover the message signal by passing the decoder output through a low-pass reconstruction filter whose cutoff frequency is equal to the message bandwidth W. Assuming that the transmission path is error free, the recovered

signal includes no noise with the exception of the initial distortion introduced by the quantization process.

Noise Considerations in PCM Systems

The performance of a PCM system is influenced by two major sources of noise:

1. *Channel noise*, which is introduced anywhere between the transmitter output and the receiver input. Channel noise is always present, once the equipment is switched on.
2. *Quantization noise*, which is introduced in the transmitter and is carried all the way along to the receiver output. Unlike channel noise, quantization noise is *signal-dependent* in the sense that it disappears when the message signal is switched off.

Naturally, these two sources of noise appear simultaneously once the PCM system is in operation. However, the traditional practice is to consider them separately, so that we may develop insight into their individual effects on the system performance.

The main effect of channel noise is to introduce *bit errors* into the received signal. In the case of a binary PCM system, the presence of a bit error causes symbol 1 to be mistaken for symbol 0, or vice versa. Clearly, the more frequently bit errors occur, the more dissimilar the receiver output becomes compared to the original message signal. The fidelity of information transmission by PCM in the presence of channel noise may be measured in terms of the *average probability of symbol error*, which is defined as the probability that the reconstructed symbol at the receiver output differs from the transmitted binary symbol, on the average. The average probability of symbol error, also referred to as the *bit error rate* (BER), assumes that all the bits in the original binary wave are of equal importance. When, however, there is more interest in reconstructuring the analog waveform of the original message signal, different symbol errors may need to be *weighted* differently; for example, an error in the most significant bit in a code word (representing a quantized sample of the message signal) is more harmful than an error in the least significant bit.

To optimize system performance in the presence of channel noise, we need to minimize the average probability of symbol error. For this evaluation, it is customary to model the channel noise as additive, white, and Gaussian. The effect of channel noise can be made practically negligible by ensuring the use of an adequate signal energy-to-noise density ratio through the provision of short-enough spacing between the regenerative repeaters in the PCM system. In such a situation, the performance of the PCM system is essentially limited by quantization noise acting alone.

From the discussion of quantization noise presented in Section 3.6, we recognize that quantization noise is essentially under the designer's control. It can be made negligibly small through the use of an adequate number of representation levels in the quantizer and the selection of a companding strategy matched to the characteristics of the type of message signal being transmitted. We thus find that the use of PCM offers the possibility of building a communication system that is *rugged* with respect to channel noise on a scale that is beyond the capability of any CW modulation or analog pulse modulation system.

Error Threshold

The underlying theory of bit error rate calculation in a PCM system is deferred until Chapter 4. For the present, it suffices to say that the average probability of symbol error in a binary encoded PCM receiver due to additive white Gaussian noise depends solely on

E_b/N_0, which is defined as *the ratio of the transmitted signal energy per bit, E_b, to the noise spectral density, N_0*. Note that the ratio E_b/N_0 is dimensionless even though the quantities E_b and N_0 have different physical meaning. In Table 3.3 we present a summary of this dependence for the case of a binary PCM system using polar nonreturn-to-zero signaling. The results presented in the last column of the table assume a bit rate of 10^5 b/s.

From Table 3.3 it is clear that there is an *error threshold* (at about 11 dB). For E_b/N_0 below the error threshold the receiver performance involves significant numbers of errors, and above it the effect of channel noise is practically negligible. In other words, provided that the ratio E_b/N_0 exceeds the error threshold, channel noise has virtually no effect on the receiver performance, which is precisely the goal of PCM. When, however, E_b/N_0 drops below the error threshold, there is a sharp increase in the rate at which errors occur in the receiver. Because decision errors result in the construction of incorrect code words, we find that when the errors are frequent, the reconstructed message at the receiver output bears little resemblance to the original message.

Comparing the figure of 11 dB for the error threshold in a PCM system using polar NRZ signaling with the 60–70 dB required for high-quality transmission of speech using amplitude modulation, we see that PCM requires much less power, even though the average noise power in the PCM system is increased by the R-fold increase in bandwidth, where R is the number of bits in a code word (i.e., bits per sample).

In most transmission systems, the effects of noise and distortion from the individual links accumulate. For a given quality of overall transmission, the longer the physical separation between the transmitter and the receiver, the more severe are the requirements on each link in the system. In a PCM system, however, because the signal can be regenerated as often as necessary, the effects of amplitude, phase, and nonlinear distortions in one link (if not too severe) have practically no effect on the regenerated input signal to the next link. We have also seen that the effect of channel noise can be made practically negligible by using a ratio E_b/N_0 above threshold. For all practical purposes, then, the transmission requirements for a PCM link are almost independent of the physical length of the communication channel.

Another important characteristic of a PCM system is its *ruggedness to interference*, caused by stray impulses or cross-talk. The combined presence of channel noise and interference causes the error threshold necessary for satisfactory operation of the PCM system to increase. If an adequate margin over the error threshold is provided in the first place, however, the system can withstand the presence of relatively large amounts of interference. In other words, a PCM system is *robust* to channel noise and interference.

Influence of E_b/N_0 on the probability of error

E_b/N_0	*Probability of Error* P_e	*For a Bit Rate of 10^5 b/s, This Is About One Error Every*
4.3 dB	10^{-2}	10^{-3} second
8.4	10^{-4}	10^{-1} second
10.6	10^{-6}	10 seconds
12.0	10^{-8}	20 minutes
13.0	10^{-10}	1 day
14.0	10^{-12}	3 months

Virtues, Limitations, and Modifications of PCM

In a generic sense, pulse-code modulation (PCM) has emerged as the most favored modulation scheme for the transmission of analog information-bearing signals such as voice and video signals. The advantages of PCM may all be traced to the use of *coded pulses for the digital representation of analog signals*, a feature that distinguishes it from all other analog methods of modulation. We may summarize the important advantages of PCM as follows:

1. *Robustness* to channel noise and interference.
2. Efficient *regeneration* of the coded signal along the transmission path.
3. Efficient *exchange* of increased channel bandwidth for improved signal-to-noise ratio, obeying an exponential law.
4. A *uniform format* for the transmission of different kinds of baseband signals, hence their integration with other forms of digital data in a common network.
5. Comparative *ease* with which message sources may be dropped or reinserted in a time-division multiplex system.
6. *Secure* communication through the use of special modulation schemes or encryption; the encryption and decryption of data are discussed in Appendix 5.

These advantages, however, are attained at the cost of increased system complexity and increased channel bandwidth. These two issues are considered in the sequel in turn.

Although the use of PCM involves many complex operations, today they can all be implemented in a cost-effective fashion using commercially available and/or custom-made *very-large-scale integrated* (VLSI) chips. In other words, the requisite device technology for the implementation of a PCM system is already in place. Moreover, with continuing improvements in VLSI technology, we are likely to see an ever-expanding use of PCM for the digital transmission of analog signals.

If, however, the simplicity of implementation is a necessary requirement, then we may use delta modulation as an alternative to pulse-code modulation. In delta modulation, the baseband signal is intentionally "oversampled" to permit the use of a simple quantizing strategy for constructing the encoded signal; delta modulation is discussed in Section 3.12.

Turning next to the issue of bandwidth, we do recognize that the increased bandwidth requirement of PCM may have been a reason for justifiable concern in the past. Today, however, it is of no real concern for two different reasons. First, the increasing availability of *wideband communication channels* means that bandwidth is no longer a system constraint in the traditional way it used to be. Liberation from the bandwidth constraint has been made possible by the deployment of communication satellites for broadcasting and the ever-increasing use of fiber optics for networking; a discussion of these communication channel concepts was presented in the Background and Preview chapter.

The second reason is that through the use of sophisticated *data compression* techniques, it is indeed possible to remove the redundancy inherently present in a PCM signal and thereby reduce the bit rate of the transmitted data without serious degradation in system performance. In effect, increased processing complexity (and therefore increased cost of implementation) is traded off for a reduced bit rate and therefore reduced bandwidth requirement. A major motivation for bit-rate reduction is for secure communication over radio channels that are inherently of low capacity.

Differential Pulse-Code Modulation

When a voice or video signal is sampled at a rate slightly higher than the Nyquist rate as usually done in pulse-code modulation, the resulting sampled signal is found to exhibit a high degree of correlation between adjacent samples. The meaning of this high correlation is that, in an average sense, the signal does not change rapidly from one sample to the next, and as a result, the difference between adjacent samples has a variance that is smaller than the variance of the signal itself. When these highly correlated samples are encoded, as in the standard PCM system, the resulting encoded signal contains *redundant information*. This means that symbols that are not absolutely essential to the transmission of information are generated as a result of the encoding process. By removing this redundancy before encoding, we obtain a more efficient coded signal, which is the basic idea behind differential pulse-code modulation.

Now if we know the past behavior of a signal up to a certain point in time, we may use prediction to make an estimate of a future value of the signal as described in Section 3.13. Suppose then a baseband signal $m(t)$ is sampled at the rate $f_s = 1/T_s$ to produce the sequence $\{m[n]\}$ whose samples are T_s seconds apart. The fact that it is possible to predict future values of the signal $m(t)$ provides motivation for the *differential quantization* scheme shown in Figure 3.28*a*. In this scheme, the input signal to the quantizer is defined by

$$e[n] = m[n] - \hat{m}[n] \tag{3.74}$$

which is the difference between the unquantized input sample $m[n]$ and a prediction of it, denoted by $\hat{m}[n]$. This predicted value is produced by using a linear prediction filter whose input, as we will see, consists of a quantized version of the input sample $m[n]$. The difference signal $e[n]$ is the prediction error, since it is the amount by which the prediction filter fails to predict the input exactly. By encoding the quantizer output, as in Figure 3.28*a*, we obtain a variant of PCM known as *differential pulse-code modulation*[10] (DPCM).

The quantizer output may be expressed as

$$e_q[n] = e[n] + q[n] \tag{3.75}$$

where $q[n]$ is the quantization error. According to Figure 3.28*a*, the quantizer output $e_q[n]$ is added to the predicted value $\hat{m}[n]$ to produce the prediction-filter input

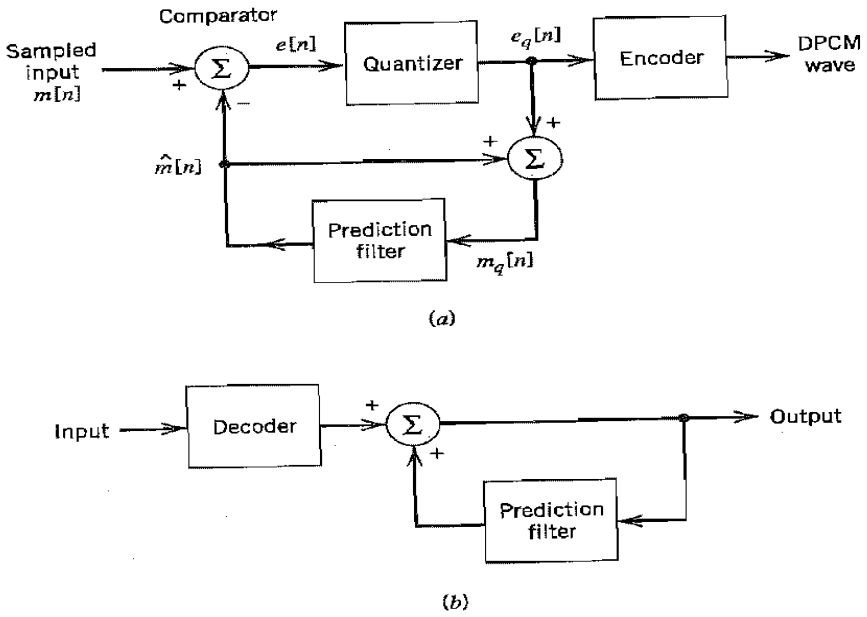
$$m_q[n] = \hat{m}[n] + e_q[n] \tag{3.76}$$

Substituting Equation (3.75) into (3.76), we get

$$m_q[n] = \hat{m}[n] + e[n] + q[n] \tag{3.77}$$

However, from Equation (3.74) we observe that the sum term $\hat{m}[n] + e[n]$ is equal to the input sample $m[n]$. Therefore, we may simplify Equation (3.77) as

$$m_q[n] = m[n] + q[n] \tag{3.78}$$



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

which represents a quantized version of the input sample $m[n]$. That is, irrespective of the properties of the prediction filter, the quantized sample $m_q[n]$ at the prediction filter input differs from the original input sample $m[n]$ by the quantization error $q[n]$. Accordingly, if the prediction is good, the variance of the prediction error $e[n]$ will be smaller than the variance of $m[n]$, so that a quantizer with a given number of levels can be adjusted to produce a quantization error with a smaller variance than would be possible if the input sample $m[n]$ were quantized directly as in a standard PCM system.

The receiver for reconstructing the quantized version of the input is shown in Figure 3.28*b*. It consists of a decoder to reconstruct the quantized error signal. The quantized version of the original input is reconstructed from the decoder output using the same prediction filter used in the transmitter of Figure 3.28*a*. In the absence of channel noise, we find that the encoded signal at the receiver input is identical to the encoded signal at the transmitter output. Accordingly, the corresponding receiver output is equal to $m_q[n]$, which differs from the original input $m[n]$ only by the quantization error $q[n]$ incurred as a result of quantizing the prediction error $e[n]$.

From the foregoing analysis we observe that, in a noise-free environment, the prediction filters in the transmitter and receiver operate on the same sequence of samples, $m_q[n]$. It is with this purpose in mind that a feedback path is added to the quantizer in the transmitter, as shown in Figure 3.28*a*.

Differential pulse-code modulation includes delta modulation as a special case. In particular, comparing the DPCM system of Figure 3.28 with the DM system of Figure 3.23, we see that they are basically similar, except for two important differences: the use of a one-bit (two-level) quantizer in the delta modulator and the replacement of the prediction filter by a single delay element (i.e., zero prediction order). Simply put, DM is the 1-bit version of DPCM. Note that unlike a standard PCM system, the transmitters of both the DPCM and DM involve the use of *feedback*.

DPCM, like DM, is subject to slope-overload distortion whenever the input signal changes too rapidly for the prediction filter to track it. Also, like PCM, DPCM suffers from quantization noise.

PROCESSING GAIN

The output signal-to-noise ratio of the DPCM system shown in Figure 3.28 is, by definition,

$$(\text{SNR})_O = \frac{\sigma_M^2}{\sigma_Q^2} \tag{3.79}$$

where σ_M^2 is the variance of the original input sample $m[n]$, assumed to be of zero mean, and σ_Q^2 is the variance of the quantization error $q[n]$. We may rewrite Equation (3.79) as the product of two factors as follows:

$$\begin{aligned}(\text{SNR})_O &= \left(\frac{\sigma_M^2}{\sigma_E^2}\right)\left(\frac{\sigma_E^2}{\sigma_Q^2}\right) \\ &= G_p(\text{SNR})_Q\end{aligned} \tag{3.80}$$

where σ_E^2 is the variance of the prediction error. The factor $(\text{SNR})_Q$ is the *signal-to-quantization noise ratio*, which is defined by

$$(\text{SNR})_Q = \frac{\sigma_E^2}{\sigma_Q^2} \tag{3.81}$$

The other factor G_p is the *processing gain* produced by the differential quantization scheme; it is defined by

$$G_p = \frac{\sigma_M^2}{\sigma_E^2} \tag{3.82}$$

The quantity G_p, when greater than unity, represents a gain in signal-to-noise ratio that is due to the differential quantization scheme of Figure 3.28. Now, for a given baseband (message) signal, the variance σ_M^2 is fixed, so that G_p is maximized by minimizing the variance σ_E^2 of the prediction error $e[n]$. Accordingly, our objective should be to design the prediction filter so as to minimize σ_E^2.

In the case of voice signals, it is found that the optimum signal-to-quantization noise advantage of DPCM over standard PCM is in the neighborhood of 4 to 11 dB. The greatest improvement occurs in going from no prediction to first-order prediction, with some additional gain resulting from increasing the order of the prediction filter up to 4 or 5, after which little additional gain is obtained. Since 6 dB of quantization noise is equivalent to 1 bit per sample by virtue of Equation (3.35), the advantage of DPCM may also be expressed in terms of bit rate. For a constant signal-to-quantization noise ratio, and assuming a sampling rate of 8 kHz, the use of DPCM may provide a saving of about 8 to 16 kb/s (i.e., 1 to 2 bits per sample) compared to the standard PCM.

Delta Modulation

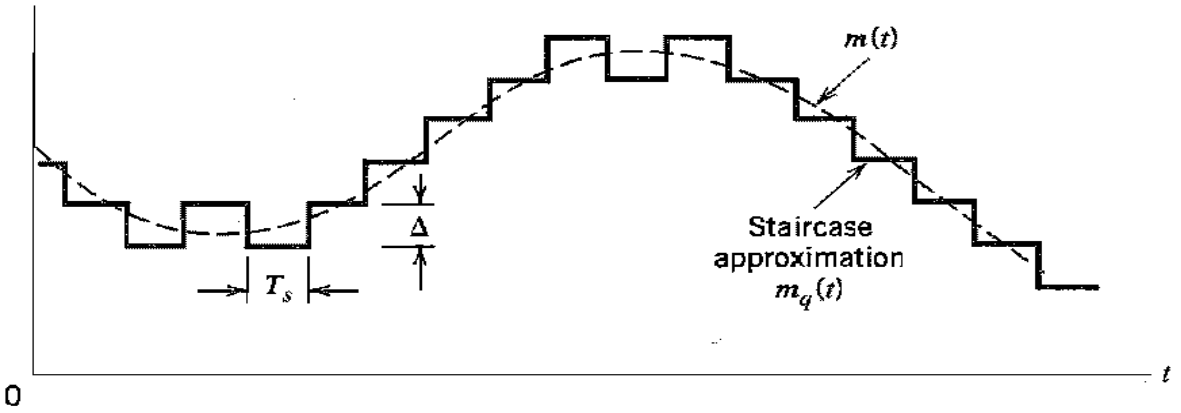
In *delta modulation*[8] (DM), an incoming message signal is oversampled (i.e., at a rate much higher than the Nyquist rate) to purposely increase the correlation between adjacent samples of the signal. This is done to permit the use of a simple quantizing strategy for constructing the encoded signal.

In its basic form, DM provides a *staircase approximation* to the oversampled version of the message signal, as illustrated in Figure 3.22*a*. The difference between the input and the approximation is quantized into only two levels, namely, $\pm\Delta$, corresponding to positive and negative differences. Thus if the approximation falls below the signal at any sampling epoch, it is increased by Δ. If on the other hand, the approximation lies above the signal, it is diminished by Δ. Provided that the signal does not change too rapidly from sample to sample, we find that the staircase approximation remains within $\pm\Delta$ of the input signal.

Let $m(t)$ denote the input (message) signal, and $m_q(t)$ denote its staircase approximation. For convenience of presentation, we adopt the following notation that is commonly used in the digital signal processing literature:

$$m[n] = m(nT_s), \qquad n = 0, \pm 1, \pm 2, \ldots$$

where T_s is the sampling period and $m(nT_s)$ is a sample of the signal $m(t)$ taken at time $t = nT_s$, and likewise for the samples of other continuous-time signals. We may then



(*a*)

Binary sequence at modulator output: 0 0 1 0 1 1 1 1 1 0 1 0 0 0 0 0 0

(*b*)

Illustration of delta modulation.

formalize the basic principles of delta modulation in the following set of discrete-time relations:

$$e[n] = m[n] - m_q[n-1] \tag{3.52}$$

$$e_q = \Delta \operatorname{sgn}(e[n]) \tag{3.53}$$

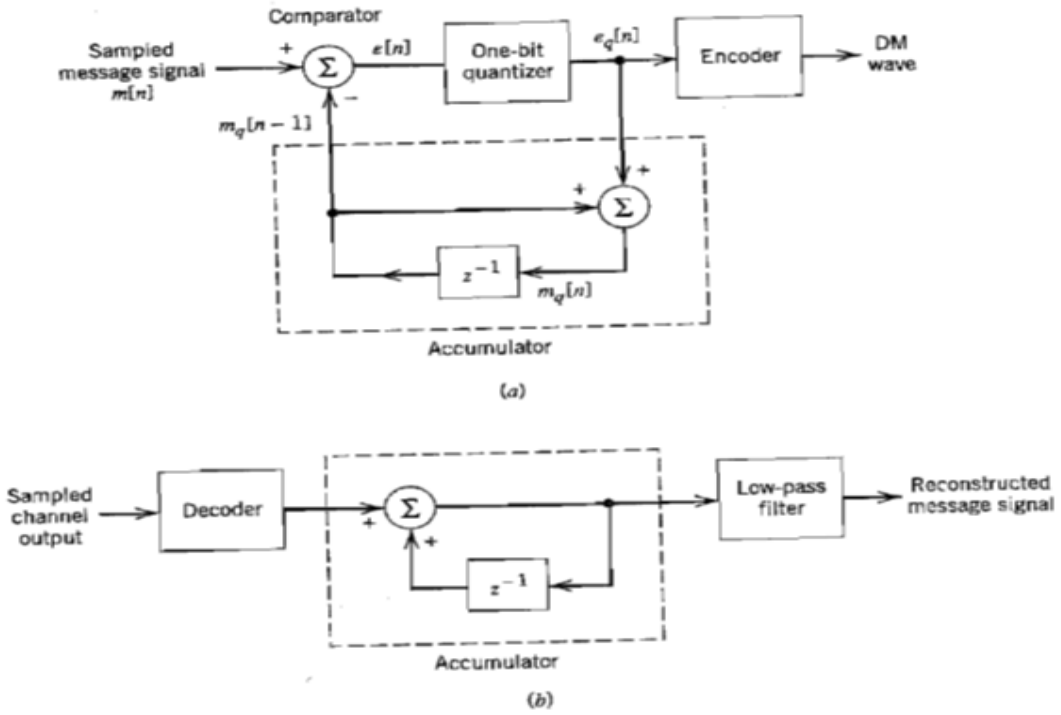
$$m_q[n] = m_q[n-1] + e_q[n] \tag{3.54}$$

where $e[n]$ is an *error signal* representing the difference between the present sample $m[n]$ of the input signal and the latest approximation $m_q[n-1]$ to it, $e_q[n]$ is the quantized version of $e[n]$, and $\operatorname{sgn}(\cdot)$ is the signum function. Finally, the quantizer output $m_q[n]$ is coded to produce the DM signal.

Figure 3.22*a* illustrates the way in which the staircase approximation $m_q(t)$ follows variations in the input signal $m(t)$ in accordance with Equations (3.52)–(3.54), and Figure 3.22*b* displays the corresponding binary sequence at the delta modulator output. It is apparent that in a delta modulation system the rate of information transmission is simply equal to the sampling rate $f_s = 1/T_s$.

The principal virtue of delta modulation is its simplicity. It may be generated by applying the sampled version of the incoming message signal to a modulator that involves a *comparator*, *quantizer*, and *accumulator* interconnected as shown in Figure 3.23*a*. The block labeled z^{-1} inside the accumulator represents a *unit delay*, that is, a delay equal to one sampling period. (The variable z is commonly used in the z-transform, which is basic to the analysis of discrete-time signals and systems.) Details of the modulator follow directly from Equations (3.52)–(3.54). The comparator computes the difference between its two inputs. The quantizer consists of a *hard limiter* with an input-output relation that is a scaled version of the signum function. The quantizer output is then applied to an accumulator, producing the result

$$\begin{aligned} m_q[n] &= \Delta \sum_{i=1}^{n} \operatorname{sgn}(e[i]) \\ &= \sum_{i=1}^{n} e_q[i] \end{aligned} \tag{3.55}$$



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

which is obtained by solving Equations (3.53) and (3.54) for $m_q[n]$. Thus, at the sampling instant nT_s, the accumulator increments the approximation by a step Δ in a positive or negative direction, depending on the algebraic sign of the error sample $e[n]$. If the input sample $m[n]$ is greater than the most recent approximation $m_q[n]$, a positive increment $+\Delta$ is applied to the approximation. If, on the other hand, the input sample is smaller, a negative increment $-\Delta$ is applied to the approximation. In this way, the accumulator does the best it can to track the input samples by one step (of amplitude $+\Delta$ or $-\Delta$) at a time. In the receiver shown in Figure 3.23*b*, the staircase approximation $m_q(t)$ is reconstructed by passing the sequence of positive and negative pulses, produced at the decoder output, through an accumulator in a manner similar to that used in the transmitter. The out-of-band quantization noise in the high-frequency staircase waveform $m_q(t)$ is rejected by passing it through a low-pass filter, as in Figure 3.23*b*, with a bandwidth equal to the original message bandwidth.

Delta modulation is subject to two types of quantization error: slope overload distortion and granular noise. We will discuss the case of slope overload distortion first.

We observe that Equation (3.54) is the digital equivalent of integration in the sense that it represents the accumulation of positive and negative increments of magnitude Δ. Also, denoting the quantization error by $q[n]$, as shown by

$$m_q[n] = m[n] + q[n] \tag{3.56}$$

we observe from Equation (3.52) that the input to the quantizer is

$$e[n] = m[n] - m[n-1] - q[n-1] \tag{3.57}$$

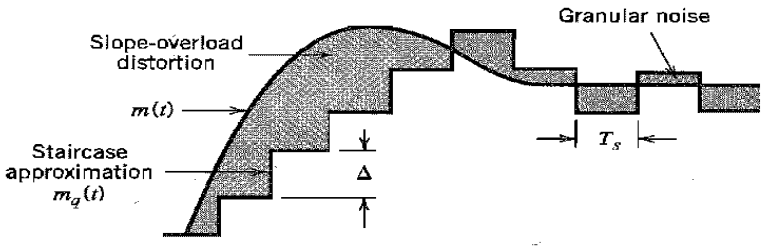


Illustration of the two different forms of quantization error in delta modulation.

Thus except for the quantization error $q[n-1]$, the quantizer input is a *first backward difference* of the input signal, which may be viewed as a digital approximation to the derivative of the input signal or, equivalently, as the inverse of the digital integration process. If we consider the maximum slope of the original input waveform $m(t)$, it is clear that in order for the sequence of samples $\{m_q[n]\}$ to increase as fast as the input sequence of samples $\{m[n]\}$ in a region of maximum slope of $m(t)$, we require that the condition

$$\frac{\Delta}{T_s} \geq \max\left|\frac{dm(t)}{dt}\right| \tag{3.58}$$

be satisfied. Otherwise, we find that the step-size Δ is too small for the staircase approximation $m_q(t)$ to follow a steep segment of the input waveform $m(t)$, with the result that $m_q(t)$ falls behind $m(t)$, as illustrated in Figure 3.24. This condition is called *slope overload*, and the resulting quantization error is called *slope-overload distortion* (*noise*). Note that since the maximum slope of the staircase approximation $m_q(t)$ is fixed by the step size Δ, increases and decreases in $m_q(t)$ tend to occur along straight lines. For this reason, a delta modulator using a fixed step size is often referred to as a *linear delta modulator*.

In contrast to slope-overload distortion, *granular noise* occurs when the step size Δ is too large relative to the local slope characteristics of the input waveform $m(t)$, thereby causing the staircase approximation $m_q(t)$ to hunt around a relatively flat segment of the input waveform; this phenomenon is also illustrated in Figure 3.24. Granular noise is analogous to quantization noise in a PCM system.

We thus see that there is a need to have a large step-size to accommodate a wide dynamic range, whereas a small step size is required for the accurate representation of relatively low-level signals. It is therefore clear that the choice of the optimum step size that minimizes the mean-square value of the quantization error in a linear delta modulator will be the result of a compromise between slope-overload distortion and granular noise. To satisfy such a requirement, we need to make the delta modulator "adaptive," in the sense that the step size is made to vary in accordance with the input signal; this issue is discussed further in a computer experiment presented in Section 3.16.

Delta-Sigma Modulation

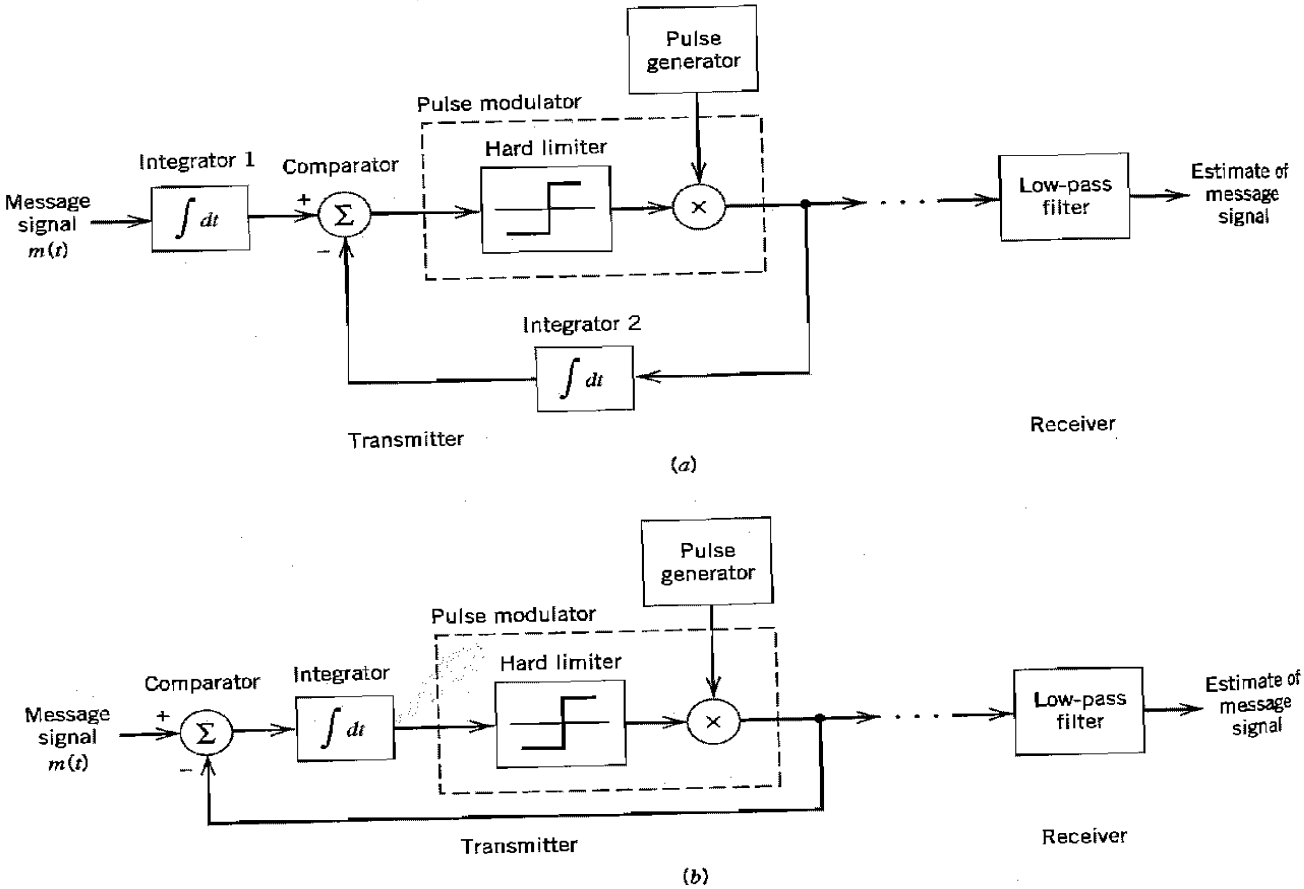
As mentioned earlier, the quantizer input in the conventional form of delta modulation may be viewed as an approximation to the *derivative* of the incoming message signal. This behavior leads to a drawback of delta modulation in that transmission disturbances such as noise result in an accumulative error in the demodulated signal. This drawback can be

overcome by *integrating* the message signal prior to delta modulation. The use of integration in the manner described here has also the following beneficial effects:

- The low-frequency content of the input signal is pre-emphasized.
- Correlation between adjacent samples of the delta modulator input is increased, which tends to improve overall system performance by reducing the variance of the error signal at the quantizer input.
- Design of the receiver is simplified.

A delta modulation scheme that incorporates integration at its input is called *delta-sigma modulation* (D-ΣM).[9] To be more precise, however, it should be called *sigma-delta modulation*, because the integration is in fact performed before the delta modulation. Nevertheless, the former terminology is the one commonly used in the literature.

Figure 3.25*a* shows the block diagram of a delta-sigma modulation system. In this diagram, the message signal $m(t)$ is defined in its continuous-time form, which means that the pulse modulator now consists of a hard-limiter followed by a multiplier; the latter component is also fed from an external pulse generator (clock) to produce a 1-bit encoded signal. The use of integration at the transmitter input clearly requires an inverse signal emphasis, namely, differentiation, at the receiver. The need for this differentiation is, however, eliminated because of its cancellation by integration in the conventional DM receiver.



Two equivalent versions of delta-sigma modulation system.

Thus the receiver of a delta-sigma modulation system consists simply of a low-pass filter,

Moreover, we note that integration is basically a linear operation. Accordingly, we may simplify the design of the transmitter by combining the two integrators 1 and 2 of Figure 3.25*a* into a single integrator placed after the comparator, as shown in Figure 3.25*b*. This latter form of the delta-sigma modulation system is not only simpler than that of Figure 3.25*a*, but it also provides an interesting interpretation of delta-sigma modulation as a "smoothed" version of 1-bit pulse-code modulation: The term *smoothness* refers to the fact that the comparator output is integrated prior to quantization, and the term *1-bit* merely restates that the quantizer consists of a hard-limiter with only two representation levels.

In delta modulation, simplicity of implementations of both the transmitter and receiver is attained by using a sampling rate far in excess of that needed for pulse-code modulation. The price paid for this benefit is a corresponding increase in the transmission and therefore channel bandwidth. There are, however, applications where channel bandwidth is at a premium, in which case we have the opposite requirement to that in delta modulation. Specifically, we may wish to trade increased system complexity for a reduced channel bandwidth. A signal-processing operation basic to the attainment of this latter design objective is prediction, the linear form of which is discussed next.

Sr. No.	Parameter	PCM	Delta modulation (DM)	Adaptive Delta Modulation (ADM)	Differential Pulse Code Modulation (DPCM)
1.	Number of bits	It can use 4, 8 or 16 bits per sample.	It uses only one bit for one sample.	Only one bit is used to encode one sample.	Bits can be more than one but are less than PCM.
2.	Levels, step size	The number of levels depend on number of bits. Level size is fixed.	Step size is fixed and cannot be varied.	According to the signal variation, step size varies (Adapted).	Fixed number of levels are used.
3.	Quantization error and distortion	Quantization error depends on number of levels used.	Slope overload distortion and granular noise is present.	Quantization error is present but other errors are absent.	Slope overload distortion and quantization noise is present.
4.	Bandwidth of transmission channel	Highest bandwidth is required since number of bits are high.	Lowest bandwidth is required.	Lowest bandwidth is required.	Bandwidth required is lower than PCM.
5	Feedback.	There is no feedback in transmitter or receiver.	Feedback exists in transmitter.	Feedback exists.	Feedback exists.
6	Complexity of notation	System is complex.	Simple.	Simple.	Simple.
7.	Signal to noise ratio	Good.	Poor.	Better than DM.	Fair.
8.	Area of applications	Audio and video Telephony.	Speech and images.	Speech and images.	Speech and video.

Comparison between PCM, Adaptive Delta Modulation and Differential Pulse Code Modulation

UNIT III BASE BAND PULSE TRANSMISSION

Maximum likelihood receiver structure – Matched filter receiver – Probability error of the Matched filter – Inter symbol interference – Nyquist criterion for distortion less baseband transmission – Correlative coding – Eye pattern.

Maximum Likelihood Receiver Structure

The decision-making criterion shown in step 2 of Figure 3.1 was described by Equation (3.7) as

$$z(T) \underset{H_2}{\overset{H_1}{\gtrless}} \gamma$$

A popular criterion for choosing the threshold level γ for the binary decision in Equation (3.7) is based on minimizing the probability of error. The computation for this *minimum error* value of $\gamma = \gamma_0$ starts with forming an inequality expression between the ratio of conditional probability density functions and the signal a priori probabilities. Since the conditional density function $p(z|s_i)$ is also called the *likelihood* of s_i, the formulation

$$\frac{p(z|s_1)}{p(z|s_2)} \underset{H_2}{\overset{H_1}{\gtrless}} \frac{P(s_2)}{P(s_1)} \tag{3.31}$$

is called the *likelihood ratio test.* (See Appendix B.) In this inequality, $P(s_1)$ and $P(s_2)$ are the a priori probabilities that $s_1(t)$ and $s_2(t)$, respectively, are transmitted, and H_1 and H_2 are the two possible hypotheses. The rule for minimizing the error probability states that we should choose hypothesis H_1 if the ratio of likelihoods is greater than the ratio of a priori probabilities, as shown in Equation (3.31).

It is shown in Section B.3.1, that if $P(s_1) = P(s_2)$, and if the likelihoods, $p(z|s_i)$ $(i = 1, 2)$, are symmetrical, the substitution of Equations (3.5) and (3.6) into (3.31) yields

$$z(T) \underset{H_2}{\overset{H_1}{\gtrless}} \frac{a_1 + a_2}{2} = \gamma_0 \tag{3.32}$$

where a_1 is the signal component of $z(T)$ when $s_1(t)$ is transmitted, and a_2 is the signal component of $z(T)$ when $s_2(t)$ is transmitted. The threshold level γ_0, represented by $(a_1 + a_2)/2$, is the *optimum threshold* for minimizing the probability of making an incorrect decision for this important special case. This strategy is known as the *minimum error criterion.*

For equally likely signals, the optimum threshold γ_0 passes through the intersection of the likelihood functions, as shown in Figure 3.2. Thus by following Equation (3.32), the decision stage effectively selects the hypothesis that corresponds to the signal with the *maximum likelihood*. For example, given an arbitrary detector output value $z_a(T)$, for which there is a nonzero likelihood that $z_a(T)$ belongs to either signal class $s_1(t)$ or $s_2(t)$, one can think of the likelihood test as a comparison of the likelihood values $p(z_a|s_1)$ and $p(z_a|s_2)$. The signal corresponding to the maximum pdf is chosen as the most likely to have been transmitted. In other words, the detector chooses $s_1(t)$ if

$$p(z_a|s_1) > p(z_a|s_2) \tag{3.33}$$

Otherwise, the detector chooses $s_2(t)$. A detector that minimizes the error probability (for the case where the signal classes are equally likely) is also known as a *maximum likelihood detector.*

Figure 3.2 illustrates that Equation (3.33) is just a "common sense" way to make a decision when there exists statistical knowledge of the classes. Given the detector output value $z_a(T)$, we see in Figure 3.2 that $z_a(T)$ intersects the likelihood of $s_1(t)$ at a value ℓ_1, and it intersects the likelihood of $s_2(t)$ at a value ℓ_2. What is the most reasonable decision for the detector to make? For this example, choosing class $s_1(t)$, which has the greater likelihood, is the most sensible choice. If this was an M-ary instead of a binary example, there would be a total of M likelihood functions representing the M signal classes to which a received signal might belong. The maximum likelihood decision would then be to choose the class that had the greatest likelihood of all M likelihoods. (Refer to Appendix B for a review of decision theory fundamentals.)

Error Probability

For the binary decision-making depicted in Figure 3.2, there are two ways errors can occur. An error e will occur when $s_1(t)$ is sent, and channel noise results in the receiver output signal $z(t)$ being less than γ_0. The probability of such an occurrence is

$$P(e|s_1) = P(H_2|s_1) = \int_{-\infty}^{\gamma_0} p(z|s_1)\,dz \tag{3.34}$$

This is illustrated by the shaded area to the left of γ_0 in Figure 3.2. Similarly, an error occurs when $s_2(t)$ is sent, and the channel noise results in $z(T)$ being greater than γ_0. The probability of this occurrence is

$$P(e|s_2) = P(H_1|s_2) = \int_{\gamma_0}^{\infty} p(z|s_2)\,dz \tag{3.35}$$

The probability of an error is the sum of the probabilities of all the ways that an error can occur. For the binary case, we can express the probability of bit error as

$$P_B = \sum_{i=1}^{2} P(e, s_i) = \sum_{i=1}^{2} P(e|s_i)\, P(s_i) \tag{3.36}$$

Combining Equations (3.34) to (3.36), we can write

$$P_B = P(e|s_1)P(s_1) + P(e|s_2)P(s_2) \tag{3.37a}$$

or equivalently,

$$P_B = P(H_2|s_1)P(s_1) + P(H_1|s_2)P(s_2) \tag{3.37b}$$

That is, given that signal $s_1(t)$ was transmitted, an error results if hypothesis H_2 is chosen; or given that signal $s_2(t)$ was transmitted, an error results if hypothesis H_1 is chosen. For the case where the a priori probabilities are equal [that is, $P(s_1) = P(s_2) = \frac{1}{2}$],

$$P_B = \tfrac{1}{2}\, P(H_2|s_1) + \tfrac{1}{2}\, P(H_1|s_2) \tag{3.38}$$

and because of the symmetry of the probability density functions,

$$P_B = P(H_2|s_1) = P(H_1|s_2) \tag{3.39}$$

The probability of a bit error, P_B, is numerically equal to the area under the "tail" of either likelihood function, $p(z|s_1)$ or $p(z|s_2)$, falling on the "incorrect" side of the threshold. We can therefore compute P_B by integrating $p(z|s_1)$ between the limits $-\infty$ and γ_0, or by integrating $p(z|s_2)$ between the limits γ_0 and ∞:

$$P_B = \int_{\gamma_0 = (a_1 + a_2)/2}^{\infty} p(z|s_2)\, dz \tag{3.40}$$

Here, $\gamma_0 = (a_1 + a_2)/2$ is the optimum threshold from Equation (3.32). Replacing the likelihood $p(z|s_2)$ with its Gaussian equivalent from Equation (3.6), we have

$$P_B = \int_{\gamma_0 = (a_1 + a_2)/2}^{\infty} \frac{1}{\sigma_0 \sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(\frac{z - a_2}{\sigma_0} \right)^2 \right] dz \tag{3.41}$$

where σ_0^2 is the variance of the noise out of the correlator.

Let $u = (z - a_2)/\sigma_0$. Then $\sigma_0\, du = dz$ and

$$P_B = \int_{u = (a_1 - a_2)/2\sigma_0}^{u = \infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2} \right) du = Q\left(\frac{a_1 - a_2}{2\sigma_0} \right) \tag{3.42}$$

where $Q(x)$, called the *complementary error function* or *co-error function*, is a commonly used symbol for the probability under the tail of the Gaussian pdf. It is defined as

$$Q(x) \approx \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{u^2}{2}\right) du \tag{3.43}$$

Note that the co-error function is defined in several ways (see Appendix B); however, all definitions are equally useful for determining probability of error in Gaussian noise. $Q(x)$ cannot be evaluated in closed form. It is presented in tabular form in Table B.1. Good approximations to $Q(x)$ by simpler functions can be found in Reference [5]. One such approximation, valid for $x > 3$, is

$$Q(x) \approx \frac{1}{x\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \tag{3.44}$$

We have optimized (in the sense of minimizing P_B) the threshold level γ, but have not optimized the receiving filter in block 1 of Figure 3.1. We next consider optimizing this filter by maximizing the argument of $Q(x)$ in Equation (3.42).

The Matched Filter

A matched filter is a linear filter designed to provide the maximum signal-to-noise power ratio at its output for a given transmitted symbol waveform. Consider that a known signal $s(t)$ plus AWGN $n(t)$ is the input to a linear, time-invariant (receiving) filter followed by a sampler, as shown in Figure 3.1. At time $t = T$, the sampler output $z(T)$ consists of a signal component a_i and a noise component n_0. The variance of the output noise (average noise power) is denoted by σ_0^2, so that the ratio of the instantaneous signal power to average noise power, $(S/N)_T$, at time $t = T$, out of the sampler in step 1, is

$$\left(\frac{S}{N}\right)_T = \frac{a_i^2}{\sigma_0^2} \tag{3.45}$$

We wish to find the filter transfer function $H_0(f)$ that *maximizes* Equation (3.45). We can express the signal $a_i(t)$ at the filter output in terms of the filter transfer function $H(f)$ (before optimization) and the Fourier transform of the input signal, as

$$a_i(t) = \int_{-\infty}^{\infty} H(f)S(f)e^{j2\pi ft}\, df$$

where $S(f)$ is the Fourier transform of the input signal, $s(t)$. If the two-sided power spectral density of the input noise is $N_0/2$ watts/hertz, then, using Equations (1.19) and (1.53), we can express the output noise power as

$$\sigma_0^2 = \frac{N_0}{2} \int_{-\infty}^{\infty} |H(f)|^2 \, df \tag{3.47}$$

We then combine Equations (3.45) to (3.47) to express $(S/N)_T$, as follows:

$$\left(\frac{S}{N}\right)_T = \frac{\left| \int_{-\infty}^{\infty} H(f) S(f) e^{j2\pi fT} \, df \right|^2}{N_0/2 \int_{-\infty}^{\infty} |H(f)|^2 \, df} \tag{3.48}$$

We next find that value of $H(f) = H_0(f)$ for which the maximum $(S/N)_T$ is achieved, by using *Schwarz's inequality*. One form of the inequality can be stated as

$$\left| \int_{-\infty}^{\infty} f_1(x) f_2(x) \, dx \right|^2 \leq \int_{-\infty}^{\infty} |f_1(x)|^2 \, dx \int_{-\infty}^{\infty} |f_2(x)|^2 \, dx \tag{3.49}$$

The equality holds if $f_1(x) = kf_2^*(x)$, where k is an arbitrary constant and * indicates complex conjugate. If we identify $H(f)$ with $f_1(x)$ and $S(f)\, e^{j2\pi fT}$ with $f_2(x)$, we can write

$$\left| \int_{-\infty}^{\infty} H(f) S(f) e^{j2\pi fT} df \right|^2 \leq \int_{-\infty}^{\infty} |H(f)|^2 \, df \int_{-\infty}^{\infty} |S(f)|^2 \, df \tag{3.50}$$

Substituting into Equation (3.48) yields

$$\left(\frac{S}{N}\right)_T \leq \frac{2}{N_0} \int_{-\infty}^{\infty} |S(f)|^2 \, df \tag{3.51}$$

or

$$\max \left(\frac{S}{N}\right)_T = \frac{2E}{N_0} \tag{3.52}$$

where the energy E of the input signal $s(t)$ is

$$E = \int_{-\infty}^{\infty} |S(f)|^2 \, df \tag{3.53}$$

Thus, the maximum output $(S/N)_T$ depends on the input *signal energy* and the power spectral density of the noise, *not on the particular shape* of the waveform that is used.

The equality in Equation (3.52) holds only if the optimum filter transfer function $H_0(f)$ is employed, such that

where $S(f)$ is the Fourier transform of the input signal, $s(t)$. If the two-sided power spectral density of the input noise is $N_0/2$ watts/hertz, then, using Equations (1.19) and (1.53), we can express the output noise power as

$$\sigma_0^2 = \frac{N_0}{2} \int_{-\infty}^{\infty} |H(f)|^2 \, df \tag{3.47}$$

We then combine Equations (3.45) to (3.47) to express $(S/N)_T$, as follows:

$$\left(\frac{S}{N}\right)_T = \frac{\left| \int_{-\infty}^{\infty} H(f) S(f) e^{j2\pi f T} \, df \right|^2}{N_0/2 \int_{-\infty}^{\infty} |H(f)|^2 \, df} \tag{3.48}$$

We next find that value of $H(f) = H_0(f)$ for which the maximum $(S/N)_T$ is achieved, by using *Schwarz's inequality*. One form of the inequality can be stated as

$$\left| \int_{-\infty}^{\infty} f_1(x) f_2(x) \, dx \right|^2 \leq \int_{-\infty}^{\infty} |f_1(x)|^2 \, dx \int_{-\infty}^{\infty} |f_2(x)|^2 \, dx \tag{3.49}$$

The equality holds if $f_1(x) = k f_2^*(x)$, where k is an arbitrary constant and * indicates complex conjugate. If we identify $H(f)$ with $f_1(x)$ and $S(f)\, e^{j2\pi fT}$ with $f_2(x)$, we can write

$$H(f) = H_0(f) = kS^*(f) e^{-j2\pi f T} \tag{3.54}$$

or

$$h(t) = \mathcal{F}^{-1}\{kS^*(f) e^{-j2\pi f T}\} \tag{3.55}$$

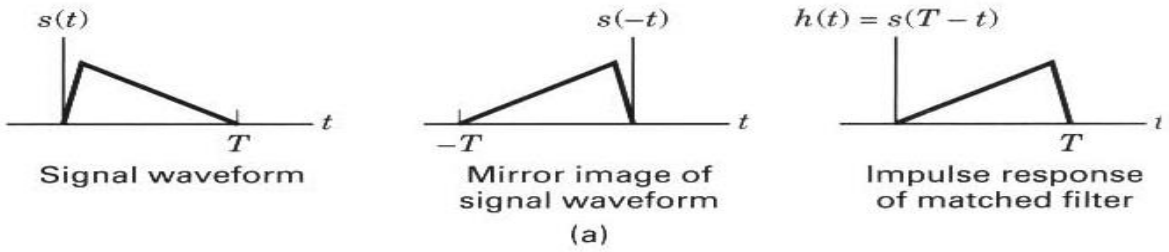
Since $s(t)$ is a real-valued signal, we can write, from Equations (A.29) and (A.31),

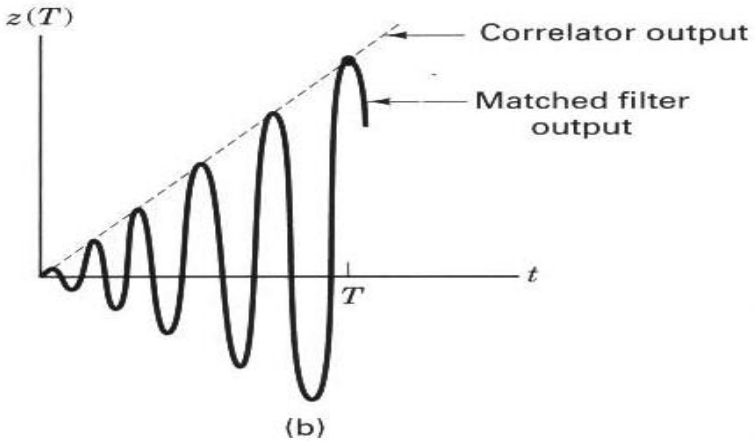
$$h(t) = \begin{cases} ks(T - t) & 0 \leq t \leq T \\ 0 & \text{elsewhere} \end{cases} \tag{3.56}$$

Thus, the impulse response of a filter that produces the maximum output signal-to-noise ratio is the mirror image of the message signal $s(t)$, *delayed* by the symbol time duration T. Note that the delay of T seconds makes Equation (3.56) *causal;* that is, the delay of T seconds makes $h(t)$ a function of positive time in the interval $0 \leq t \leq T$. Without the delay of T seconds, the response $s\ (-t)$ is unrealizable because it describes a response as a function of negative time.

Correlation Realization of the Matched Filter

Equation (3.56) and Figure 3.7a illustrate the matched filter's basic property: The impulse response of the filter is a delayed version of the mirror image (rotated on the $t = 0$ axis) of the signal waveform. Therefore, if the signal waveform is $s(t)$, its mirror image is $s(-t)$, and the mirror image delayed by T seconds is $s(T - t)$. The output $z(t)$ of a causal filter can be described in the time domain as the convolution of a received input waveform $r(t)$ with the impulse response of the filter (see Section A.5):





Correlator and matched filter. (a) Matched filter characteristic. (b) Comparison of correlator and matched filter outputs.

output $z_i(T)$ is the signal that matches $r(t)$ better than all the other $s_j(t)$, $j \neq i$. We will subsequently use this correlation characteristic for the optimum detection of signals.

Error Rate Due to Noise

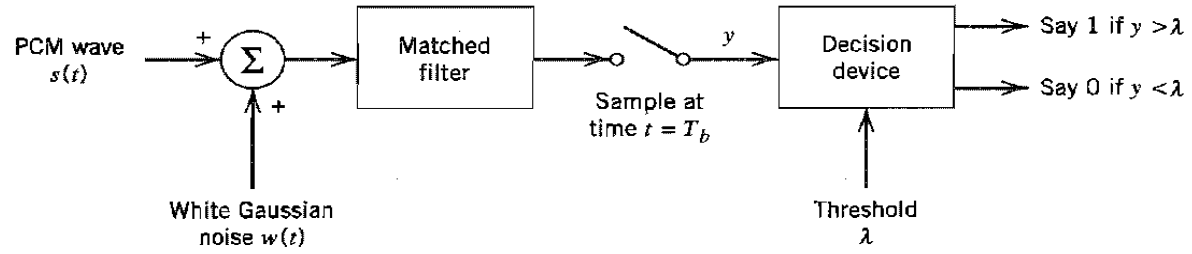
In Section 3.8 we presented a qualitative discussion of the effect of channel noise on the performance of a binary PCM system. Now that we are equipped with the matched filter as the optimum detector of a known pulse in additive white noise, we are ready to derive a formula for the error rate in such a system due to noise.

To proceed with the analysis, consider a binary PCM system based on *polar nonreturn-to-zero (NRZ) signaling*. In this form of signaling, symbols 1 and 0 are represented by positive and negative rectangular pulses of equal amplitude and equal duration. The channel noise is modeled as *additive white Gaussian noise* $w(t)$ of zero mean and power spectral density $N_0/2$; the Gaussian assumption is needed for later calculations. In the signaling interval $0 \leq t \leq T_b$, the received signal is thus written as follows:

$$x(t) = \begin{cases} +A + w(t), & \text{symbol 1 was sent} \\ -A + w(t), & \text{symbol 0 was sent} \end{cases} \tag{4.21}$$

where T_b is the *bit duration*, and A is the *transmitted pulse amplitude*. It is assumed that the receiver has acquired knowledge of the starting and ending times of each transmitted pulse; in other words, the receiver has prior knowledge of the pulse shape, but not its polarity. Given the noisy signal $x(t)$, the receiver is required to make a decision in each signaling interval as to whether the transmitted symbol is a 1 or a 0.

The structure of the receiver used to perform this decision-making process is shown in Figure . It consists of a matched filter followed by a sampler, and then finally a



Receiver for baseband transmission of binary-encoded PCM wave using polar NRZ signaling.

decision device. The filter is matched to a rectangular pulse of amplitude A and duration T_b, exploiting the bit-timing information available to the receiver. The resulting matched filter output is sampled at the end of each signaling interval. The presence of channel noise $w(t)$ adds randomness to the matched filter output.

Let y denote the sample value obtained at the end of a signaling interval. The sample value y is compared to a preset *threshold* λ in the decision device. If the threshold is exceeded, the receiver makes a decision in favor of symbol 1; if not, a decision is made in favor of symbol 0. We adopt the convention that when the sample value y is exactly equal to the threshold λ, the receiver just makes a guess as to which symbol was transmitted; such a decision is the same as that obtained by flipping a fair coin, the outcome of which will not alter the average probability of error.

There are two possible kinds of error to be considered:

1. Symbol 1 is chosen when a 0 was actually transmitted; we refer to this error as an *error of the first kind.*
2. Symbol 0 is chosen when a 1 was actually transmitted; we refer to this error as an *error of the second kind.*

To determine the average probability of error, we consider these two situations separately.

Suppose that symbol 0 was sent. Then, according to Equation (4.21), the received signal is

$$x(t) = -A + w(t), \qquad 0 \le t \le T_b \tag{4.22}$$

Correspondingly, the matched filter output, sampled at time $t = T_b$, is given by (in light of Example 4.1 with kAT_b set equal to unity for convenience of presentation)

$$\begin{aligned} y &= \int_0^{T_b} x(t)\, dt \\ &= -A + \frac{1}{T_b}\int_0^{T_b} w(t)\, dt \end{aligned} \tag{4.23}$$

which represents the sample value of a random variable Y. By virtue of the fact that the noise $w(t)$ is white and Gaussian, we may characterize the random variable Y as follows:

- The random variable Y is Gaussian distributed with a mean of $-A$.
- The variance of the random variable Y is

$$\begin{aligned}\sigma_Y^2 &= E[(Y + A)^2] \\ &= \frac{1}{T_b^2} E\left[\int_0^{T_b}\int_0^{T_b} w(t)w(u)\, dt\, du\right] \\ &= \frac{1}{T_b^2}\int_0^{T_b}\int_0^{T_b} E[w(t)w(u)]\, dt\, du \\ &= \frac{1}{T_b^2}\int_0^{T_b}\int_0^{T_b} R_W(t, u)\, dt\, du\end{aligned} \tag{4.24}$$

where $R_W(t, u)$ is the autocorrelation function of the white noise $w(t)$. Since $w(t)$ is white with a power spectral density $N_0/2$, we have

$$R_W(t, u) = \frac{N_0}{2}\delta(t - u) \tag{4.25}$$

where $\delta(t - u)$ is a time-shifted delta function. Hence, substituting Equation (4.25) into (4.24) yields

$$\begin{aligned}\sigma_Y^2 &= \frac{1}{T_b^2}\int_0^{T_b}\int_0^{T_b}\frac{N_0}{2}\delta(t - u)\, dt\, du \\ &= \frac{N_0}{2T_b}\end{aligned} \tag{4.26}$$

where we have used the sifting property of the delta function and the fact that its area is unity. The conditional probability density function of the random variable Y, given that symbol 0 was sent, is therefore

$$f_Y(y|0) = \frac{1}{\sqrt{\pi N_0/T_b}}\exp\left(-\frac{(y + A)^2}{N_0/T_b}\right) \tag{4.27}$$

This function is plotted in Figure 4.5(*a*). Let p_{10} denote the *conditional probability of error, given that symbol 0 was sent*. This probability is defined by the shaded area under the curve of $f_Y(y|0)$ from the threshold λ to infinity, which corresponds to the range of values assumed by y for a decision in favor of symbol 1. In the absence of noise, the matched filter output y sampled at time $t = T_b$ is equal to $-A$. When noise is present, y occasionally assumes a value greater than λ, in which case an error is made. The probability of this error, conditional on sending symbol 0, is defined by

$$\begin{aligned}p_{10} &= P(y > \lambda \,|\, \text{symbol 0 was sent}) \\ &= \int_\lambda^\infty f_Y(y|0)\, dy \\ &= \frac{1}{\sqrt{\pi N_0/T_b}}\int_\lambda^\infty \exp\left(-\frac{(y + A)^2}{N_0/T_b}\right) dy\end{aligned} \tag{4.28}$$

To reformulate the conditional probability of error p_{10} in terms of the complementary error function, we first define a new variable

$$z = \frac{y + A}{\sqrt{N_0/T_b}}$$

Accordingly, we may rewrite Equation (4.28) in the compact form

$$\begin{aligned} p_{10} &= \frac{1}{\sqrt{\pi}} \int_{(A+\lambda)/\sqrt{N_0/T_b}}^{\infty} \exp(-z^2)\, dz \\ &= \frac{1}{2} \operatorname{erfc}\left(\frac{A + \lambda}{\sqrt{N_0/T_b}}\right) \end{aligned} \tag{4.31}$$

Assume next that symbol 1 was transmitted. This time the Gaussian random variable Y represented by the sample value y of the matched filter output has a mean $+A$ and variance $N_0/2T_b$. Note that, compared to the situation when symbol 0 was sent, the mean of the random variable Y has changed, but its variance is exactly the same as before. The conditional probability density function of Y, given that symbol 1 was sent, is therefore

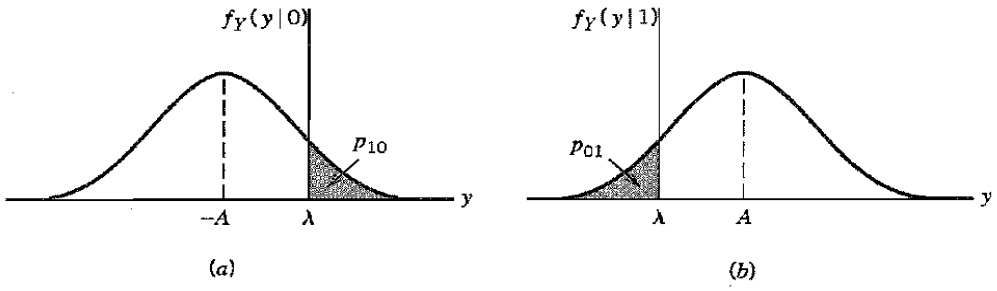
$$f_Y(y \mid 1) = \frac{1}{\sqrt{\pi N_0/T_b}} \exp\left(-\frac{(y - A)^2}{N_0/T_b}\right) \tag{4.32}$$

At this point in the discussion we digress briefly and introduce the definition of the so-called *complementary error function*:[3]

$$\operatorname{erfc}(u) = \frac{2}{\sqrt{\pi}} \int_u^{\infty} \exp(-z^2)\, dz \tag{4.29}$$

which is closely related to the Gaussian distribution. For large positive values of u, we have the following *upper bound* on the complementary error function:

$$\operatorname{erfc}(u) < \frac{\exp(-u^2)}{\sqrt{\pi u}} \tag{4.30}$$



Noise analysis of PCM system. (*a*) Probability density function of random variable Y at matched filter output when 0 is transmitted. (*b*) Probability density function of Y when 1 is transmitted.

which is plotted in Figure 4.5*b*. Let p_{01} denote the *conditional probability of error, given that symbol 1 was sent.* This probability is defined by the shaded area under the curve of $f_Y(y|1)$ extending from $-\infty$ to the threshold λ, which corresponds to the range of values assumed by y for a decision in favor of symbol 0. In the absence of noise, the matched filter output y sampled at time $t = T_b$ is equal to $+A$. When noise is present, y occasionally assumes a value less than λ, and an error is then made. The probability of this error, conditional on sending symbol 1, is defined by

$$\begin{aligned} p_{01} &= P(y < \lambda \,|\, \text{symbol 1 was sent}) \\ &= \int_{-\infty}^{\lambda} f_Y(y|1)\, dy \\ &= \frac{1}{\sqrt{\pi N_0/T_b}} \int_{-\infty}^{\lambda} \exp\left(-\frac{(y-A)^2}{N_0/T_b}\right) dy \end{aligned} \tag{4.33}$$

To express p_{01} in terms of the complementary error function, this time we define a new variable

$$z = \frac{A - y}{\sqrt{N_0/T_b}}$$

Accordingly, we may reformulate Equation (4.33) in the compact form

$$\begin{aligned} p_{01} &= \frac{1}{\sqrt{\pi}} \int_{(A-\lambda)/\sqrt{N_0/T_b}}^{\infty} \exp(-z^2)\, dz \\ &= \frac{1}{2} \operatorname{erfc}\left(\frac{A-\lambda}{\sqrt{N_0/T_b}}\right) \end{aligned} \tag{4.34}$$

Having determined the conditional probabilities of error, p_{10} and p_{01}, our next task is to derive the formula for the *average probability of symbol error*, denoted by P_e. Here we note that these two possible kinds of error are mutually exclusive events in that if the receiver, at a particular sampling instant, chooses symbol 1, then symbol 0 is excluded from appearing, and vice versa. Let p_0 and p_1 denote the *a priori* probabilities of transmitting symbols 0 and 1, respectively. Hence, the *average probability of symbol error* P_e in the receiver is given by

$$\begin{aligned} P_e &= p_0 p_{10} + p_1 p_{01} \\ &= \frac{p_0}{2} \operatorname{erfc}\left(\frac{A+\lambda}{\sqrt{N_0/T_b}}\right) + \frac{p_1}{2} \operatorname{erfc}\left(\frac{A-\lambda}{\sqrt{N_0/T_b}}\right) \end{aligned} \tag{4.35}$$

From Equation (4.35) we see that P_e is in fact a function of the threshold λ, which immediately suggests the need for formulating an *optimum threshold* that minimizes P_e. For this optimization we use *Leibniz's rule.*

Consider the integral

$$\int_{a(u)}^{b(u)} f(z, u)\, dz$$

Leibniz's rule states that the derivative of this integral with respect to u is

$$\frac{d}{du}\int_{a(u)}^{b(u)} f(z, u)\, dz = f(b(u), u)\frac{db(u)}{du} - f(a(u), u)\frac{d(a(u)}{du} + \int_{a(u)}^{b(u)} \frac{\partial f(z, u)}{\partial u}\, dz$$

For the problem at hand, we note from the definition of the complementary error function in Equation (4.29) that

$$f(z, u) = \frac{2}{\sqrt{\pi}} \exp(-z^2)$$
$$a(u) = u$$
$$b(u) = \infty$$

The application of Leibniz's rule to the complementary error function thus yields

$$\frac{d}{du}\operatorname{erfc}(u) = -\frac{1}{\sqrt{\pi}} \exp(-u^2) \tag{4.36}$$

Hence, differentiating Equation (4.35) with respect to λ by making use of the formula in Equation (4.36), then setting the result equal to zero and simplifying terms, we obtain the optimum threshold as

$$\lambda_{\text{opt}} = \frac{N_0}{4AT_b} \log\left(\frac{p_0}{p_1}\right) \tag{4.37}$$

For the special case when symbols 1 and 0 are equiprobable, we have

$$p_1 = p_0 = \frac{1}{2}$$

in which case Equation (4.37) reduces to

$$\lambda_{\text{opt}} = 0$$

This result is intuitively satisfying as it states that, for the transmission of equiprobable binary symbols, we should choose the threshold at the midpoint between the pulse heights $-A$ and $+A$ representing the two symbols 0 and 1. Note that for this special case we also have

$$p_{01} = p_{10}$$

A channel for which the conditional probabilities of error p_{01} and p_{10} are equal is said to be *binary symmetric*. Correspondingly, the average probability of symbol error in Equation (4.35) reduces to

$$P_e = \frac{1}{2}\operatorname{erfc}\left(\frac{A}{\sqrt{N_0/T_b}}\right) \tag{4.38}$$

Now the *transmitted signal energy per bit* is defined by

$$E_b = A^2 T_b \tag{4.39}$$

Accordingly, we may finally formulate the average probability of symbol error for the receiver in Figure 4.4 as

$$P_e = \frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right) \tag{4.40}$$

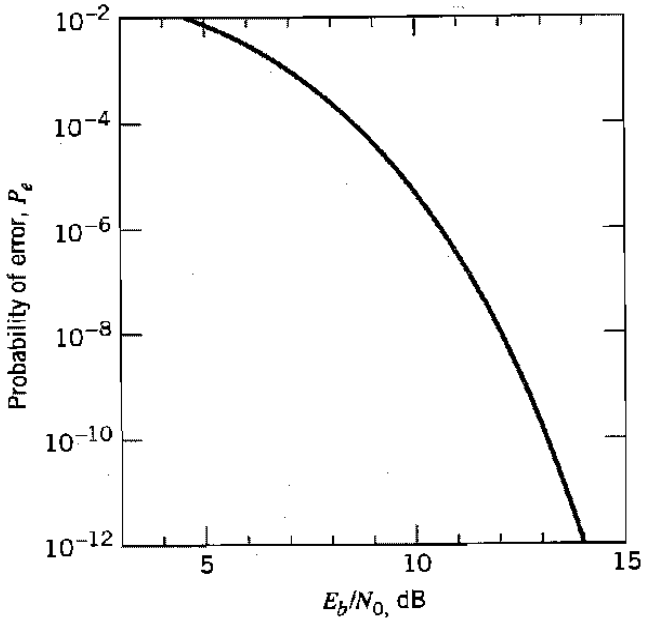
which shows that *the average probability of symbol error in a binary symmetric channel depends solely on E_b/N_0, the ratio of the transmitted signal energy per bit to the noise spectral density.*

Using the upper bound of Equation (4.30) on the complementary error function, we may correspondingly bound the average probability of symbol error for the PCM receiver as

$$P_e < \frac{\exp(-E_b/N_0)}{2\sqrt{\pi E_b/N_0}} \tag{4.41}$$

The PCM receiver of Figure 4.4 therefore exhibits an *exponential* improvement in the average probability of symbol error with increase in E_b/N_0.

This important result is further illustrated in Figure 4.6 where the average probability of symbol error P_e is plotted versus the dimensionless ratio E_b/N_0. In particular, we see that P_e decreases very rapidly as the ratio E_b/N_0 is increased, so that eventually a very "small increase" in transmitted signal energy will make the reception of binary pulses almost error free, as discussed previously in Section 3.8. Note, however, that in practical terms the increase in signal energy has to be viewed in the context of the bias



Probability of error in a PCM receiver.

Intersymbol Interference

The next source of bit errors in a baseband-pulse transmission system that we wish to study is intersymbol interference (ISI), which arises when the communication channel is *dispersive*. First of all, however, we need to address a key question: Given a pulse shape of interest, how do we use it to transmit data in M-ary form? The answer lies in the use of *discrete pulse modulation*, in which the amplitude, duration, or position of the transmitted pulses is varied in a discrete manner in accordance with the given data stream. However, for the baseband transmission of digital data, the use of *discrete pulse-amplitude modulation* (PAM) is one of the most efficient schemes in terms of power and bandwidth utilization. Accordingly, we confine our attention to discrete PAM systems. We begin the study by first considering the case of binary data; later in the chapter, we consider the more general case of M-ary data.

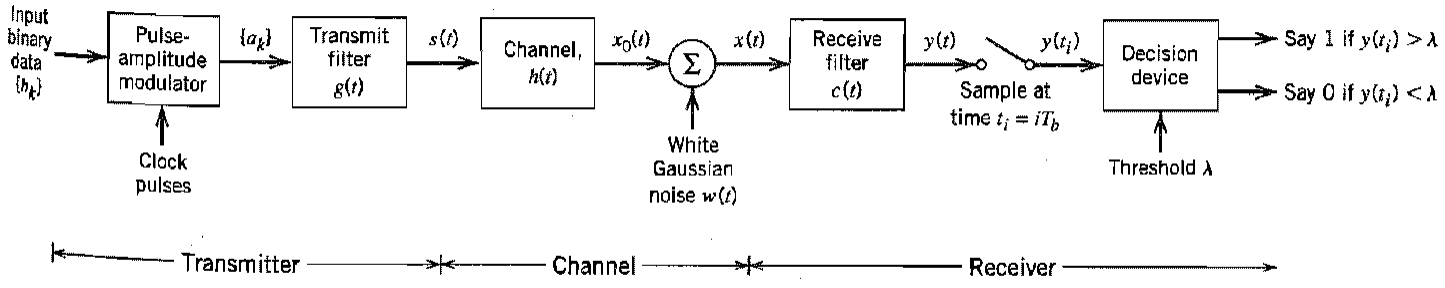
Consider then a *baseband binary PAM system*, a generic form of which is shown in Figure 4.7. The incoming binary sequence $\{b_k\}$ consists of symbols 1 and 0, each of duration T_b. The *pulse-amplitude modulator* modifies this binary sequence into a new sequence of short pulses (approximating a unit impulse), whose amplitude a_k is represented in the polar form

$$a_k = \begin{cases} +1 & \text{if symbol } b_k \text{ is 1} \\ -1 & \text{if symbol } b_k \text{ is 0} \end{cases} \tag{4.42}$$

The sequence of short pulses so produced is applied to a *transmit filter* of impulse response $g(t)$, producing the transmitted signal

$$s(t) = \sum_k a_k g(t - kT_b) \tag{4.43}$$

The signal $s(t)$ is modified as a result of transmission through the *channel* of impulse response $h(t)$. In addition, the channel adds random noise to the signal at the receiver input. The noisy signal $x(t)$ is then passed through a *receive filter* of impulse response $c(t)$. The resulting filter output $y(t)$ is sampled *synchronously* with the transmitter, with the sampling instants being determined by a *clock* or *timing signal* that is usually extracted from the receive filter output. Finally, the sequence of samples thus obtained is used to reconstruct the original data sequence by means of a *decision device*. Specifically, the amplitude of each sample is compared to a *threshold* λ. If the threshold λ is exceeded, a decision is made in favor of symbol 1. If the threshold λ is not exceeded, a decision is made in favor of symbol 0. If the sample amplitude equals the threshold exactly, the flip of a



Baseband binary data transmission system.

fair coin will determine which symbol was transmitted (i.e., the receiver simply makes a random guess).

The receive filter output is written as

$$y(t) = \mu \sum_k a_k p(t - kT_b) + n(t) \tag{4.44}$$

where μ is a scaling factor, and the pulse $p(t)$ is to be defined. To be precise, an arbitrary time delay t_0 should be included in the argument of the pulse $p(t - kT_b)$ in Equation (4.44) to represent the effect of transmission delay through the system. To simplify the exposition, we have put this delay equal to zero in Equation (4.44) without loss of generality.

The scaled pulse $\mu p(t)$ is obtained by a double convolution involving the impulse response $g(t)$ of the transmit filter, the impulse response $h(t)$ of the channel, and the impulse response $c(t)$ of the receive filter, as shown by

$$\mu p(t) = g(t) \bigstar h(t) \bigstar c(t) \tag{4.45}$$

where the star denotes convolution. We assume that the pulse $p(t)$ is *normalized* by setting

$$p(0) = 1 \tag{4.46}$$

which justifies the use of μ as a scaling factor to account for amplitude changes incurred in the course of signal transmission through the system.

Since convolution in the time domain is transformed into multiplication in the frequency domain, we may use the Fourier transform to change Equation (4.45) into the equivalent form

$$\mu P(f) = G(f)H(f)C(f) \tag{4.47}$$

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

Finally, the term $n(t)$ in Equation (4.44) is the noise produced at the output of the receive filter due to the channel noise $w(t)$. It is customary to model $w(t)$ as a white Gaussian noise of zero mean.

The receive filter output $y(t)$ is sampled at time $t_i = iT_b$ (with i taking on integer values), yielding [in light of Equation (4.46)]

$$\begin{aligned} y(t_i) &= \mu \sum_{k=-\infty}^{\infty} a_k p[(i - k)T_b] + n(t_i) \\ &= \mu a_i + \mu \sum_{\substack{k=-\infty \\ k \neq i}}^{\infty} a_k p[(i - k)T_b] + n(t_i) \end{aligned} \tag{4.48}$$

In Equation (4.48), the first term μa_i represents the contribution of the ith transmitted bit. The second term represents the residual effect of all other transmitted bits on the decoding of the ith bit; this residual effect due to the occurrence of pulses before and after the sampling instant t_i is called intersymbol interference (ISI). The last term $n(t_i)$ represents the noise sample at time t_i.

In the absence of both ISI and noise, we observe from Equation (4.48) that

$$y(t_i) = \mu a_i$$

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

When the signal-to-noise ratio is high, as is the case in a telephone system, for example, the operation of the system is largely limited by ISI rather than noise; in other words, we may ignore $n(t_i)$. In the next couple of sections, we assume that this condition holds so that we may focus our attention on ISI and the techniques for its control. In particular, the issue we wish to consider is to determine the pulse waveform $p(t)$ for which the ISI is completely eliminated.

Nyquist's Criterion for Distortionless Baseband Binary Transmission

Typically, the frequency response of the channel and the transmitted pulse shape are specified, and the problem is to determine the frequency responses of the transmit and receive filters so as to reconstruct the original binary data sequence $\{b_k\}$. The receiver does this by *extracting* and then *decoding* the corresponding sequence of coefficients, $\{a_k\}$, from the output $y(t)$. The *extraction* involves sampling the output $y(t)$ at time $t = iT_b$. The *decoding* requires that the weighted pulse contribution $a_k p(iT_b - kT_b)$ for $k = i$ be *free* from ISI due to the overlapping tails of all other weighted pulse contributions represented by $k \neq i$. This, in turn, requires that we *control* the overall pulse $p(t)$, as shown by

$$p(iT_b - kT_b) = \begin{cases} 1, & i = k \\ 0, & i \neq k \end{cases} \tag{4.49}$$

where $p(0) = 1$, by normalization. If $p(t)$ satisfies the conditions of Equation (4.49), the receiver output $y(t_i)$ given in Equation (4.48) simplifies to (ignoring the noise term)

$$y(t_i) = \mu a_i \qquad \text{for all } i$$

which implies zero intersymbol interference. Hence, the two conditions of Equation (4.49) ensure *perfect reception in the absence of noise.*

From a design point of view, it is informative to transform the conditions of Equation (4.49) into the frequency domain. Consider then the sequence of samples $\{p(nT_b)\}$, where $n = 0, \pm 1, \pm 2, \cdots$. From the discussion presented in Chapter 3 on the sampling process, we recall that sampling in the time domain produces periodicity in the frequency domain. In particular, we may write

$$P_\delta(f) = R_b \sum_{n=-\infty}^{\infty} P(f - nR_b) \tag{4.50}$$

where $R_b = 1/T_b$ is the *bit rate* in bits per second (b/s); $P_\delta(f)$ is the Fourier transform of an infinite periodic sequence of delta functions of period T_b, whose individual areas are weighted by the respective sample values of $p(t)$. That is, $P_\delta(f)$ is given by

$$P_\delta(f) = \int_{-\infty}^{\infty} \sum_{m=-\infty}^{\infty} [p(mT_b)\ \delta(t - mT_b)] \exp(-j2\pi ft)\ dt \tag{4.51}$$

Let the integer $m = i - k$. Then, $i = k$ corresponds to $m = 0$, and likewise $i \neq k$ corresponds to $m \neq 0$. Accordingly, imposing the conditions of Equation (4.49) on the sample values of $p(t)$ in the integral of Equation (4.51), we get

$$\begin{aligned} P_\delta(f) &= \int_{-\infty}^{\infty} p(0)\ \delta(t) \exp(-j2\pi ft)\ dt \\ &= p(0) \end{aligned} \tag{4.52}$$

where we have made use of the sifting property of the delta function. Since from Equation (4.46) we have $p(0) = 1$, it follows from Equations (4.50) and (4.52) that the condition for zero intersymbol interference is satisfied if

$$\sum_{n=-\infty}^{\infty} P(f - nR_b) = T_b \tag{4.53}$$

We may now state the *Nyquist criterion*[4] *for distortionless baseband transmission* in the absence of noise: *The frequency function $P(f)$ eliminates intersymbol interference for samples taken at intervals T_b provided that it satisfies Equation (4.53).* Note that $P(f)$ refers to the overall system, incorporating the transmit filter, the channel, and the receive filter in accordance with Equation (4.47).

IDEAL NYQUIST CHANNEL

The simplest way of satisfying Equation (4.53) is to specify the frequency function $P(f)$ to be in the form of a *rectangular function*, as shown by

$$\begin{aligned} P(f) &= \begin{cases} \dfrac{1}{2W}, & -W < f < W \\ 0, & |f| > W \end{cases} \\ &= \frac{1}{2W}\,\mathrm{rect}\left(\frac{f}{2W}\right) \end{aligned} \tag{4.54}$$

where $\mathrm{rect}(f)$ stands for a *rectangular function* of unit amplitude and unit support centered on $f = 0$, and the overall system bandwidth W is defined by

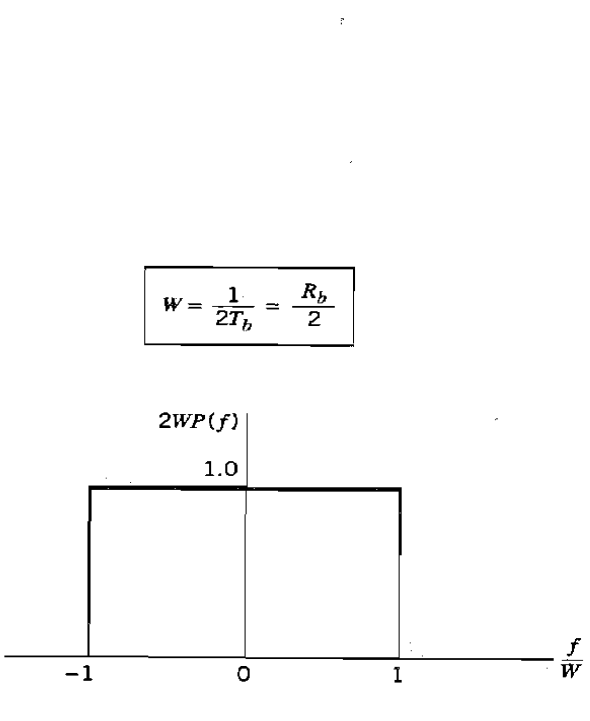
$$W = \frac{R_b}{2} = \frac{1}{2T_b} \tag{4.55}$$

According to the solution described by Equations (4.54) and (4.55), no frequencies of absolute value exceeding half the bit rate are needed. Hence, from Fourier-transform pair 2 of Table A6.3 we find that a signal waveform that produces zero intersymbol interference is defined by the *sinc function*:

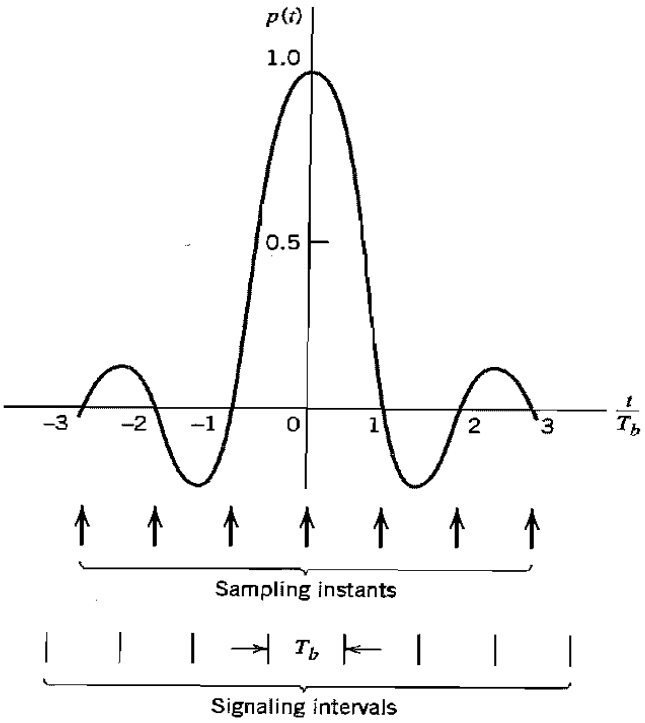
$$\begin{aligned} p(t) &= \frac{\sin(2\pi Wt)}{2\pi Wt} \\ &= \mathrm{sinc}(2Wt) \end{aligned} \tag{4.56}$$

The special value of the bit rate $R_b = 2W$ is called the *Nyquist rate*, and W is itself called the *Nyquist bandwidth*. Correspondingly, the ideal baseband pulse transmission system described by Equation (4.54) in the frequency domain or, equivalently, Equation (4.56) in the time domain, is called the *ideal Nyquist channel*.

Figures 4.8*a* and 4.8*b* show plots of $P(f)$ and $p(t)$, respectively. In Figure 4.8*a*, the normalized form of the frequency function $P(f)$ is plotted for positive and negative frequencies. In Figure 4.8*b*, we have also included the signaling intervals and the corresponding centered sampling instants. The function $p(t)$ can be regarded as the impulse response of an ideal low-pass filter with passband magnitude response $1/2W$ and bandwidth W. The function $p(t)$ has its peak value at the origin and goes through zero at integer multiples of the bit duration T_b. It is apparent that if the received waveform $y(t)$ is sampled at the



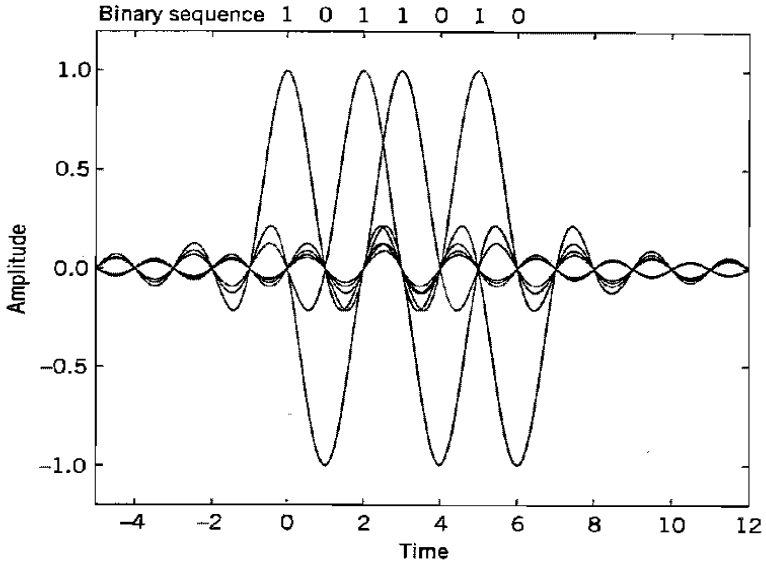
Ideal magnitude response.



Ideal basic pulse shape.

instants of time $t = 0, \pm T_b, \pm 2T_b, \cdots$, then the pulses defined by $\mu p(t - iT_b)$ with arbitrary amplitude μ and index $i = 0, \pm 1, \pm 2, \cdots$, will not interfere with each other. This condition is illustrated in Figure 4.9 for the binary sequence 1011010.

Although the use of the ideal Nyquist channel does indeed achieve economy in bandwidth in that it solves the problem of zero intersymbol interference with the minimum



A series of sinc pulses corresponding to the sequence 1011010.

bandwidth possible, there are two practical difficulties that make it an undesirable objective for system design:

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

To evaluate the effect of this *timing error*, consider the sample of $y(t)$ at $t = \Delta t$, where Δt is the timing error. To simplify the exposition, we may put the correct sampling time t_i equal to zero. In the absence of noise, we thus have (from Equation (4.48))

$$\begin{aligned} y(\Delta t) &= \mu \sum_k a_k p(\Delta t - kT_b) \\ &= \mu \sum_k a_k \frac{\sin[2\pi W(\Delta t - kT_b)]}{2\pi W(\Delta t - kT_b)} \end{aligned} \tag{4.57}$$

Since $2WT_b = 1$, by definition, we may rewrite Equation (4.57) as

$$y(\Delta t) = \mu a_0 \operatorname{sinc}(2W\,\Delta t) + \frac{\mu \sin(2\pi W\,\Delta t)}{\pi} \sum_{\substack{k \\ k \neq 0}} \frac{(-1)^k a_k}{(2W\,\Delta t - k)} \tag{4.58}$$

The first term on the right-hand side of Equation (4.58) defines the desired symbol, whereas the remaining series represents the intersymbol interference caused by the timing error Δt in sampling the output $y(t)$. Unfortunately, it is possible for this series to diverge, thereby causing erroneous decisions in the receiver.

Raised Cosine Spectrum

We may overcome the practical difficulties encountered with the ideal Nyquist channel by extending the bandwidth from the minimum value $W = R_b/2$ to an adjustable value between W and $2W$. We now specify the overall frequency response $P(f)$ to satisfy a condition more elaborate than that for the ideal Nyquist channel; specifically, we retain three terms of Equation (4.53) and restrict the frequency band of interest to $[-W, W]$, as shown by

$$P(f) + P(f - 2W) + P(f + 2W) = \frac{1}{2W}, \qquad -W \leq f \leq W \tag{4.59}$$

We may devise several band-limited functions that satisfy Equation (4.59). A particular form of $P(f)$ that embodies many desirable features is provided by a *raised cosine spectrum*. This frequency response consists of a *flat* portion and a *rolloff* portion that has a sinusoidal form, as follows:

$$P(f) = \begin{cases} \dfrac{1}{2W}, & 0 \leq |f| < f_1 \\ \dfrac{1}{4W}\left\{1 - \sin\left[\dfrac{\pi(|f| - W)}{2W - 2f_1}\right]\right\}, & f_1 \leq |f| < 2W - f_1 \\ 0, & |f| \geq 2W - f_1 \end{cases} \tag{4.60}$$

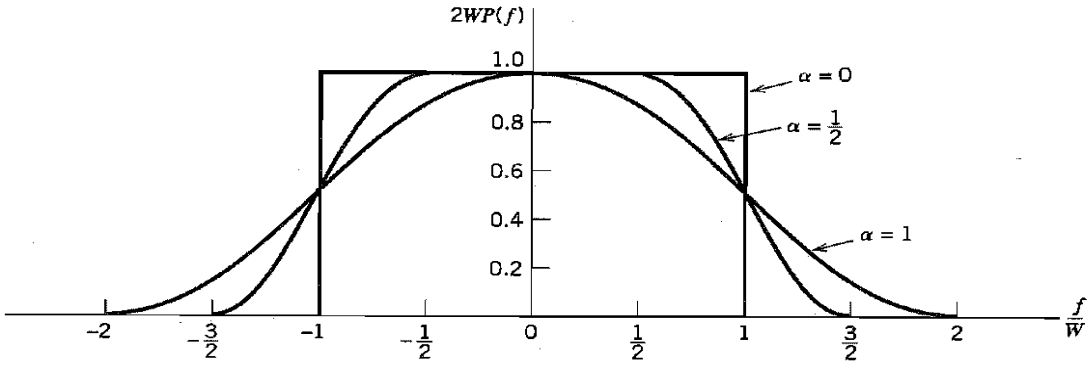
The frequency parameter f_1 and bandwidth W are related by

$$\alpha = 1 - \frac{f_1}{W} \tag{4.61}$$

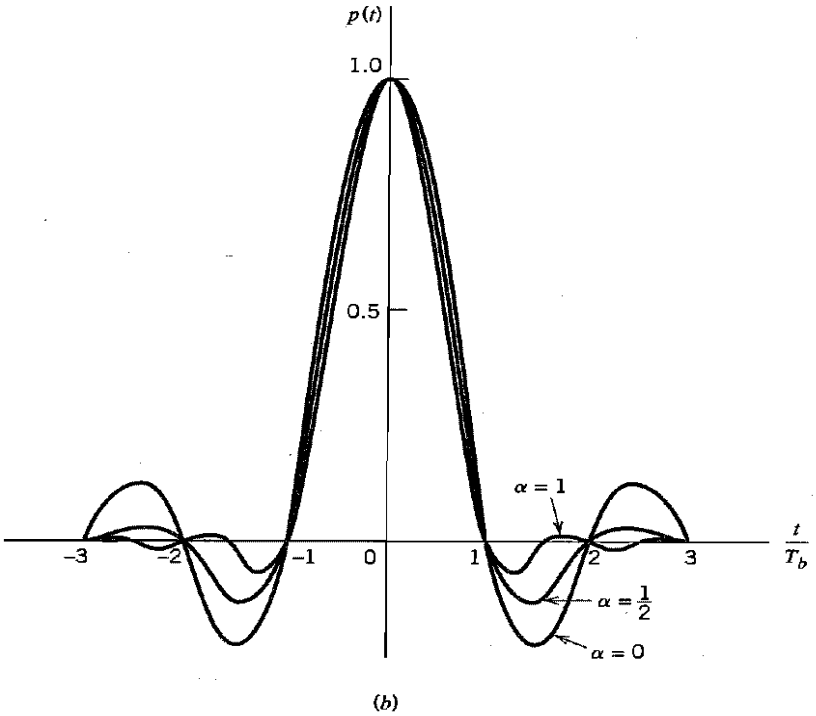
The parameter α is called the *rolloff factor*; it indicates the *excess bandwidth* over the ideal solution, W. Specifically, the transmission bandwidth B_T is defined by

$$\begin{aligned} B_T &= 2W - f_1 \\ &= W(1 + \alpha) \end{aligned}$$

The frequency response $P(f)$, normalized by multiplying it by $2W$, is plotted in Figure 4.10*a* for three values of α, namely, 0, 0.5, and 1. We see that for $\alpha = 0.5$ or 1, the



(*a*)



(*b*)

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

function $P(f)$ cuts off gradually as compared with the ideal Nyquist channel (i.e., $\alpha = 0$) and is therefore easier to implement in practice. Also the function $P(f)$ exhibits odd symmetry with respect to the Nyquist bandwidth W, making it possible to satisfy the condition of Equation (4.59).

The time response $p(t)$ is the inverse Fourier transform of the frequency response $P(f)$. Hence, using the $P(f)$ defined in Equation (4.60), we obtain the result (see Problem 4.13)

$$p(t) = (\text{sinc}(2Wt))\left(\frac{\cos(2\pi\alpha Wt)}{1 - 16\alpha^2 W^2 t^2}\right) \tag{4.62}$$

which is plotted in Figure 4.10*b* for $\alpha = 0$, 0.5, and 1.

The time response $p(t)$ consists of the product of two factors: the factor $\text{sinc}(2Wt)$ characterizing the ideal Nyquist channel and a second factor that decreases as $1/|t|^2$ for large $|t|$. The first factor ensures zero crossings of $p(t)$ at the desired sampling instants of time $t = iT$ with i an integer (positive and negative). The second factor reduces the tails of the pulse considerably below that obtained from the ideal Nyquist channel, so that the transmission of binary waves using such pulses is relatively insensitive to sampling time errors. In fact, for $\alpha = 1$ we have the most gradual rolloff in that the amplitudes of the oscillatory tails of $p(t)$ are smallest. Thus the amount of intersymbol interference resulting from timing error decreases as the rolloff factor α is increased from zero to unity.

The special case with $\alpha = 1$ (i.e., $f_1 = 0$) is known as the *full-cosine rolloff* characteristic, for which the frequency response of Equation (4.60) simplifies to

$$P(f) = \begin{cases} \dfrac{1}{4W}\left[1 + \cos\left(\dfrac{\pi f}{2W}\right)\right], & 0 < |f| < 2W \\ 0, & |f| \geq 2W \end{cases} \tag{4.63}$$

Correspondingly, the time response $p(t)$ simplifies to

$$p(t) = \frac{\text{sinc}(4Wt)}{1 - 16W^2 t^2} \tag{4.64}$$

This time response exhibits two interesting properties:

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

These two properties are extremely useful in extracting a timing signal from the received signal for the purpose of synchronization. However, the price paid for this desirable property is the use of a channel bandwidth double that required for the ideal Nyquist channel corresponding to $\alpha = 0$.

Correlative-Level Coding

Thus far we have treated intersymbol interference as an undesirable phenomenon that produces a degradation in system performance. Indeed, its very name connotes a nuisance effect. Nevertheless, by adding intersymbol interference to the transmitted signal in a controlled manner, it is possible to achieve a signaling rate equal to the Nyquist rate of $2W$ symbols per second in a channel of bandwidth W Hertz. Such schemes are called *correlative-level coding* or *partial-response signaling* schemes.[5] The design of these schemes is based on the following premise: Since intersymbol interference introduced into the transmitted signal is known, its effect can be interpreted at the receiver in a deterministic way. Thus correlative-level coding may be regarded as a practical method of achieving the theoretical maximum signaling rate of $2W$ symbols per second in a bandwidth of W Hertz, as postulated by Nyquist, using realizable and perturbation-tolerant filters.

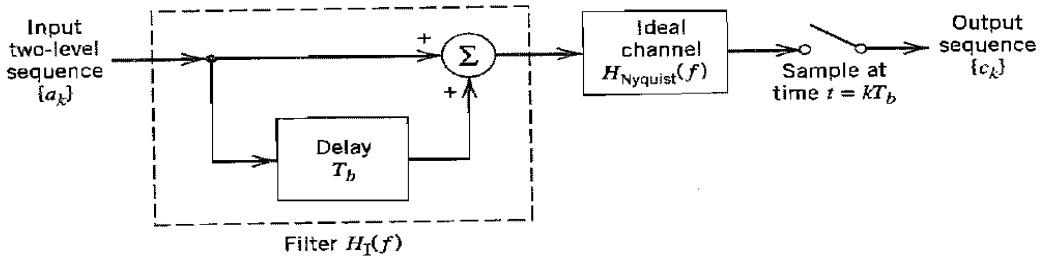
DUOBINARY SIGNALING

The basic idea of correlative-level coding will now be illustrated by considering the specific example of *duobinary signaling*, where "duo" implies doubling of the transmission capacity of a straight binary system. This particular form of correlative-level coding is also called *class I partial response*.

Consider a binary input sequence $\{b_k\}$ consisting of uncorrelated binary symbols 1 and 0, each having duration T_b. As before, this sequence is applied to a pulse-amplitude modulator producing a two-level sequence of short pulses (approximating a unit impulse), whose amplitude a_k is defined by

$$a_k = \begin{cases} +1 & \text{if symbol } b_k \text{ is 1} \\ -1 & \text{if symbol } b_k \text{ is 0} \end{cases} \tag{4.65}$$

When this sequence is applied to a *duobinary encoder*, it is converted into a *three-level output*, namely, −2, 0, and +2. To produce this transformation, we may use the scheme shown in Figure 4.11. The two-level sequence $\{a_k\}$ is first passed through a simple filter involving a single delay element and summer. For every unit impulse applied to the



Duobinary signaling scheme.

input of this filter, we get two unit impulses spaced T_b seconds apart at the filter output. We may therefore express the duobinary coder output c_k as the sum of the present input pulse a_k and its previous value a_{k-1}, as shown by

$$c_k = a_k + a_{k-1} \tag{4.66}$$

One of the effects of the transformation described by Equation (4.66) is to change the input sequence $\{a_k\}$ of uncorrelated two-level pulses into a sequence $\{c_k\}$ of correlated three-level pulses. This correlation between the adjacent pulses may be viewed as introducing intersymbol interference into the transmitted signal in an artificial manner. However, the intersymbol interference so introduced is under the designer's control, which is the basis of correlative coding.

An ideal delay element, producing a delay of T_b seconds, has the frequency response $\exp(-j2\pi fT_b)$, so that the frequency response of the simple delay-line filter in Figure 4.11 is $1 + \exp(-j2\pi fT_b)$. Hence, the overall frequency response of this filter connected in cascade with an ideal Nyquist channel is

$$\begin{aligned} H_I(f) &= H_{Nyquist}(f)[1 + \exp(-j2\pi fT_b)] \\ &= H_{Nyquist}(f)[\exp(j\pi fT_b) + \exp(-j\pi fT_b)] \exp(-j\pi fT_b) \\ &= 2H_{Nyquist}(f) \cos(\pi fT_b) \exp(-j\pi fT_b) \end{aligned} \tag{4.67}$$

where the subscript I in $H_I(f)$ indicates the pertinent class of partial response. For an ideal Nyquist channel of bandwidth $W = 1/2T_b$, we have (ignoring the scaling factor T_b)

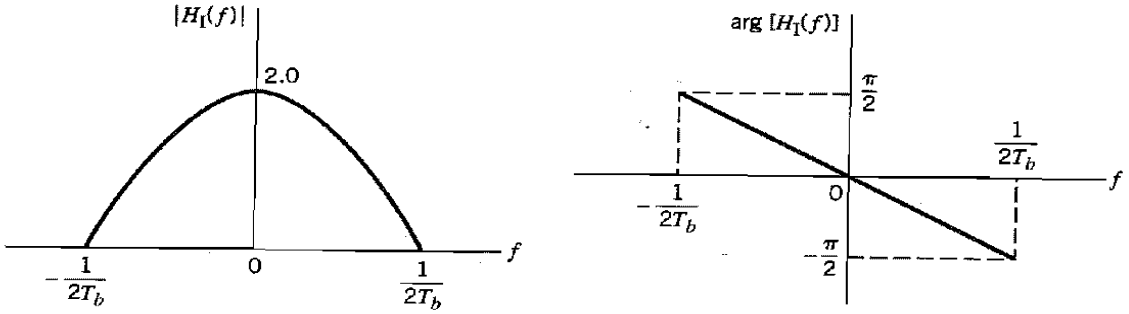
$$H_{Nyquist}(f) = \begin{cases} 1, & |f| \le 1/2T_b \\ 0, & \text{otherwise} \end{cases} \tag{4.68}$$

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

$$H_I(f) = \begin{cases} 2\cos(\pi fT_b) \exp(-j\pi fT_b), & |f| \le 1/2T_b \\ 0, & \text{otherwise} \end{cases} \tag{4.69}$$

for which the magnitude response and phase response are as shown in Figures 4.12*a* and 4.12*b*, respectively. An advantage of this frequency response is that it can be easily approximated, in practice, by virtue of the fact that there is continuity at the band edges.

From the first line in Equation (4.67) and the definition of $H_{Nyquist}(f)$ in Equation (4.68), we find that the impulse response corresponding to the frequency response $H_I(f)$



Magnitude response. Phase response

Frequency response of the duobinary conversion filter.

consists of two sinc (Nyquist) pulses that are time-displaced by T_b seconds with respect to each other, as shown by (except for a scaling factor)

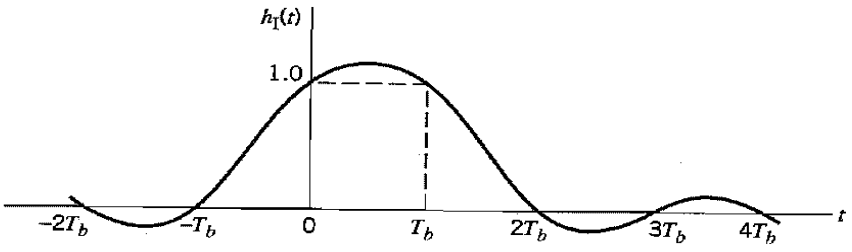
$$\begin{aligned} h_I(t) &= \frac{\sin(\pi t/T_b)}{\pi t/T_b} + \frac{\sin[\pi(t - T_b)/T_b]}{\pi(t - T_b)/T_b} \\ &= \frac{\sin(\pi t/T_b)}{\pi t/T_b} - \frac{\sin(\pi t/T_b)}{\pi(t - T_b)/T_b} \\ &= \frac{T_b^2 \sin(\pi t/T_b)}{\pi t(T_b - t)} \end{aligned} \tag{4.70}$$

The impulse response $h_I(t)$ is plotted in Figure 4.13, where we see that it has only *two* distinguishable values at the sampling instants. The form of $h_I(t)$ shown here explains why we also refer to this type of correlative coding as partial-response signaling. The response to an input pulse is spread over more than one signaling interval; stated in another way, the response in any signaling interval is "partial." Note also that the tails of $h_I(t)$ decay as $1/|t|^2$, which is a faster rate of decay than the $1/|t|$ encountered in the ideal Nyquist channel.

The original two-level sequence $\{a_k\}$ may be detected from the duobinary-coded sequence $\{c_k\}$ by invoking the use of Equation (4.66). Specifically, let $\hat{a}_k$ represent the *estimate* of the original pulse a_k as conceived by the receiver at time $t = kt_b$. Then, subtracting the previous estimate $\hat{a}_{k-1}$ from c_k, we get

$$\hat{a}_k = c_k - \hat{a}_{k-1} \tag{4.71}$$

It is apparent that if c_k is received without error and if also the previous estimate $\hat{a}_{k-1}$ at time $t = (k - 1)T_b$ corresponds to a correct decision, then the current estimate $\hat{a}_k$ will be



Impulse response of the duobinary conversion filter.

correct too. The technique of using a stored estimate of the previous symbol is called *decision feedback*.

We observe that the detection procedure just described is essentially an inverse of the operation of the simple delay-line filter at the transmitter. However, a major drawback of this detection procedure is that once errors are made, they tend to *propagate* through the output because a decision on the current input a_k depends on the correctness of the decision made on the previous input a_{k-1}.

A practical means of avoiding the error-propagation phenomenon is to use *precoding* before the duobinary coding, as shown in Figure 4.14. The precoding operation performed on the binary data sequence $\{b_k\}$ converts it into another binary sequence $\{d_k\}$ defined by

$$d_k = b_k \oplus d_{k-1} \tag{4.72}$$

where the symbol $\oplus$ denotes *modulo-two addition* of the binary digits b_k and d_{k-1}. This addition is equivalent to a two-input EXCLUSIVE OR operation, which is performed as follows:

$$d_k = \begin{cases} \text{symbol 1} & \text{if either symbol } b_k \text{ or symbol } d_{k-1} \text{ (but not both) is 1} \\ \text{symbol 0} & \text{otherwise} \end{cases} \tag{4.73}$$

The precoded binary sequence $\{d_k\}$ is applied to a pulse-amplitude modulator, producing a corresponding two-level sequence of short pulses $\{a_k\}$, where $a_k = \pm 1$ as before. This sequence of short pulses is next applied to the duobinary coder, thereby producing the sequence $\{c_k\}$ that is related to $\{a_k\}$ as follows:

$$c_k = a_k + a_{k-1} \tag{4.74}$$

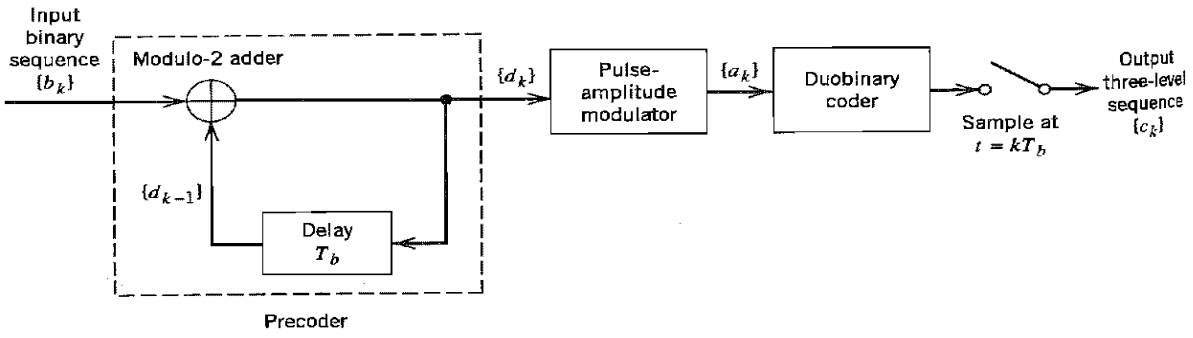
Note that unlike the linear operation of duobinary coding, the precoding described by Equation (4.72) is a *nonlinear* operation.

The combined use of Equations (4.72) and (4.74) yields

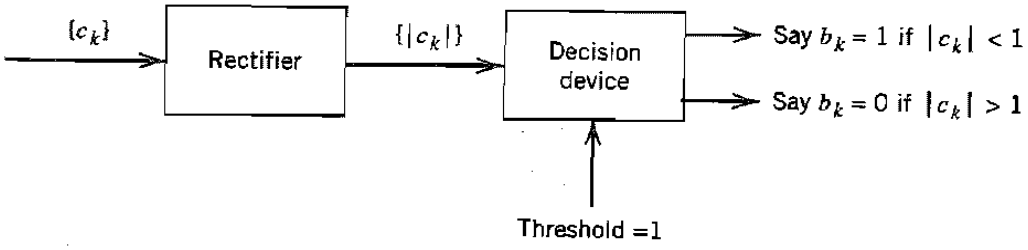
$$c_k = \begin{cases} 0 & \text{if data symbol } b_k \text{ is 1} \\ \pm 2 & \text{if data symbol } b_k \text{ is 0} \end{cases} \tag{4.75}$$

which is illustrated in Example 4.3. From Equation (4.75) we deduce the following decision rule for detecting the original binary sequence $\{b_k\}$ from $\{c_k\}$:

$$\begin{aligned} &\text{If } |c_k| < 1, && \text{say symbol } b_k \text{ is 1} \\ &\text{If } |c_k| > 1, && \text{say symbol } b_k \text{ is 0} \end{aligned} \tag{4.76}$$



A precoded duobinary scheme; details of the duobinary coder



Detector for recovering original binary sequence from the precoded duobinary coder output.

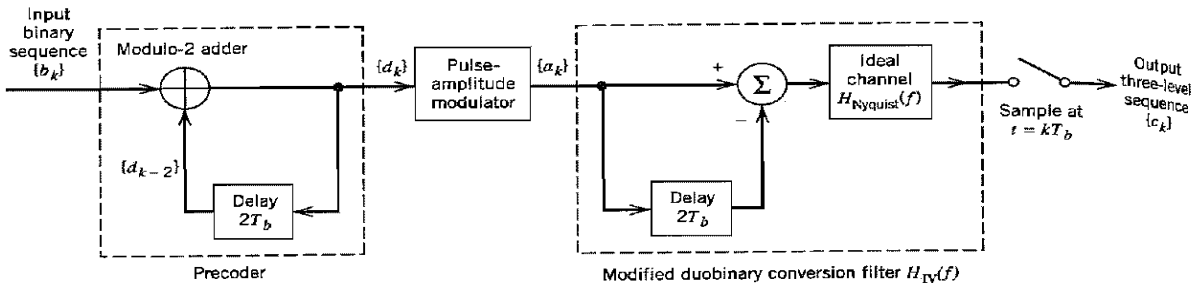
When $|c_k| = 1$, the receiver simply makes a random guess in favor of symbol 1 or 0. According to this decision rule, the detector consists of a rectifier, the output of which is compared in a decision device to a threshold of 1. A block diagram of the detector is shown in Figure 4.15. A useful feature of this detector is that no knowledge of any input sample other than the present one is required. Hence, error propagation cannot occur in the detector .

MODIFIED DUOBINARY SIGNALING

In the duobinary signaling technique the frequency response $H(f)$, and consequently the power spectral density of the transmitted pulse, is nonzero at the origin. This is considered to be an undesirable feature in some applications, since many communications channels cannot transmit a DC component. We may correct for this deficiency by using the *class IV partial response* or *modified duobinary* technique, which involves a correlation span of two binary digits. This special form of correlation is achieved by subtracting amplitude-modulated pulses spaced $2T_b$ seconds apart, as indicated in the block diagram of Figure

Illustrating on duobinary coding

Binary sequence $\{b_k\}$		0	0	1	0	1	1	0
Precoded sequence $\{d_k\}$	1	1	1	0	0	1	0	0
Two-level sequence $\{a_k\}$	+1	+1	+1	−1	−1	+1	−1	−1
Duobinary coder output $\{c_k\}$		+2	+2	0	−2	0	0	−2
Binary sequence obtained by applying decision rule of Eq. (4.76)		0	0	1	0	1	1	0



Modified duobinary signaling scheme.

4.16. The precoder involves a delay of $2T_b$ seconds. The output of the modified duobinary conversion filter is related to the input two-level sequence $\{a_k\}$ at the pulse-amplitude modulator output as follows:

$$c_k = a_k - a_{k-2} \tag{4.77}$$

Here, again, we find that a three-level signal is generated. With $a_k = \pm 1$, we find that c_k takes on one of three values: $+2$, 0, and -2.

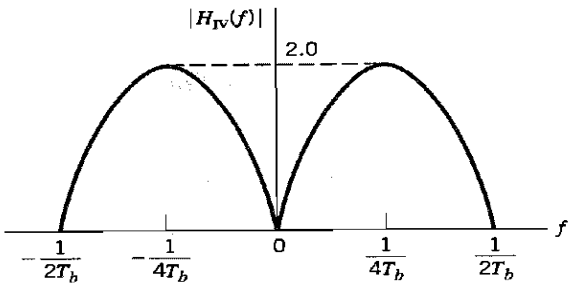
The overall frequency response of the delay-line filter connected in cascade with an ideal Nyquist channel, as in Figure 4.16, is given by

$$\begin{aligned} H_{IV}(f) &= H_{Nyquist}(f)[1 - \exp(-j4\pi fT_b)] \\ &= 2jH_{Nyquist}(f)\sin(2\pi fT_b)\exp(-j2\pi fT_b) \end{aligned} \tag{4.78}$$

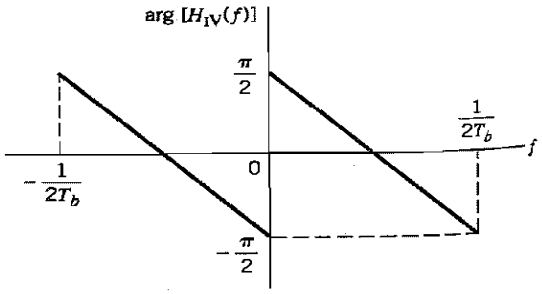
where the subscript IV in $H_{IV}(f)$ indicates the pertinent class of partial response and $H_{Nyquist}(f)$ is as defined in Equation (4.68). We therefore have an overall frequency response in the form of a half-cycle sine function, as shown by

$$H_{IV}(f) = \begin{cases} 2j\sin(2\pi fT_b)\exp(-j2\pi fT_b), & |f| \le 1/2T_b \\ 0, & \text{elsewhere} \end{cases} \tag{4.79}$$

The corresponding magnitude response and phase response of the modified duobinary coder are shown in Figures 4.17*a* and 4.17*b*, respectively. A useful feature of the modified duobinary coder is the fact that its output has no DC component. Note also that this



Magnitude Response



Phase response.

Frequency response of the modified duobinary conversion filter.

second form of correlative-level coding exhibits the same continuity at the band edges as in duobinary signaling.

From the first line of Equation (4.78) and the definition of $H_{\text{Nyquist}}(f)$ in Equation (4.68), we find that the impulse response of the modified duobinary coder consists of two sinc (Nyquist) pulses that are time-displaced by $2T_b$ seconds with respect to each other, as shown by (except for a scaling factor)

$$\begin{aligned} h_{\text{IV}}(t) &= \frac{\sin(\pi t/T_b)}{\pi t/T_b} - \frac{\sin[\pi(t-2T_b)/T_b)]}{\pi(t-2T_b)/T_b} \\ &= \frac{\sin(\pi t/T_b)}{\pi t/T_b} - \frac{\sin(\pi t/T_b)}{\pi(t-2T_b)/T_b} \\ &= \frac{2T_b^2 \sin(\pi t/T_b)}{\pi t(2T_b - t)} \end{aligned} \tag{4.80}$$

This impulse response is plotted in Figure 4.18, which shows that it has *three* distinguishable levels at the sampling instants. Note also that, as with duobinary signaling, the tails of $h_{\text{IV}}(t)$ for the modified duobinary signaling decay as $1/|t|^2$.

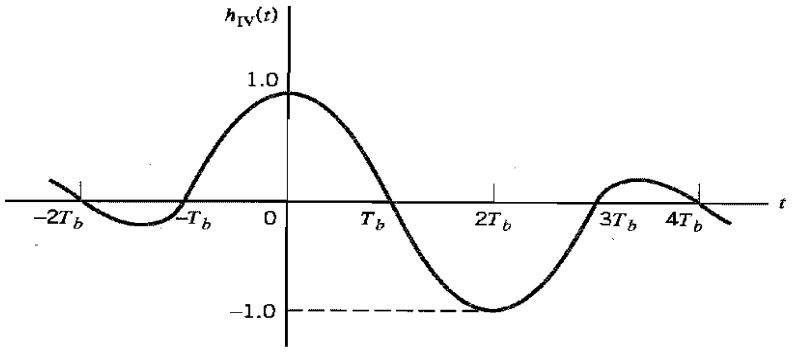
To eliminate the possibility of error propagation in the modified duobinary system, we use a precoding procedure similar to that used for the duobinary case. Specifically, prior to the generation of the modified duobinary signal, a modulo-two logical addition is used on signals $2T_b$ seconds apart, as shown by (see the front end of Figure 4.16)

$$\begin{aligned} d_k &= b_k \oplus d_{k-2} \\ &= \begin{cases} \text{symbol 1} & \text{if either symbol } b_k \text{ or symbol } d_{k-2} \text{ (but not both) is 1} \\ \text{symbol 0} & \text{otherwise} \end{cases} \end{aligned} \tag{4.81}$$

where $\{b_k\}$ is the incoming binary data sequence and $\{d_k\}$ is the sequence at the precoder output. The precoded sequence $\{d_k\}$ thus produced is then applied to a pulse-amplitude modulator and then to the modified duobinary conversion filter.

In Figure 4.16, the output digit c_k equals -2, 0, or $+2$, assuming that the pulse-amplitude modulator uses a polar representation for the precoded sequence $\{d_k\}$. Also we find that the detected digit $\hat{b}_k$ at the receiver output may be extracted from c_k by disregarding the polarity of c_k. Specifically, we may formulate the following decision rule:

$$\begin{aligned} &\text{If } |c_k| > 1, \quad \text{say symbol } b_k \text{ is 1} \\ &\text{If } |c_k| < 1, \quad \text{say symbol } b_k \text{ is 0} \end{aligned} \tag{4.82}$$



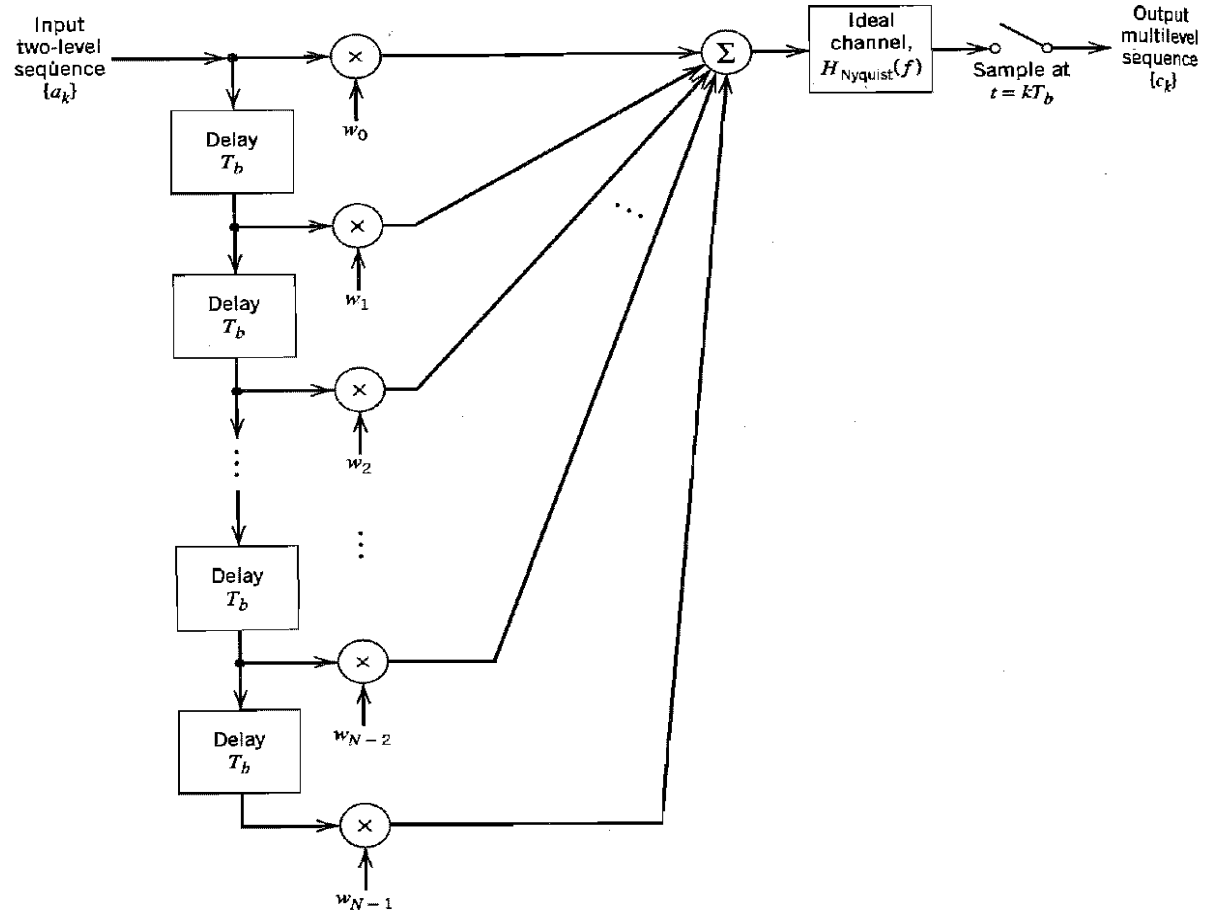
Impulse response of the modified duobinary conversion filter.

When $|c_k| = 1$, the receiver makes a random guess in favor of symbol 1 or 0. As with the duobinary signaling, we may note the following:

- In the absence of channel noise, the detected binary sequence $\{\hat{b}_k\}$ is exactly the same as the original binary sequence $\{b_k\}$ at the transmitter input.
- The use of Equation (4.81) requires the addition of two extra bits to the precoded sequence $\{a_k\}$. The composition of the decoded sequence $\{\hat{b}_k\}$ using Equation (4.82) is invariant to the selection made for these two bits.

Generalized Form of Correlative-Level Coding (Partial-Response Signaling)

The duobinary and modified duobinary techniques have correlation spans of 1 binary digit and 2 binary digits, respectively. It is a straightforward matter to generalize these two techniques to other schemes, which are known collectively as *correlative-level coding* or *partial-response signaling* schemes. This generalization is shown in Figure 4.19, where $H_{\text{Nyquist}}(f)$ is defined in Equation (4.68). It involves the use of a tapped-delay-line filter with tap-weights $w_0, w_1, \cdots, w_{N-1}$. Specifically, different classes of partial-response sig-



Generalized correlative coding scheme.

Different classes of partial-response signaling schemes

Type of Class	N	w_0	w_1	w_2	w_3	w_4	Comments
I	2	1	1				Duobinary coding
II	3	1	2	1			
III	3	2	1	−1			
IV	3	1	0	−1			Modified duobinary coding
V	5	−1	0	2	0	−1	

naling schemes may be achieved by using a weighted linear combination of N ideal Nyquist (sinc) pulses, as shown by

$$h(t) = \sum_{n=0}^{N-1} w_n \operatorname{sinc}\left(\frac{t}{T_b} - n\right) \tag{4.83}$$

An appropriate choice of the tap-weights in Equation (4.83) results in a variety of spectral shapes designed to suit individual applications. Table 4.2 presents the specific details of five different classes of partial-response signaling schemes. For example, in the duobinary case (class I partial response), we have

$$w_0 = +1$$
$$w_1 = +1$$

and $w_n = 0$ for $n \geq 2$. In the modified duobinary case (class IV partial response), we have

$$w_0 = +1$$
$$w_1 = 0$$
$$w_2 = -1$$

and $w_n = 0$ for $n \geq 3$.

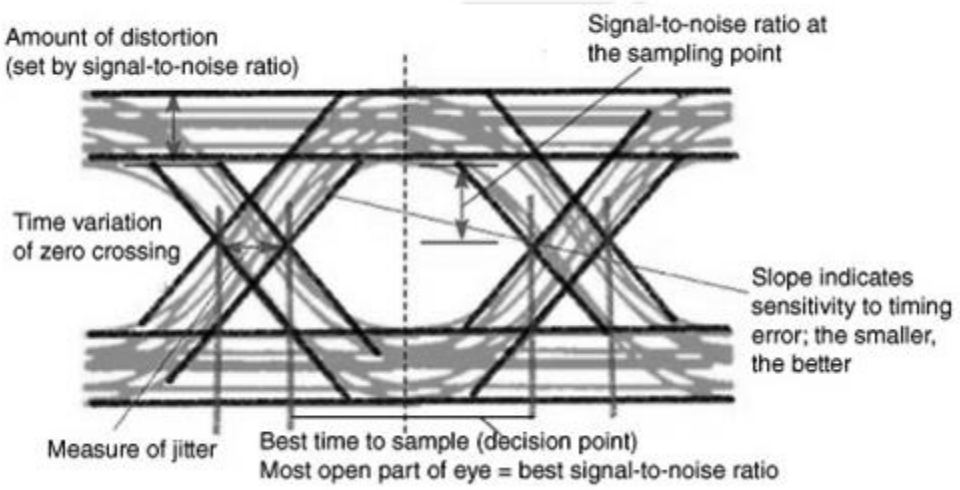
The useful characteristics of partial-response signaling schemes may now be summarized as follows:

- Binary data transmission over a physical baseband channel can be accomplished at a rate close to the Nyquist rate, using realizable filters with gradual cutoff characteristics.
- Different spectral shapes can be produced, appropriate for the application at hand.

However, these desirable characteristics are achieved at a price: A larger signal-to-noise ratio is required to yield the same average probability of symbol error in the presence of noise as in the corresponding binary PAM systems because of an increase in the number of signal levels used.

Eye Pattern

An effective way to study the effects of ISI is the Eye Pattern. The name Eye Pattern was given from its resemblance to the human eye for binary waves. The interior region of the eye pattern is called the eye opening. The following figure shows the image of an eye-pattern.



Jitter is the short-term variation of the instant of digital signal, from its ideal position, which may lead to data errors.

When the effect of ISI increases, traces from the upper portion to the lower portion of the eye opening increases and the eye gets completely closed, if ISI is very high.

An eye pattern provides the following information about a particular system.

- **Actual eye patterns are used to estimate the bit error rate and the signal-to-noise ratio.**
- **The width of the eye opening defines the time interval over which the received wave can be sampled without error from ISI.**
- **The instant of time when the eye opening is wide, will be the preferred time for sampling.**
- **The rate of the closure of the eye, according to the sampling time, determines how sensitive the system is to the timing error.**
- **The height of the eye opening, at a specified sampling time, defines the margin over noise.**

Hence, the interpretation of eye pattern is an important consideration.

UNIT IV-PASS BAND DATA TRANSMISSION

Pass Band Transmission Model – Generation, Detection, Signal Space Diagram, Probability of Error for BFSK, BPSK, QPSK, DPSK, and Schemes– Comparison.

Passband Transmission Model

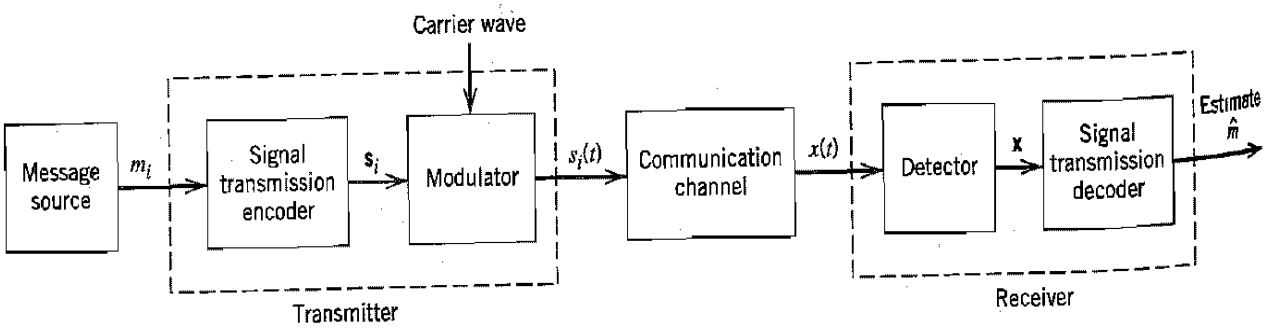
In a functional sense, we may model a passband data transmission system as shown in Figure 6.2. First, there is assumed to exist a *message source* that emits one *symbol* every T seconds, with the symbols belonging to an alphabet of M symbols, which we denote by $m_1, m_2, \ldots, m_M$. The *a priori probabilities* $P(m_1), P(m_2), \ldots, P(m_M)$ specify the message source output. When the M symbols of the alphabet are *equally likely*, we write

$$\begin{aligned} p_i &= P(m_i) \\ &= \frac{1}{M} \qquad \text{for all } i \end{aligned} \tag{6.6}$$

The M-ary output of the message source is presented to a *signal transmission encoder*, producing a corresponding vector $\mathbf{s}_i$ made up of N real elements, one such set for each of the M symbols of the source alphabet; the dimension N is less than or equal to M. With the vector $\mathbf{s}_i$ as input, the *modulator* then constructs a *distinct* signal $s_i(t)$ of duration T seconds as the representation of the symbol m_i generated by the message source. The signal $s_i(t)$ is necessarily an energy signal, as shown by

$$E_i = \int_0^T s_i^2(t)\, dt, \qquad i = 1, 2, \ldots, M \tag{6.7}$$

Note that $s_i(t)$ is real valued. One such signal is transmitted every T seconds. The particular signal chosen for transmission depends in some fashion on the incoming message and possibly on the signals transmitted in preceding time slots. With a sinusoidal carrier, the feature that is used by the modulator to distinguish one signal from another is a *step* change in the amplitude, frequency, or phase of the carrier. (Sometimes, a hybrid form of modulation that combines changes in both amplitude and phase or amplitude and frequency is used.)



Functional model of passband data transmission system.

Returning to the functional model of Figure 6.2, the bandpass communication channel, coupling the transmitter to the receiver, is assumed to have two characteristics:

1. The channel is linear, with a bandwidth that is wide enough to accommodate the transmission of the modulated signal $s_i(t)$ with negligible or no distortion.
2. The channel noise $w(t)$ is the sample function of a white Gaussian noise process of zero mean and power spectral density $N_0/2$.

The assumptions made herein are basically the same as those invoked in Chapter 5 dealing with signal-space analysis.

The receiver, which consists of a *detector* followed by a *signal transmission decoder*, performs two functions:

1. It reverses the operations performed in the transmitter.
2. It minimizes the effect of channel noise on the estimate $\hat{m}$ computed for the transmitted symbol m_i.

BINARY FSK

In a *binary FSK system*, symbols 1 and 0 are distinguished from each other by transmitting one of two sinusoidal waves that differ in frequency by a fixed amount. A typical pair of sinusoidal waves is described by

$$s_i(t) = \begin{cases} \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_i t), & 0 \le t \le T_b \\ 0, & \text{elsewhere} \end{cases} \tag{6.86}$$

where $i = 1, 2$, and E_b is the transmitted signal energy per bit; the transmitted frequency is

$$f_i = \frac{n_c + i}{T_b} \quad \text{for some fixed integer } n_c \text{ and } i = 1, 2 \tag{6.87}$$

Thus symbol 1 is represented by $s_1(t)$, and symbol 0 by $s_2(t)$. The FSK signal described here is known as *Sunde's FSK*. It is a *continuous-phase signal* in the sense that phase continuity is always maintained, including the inter-bit switching times. This form of digital modulation is an example of *continuous-phase frequency-shift keying* (CPFSK), on which we have more to say later on in the section.

From Equations (6.86) and (6.87), we observe directly that the signals $s_1(t)$ and $s_2(t)$ are orthogonal, but not normalized to have unit energy. We therefore deduce that the most useful form for the set of orthonormal basis functions is

$$\phi_i(t) = \begin{cases} \sqrt{\frac{2}{T_b}} \cos(2\pi f_i t), & 0 \le t \le T_b \\ 0, & \text{elsewhere} \end{cases} \tag{6.88}$$

where $i = 1, 2$. Correspondingly, the coefficient s_{ij} for $i = 1, 2$, and $j = 1, 2$ is defined by

$$\begin{aligned} s_{ij} &= \int_0^{T_b} s_i(t)\phi_j(t)\, dt \\ &= \int_0^{T_b} \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_i t) \sqrt{\frac{2}{T_b}} \cos(2\pi f_j t)\, dt \\ &= \begin{cases} \sqrt{E_b}, & i = j \\ 0, & i \ne j \end{cases} \end{aligned} \tag{6.89}$$

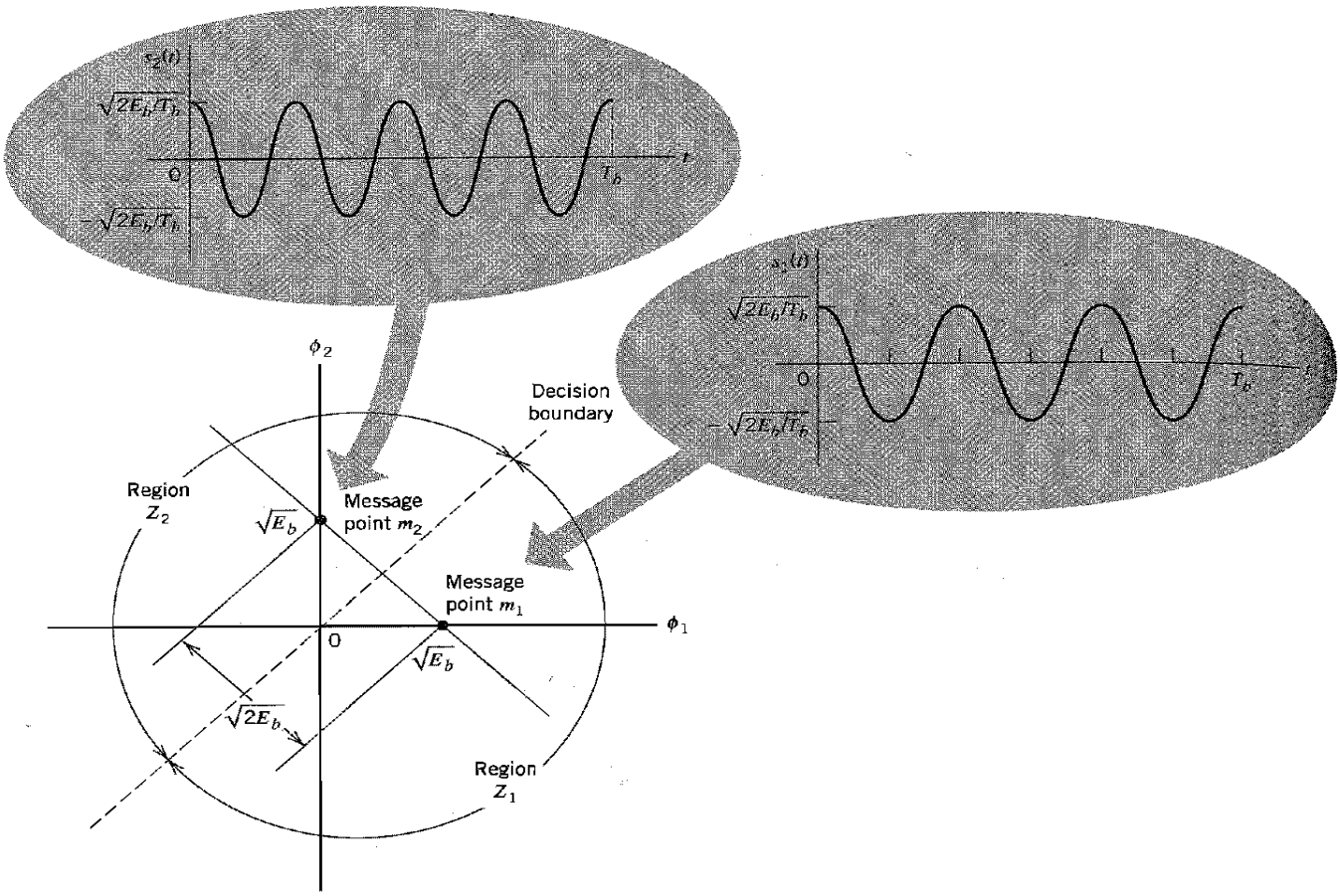
Thus, unlike coherent binary PSK, a coherent binary FSK system is characterized by having a signal space that is two-dimensional (i.e., $N = 2$) with two message points (i.e., $M = 2$), as shown in Figure 6.25. The two message points are defined by the

$$\mathbf{s}_1 = \begin{bmatrix} \sqrt{E_b} \\ 0 \end{bmatrix} \tag{6.90}$$

and

$$\mathbf{s}_2 = \begin{bmatrix} 0 \\ \sqrt{E_b} \end{bmatrix} \tag{6.91}$$

with the Euclidean distance between them equal to $\sqrt{2E_b}$. Figure 6.25 also includes a couple of inserts, which show waveforms representative of signals $s_1(t)$ and $s_2(t)$.



Signal-space diagram for binary FSK system. The diagram also includes two inserts showing example waveforms of the two modulated signals $s_1(t)$ and $s_2(t)$.

Error Probability of Binary FSK

The observation vector $\mathbf{x}$ has two elements x_1 and x_2 that are defined by, respectively,

$$x_1 = \int_0^{T_b} x(t)\phi_1(t)\, dt \tag{6.92}$$

and

$$x_2 = \int_0^{T_b} x(t)\phi_2(t)\, dt \tag{6.93}$$

where $x(t)$ is the received signal, the form of which depends on which symbol was transmitted. Given that symbol 1 was transmitted, $x(t)$ equals $s_1(t) + w(t)$, where $w(t)$ is the sample function of a white Gaussian noise process of zero mean and power spectral density $N_0/2$. If, on the other hand, symbol 0 was transmitted, $x(t)$ equals $s_2(t) + w(t)$.

Now, applying the decision rule of Equation (5.59), we find that the observation space is partitioned into two decision regions, labeled Z_1 and Z_2 in Figure 6.25. The decision boundary, separating region Z_1 from region Z_2 is the perpendicular bisector of

the line joining the two message points. The receiver decides in favor of symbol 1 if the received signal point represented by the observation vector **x** falls inside region Z_1. This occurs when $x_1 > x_2$. If, on the other hand, we have $x_1 < x_2$, the received signal point falls inside region Z_2, and the receiver decides in favor of symbol 0. On the decision boundary, we have $x_1 = x_2$, in which case the receiver makes a random guess in favor of symbol 1 or 0.

Define a new Gaussian random variable Y whose sample value y is equal to the difference between x_1 and x_2; that is,

$$y = x_1 - x_2 \tag{6.94}$$

The mean value of the random variable Y depends on which binary symbol was transmitted. Given that symbol 1 was transmitted, the Gaussian random variables X_1 and X_2, whose sample values are denoted by x_1 and x_2, have mean values equal to $\sqrt{E_b}$ and zero, respectively. Correspondingly, the conditional mean of the random variable Y, given that symbol 1 was transmitted, is

$$\begin{aligned} E[Y|1] &= E[X_1|1] - E[X_2|1] \\ &= +\sqrt{E_b} \end{aligned} \tag{6.95}$$

On the other hand, given that symbol 0 was transmitted, the random variables X_1 and X_2 have mean values equal to zero and $\sqrt{E_b}$, respectively. Correspondingly, the conditional mean of the random variable Y, given that symbol 0 was transmitted, is

$$\begin{aligned} E[Y|0] &= E[X_1|0] - E[X_2|0] \\ &= -\sqrt{E_b} \end{aligned} \tag{6.96}$$

The variance of the random variable Y is independent of which binary symbol was transmitted. Since the random variables X_1 and X_2 are statistically independent, each with a variance equal to $N_0/2$, it follows that

$$\begin{aligned} \text{var}[Y] &= \text{var}[X_1] + \text{var}[X_2] \\ &= N_0 \end{aligned} \tag{6.97}$$

Suppose we know that symbol 0 was transmitted. The conditional probability density function of the random variable Y is then given by

$$f_Y(y|0) = \frac{1}{\sqrt{2\pi N_0}} \exp\left[-\frac{(y + \sqrt{E_b})^2}{2N_0}\right] \tag{6.98}$$

Since the condition $x_1 > x_2$, or equivalently, $y > 0$, corresponds to the receiver making a decision in favor of symbol 1, we deduce that the conditional probability of error, given that symbol 0 was transmitted, is

$$\begin{aligned} p_{10} &= P(y > 0 | \text{symbol 0 was sent}) \\ &= \int_0^\infty f_Y(y|0)\, dy \\ &= \frac{1}{\sqrt{2\pi N_0}} \int_0^\infty \exp\left[-\frac{(y + \sqrt{E_b})^2}{2N_0}\right] dy \end{aligned} \tag{6.99}$$

Put

$$\frac{y + \sqrt{E_b}}{\sqrt{2N_0}} = z \tag{6.100}$$

Then, changing the variable of integration from y to z, we may rewrite Equation (6.99) as follows:

$$p_{10} = \frac{1}{\sqrt{\pi}} \int_{\sqrt{E_b/2N_0}}^{\infty} \exp(-z^2)\, dz = \frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{E_b}{2N_0}}\right) \tag{6.101}$$

Similarly, we may show the p_{01}, the conditional probability of error given that symbol 1 was transmitted, has the same value as in Equation (6.101). Accordingly, averaging p_{10} and p_{01}, we find that the *average probability of bit error* or, equivalently, the *bit error rate for coherent binary FSK* is (assuming equiprobable symbols)

$$P_e = \frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{E_b}{2N_0}}\right) \tag{6.102}$$

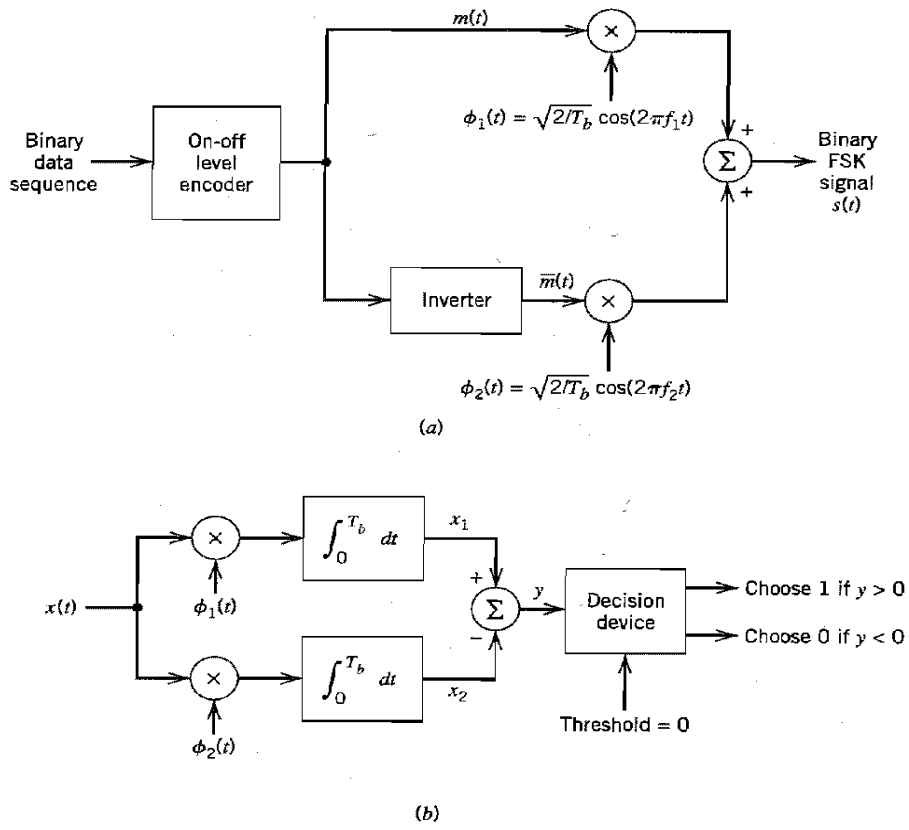
Comparing Equations (6.20) and (6.102), we see that, in a coherent binary FSK system, we have to double the *bit energy-to-noise density ratio*, E_b/N_0, to maintain the same bit error rate as in a coherent binary PSK system. This result is in perfect accord with the signal-space diagrams of Figures 6.3 and 6.25, where we see that in a binary PSK system the Euclidean distance between the two message points is equal to $2\sqrt{E_b}$, whereas in a binary FSK system the corresponding distance is $\sqrt{2E_b}$. For a prescribed E_b, the minimum distance $d_{\min}$ in binary PSK is therefore $\sqrt{2}$ times that in binary FSK. Recall from Chapter 5 that the probability of error decreases exponentially as $d_{\min}^2$, hence the difference between the formulas of Equations (6.20) and (6.102).

Generation and Detection of Coherent Binary FSK Signals

To generate a binary FSK signal, we may use the scheme shown in Figure 6.26*a*. The incoming binary data sequence is first applied to an *on–off level encoder*, at the output of which symbol 1 is represented by a constant amplitude of $\sqrt{E_b}$ volts and symbol 0 is represented by zero volts. By using an *inverter* in the lower channel in Figure 6.26*a*, we in effect make sure that when we have symbol 1 at the input, the oscillator with frequency f_1 in the upper channel is switched on while the oscillator with frequency f_2 in the lower channel is switched off, with the result that frequency f_1 is transmitted. Conversely, when we have symbol 0 at the input, the oscillator in the upper channel is switched off and the oscillator in the lower channel is switched on, with the result that frequency f_2 is transmitted. The two frequencies f_1 and f_2 are chosen to equal different integer multiples of the bit rate $1/T_b$, as in Equation (6.87).

In the transmitter of Figure 6.26*a*, we assume that the two oscillators are synchronized, so that their outputs satisfy the requirements of the two orthonormal basis functions $\phi_1(t)$ and $\phi_2(t)$, as in Equation (6.88). Alternatively, we may use a single keyed (voltage-controlled) oscillator. In either case, the frequency of the modulated wave is shifted with a continuous phase, in accordance with the input binary wave.

To detect the original binary sequence given the noisy received signal $x(t)$, we may use the receiver shown in Figure 6.26*b*. It consists of two correlators with a common input, which are supplied with locally generated coherent reference signals $\phi_1(t)$ and $\phi_2(t)$. The correlator outputs are then subtracted, one from the other, and the resulting difference, y, is compared with a threshold of zero volts. If $y > 0$, the receiver decides in favor of 1. On the other hand, if $y < 0$, it decides in favor of 0. If y is exactly zero, the receiver makes a random guess in favor of 1 or 0.



Block diagrams for (*a*) binary FSK transmitter and (*b*) coherent binary FSK receiver.

Power Spectra of Binary FSK Signals

Consider the case of Sunde's FSK, for which the two transmitted frequencies f_1 and f_2 differ by an amount equal to the bit rate $1/T_b$, and their arithmetic mean equals the nominal carrier frequency f_c; phase continuity is always maintained, including inter-bit switching times. We may express this special binary FSK signal as follows:

$$s(t) = \sqrt{\frac{2E_b}{T_b}}\cos\left(2\pi f_c t \pm \frac{\pi t}{T_b}\right), \qquad 0 \le t \le T_b \tag{6.103}$$

and using a well-known trigonometric identity, we get

$$\begin{aligned} s(t) &= \sqrt{\frac{2E_b}{T_b}}\cos\left(\pm\frac{\pi t}{T_b}\right)\cos(2\pi f_c t) - \sqrt{\frac{2E_b}{T_b}}\sin\left(\pm\frac{\pi t}{T_b}\right)\sin(2\pi f_c t) \\ &= \sqrt{\frac{2E_b}{T_b}}\cos\left(\frac{\pi t}{T_b}\right)\cos(2\pi f_c t) \mp \sqrt{\frac{2E_b}{T_b}}\sin\left(\frac{\pi t}{T_b}\right)\sin(2\pi f_c t) \end{aligned} \tag{6.104}$$

In the last line of Equation (6.104), the plus sign corresponds to transmitting symbol 0, and the minus sign corresponds to transmitting symbol 1. As before, we assume that the symbols 1 and 0 in the random binary wave at the modulator input are equally likely, and that the symbols transmitted in adjacent time slots are statistically independent. Then,

based on the representation of Equation (6.104), we may make the following observations pertaining to the in-phase and quadrature components of a binary FSK signal with continuous phase:

1. The in-phase component is completely independent of the input binary wave. It equals $\sqrt{2E_b/T_b}\cos(\pi t/T_b)$ for all values of time t. The power spectral density of this component therefore consists of two delta functions, weighted by the factor $E_b/2T_b$, and occurring at $f = \pm 1/2T_b$.
2. The quadrature component is directly related to the input binary wave. During the signaling interval $0 \leq t \leq T_b$, it equals $-g(t)$ when we have symbol 1, and $+g(t)$ when we have symbol 0. The symbol shaping function $g(t)$ is defined by

$$g(t) = \begin{cases} \sqrt{\dfrac{2E_b}{T_b}} \sin\left(\dfrac{\pi t}{T_b}\right), & 0 \leq t \leq T_b \\ 0, & \text{elsewhere} \end{cases} \tag{6.105}$$

The energy spectral density of this symbol shaping function equals

$$\Psi_g(f) = \frac{8E_bT_b\cos^2(\pi T_bf)}{\pi^2(4T_b^2f^2 - 1)^2} \tag{6.106}$$

The power spectral density of the quadrature component equals $\Psi_g(f)/T_b$. It is also apparent that the in-phase and quadrature components of the binary FSK signal are independent of each other. Accordingly, the baseband power spectral density of Sunde's FSK signal equals the sum of the power spectral densities of these two components, as shown by

$$S_B(f) = \frac{E_b}{2T_b}\left[\delta\left(f - \frac{1}{2T_b}\right) + \delta\left(f + \frac{1}{2T_b}\right)\right] + \frac{8E_b\cos^2(\pi T_bf)}{\pi^2(4T_b^2f^2 - 1)^2} \tag{6.107}$$

Substituting Equation (6.107) in Equation (6.4), we find that the power spectrum of the binary FSK signal contains two discrete frequency components located at $(f_c + 1/2T_b) = f_1$ and $(f_c - 1/2T_b) = f_2$, with their average powers adding up to one-half the total power of the binary FSK signal. The presence of these two discrete frequency components provides a means of synchronizing the receiver with the transmitter.

Note also that the baseband power spectral density of a binary FSK signal with continuous phase ultimately falls off as the inverse fourth power of frequency. This is readily established by taking the limit in Equation (6.107) as f approaches infinity. If, however, the FSK signal exhibits phase discontinuity at the inter-bit switching instants (this arises when the two oscillators applying frequencies f_1 and f_2 operate independently of each other), the power spectral density ultimately falls off as the inverse square of frequency; see Problem 6.23. Accordingly, an FSK signal with continuous phase does not produce as much interference outside the signal band of interest as an FSK signal with discontinuous phase.

In Figure 6.5, we have plotted the baseband power spectra of Equations (6.22) and (6.107). (To simplify matters, we have only plotted the results for positive frequency.) In both cases, $S_B(f)$ is shown normalized with respect to $2E_b$, and the frequency is normalized with respect to the bit rate $R_b = 1/T_b$. The difference in the falloff rates of these spectra can be explained on the basis of the pulse shape $g(t)$. The smoother the pulse, the faster the drop of spectral tails to zero. Thus, since binary FSK (with continuous phase) has a smoother pulse shape, it has lower sidelobes than binary PSK.

BINARY PHASE-SHIFT KEYING

In a coherent binary PSK system, the pair of signals $s_1(t)$ and $s_2(t)$ used to represent binary symbols 1 and 0, respectively, is defined by

$$s_1(t) = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t) \tag{6.8}$$

$$s_2(t) = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t + \pi) = -\sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t) \tag{6.9}$$

where $0 \leq t \leq T_b$, and E_b is the *transmitted signal energy per bit*. To ensure that each transmitted bit contains an integral number of cycles of the carrier wave, the carrier frequency f_c is chosen equal to n_c/T_b for some fixed integer n_c. A pair of sinusoidal waves that differ only in a relative phase-shift of 180 degrees, as defined in Eqautions (6.8) and (6.9), are referred to as *antipodal signals*.

From this pair of equations it is clear that, in the case of binary PSK, there is only one basis function of unit energy, namely,

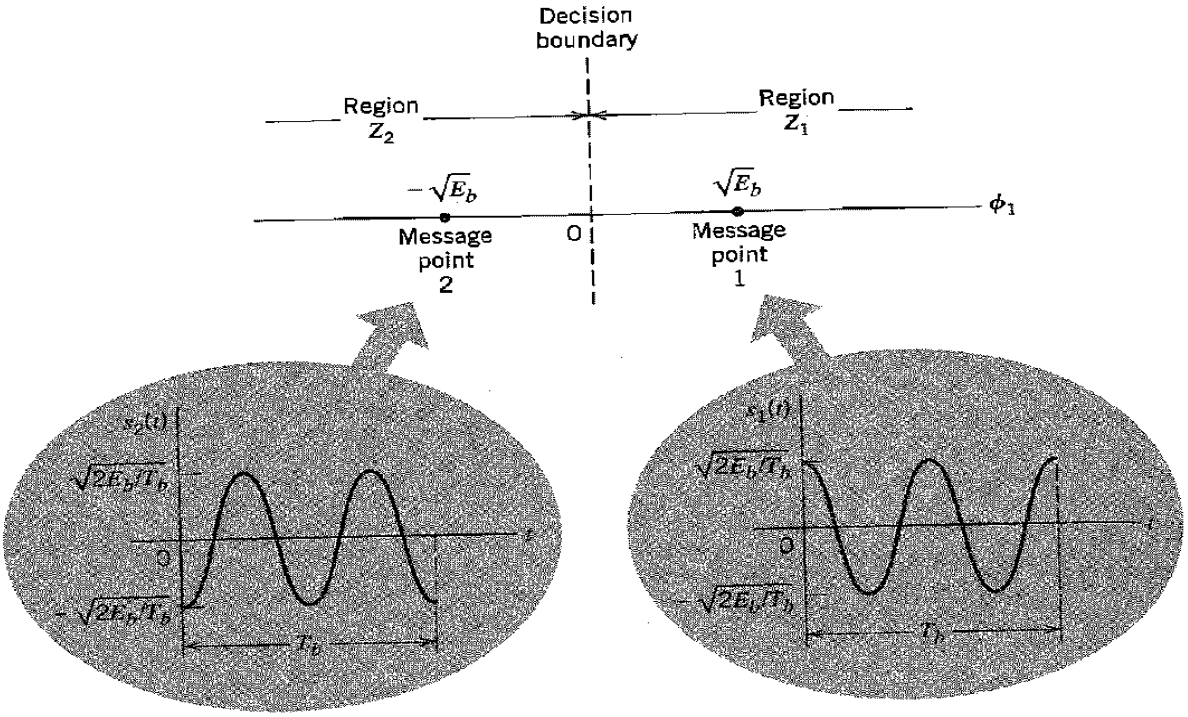
$$\phi_1(t) = \sqrt{\frac{2}{T_b}} \cos(2\pi f_c t), \qquad 0 \leq t < T_b \tag{6.10}$$

Then we may express the transmitted signals $s_1(t)$ and $s_2(t)$ in terms of $\phi_1(t)$ as follows:

$$s_1(t) = \sqrt{E_b}\phi_1(t), \qquad 0 \leq t < T_b \tag{6.11}$$

and

$$s_2(t) = -\sqrt{E_b}\phi_1(t), \qquad 0 \leq t < T_b \tag{6.12}$$



Signal-space diagram for coherent binary PSK system. The waveforms depicting the transmitted signals $s_1(t)$ and $s_2(t)$, displayed in the inserts, assume $n_c = 2$.

A coherent binary PSK system is therefore characterized by having a signal space that is one-dimensional (i.e., $N = 1$), with a signal constellation consisting of two message points (i.e., $M = 2$). The coordinates of the message points are

$$s_{11} = \int_0^{T_b} s_1(t)\phi_1(t)\,dt = +\sqrt{E_b} \tag{6.13}$$

and

$$s_{21} = \int_0^{T_b} s_2(t)\phi_1(t)\,dt = -\sqrt{E_b} \tag{6.14}$$

The message point corresponding to $s_1(t)$ is located at $s_{11} = +\sqrt{E_b}$, and the message point corresponding to $s_2(t)$ is located at $s_{21} = -\sqrt{E_b}$. Figure 6.3 displays the signal-space diagram for binary PSK. This figure also includes two inserts, showing example waveforms of antipodal signals representing $s_1(t)$ and $s_2(t)$. Note that the constellation of Figure 6.3 has minimum average energy.

Error Probability of Binary PSK

To realize a *rule for making a decision* in favor of symbol 1 or symbol 0, we apply Equation (5.59) of Chapter 5. Specifically, we partition the signal space of Figure 6.3 into two regions:

- The set of points closest to message point 1 at $+\sqrt{E_b}$.
- The set of points closest to message point 2 at $-\sqrt{E_b}$.

This is accomplished by constructing the midpoint of the line joining these two message points, and then marking off the appropriate decision regions. In Figure 6.3 these decision regions are marked Z_1 and Z_2, according to the message point around which they are constructed.

The decision rule is now simply to decide that signal $s_1(t)$ (i.e., binary symbol 1) was transmitted if the received signal point falls in region Z_1, and decide that signal $s_2(t)$ (i.e., binary symbol 0) was transmitted if the received signal point falls in region Z_2. Two kinds of erroneous decisions may, however, be made. Signal $s_2(t)$ is transmitted, but the noise is such that the received signal point falls inside region Z_1 and so the receiver decides in favor of signal $s_1(t)$. Alternatively, signal $s_1(t)$ is transmitted, but the noise is such that the received signal point falls inside region Z_2 and so the receiver decides in favor of signal $s_2(t)$.

To calculate the probability of making an error of the first kind, we note from Figure 6.3 that the decision region associated with symbol 1 or signal $s_1(t)$ is described by

$$Z_1: 0 < x_1 < \infty$$

where the observable element x_1 is related to the received signal $x(t)$ by

$$x_1 = \int_0^{T_b} x(t)\phi_1(t)\, dt \tag{6.15}$$

The conditional probability density function of random variable X_1, given that symbol 0 [i.e., signal $s_2(t)$] was transmitted, is defined by

$$\begin{aligned} f_{X_1}(x_1|0) &= \frac{1}{\sqrt{\pi N_0}} \exp\left[-\frac{1}{N_0}(x_1 - s_{21})^2\right] \\ &= \frac{1}{\sqrt{\pi N_0}} \exp\left[-\frac{1}{N_0}(x_1 + \sqrt{E_b})^2\right] \end{aligned} \tag{6.16}$$

The conditional probability of the receiver deciding in favor of symbol 1, given that symbol 0 was transmitted, is therefore

$$\begin{aligned} p_{10} &= \int_0^{\infty} f_{X_1}(x_1|0)\, dx_1 \\ &= \frac{1}{\sqrt{\pi N_0}} \int_0^{\infty} \exp\left[-\frac{1}{N_0}(x_1 + \sqrt{E_b})^2\right] dx_1 \end{aligned} \tag{6.17}$$

Putting

$$z = \frac{1}{\sqrt{N_0}}(x_1 + \sqrt{E_b}) \tag{6.18}$$

and changing the variable of integration from x_1 to z, we may rewrite Equation (6.17) in the compact form

$$\begin{aligned} p_{10} &= \frac{1}{\sqrt{\pi}} \int_{\sqrt{E_b/N_0}}^{\infty} \exp(-z^2)\, dz \\ &= \frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right) \end{aligned} \tag{6.19}$$

where erfc(·) is the complementary error function.

Consider next an error of the second kind. We note that the signal space of Figure 6.3 is symmetric with respect to the origin. It follows therefore that p_{01}, the conditional probability of the receiver deciding in favor of symbol 0, given that symbol 1 was transmitted, also has the same value as in Equation (6.19).

Thus, averaging the conditional error probabilities p_{10} and p_{01}, we find that the *average probability of symbol error* or, equivalently, the *bit error rate for coherent binary PSK* is (assuming equiprobable symbols)

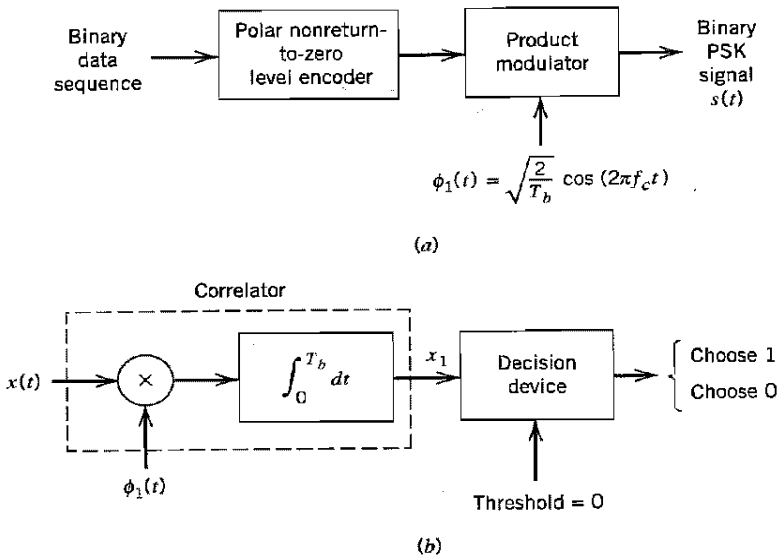
$$P_e = \frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right) \tag{6.20}$$

As we increase the transmitted signal energy per bit, E_b, for a specified noise spectral density N_0, the message points corresponding to symbols 1 and 0 move further apart, and the average probability of error P_e is correspondingly reduced in accordance with Equation (6.20), which is intuitively satisfying.

Generation and Detection of Coherent Binary PSK Signals

To generate a binary PSK signal, we see from Equations (6.8)–(6.10) that we have to represent the input binary sequence in polar form with symbols 1 and 0 represented by constant amplitude levels of $+\sqrt{E_b}$ and $-\sqrt{E_b}$, respectively. This signal transmission encoding is performed by a polar nonreturn-to-zero (NRZ) level encoder. The resulting binary wave and a sinusoidal carrier $\phi_1(t)$, whose frequency $f_c = (n_c/T_b)$ for some fixed integer n_c, are applied to a product modulator, as in Figure 6.4*a*. The carrier and the timing pulses used to generate the binary wave are usually extracted from a common master clock. The desired PSK wave is obtained at the modulator output.

To detect the original binary sequence of 1s and 0s, we apply the noisy PSK signal $x(t)$ (at the channel output) to a correlator, which is also supplied with a locally generated coherent reference signal $\phi_1(t)$, as in Figure 6.4*b*. The correlator output, x_1, is compared with a threshold of zero volts. If $x_1 > 0$, the receiver decides in favor of symbol 1. On the



Block diagrams for (*a*) binary PSK transmitter and (*b*) coherent binary PSK receiver.

other hand, if $x_1 < 0$, it decides in favor of symbol 0. If x_1 is exactly zero, the receiver makes a random guess in favor of 0 or 1.

Power Spectra of Binary PSK Signals

From the modulator of Figure 6.4*a*, we see that the complex envelope of a binary PSK wave consists of an in-phase component only. Furthermore, depending on whether we have symbol 1 or symbol 0 at the modulator input during the signaling interval $0 \le t \le T_b$, we find that this in-phase component equals $+g(t)$ or $-g(t)$, respectively, where $g(t)$ is the *symbol shaping function* defined by

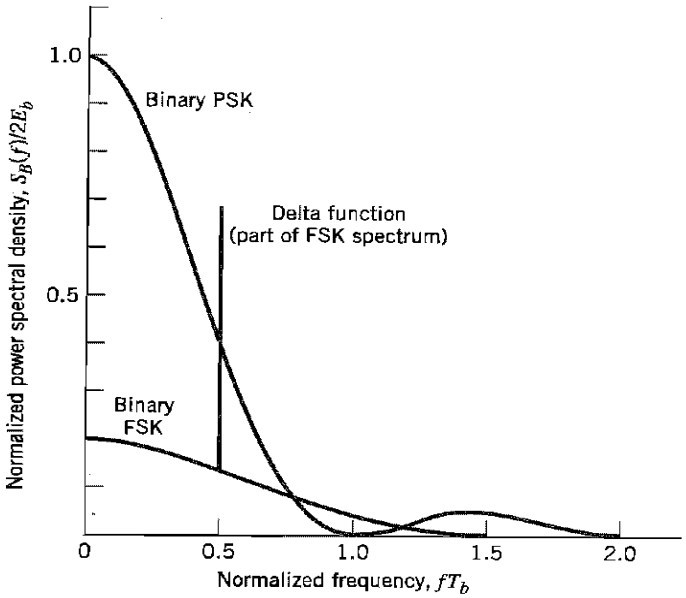
$$g(t) = \begin{cases} \sqrt{\dfrac{2E_b}{T_b}}, & 0 \le t \le T_b \\ 0, & \text{otherwise} \end{cases} \tag{6.21}$$

We assume that the input binary wave is random, with symbols 1 and 0 equally likely and the symbols transmitted during the different time slots being statistically independent. In Example 1.6 of Chapter 1 it is shown that the power spectral density of a random binary wave so described is equal to the energy spectral density of the symbol shaping function divided by the symbol duration. The energy spectral density of a Fourier transformable signal $g(t)$ is defined as the squared magnitude of the signal's Fourier transform. Hence, the baseband power spectral density of a binary PSK signal equals

$$\begin{aligned} S_B(f) &= \frac{2E_b \sin^2(\pi T_b f)}{(\pi T_b f)^2} \\ &= 2E_b \operatorname{sinc}^2(T_b f) \end{aligned} \tag{6.22}$$

This power spectrum falls off as the inverse square of frequency, as shown in Figure 6.5.

Figure 6.5 also includes a plot of the baseband power spectral density of a binary FSK signal, details of which are presented in Section 6.5. Comparison of these two spectra is deferred to that section.



Power spectra of binary PSK and FSK signals.

QUADRIPHASE-SHIFT KEYING

The provision of reliable performance, exemplified by a very low probability of error, is one important goal in the design of a digital communication system. Another important goal is the efficient utilization of channel bandwidth. In this subsection, we study a bandwidth-conserving modulation scheme known as coherent quadriphase-shift keying, which is an example of *quadrature-carrier multiplexing*.

In *quadriphase-shift keying* (QPSK), as with binary PSK, information carried by the transmitted signal is contained in the phase. In particular, the phase of the carrier takes on one of four equally spaced values, such as $\pi/4$, $3\pi/4$, $5\pi/4$, and $7\pi/4$. For this set of values we may define the transmitted signal as

$$s_i(t) = \begin{cases} \sqrt{\frac{2E}{T}} \cos\left[2\pi f_c t + (2i-1)\frac{\pi}{4}\right], & 0 \le t \le T \\ 0, & \text{elsewhere} \end{cases} \tag{6.23}$$

where $i = 1, 2, 3, 4$; E is the transmitted signal energy per symbol, and T is rhe symbol duration. The carrier frequency f_c equals n_c/T for some fixed integer n_c. Each possible value of the phase corresponds to a unique dibit. Thus, for example, we may choose the foregoing set of phase values to represent the *Gray-encoded* set of dibits: 10, 00, 01, and 11, where only a single bit is changed from one dibit to the next.

Signal-Space Diagram of QPSK

Using a well-known trigonometric identity, we may use Equation (6.23) to redefine the transmitted signal $s_i(t)$ for the interval $0 \le t \le T$ in the equivalent form:

$$s_i(t) = \sqrt{\frac{2E}{T}} \cos\left[(2i-1)\frac{\pi}{4}\right] \cos(2\pi f_c t) - \sqrt{\frac{2E}{T}} \sin\left[(2i-1)\frac{\pi}{4}\right] \sin(2\pi f_c t) \tag{6.24}$$

where $i = 1, 2, 3, 4$. Based on this representation, we can make the following observations:

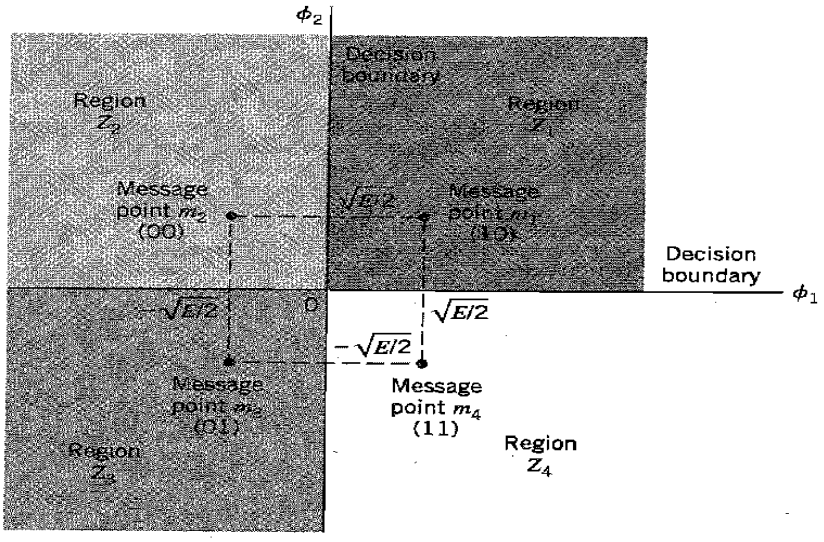
- There are two orthonormal basis functions, $\phi_1(t)$ and $\phi_2(t)$, contained in the expansion of $s_i(t)$. Specifically, $\phi_1(t)$ and $\phi_2(t)$ are defined by a pair of *quadrature carriers*:

$$\phi_1(t) = \sqrt{\frac{2}{T}} \cos(2\pi f_c t), \qquad 0 \le t \le T \tag{6.25}$$

$$\phi_2(t) = \sqrt{\frac{2}{T}} \sin(2\pi f_c t), \qquad 0 \le t \le T \tag{6.26}$$

Signal-space characterization of QPSK

Gray-encoded Input Dibit	*Phase of QPSK Signal (radians)*	*Coordinates of Message Points*	
		s_{i1}	s_{i2}
10	$\pi/4$	$+\sqrt{E/2}$	$-\sqrt{E/2}$
00	$3\pi/4$	$-\sqrt{E/2}$	$-\sqrt{E/2}$
01	$5\pi/4$	$-\sqrt{E/2}$	$+\sqrt{E/2}$
11	$7\pi/4$	$+\sqrt{E/2}$	$+\sqrt{E/2}$



Signal-space diagram of coherent QPSK system.

- There are four message points, and the associated signal vectors are defined by

$$\mathbf{s}_i = \begin{bmatrix} \sqrt{E}\cos\left((2i-1)\frac{\pi}{4}\right) \\ -\sqrt{E}\sin\left((2i-1)\frac{\pi}{4}\right) \end{bmatrix}, \qquad i = 1, 2, 3, 4 \tag{6.27}$$

The elements of the signal vectors, namely, s_{i1} and s_{i2}, have their values summarized in Table 6.1. The first two columns of this table give the associated dibit and phase of the QPSK signal.

Accordingly, a QPSK signal has a two-dimensional signal constellation (i.e., $N = 2$) and four message points (i.e., $M = 4$) whose phase angles increase in a counterclockwise direction, as illustrated in Figure 6.6. As with binary PSK, the QPSK signal has minimum average energy.

Error Probability of QPSK

In a coherent QPSK system, the received signal $x(t)$ is defined by

$$x(t) = s_i(t) + w(t), \qquad \begin{cases} 0 \le t \le T \\ i = 1, 2, 3, 4 \end{cases} \tag{6.28}$$

where $w(t)$ is the sample function of a white Gaussian noise process of zero mean and power spectral density $N_0/2$. Correspondingly, the observation vector $\mathbf{x}$ has two elements, x_1 and x_2, defined by

$$\begin{aligned} x_1 &= \int_0^T x(t)\phi_1(t)\,dt \\ &= \sqrt{E}\cos\left[(2i-1)\frac{\pi}{4}\right] + w_1 \\ &= \pm\sqrt{\frac{E}{2}} + w_1 \end{aligned} \tag{6.29}$$

$$
\begin{aligned}
x_2 &= \int_0^T x(t)\phi_2(t)\,dt \\
&= -\sqrt{E}\sin\left[(2i-1)\frac{\pi}{4}\right] + w_2 \\
&= \mp\sqrt{\frac{E}{2}} + w_2
\end{aligned}
\tag{6.30}
$$

Thus the observable elements x_1 and x_2 are sample values of independent Gaussian random variables with mean values equal to $\pm\sqrt{E/2}$ and $\mp\sqrt{E/2}$, respectively, and with a common variance equal to $N_0/2$.

The decision rule is now simply to decide that $s_1(t)$ was transmitted if the received signal point associated with the observation vector $\mathbf{x}$ falls inside region Z_1, decide that $s_2(t)$ was transmitted if the received signal point falls inside region Z_2, and so on. An erroneous decision will be made if, for example, signal $s_4(t)$ is transmitted but the noise $w(t)$ is such that the received signal point falls outside region Z_4.

To calculate the average probability of symbol error, we note from Equation (6.24) that a coherent QPSK system is in fact equivalent to two coherent binary PSK systems working in parallel and using two carriers that are in phase quadrature; this is merely a statement of the quadrature-carrier multiplexing property of coherent QPSK. The in-phase channel output x_1 and the quadrature channel output x_2 (i.e., the two elements of the observation vector $\mathbf{x}$) may be viewed as the individual outputs of the two coherent binary PSK systems. Thus, according to Equations (6.29) and (6.30), these two binary PSK systems may be characterized as follows:

- The signal energy per bit is $E/2$.
- The noise spectral density is $N_0/2$.

Hence, using Equation (6.20) for the average probability of bit error of a coherent binary PSK system, we may now state that the average probability of bit error in *each* channel of the coherent QPSK system is

$$
\begin{aligned}
P' &= \frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{E/2}{N_0}}\right) \\
&= \frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{E}{2N_0}}\right)
\end{aligned}
\tag{6.31}
$$

Another important point to note is that the bit errors in the in-phase and quadrature channels of the coherent QPSK system are statistically independent. The in-phase channel makes a decision on one of the two bits constituting a symbol (dibit) of the QPSK signal, and the quadrature channel takes care of the other bit. Accordingly, the *average probability of a correct decision* resulting from the combined action of the two channels working together is

$$
\begin{aligned}
P_c &= (1-P')^2 \\
&= \left[1 - \frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{E}{2N_0}}\right)\right]^2 \\
&= 1 - \operatorname{erfc}\left(\sqrt{\frac{E}{2N_0}}\right) + \frac{1}{4}\operatorname{erfc}^2\left(\sqrt{\frac{E}{2N_0}}\right)
\end{aligned}
\tag{6.32}
$$

The average probability of symbol error for coherent QPSK is therefore

$$P_e = 1 - P_c = \operatorname{erfc}\left(\sqrt{\frac{E}{2N_0}}\right) - \frac{1}{4}\operatorname{erfc}^2\left(\sqrt{\frac{E}{2N_0}}\right) \tag{6.33}$$

In the region where $(E/2N_0) \gg 1$, we may ignore the quadratic term on the right-hand side of Equation (6.33), so we approximate the formula for the average probability of symbol error for coherent QPSK as

$$P_e \simeq \operatorname{erfc}\left(\sqrt{\frac{E}{2N_0}}\right) \tag{6.34}$$

The formula of Equation (6.34) may also be derived in another insightful way, using the signal-space diagram of Figure 6.6. Since the four message points of this diagram are circularly symmetric with respect to the origin, we may apply Equation (5.92), reproduced here in the form

$$P_e \leq \frac{1}{2}\sum_{\substack{k=1 \\ k\neq i}}^{4} \operatorname{erfc}\left(\frac{d_{ik}}{2\sqrt{N_0}}\right) \qquad \text{for all } i \tag{6.35}$$

Consider, for example, message point m_1 (corresponding to dibit 10) chosen as the transmitted message point. The message points m_2 and m_4 (corresponding to dibits 00 and 11) are the *closest* to m_1. From Figure 6.6 we readily find that m_1 is equidistant from m_2 and m_4 in a Euclidean sense, as shown by

$$d_{12} = d_{14} = \sqrt{2E}$$

Assuming that E/N_0 is large enough to ignore the contribution of the most distant message point m_3 (corresponding to dibit 01) relative to m_1, we find that the use of Equation (6.35) yields an approximate expression for P_e that is the same as Equation (6.34). Note that in mistaking either m_2 or m_4 for m_1, a single bit error is made; on the other hand, in mistaking m_3 for m_1, two bit errors are made. For a high enough E/N_0, the likelihood of both bits of a symbol being in error is much less than a single bit, which is a further justification for ignoring m_3 in calculating P_e when m_1 is sent.

In a QPSK system, we note that since there are two bits per symbol, the transmitted signal energy per symbol is twice the signal energy per bit, as shown by

$$E = 2E_b \tag{6.36}$$

Thus expressing the average probability of symbol error in terms of the ratio E_b/N_0, we may write

$$P_e \simeq \operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right) \tag{6.37}$$

With Gray encoding used for the incoming symbols, we find from Equations (6.31) and (6.36) that the *bit error rate* of QPSK is exactly

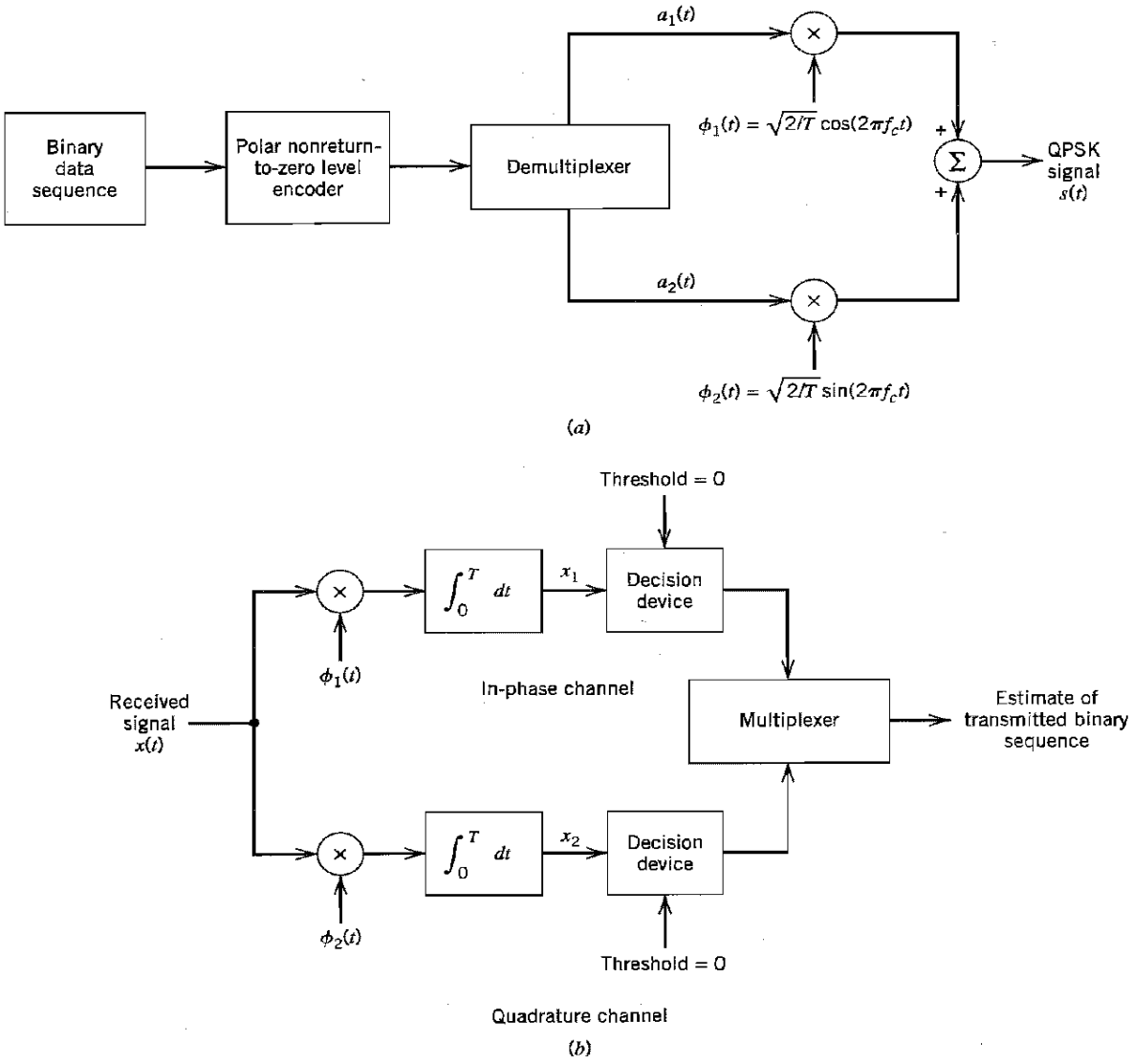
$$\text{BER} = \frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right) \tag{6.38}$$

We may therefore state that a coherent QPSK system achieves the same average probability of bit error as a coherent binary PSK system for the same bit rate and the same E_b/N_0, but uses only half the channel bandwidth. Stated in a different way, for the same E_b/N_0 and therefore the same average probability of bit error, a coherent QPSK system transmits information at twice the bit rate of a coherent binary PSK system for the same channel

bandwidth. For a prescribed performance, QPSK uses channel bandwidth better than binary PSK, which explains the preferred use of QPSK over binary PSK in practice.

Generation and Detection of Coherent QPSK Signals

Consider next the generation and detection of QPSK signals. Figure 6.8*a* shows a block diagram of a typical QPSK transmitter. The incoming binary data sequence is first transformed into polar form by a *nonreturn-to-zero level* encoder. Thus, symbols 1 and 0 are represented by $+\sqrt{E_b}$ and $-\sqrt{E_b}$, respectively. This binary wave is next divided by means of a *demultiplexer* into two separate binary waves consisting of the odd- and even-numbered input bits. These two binary waves are denoted by $a_1(t)$ and $a_2(t)$. We note that in any signaling interval, the amplitudes of $a_1(t)$ and $a_2(t)$ equal s_{i1} and s_{i2}, respectively, depending on the particular dibit that is being transmitted. The two binary waves $a_1(t)$ and $a_2(t)$ are used to modulate a pair of quadrature carriers or orthonormal basis functions: $\phi_1(t)$ equal to $\sqrt{2/T}\cos(2\pi f_c t)$ and $\phi_2(t)$ equal to $\sqrt{2/T}\sin(2\pi f_c t)$. The result is a pair of



Block diagrams of (*a*) QPSK transmitter and (*b*) coherent QPSK receiver.

binary PSK signals, which may be detected independently due to the orthogonality of $\phi_1(t)$ and $\phi_2(t)$. Finally, the two binary PSK signals are added to produce the desired QPSK signal.

The QPSK receiver consists of a pair of correlators with a common input and supplied with a locally generated pair of coherent reference signals $\phi_1(t)$ and $\phi_2(t)$, as in Figure 6.8*b*. The correlator outputs x_1 and x_2, produced in response to the received signal $x(t)$, are each compared with a threshold of zero. If $x_1 > 0$, a decision is made in favor of symbol 1 for the in-phase channel output, but if $x_1 < 0$, a decision is made in favor of symbol 0. Similarly, if $x_2 > 0$, a decision is made in favor of symbol 1 for the quadrature channel output, but if $x_2 < 0$, a decision is made in favor of symbol 0. Finally, these two binary sequences at the in-phase and quadrature channel outputs are combined in a *multiplexer* to reproduce the original binary sequence at the transmitter input with the minimum probability of symbol error in an AWGN channel.

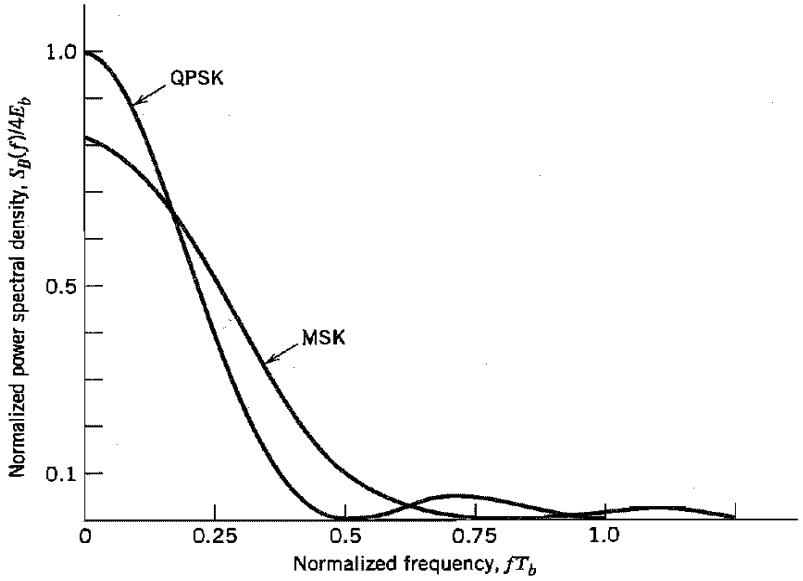
Power Spectra of QPSK Signals

Assume that the binary wave at the modulator input is random, with symbols 1 and 0 being equally likely, and with the symbols transmitted during adjacent time slots being statistically independent. We make the following observations pertaining to the in-phase and quadrature components of a QPSK signal:

1. Depending on the dibit sent during the signaling interval $-T_b \le t \le T_b$, the in-phase component equals $+g(t)$ or $-g(t)$, and similarly for the quadrature component. The $g(t)$ denotes the symbol shaping function, defined by

$$g(t) = \begin{cases} \sqrt{\frac{E}{T}}, & 0 \le t \le T \\ 0, & \text{otherwise} \end{cases} \tag{6.39}$$

Hence, the in-phase and quadrature components have a common power spectral density, namely, $E\,\text{sinc}^2(Tf)$.



Power spectra of QPSK and MSK signals.

2. The in-phase and quadrature components are statistically independent. Accordingly, the baseband power spectral density of the QPSK signal equals the sum of the individual power spectral densities of the in-phase and quadrature components, so we may write

$$\begin{aligned} S_B(f) &= 2E\ \text{sinc}^2(Tf) \\ &= 4E_b\ \text{sinc}^2(2T_bf) \end{aligned} \tag{6.40}$$

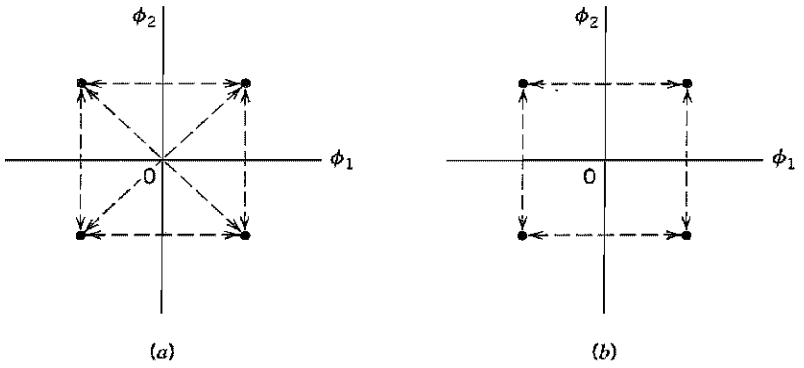
Figure 6.9 plots $S_B(f)$, normalized with respect to $4E_b$, versus the normalized frequency fT_b. This figure also includes a plot of the baseband power spectral density of a certain form of binary FSK called minimum shift keying, the evaluation of which is presented in Section 6.5. Comparison of these two spectra is deferred to that section.

OFFSET QPSK

The signal space diagram of Figure 6.10*a* embodies all the possible phase transitions that can arise in the generation of a QPSK signal. More specifically, examining the QPSK waveform illustrated in Figure 6.7 for Example 6.1, we may make the following observations:

1. The carrier phase changes by ±180 degrees whenever both the in-phase and quadrature components of the QPSK signal changes sign. An example of this situation is illustrated in Figure 6.7 when the input binary sequence switches from dibit 01 to dibit 10.
2. The carrier phase changes by ±90 degrees whenever the in-phase or quadrature component changes sign. An example of this second situation is illustrated in Figure 6.7 when the input binary sequence switches from dibit 10 to dibit 00, during which the in-phase component changes sign, whereas the quadrature component is unchanged.
3. The carrier phase is unchanged when neither the in-phase component nor the quadrature component changes sign. This last situation is illustrated in Figure 6.7 when dibit 10 is transmitted in two successive symbol intervals.

Situation 1 and, to a much lesser extent, situation 2 can be of a particular concern when the QPSK signal is filtered during the course of transmission, prior to detection. Specifically, the 180- and 90-degree shifts in carrier phase can result in changes in the carrier amplitude (i.e., envelope of the QPSK signal), thereby causing additional symbol errors on detection.



Possible paths for switching between the message points in (*a*) QPSK and (*b*) offset QPSK.

The extent of amplitude fluctuations exhibited by QPSK signals may be reduced by using *offset QPSK*.[2] In this variant of QPSK, the bit stream responsible for generating the quadrature component is delayed (i.e., offset) by half a symbol interval with respect to the bit stream responsible for generating the in-phase component. Specifically, the two basis functions of offset QPSK are defined by

$$\phi_1(t) = \sqrt{\frac{2}{T}}\cos(2\pi f_c t), \qquad 0 \le t \le T \tag{6.41}$$

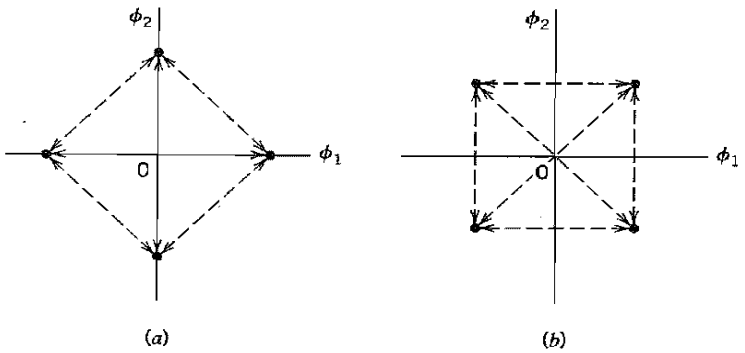
$$\phi_2(t) = \sqrt{\frac{2}{T}}\sin(2\pi f_c t), \qquad \frac{T}{2} \le t \le \frac{3T}{2} \tag{6.42}$$

Accordingly, unlike QPSK, the phase transitions likely to occur in offset QPSK are confined to ± 90 degrees, as indicated in the signal space diagram of Figure 6.10*b*. However, ± 90 degree phase transitions in offset QPSK occur twice as frequently but with half the intensity encountered in QPSK. Since, in addition to ± 90-degree phase transitions, ± 180-degree phase transitions also occur in QPSK, we find that amplitude fluctuations in offset QPSK due to filtering have a smaller amplitude than in the case of QPSK.

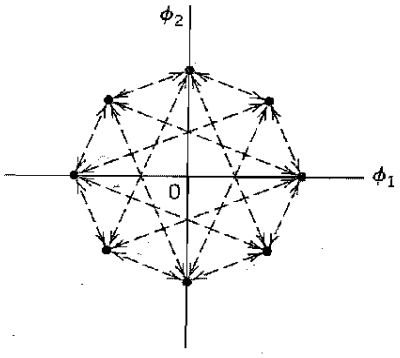
Despite the delay $T/2$ applied to the basis function $\phi_2(t)$ in Equation (6.42) compared to that in Equation (6.26), the offset QPSK has exactly the same probability of symbol error in an AWGN channel as QPSK. The equivalence in noise performance between these phase-shift keying schemes assumes the use of coherent detection. The reason for the equivalence is that the statistical independence of the in-phase and quadrature components applies to both QPSK and offset QPSK. We may therefore say that the error probability in the in-phase or quadrature channel of a coherent offset QPSK receiver is still equal to $(1/2)\,\text{erfc}(\sqrt{E/2N_0})$. Hence the formula of Equation (6.34) applies equally well to the offset QPSK.

$\pi/4$-Shifted QPSK

An ordinary QPSK signal may reside in either one of the two commonly used constellations shown in Figures 6.11*a* and 6.11*b*, which are shifted by $\pi/4$ radians with respect to each other. In another variant of QPSK known as *$\pi/4$-shifted QPSK*,[3] the carrier phase used for the transmission of successive symbols (i.e., dibits) is alternately picked from one of the two QPSK constellations in Figure 6.11 and then the other. It follows therefore that a $\pi/4$-shifted QPSK signal may reside in any one of eight possible phase states, as indicated



Two commonly used signal constellations for QPSK; the arrows indicate the paths along which the QPSK modulator can change its state.



Eight possible phase states for the $\pi/4$-shifted QPSK modulator.

in Figure 6.12. The four dashed lines emanating from each possible message point in Figure 6.12 define the phase transitions that are feasible in $\pi/4$-shifted QPSK.

Table 6.2 summarizes a possible set of relationships between the phase transitions in this new digital modulation scheme and the incoming Gray-encoded dibits. For example, if the modulator is in one of the phase states portrayed in Figure 6.11*b*, then on receiving the dibit 00 it shifts into a phase state portrayed in Figure 6.11*a* by rotating through $\pi/4$ radians in a counterclockwise direction.

Attractive features of the $\pi/4$-shifted QPSK scheme include the following:

- The phase transitions from one symbol to the next are restricted to $\pm\pi/4$ and $\pm 3\pi/4$ radians, which is to be contrasted with the $\pm\pi/2$ and $\pm\pi$ phase transitions in QPSK. Consequently, envelope variations of $\pi/4$-shifted QPSK signals due to filtering are significantly reduced, compared to those in QPSK.
- Unlike offset QPSK signals, $\pi/4$-shifted QPSK signals can be noncoherently detected, thereby considerably simplifying the receiver design. Moreover, like QPSK signals, $\pi/4$-shifted QPSK can be differently encoded, in which case we should really speak of *$\pi/4$-shifted DQPSK*.

The generation of $\pi/4$-shifted DQPSK symbols, represented by the symbol pair (I, Q), is described by the following pair of relationships (see Problem 6.13):

$$I_k = \cos(\theta_{k-1} + \Delta\theta_k) = \cos\theta_k \tag{6.43}$$

$$Q_k = \sin(\theta_{k-1} + \Delta\theta_k) = \sin\theta_k \tag{6.44}$$

Correspondence between input dibit and phase change for $\pi/4$-shifted DQPSK

Gray-Encoded Input Dibit	*Phase Change, $\Delta\theta$ (radians)*
00	$\pi/4$
01	$3\pi/4$
11	$-3\pi/4$
10	$-\pi/4$

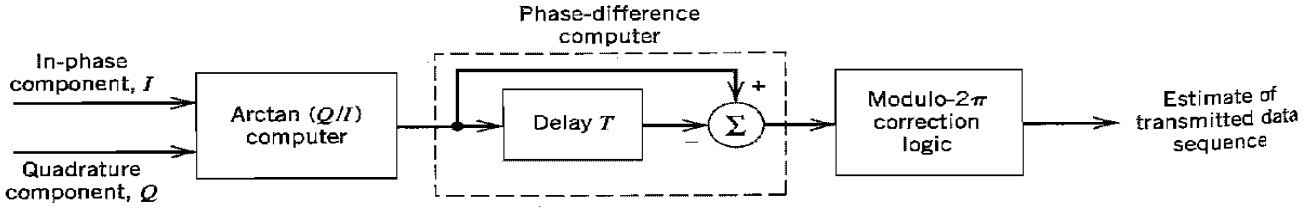
Detection of π/4-Shifted DQPSK Signals

Having familiarized ourselves with the generation of $\pi/4$-shifted DQPSK signals, we go on to consider their differential detection. Given the noisy channel output $x(t)$, the receiver first computes the projections of $x(t)$ onto the basis functions $\phi_1(t)$ and $\phi_2(t)$. The resulting outputs, denoted by I and Q, respectively, are applied to a *differential detector* that consists of the following components, as indicated in Figure 6.13:

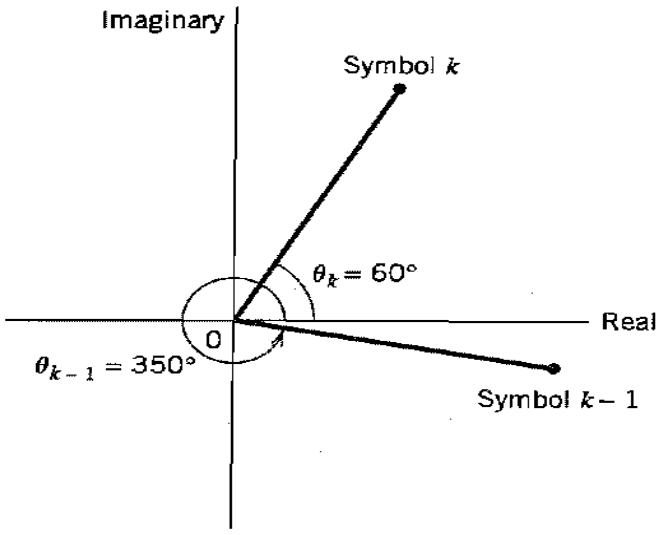
- *Arctangent computer* for extracting the phase angle θ of the channel output (received signal).
- *Phase-difference computer* for determining the change in the phase θ occurring over one symbol interval.
- *Modulo-2π correction logic* for correcting errors due to the possibility of phase angles wrapping around the real axis.

Elaborating further on the latter point, let $\Delta\theta_k$ denote the computed phase difference between θ_k and θ_{k-1} representing the phase angles of the channel output for symbols k and $k-1$, respectively. Then the modulo-2π correction logic operates as follows:

$$\begin{aligned}&\text{IF } \Delta\theta_k < -180 \text{ degrees THEN } \Delta\theta_k = \Delta\theta_k + 360 \text{ degrees}\\ &\text{IF } \Delta\theta_k > 180 \text{ degrees THEN } \Delta\theta_k = \Delta\theta_k - 360 \text{ degrees}\end{aligned} \tag{6.45}$$



Block diagram of the $\pi/4$-shifted DQPSK detector.



Illustrating the possibility of phase angles wrapping around the positive real axis.

To illustrate the need for this phase correction, consider the situation depicted in Figure 6.14, where $\theta_{k-1} = 350$ degrees and $\theta_k = 60$ degrees, both phase angles being measured in a counterclockwise direction. From this figure we readily see that the phase change $\Delta\theta_k$, measured in a counterclockwise direction, is 70 degrees. However, without correction the phase change $\Delta\theta_k$ is computed as 60 degrees $-$ 350 degrees $= -290$ degrees. Applying the first line of Equation (6.45), the modulo-2π correction logic compensates for the wrap-around the positive real axis, yielding the corrected result

$$\Delta\theta_k = -290 \text{ degrees} + 360 \text{ degrees} = 70 \text{ degrees}$$

The tangent type differential detector of Figure 6.13 for the demodulation of $\pi/4$-shifted DQPSK signals is relatively simple to implement. It offers a satisfactory performance in a Rayleigh fading channel as in a static multipath environment. However, when the multipath environment is time varying as experienced in a commercial digital wireless communication system, computer simulation results appear to show that the receiver performance degrades very rapidly.

Differential Phase-Shift Keying

As remarked earlier in Section 6.1, we may view *differential phase-shift keying* (DPSK) as the noncoherent version of PSK. It eliminates the need for a coherent reference signal at the receiver by combining two basic operations at the transmitter: (1) *differential encoding* of the input binary wave and (2) *phase-shift keying*—hence, the name, *differential phase-shift keying* (DPSK). In effect, to send symbol 0, we phase advance the current signal waveform by 180 degrees, and to send symbol 1 we leave the phase of the current signal waveform unchanged. The receiver is equipped with a *storage* capability, so that it can measure the *relative phase difference* between the waveforms received during two successive bit intervals. Provided that the unknown phase θ contained in the received wave varies slowly (that is, slow enough for it to be considered essentially constant over two bit intervals), the phase difference between waveforms received in two successive bit intervals will be independent of θ.

DPSK is another example of noncoherent orthogonal modulation, when it is considered over two bit intervals. Suppose the transmitted DPSK signal equals $\sqrt{E_b/2T_b}\cos(2\pi f_c t)$ for $0 \le t \le T_b$, where T_b is the bit duration and E_b is the signal energy per bit. Let $s_1(t)$ denote the transmitted DPSK signal for $0 \le t \le 2T_b$ for the case when we have binary symbol 1 at the transmitter input for the second part of this interval, namely, $T_b \le t \le 2T_b$. The transmission of symbol 1 leaves the carrier phase unchanged over the interval $0 \le t \le 2T_b$, and so we define $s_1(t)$ as

$$s_1(t) = \begin{cases} \sqrt{\dfrac{E_b}{2T_b}}\cos(2\pi f_c t), & 0 \le t \le T_b \\ \sqrt{\dfrac{E_b}{2T_b}}\cos(2\pi f_c t), & T_b \le t \le 2T_b \end{cases} \tag{6.182}$$

Let $s_2(t)$ denote the transmitted DPSK signal for $0 \le t \le 2T_b$ for the case when we have binary symbol 0 at the transmitter input for $T_b \le t \le 2T_b$. The transmission of 0 advances the carrier phase by 180 degrees, and so we define $s_2(t)$ as

$$s_2(t) = \begin{cases} \sqrt{\dfrac{E_b}{2T_b}}\cos(2\pi f_c t), & 0 \le t \le T_b \\ \sqrt{\dfrac{E_b}{2T_b}}\cos(2\pi f_c t + \pi), & T_b \le t \le 2T_b \end{cases} \tag{6.183}$$

We readily see from Equations (6.182) and (6.183) that $s_1(t)$ and $s_2(t)$ are indeed orthogonal over the two-bit interval $0 \leq t \leq 2T_b$. In other words, DPSK is a special case of noncoherent orthogonal modulation with $T = 2T_b$ and $E = 2E_b$. Hence, using Equation (6.163), we find that the *bit error rate for DPSK* is given by

$$P_e = \frac{1}{2}\exp\left(-\frac{E_b}{N_0}\right) \tag{6.184}$$

which provides a gain of 3 dB over noncoherent FSK for the same E_b/N_0.

Generation and Detection of DPSK

The next issue to be considered is the generation of DPSK signals. The differential encoding process at the transmitter input starts with an arbitrary first bit, serving as reference. Let $\{d_k\}$ denote the differentially encoded sequence with this added reference bit. We now introduce the following definitions in the generation of this sequence:

- If the incoming binary symbol b_k is 1, leave the symbol d_k unchanged with respect to the previous bit.
- If the incoming binary symbol b_k is 0, change the symbol d_k with respect to the previous bit.

The differentially encoded sequence $\{d_k\}$ thus generated is used to phase-shift a carrier with phase angles 0 and π radians representing symbols 1 and 0, respectively. The differential-phase encoding process is illustrated in Table 6.7. Note that d_k is the complement of the modulo-2 sum of b_k and d_{k-1}.

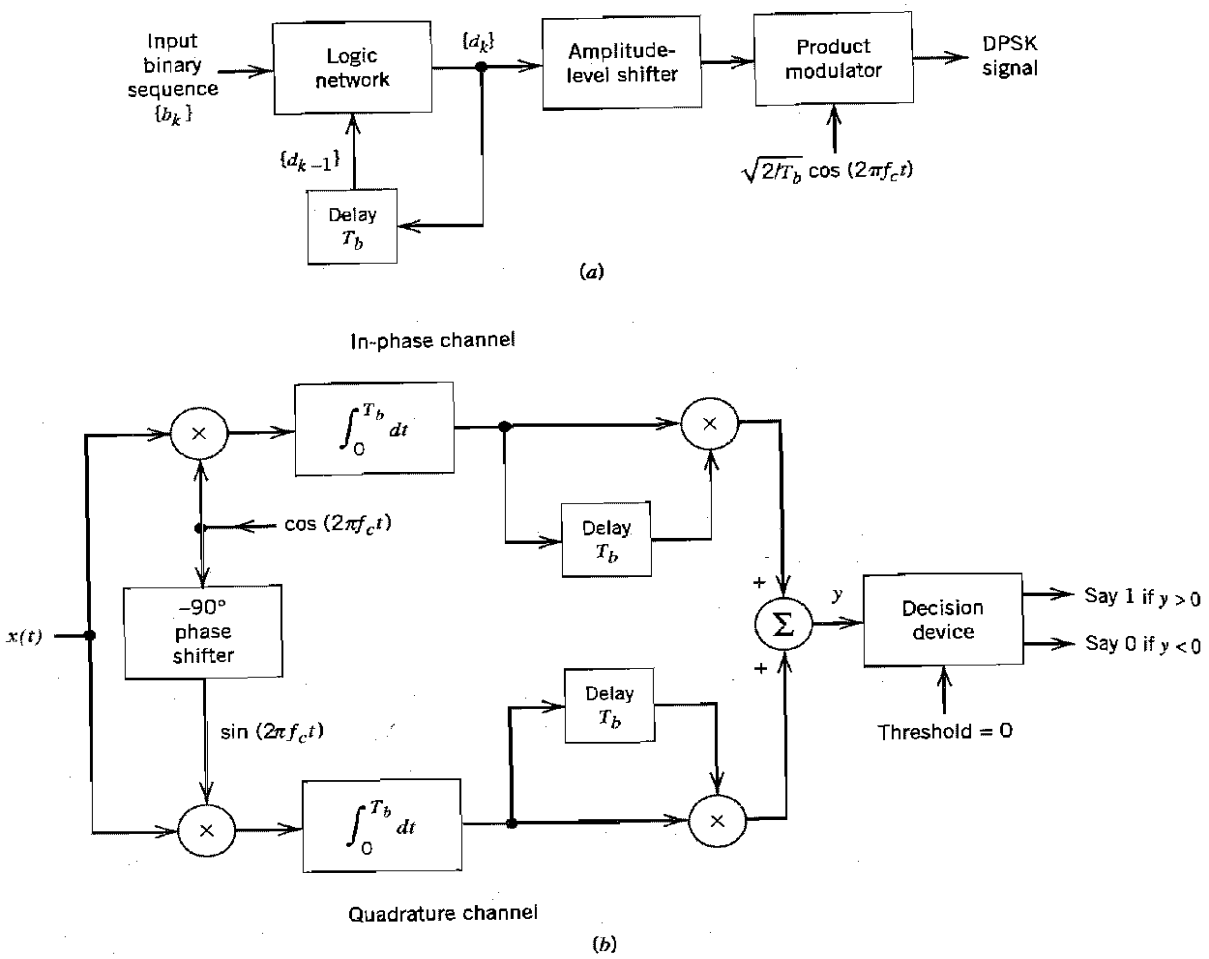
The block diagram of a DPSK transmitter is shown in Figure 6.43*a*. It consists, in part, of a logic network and a one-bit delay element interconnected so as to convert the raw binary sequence $\{b_k\}$ into a differentially encoded sequence $\{d_k\}$. This sequence is amplitude-level encoded and then used to modulate a carrier wave of frequency f_c, thereby producing the desired DPSK signal.

Suppose next, in differentially coherent detection of binary DPSK, the carrier phase is unknown. Then, in light of the receiver being equipped with an in-phase and a quadrature channel, we have a signal space diagram where the received signal points are $(A\cos\theta, A\sin\theta)$ and $(-A\cos\theta, -A\sin\theta)$, with θ denoting the unknown phase and A denoting the amplitude. This geometry of possible signals is illustrated in Figure 6.44. The receiver measures the coordinates (x_{I_0}, x_{Q_0}) at time $t = T_b$ and (x_{I_1}, x_{Q_1}) at time $t = 2T_b$. The issue to be resolved is whether these two points map to the same signal point or different ones. Recognizing that the two vectors $\mathbf{x}_0$ and $\mathbf{x}_1$, with end points (x_{I_0}, x_{Q_0}) and (x_{I_1}, x_{Q_1}) are pointed roughly in the same direction if their inner product is positive, we may formulate the hypothesis test as follows:

Is the inner product $\mathbf{x}_0^T\mathbf{x}_1$ positive or negative?

Illustrating the generation of DPSK signal

$\{b_k\}$		1	0	0	1	0	0	1	1
$\{d_{k-1}\}$		1	1	0	1	1	0	1	1
Differentially encoded sequence $\{d_k\}$	1	1	0	1	1	0	1	1	1
Transmitted phase (radians)	0	0	π	0	0	π	0	0	0



Block diagrams of (*a*) DPSK transmitter and (*b*) DPSK receiver.

Accordingly, we may write

$$x_{I_0}x_{I_1} + x_{Q_0}x_{Q_1} \underset{\text{say 0}}{\overset{\text{say 1}}{\gtrless}} 0 \tag{6.185}$$

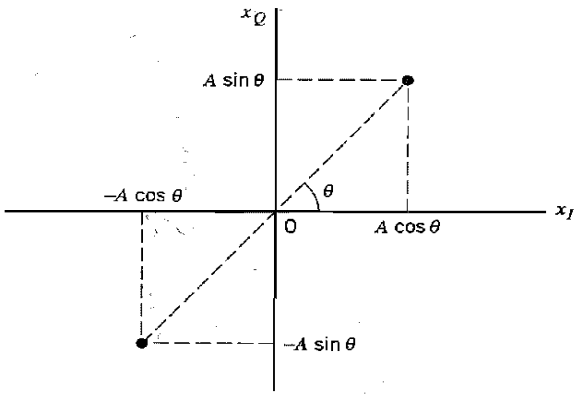
We now note the following identity:

$$x_{I_0}x_{I_1} + x_{Q_0}x_{Q_1} = \frac{1}{4}\left[(x_{I_0} + x_{I_1})^2 - (x_{I_0} - x_{I_1})^2 + (x_{Q_0} + x_{Q_1})^2 - (x_{Q_0} - x_{Q_1})^2\right]$$

Hence substituting this identity into Equation (6.185) and multiplying both sides of the test by 4, we get the equivalent test:

$$(x_{I_0} + x_{I_1})^2 + (x_{Q_0} + x_{Q_1})^2 - (x_{I_0} - x_{I_1})^2 - (x_{Q_0} - x_{I_1})^2 \underset{\text{say 0}}{\overset{\text{say 1}}{\gtrless}} 0 \tag{6.186}$$

The decision-making process may therefore be thought of as testing whether the point (x_{I_0}, x_{Q_0}) is closer to (x_{I_1}, x_{Q_1}) or its image $(-x_{I_1}, -x_{Q_1})$.



Signal-space diagram of received DPSK signal.

Thus, the *optimum receiver*[12] for differentially coherent detection of binary DPSK is as shown in Figure 6.43*b*, which follows directly from Equation (6.185). This implementation merely requires that *sample* values be stored, thereby avoiding the need for fancy delay lines that may be needed otherwise. The equivalent receiver implementation that tests squared elements as in Equation (6.186) is more complicated, but its use makes the *analysis* easier to handle in that the two signals to be considered are orthogonal over the interval $(0, 2T_b)$; hence, the noncoherent orthogonal demodulation analysis applies.

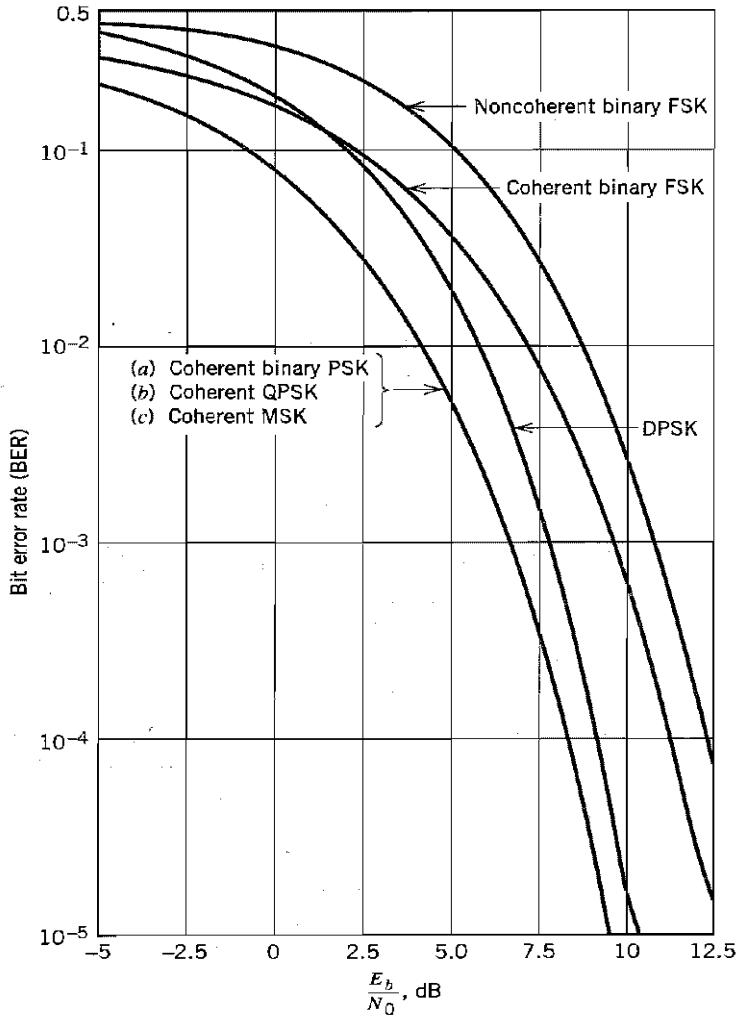
Comparison of Digital Modulation Schemes Using a Single Carrier

PROBABILITY OF ERROR

we have summarized the expressions for the bit error rate (BER) for coherent binary PSK, conventional coherent binary FSK with one-bit decoding, DPSK, noncoherent binary FSK, coherent QPSK, and coherent MSK, when operating over an AWGN channel. In Figure 6.45 we have used the expressions summarized in Table 6.8 to plot the BER as a function of the signal energy per bit-to-noise spectral density ratio, E_b/N_0.

Summary of formulas for the bit error rate of different digital modulation schemes

Signaling Scheme	*Bit Error Rate*
(a) Coherent binary PSK Coherent QPSK Coherent MSK	$\frac{1}{2}\operatorname{erfc}(\sqrt{E_b/N_0})$
(b) Coherent binary FSK	$\frac{1}{2}\operatorname{erfc}(\sqrt{E_b/2N_0})$
(c) DPSK	$\frac{1}{2}\exp(-E_b/N_0)$
(d) Noncoherent binary FSK	$\frac{1}{2}\exp(-E_b/2N_0)$



Comparison of the noise performance of different PSK and FSK schemes.

Based on the performance curves shown in Figure 6.45, the summary of formulas given in Table 6.8, and the defining equations for the pertinent modulation formats, we can make the following statements:

1. The bit error rates for all the systems decrease monotonically with increasing values of E_b/N_0; the defining curves have a similar shape in the form of a *waterfall*.
2. For any value of E_b/N_0, coherent binary PSK, QPSK, and MSK produce a smaller bit error rate than any of the other modulation schemes.
3. Coherent binary PSK and DPSK require an E_b/N_0 that is 3 dB less than the corresponding values for conventional coherent binary FSK and noncoherent binary FSK, respectively, to realize the same bit error rate.
4. At high values of E_b/N_0, DPSK and noncoherent binary FSK perform almost as well (to within about 1 dB) as coherent binary PSK and conventional coherent binary FSK, respectively, for the same bit rate and signal energy per bit.
5. In coherent QPSK, two orthogonal carriers $\sqrt{2/T}\cos(2\pi f_c t)$ and $\sqrt{2/T}\sin(2\pi f_c t)$ are used, where the carrier frequency f_c is an integer multiple of the symbol rate

$1/T$, with the result that two independent bit streams can be transmitted simultaneously and subsequently detected in the receiver.

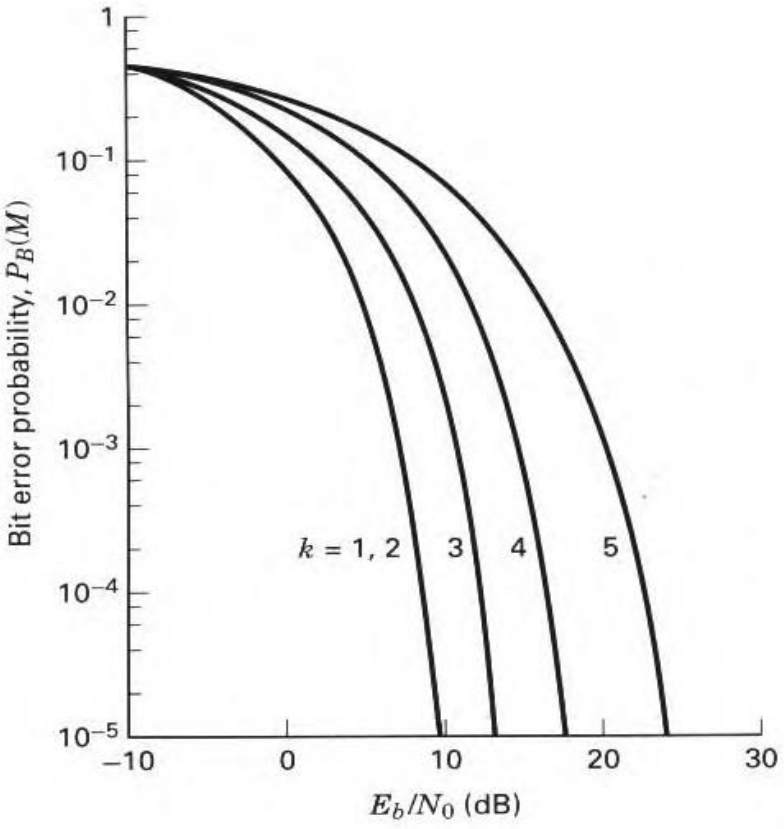
6. In the case of coherent MSK, there are two orthogonal carriers, namely, $\sqrt{2/T_b}\cos(2\pi f_c t)$ and $\sqrt{2/T_b}\sin(2\pi f_c t)$, which are modulated by the two antipodal symbol shaping pulses $\cos(\pi t/2T_b)$ and $\sin(\pi t/2T_b)$, respectively, over $2T_b$ intervals, where T_b is the bit duration. Correspondingly, the receiver uses a coherent phase decoding process over two successive bit intervals to recover the original bit stream.
7. The MSK scheme differs from its counterpart, the QPSK, in that its receiver has *memory*. In particular, the MSK receiver makes decisions based on observations over two successive bit intervals. Thus, although the transmitted signal has a binary format represented by the transmission of two distinct frequencies, the presence of memory in the receiver makes it assume a two-dimensional signal space diagram. There are four message points, depending on which binary symbol (0 or 1) was sent and the past phase history of the FSK signal.

UNIT V M-ARY SIGNALING AND INTRODUCTION TO SPREAD SPECTRUM TECHNIQUES
M-ary signaling, vectoral view of MPSK and MFSK signaling, symbol error performance of M-ary systems – Introduction – Discrete Sequence Spread Spectrum technique – Use of Spread Spectrum with CDMA-Ranging Using Discrete Sequence Spread Spectrum – Frequency Hopping Spread Spectrum – Generation & Characteristics of PN Sequence.

M-ary Signaling

Let us review M-ary signaling. The processor considers k bits at a time. It instructs the modulator to produce one of $M = 2^k$ waveforms; binary signaling is the special case where $k = 1$. Does M-ary signaling improve or degrade error performance? (Be careful with your answer—the question is a loaded one.) Figure 4.28 illustrates the probability of bit error $P_B(M)$ versus E_b/N_0 for coherently detected *orthogonal* M-ary signaling over a Gaussian channel. Figure 4.29 similarly illustrates $P_B(M)$ versus E_b/N_0 for coherently detected *multiple phase* M-ary signaling over a Gaussian channel. In which direction do the curves move as the value of k (or M) increases? From Figure 4.27 we know the directions of curve movement for improved and degraded error performance. In Figure 4.28, as k increases, the curves move in the direction of improved error performance. In Figure 4.29, as k increases, the curve move in the direction of degraded error performance. Such movement tells us that M-ary signaling produces improved error performance with orthogonal signaling and degraded error performance with multiple phase signaling. Can that be true? Why would anyone ever use multiple phase PSK signaling if it provides degraded error performance compared to binary PSK signaling? It *is* true, and many systems do use multiple phase signaling. The question, as stated, is loaded because it implies that error probability versus E_b/N_0 is the *only* performance criterion; there are many others (e.g., bandwidth, throughput, complexity, cost), but in Figures 4.28 and 4.29 error performance is the characteristic that stands out explicitly.

A performance characteristic that is not explicitly seen in Figures 4.28 and 4.29 is the required system bandwidth. For the curves characterizing M-ary orthogonal signals in Figure 4.28, as k increases, the required bandwidth also increases. For the M-ary multiple phase curves in Figure 4.29, as k increases, a larger bit rate can be transmitted within the same bandwidth. In other words, for a fixed data rate, the required bandwidth is decreased. Therefore, *both* the orthogonal and multiple phase error performance curves tell us that M-ary signaling represents a vehicle for performing a system trade-off. In the case of orthogonal signaling, error performance improvement can be achieved at the expense of bandwidth. In the case of multiple phase signaling, bandwidth performance can be achieved at the expense of error performance. Error performance versus bandwidth performance, a fundamental communications trade-off,



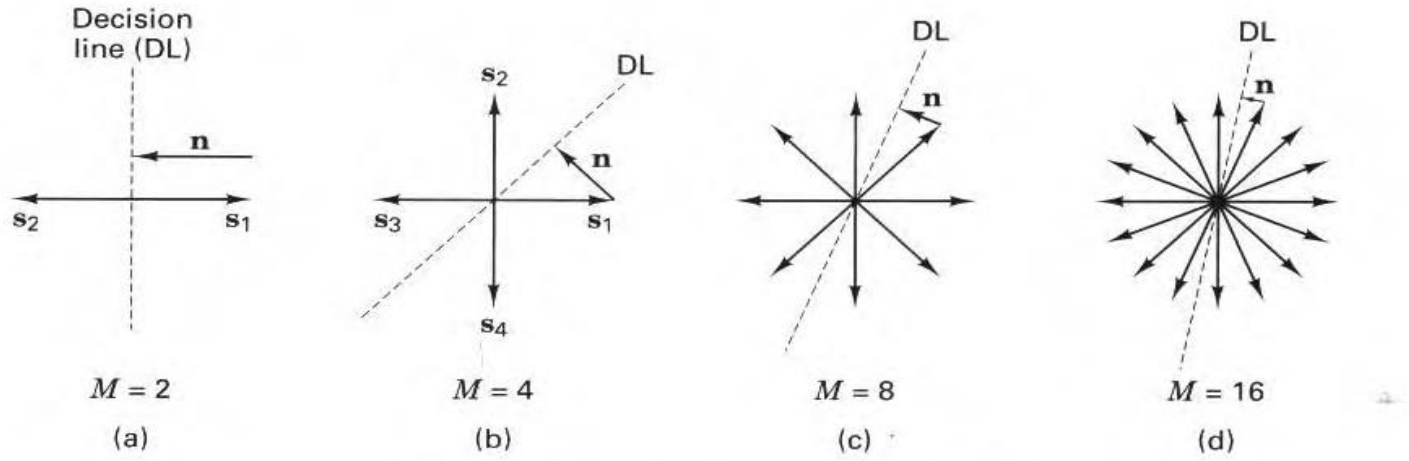
Bit error probability for coherently detected multiple phase signaling.

Vectorial View of MPSK Signaling

MPSK signal sets for $M = 2$, 4, 8, and 16. In Figure 4.30a we see the binary ($k = 2$, $M = 2$) antipodal vectors $\mathbf{s}_1$ and $\mathbf{s}_2$ positioned 180° apart. The decision boundary is drawn so as to partition the signal space into two regions. On the figure is also shown a noise vector $\mathbf{n}$ equal in magnitude to $\mathbf{s}_1$. The figure establishes the magnitude and orientation of the minimum energy noise vector that would cause the detector to make a symbol error.

In Figure 4.30b we see the 4-ary ($k = 2$, $M = 4$) vectors positioned 90° apart. The decision boundaries (only one line is drawn) divide the signal space into four regions. Again a noise vector $\mathbf{n}$ is drawn (from the head of a signal vector, normal to the closest decision boundary) to illustrate the minimum energy noise vector that would cause the detector to make a symbol error. Notice that the minimum energy noise vector of Figure 4.30b is smaller than that of Figure 4.30a, illustrating that the 4-ary system is more vulnerable to noise than the 2-ary system (signal energy being equal for each case). As we move on to Figure 4.30c for the 8-ary case and Figure 4.30d for the 16-ary case, it should be clear that for multiple phase signaling, as M increases, we are crowding more signal vectors into the signal plane. As the vectors are moved closer together, a smaller amount of noise energy is required to cause an error.

Figure 4.30 adds some insight as to why the curves of Figure 4.29 behave as they do as k is increased. Figure 4.30 also provides some insight into a basic trade-



MPSK signal sets for $M = 2, 4, 8, 16$.

off in multiple phase signaling. Crowding more signal vectors into the signal space is tantamount to increasing the data rate without increasing the system bandwidth (the vectors are all confined to the same plane). In other words, we have increased the bandwidth utilization at the expense of error performance. Look at Figure 4.30d, where the error performance is worse than any of the other examples in Figure 4.30. How might we "buy back" the degraded error performance? In other words, what can we trade-off so that the distance between neighboring signal vectors in Figure 4.30d is increased to that in Figure 4.30a? We can increase the signal strength (make the signal vectors larger) until the minimum distance from the head of a signal vector to a decision line equals the length of the noise vector in Figure 4.30a. Therefore, in a multiple phase system, as M is increased, we can either achieve improved bandwidth performance at the expense of error performance, or if we increase the E_b/N_0 so that the error probability is not degraded, we can achieve improved bandwidth performance at the expense of increasing E_b/N_0.

Note that Figure 4.30 has been sketched so that all phasors have the same length for any of the M-ary cases. This is tantamount to saying that the comparisons are being considered for a fixed E_s/N_0, where E_s is symbol energy. The figure can also be drawn for a fixed E_b/N_0, in which case the phasor magnitudes would increase with increasing M. The phasors for $M = 4$, 8, and 16 would then have lengths greater than the $M = 2$ case by the factors $\sqrt{2}$, $\sqrt{3}$, and 2 respectively. We would still see crowding and increased vulnerability to noise, with increasing M, but the appearance would not be as pronounced as it is in Figure 4.30.

BPSK and QPSK Have the Same Bit Error Probability

In Equation (3.30) we stated the general relationship between E_b/N_0 and S/N which is rewritten

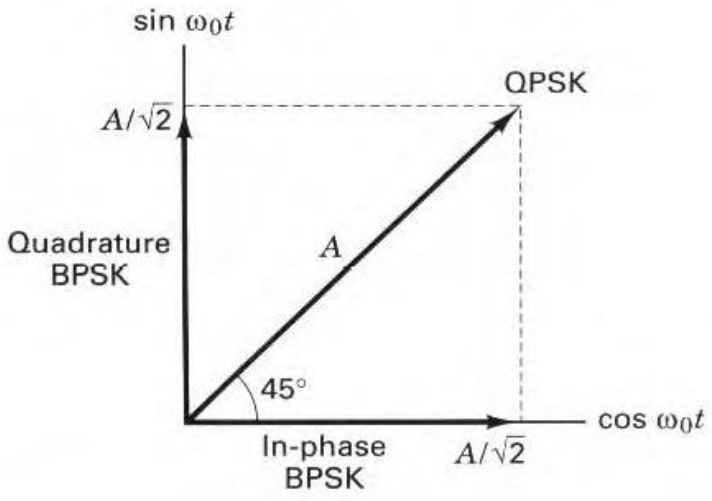
$$\frac{E_b}{N_0} = \frac{S}{N}\left(\frac{W}{R}\right) \tag{4.101}$$

where S is the average signal power and R is the bit rate. A BPSK signal with the available E_b/N_0 found from Equation (4.101) will perform with a P_B that can be read from the $k = 1$ curve in Figure 4.29. QPSK can be characterized as two orthogonal BPSK channels. The QPSK bit stream is usually partitioned into an even and odd (I and Q) stream; each new stream modulates an orthogonal component of the carrier at half the bit rate of the original stream. The I stream modulates the $\cos \omega_0 t$ term and the Q stream modulates the $\sin \omega_0 t$ term. If the magnitude of the original QPSK vector has the value A, the magnitude of the I and Q component vectors each has a value of $A/\sqrt{2}$, as shown in Figure 4.31. Thus, each of the quadrature BPSK signals has half of the average power of the original QPSK signal. Hence if the original QPSK waveform has a bit rate of R bits/s and an average power of S watts, the quadrature partitioning results in each of the BPSK waveforms having a bit rate of $R/2$ bits/s and an average power of $S/2$ watts.

Therefore, the E_b/N_0 characterizing each of the orthogonal BPSK channels, making up the QPSK signal, is equivalent to the E_b/N_0 in Equation (4.101), since it can be written as

$$\frac{E_b}{N_0} = \frac{S/2}{N_0}\left(\frac{W}{R/2}\right) = \frac{S}{N_0}\left(\frac{W}{R}\right) \tag{4.102}$$

Thus each of the orthogonal BPSK channels, and hence the composite QPSK signal, is characterized by the same E_b/N_0 and hence the same P_B performance as a BPSK signal. The natural orthogonality of the 90° phase shifts between adjacent QPSK symbols results in the *bit error probabilities* being equal for both BPSK and QPSK signaling. It is important to note that the symbol *error probabilities* are *not* equal for BPSK and QPSK signaling. The relationship between bit error probability and symbol error probability is treated in Sections 4.9.3 and 4.9.4. We see that, in effect, QPSK is the equivalent of two BPSK channels in quadrature. This same idea can be extended to any symmetrical M-ary amplitude/phase signaling, such as quadrature amplitude modulation (QAM) described in Section 9.8.3.



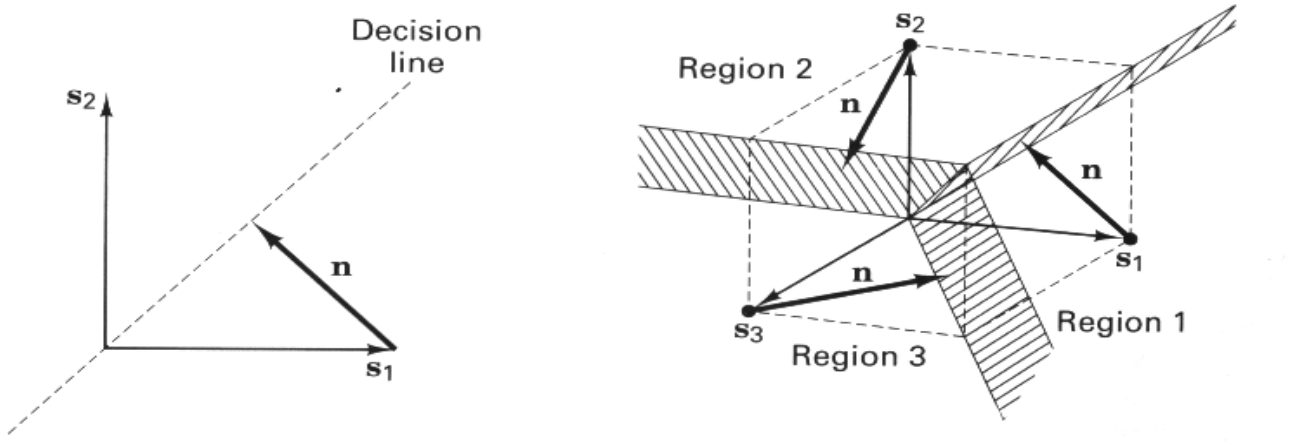
In-phase and quadrature BPSK components of QPSK signaling.

Vectorial View of MFSK Signaling

In Section 4.8.3, Figure 4.30 provides some insight as to why the error performance of MPSK signaling degrades as k (or M) increases. It would be useful to have a similar vectorial illustration for the error performance of orthogonal MFSK signaling as seen in the curves of Figure (4.28). Since the MFSK signal space is characterized by M mutually perpendicular axes, we can only conveniently illustrate the cases $M = 2$ and $M = 3$. In Figure 4.32a we see the binary orthogonal vectors $\mathbf{s}_1$ and $\mathbf{s}_2$ positioned 90° apart. The decision boundary is drawn so as to partition the signal space into two regions. On the figure is also shown a noise vector $\mathbf{n}$, which represents the minimum noise vector that would cause the detector to make an error.

In Figure 4.32b we see a 3-ary signal space with axes positioned 90° apart. Here decision planes partition the signal space into three regions. Noise vectors $\mathbf{n}$ are shown added to each of the prototype signal vectors $\mathbf{s}_1$, $\mathbf{s}_2$, and $\mathbf{s}_3$; each noise vector illustrates an example of the minimum noise vector that would cause the detector to make a symbol error. The minimum noise vectors in Figure 4.32b are the same length as the noise vector in Figure 4.32a. In Section 4.4.4 we stated that for a given level of received energy, the distance between any two prototype signal vectors $\mathbf{s}_i$ and $\mathbf{s}_j$ in an M-ary orthogonal space is constant. It follows that the minimum distance between a prototype signal vector and any of the decision boundaries remains fixed as M increases. Unlike the case of MPSK signaling, where adding new signals to the signal set makes the signals vulnerable to smaller noise vectors, here, in the case of MFSK signaling, adding new signals to the signal set does *not* make the signals vulnerable to smaller noise vectors.

It would be convenient to illustrate the point by drawing higher dimensional orthogonal spaces, but of course this is not possible. We can only use our "mind's eye" to understand that increasing the signal set M—by adding additional axes, where each new axis is mutually perpendicular to all the others—does not crowd



MFSK signal sets for $M = 2, 3$.

the signal set more closely together. Thus, a transmitted signal from an orthogonal set is *not* more vulnerable to a noise vector when the set is increased in size. In fact, we see from Figure 4.28 that as k increases, the bit error performance improves.

Understanding the error performance improvement of orthogonal signaling, as illustrated in Figure 4.28, is facilitated by comparing the probability of symbol error (P_E) versus unnormalized SNR, with P_E versus E_b/N_0. Figure 4.33 represents a set of P_E performance curves plotted against unnormalized SNR for coherent FSK signaling. Here we see that P_E degrades as M is increased. Didn't we say that a signal from an orthogonal set is *not* made more vulnerable to a given noise vector, as the orthogonal set is increased in size? It is correct that for orthogonal signaling, with a given SNR it takes a fixed size noise vector to perturb a transmitted signal into an error region; the signals do not become vulnerable to smaller noise vectors as M increases. However, as M increases, more neighboring decision regions are introduced; thus the number of ways in which a symbol error can be made increases. Figure 4.33 reflects the degradation in P_E versus unnormalized SNR as M is increased; there are $(M - 1)$ ways to make an error. Examining performance under the condition of a fixed SNR (as M increases) is not very useful for digital communications. A fixed SNR means a fixed amount of energy per symbol; thus as M increases, there is a fixed amount of energy to be apportioned over a larger number of bits, or there is less energy per bit. The most useful way of comparing one digital system with another is on the basis of *bit-normalized SNR* or E_b/N_0. The error performance improvement with increasing M (see Figure 4.28) manifests itself only when error probability is plotted against E_b/N_0. For this case, as M increases, the required E_b/N_0 (to meet a given error probability) is reduced for a fixed SNR; therefore, we need to map the plot shown in Figure 4.33 into a new plot, similar to that shown in Figure 4.28, where the abscissa represents E_b/N_0 instead of SNR. Figure 4.34 illustrates such a mapping; it demonstrates that curves manifesting degraded P_E with increasing M (such as Figure 4.33) are transformed into curves manifesting improved P_E with increasing M. The basic mapping relationship is expressed in Equation (4.101), repeated here as

$$\frac{E_b}{N_0} = \frac{S}{N}\left(\frac{W}{R}\right)$$

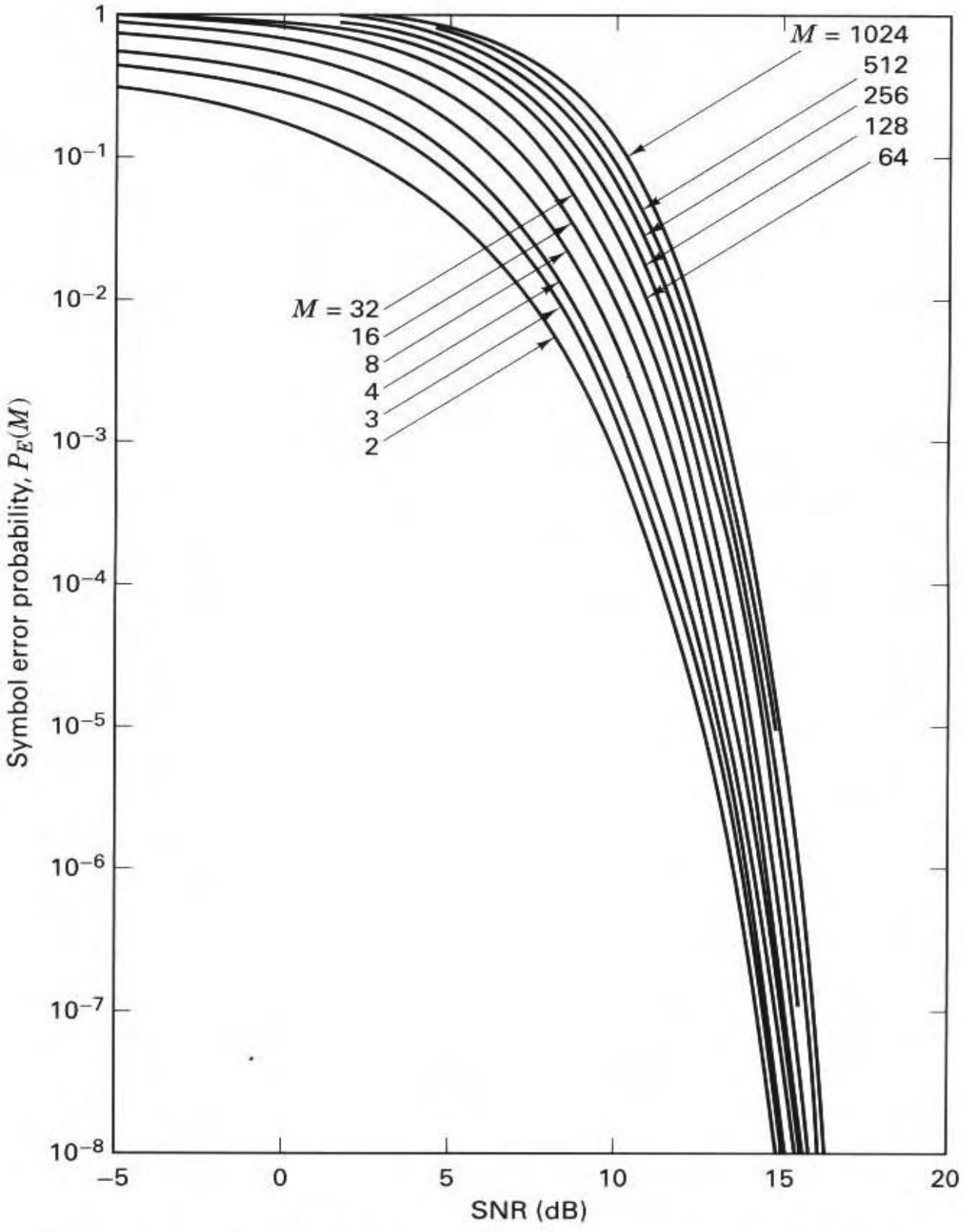
where W is the detection bandwidth. Since

$$R = \frac{\log_2 M}{T} = \frac{k}{T}$$

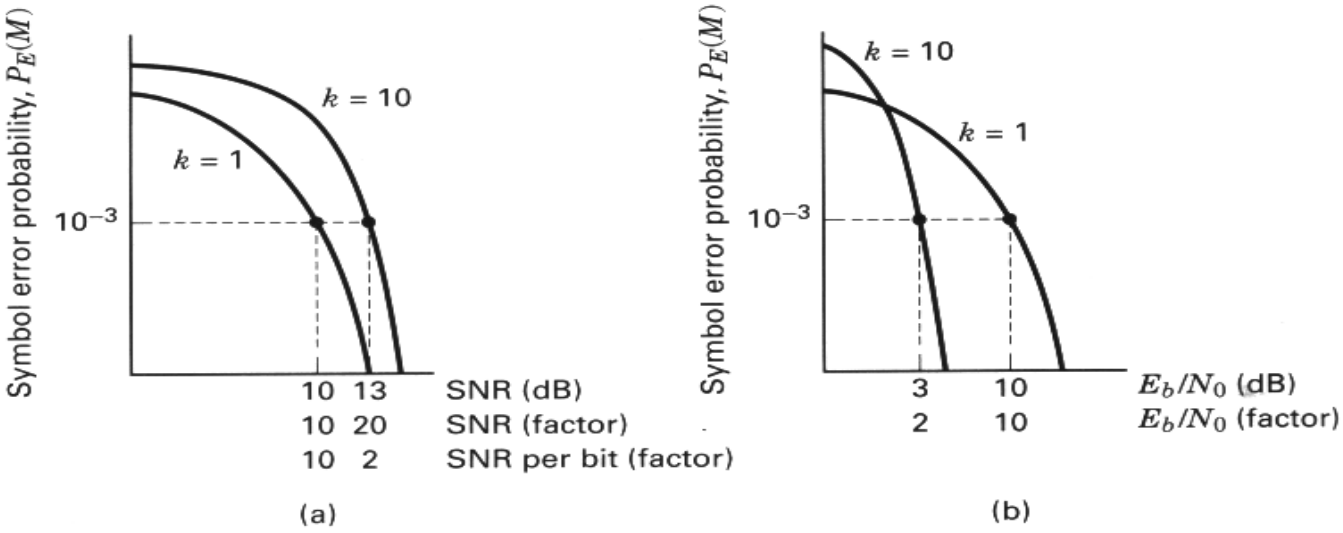
where T is the symbol duration, we can then write

$$\frac{E_b}{N_0} = \frac{S}{N}\left(\frac{WT}{\log_2 M}\right) = \frac{S}{N}\left(\frac{WT}{k}\right) \tag{4.103}$$

For FSK signaling, the detection bandwidth W (in hertz) is typically equal in value to the symbol rate $1/T$; in other words, $WT \approx 1$. Therefore,



Symbol error probability versus SNR for coherent FSK signaling.



Mapping P_E versus SNR into P_E versus E_b/N_0 for orthogonal signaling. (a) Unnormalized. (b) Normalized.

$$\frac{E_b}{N_0} \approx \frac{S}{N}\left(\frac{1}{k}\right) \tag{4.104}$$

Figure 4.34 illustrates the mapping from P_E versus SNR to P_E versus E_b/N_0 for coherently detected M-ary orthogonal signaling, with "ballpark" numbers on the axes. In Figure 4.34a, on the $k = 1$ curve is shown an operating point corresponding to $P_E = 10^{-3}$ and SNR = 10 dB. On the $k = 10$ curve is shown an operating point at the same $P_E = 10^{-3}$ but with SNR = 13 dB (approximate values taken from Figure 4.33). Here we clearly see the degradation in error performance as k increases. To appreciate where the performance improvement comes from, let us convert the abscissa from the nonlinear scale of SNR in decibels to a linear one—SNR expressed as a factor. This is shown in Figure 4.34a as the factors 10 and 20 for the $k = 1$ and $k = 10$ cases, respectively. Next, we further convert the abscissa scale to SNR per bit (expressed as a factor). This is shown in Figure 4.34a as the factors 10 and 2 for the $k = 1$ and $k = 10$ cases, respectively. It is convenient to think of the 1024-ary symbol or waveform ($k = 10$ case) as being interchangeable with its 10-bit meaning. Thus, if the symbol requires 20 units of SNR then the 10 bits belonging to that symbol require that same 20 units; or, in other words, each bit requires 2 units.

Rather than performing such computations, we can simply map these same $k = 1$ and $k = 10$ cases onto the Figure 4.34b plane, representing P_E versus E_b/N_0. The $k = 1$ case looks exactly the same as it does in Figure 4.34a. But for the $k = 10$ case, there is a dramatic change. We can immediately see that signaling with the $k = 10$-bit symbol requires only 2 units (3 dB) of E_b/N_0 compared with 10 units (10 dB) for the binary symbol. The mapping that gives rise to the required E_b/N_0 for the $k = 10$ case is obtained from Equation (4.104) as follows: $E_b/N_0 = 20\ (1/10) = 2$ (or 3-dB), which shows the error performance improvement as k is increased. In digital communication systems, error performance is almost always considered in terms of

E_b/N_0, since such a measurement makes for a meaningful comparison between one system's performance and another. Therefore, the curves shown in Figures 4.33 and 4.34a are hardly ever seen.

Although Figure 4.33 is not often seen, we can still use it for gaining insight into why orthogonal signaling provides improved error performance as M or k increases. Let us consider the analogy of purchasing a commodity—say, grade A cottage cheese. The choice of the grade corresponds to some point on the P_E axis of Figure 4.33—say, 10^{-3}. From this point, construct a horizontal line through all of the curves (from $M = 2$ through $M = 1024$). At the grocery store we buy the very smallest container of cottage cheese, containing 2 ounces and costing \$1. On Figure 4.33 we can say that this purchase corresponds to our horizontal construct intercepting the $M = 2$ curve. We look down at the corresponding SNR and call the intercept on this axis our cost of \$1. The next time we purchase cottage cheese, we remember that the first purchase seemed expensive at 50 cents an ounce. So, we decide to buy a larger carton, containing 8 ounces and costing \$2. On Figure 4.33, we can say that this purchase corresponds to the point at which our horizontal construct intercepts the $M = 8$ curve. We look down at the corresponding SNR, and call this intercept our cost of \$2. Notice that we bought a larger container so the price went up, but because we bought a greater quantity, the price per ounce went down (the unit cost is now only 25 cents per ounce). We can continue this analogy by purchasing larger and larger containers so that the price of the container (SNR) keeps going up, but the price per ounce keeps going down. This is the age-old story called the *economy of scale*. Buying larger quantities at a time is commensurate with purchasing at the wholesale level; it makes for a lower unit price. Similarly, when we use orthogonal signaling with symbols that contain more bits, we need more power (more SNR), but the requirement per bit (E_b/N_0) is reduced.

SYMBOL ERROR PERFORMANCE FOR *M*-ARY SYSTEMS ($M > 2$)

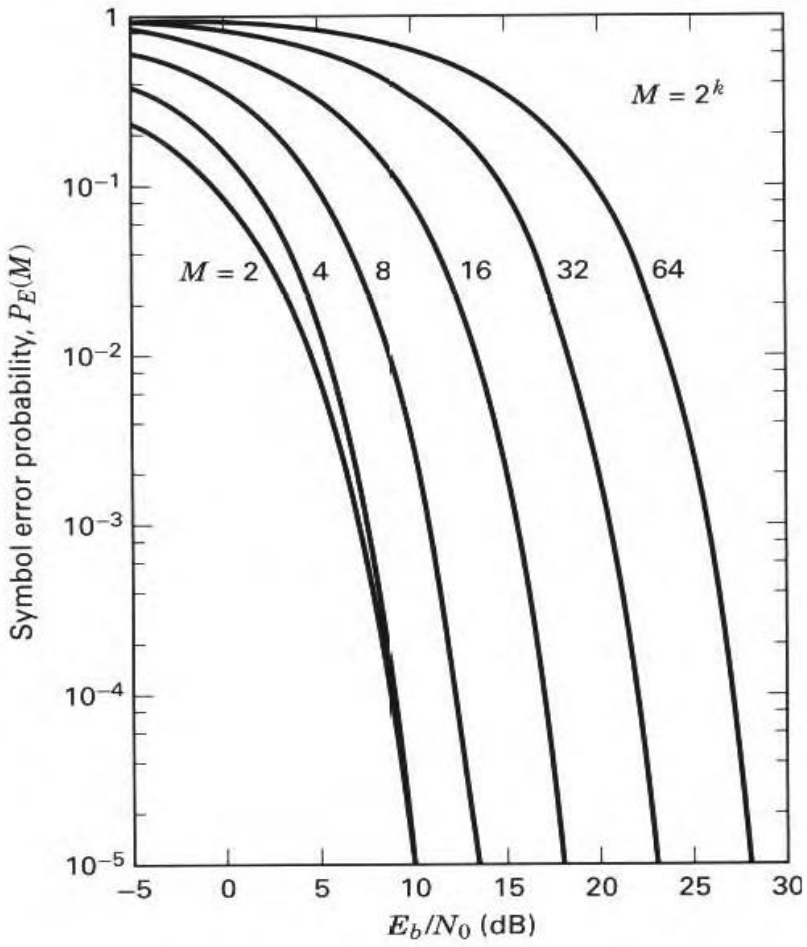
Probability of Symbol Error for MPSK

For large energy-to-noise ratios, the symbol error performance $P_E(M)$, for equally likely, coherently detected M-ary PSK signaling, can be expressed [7] as

$$P_E(M) \approx 2Q\left(\sqrt{\frac{2E_s}{N_0}}\sin\frac{\pi}{M}\right) \tag{4.105}$$

where $P_E(M)$ is the probability of symbol error, $E_s = E_b(\log_2 M)$ is the energy per symbol, and $M = 2^k$ is the size of the symbol set. The $P_E(M)$ performance curves for coherently detected MPSK signaling are plotted versus E_b/N_0 in Figure 4.35.

The symbol error performance for differentially coherent detection of M-ary DPSK (for large E_s/N_0) is similarly expressed [7] as



Symbol error probability for coherently detected multiple phase signaling. (Reprinted from W. C. Lindsey and M. K. Simon, *Telecommunication Systems Engineering*, Prentice-Hall, Inc. Englewood Cliffs, N.J., 1973, courtesy of W. C. Lindsey and Marvin K. Simon.)

$$P_E(M) \approx 2Q\left(\sqrt{\frac{2E_s}{N_0}}\sin\frac{\pi}{\sqrt{2}M}\right) \tag{4.106}$$

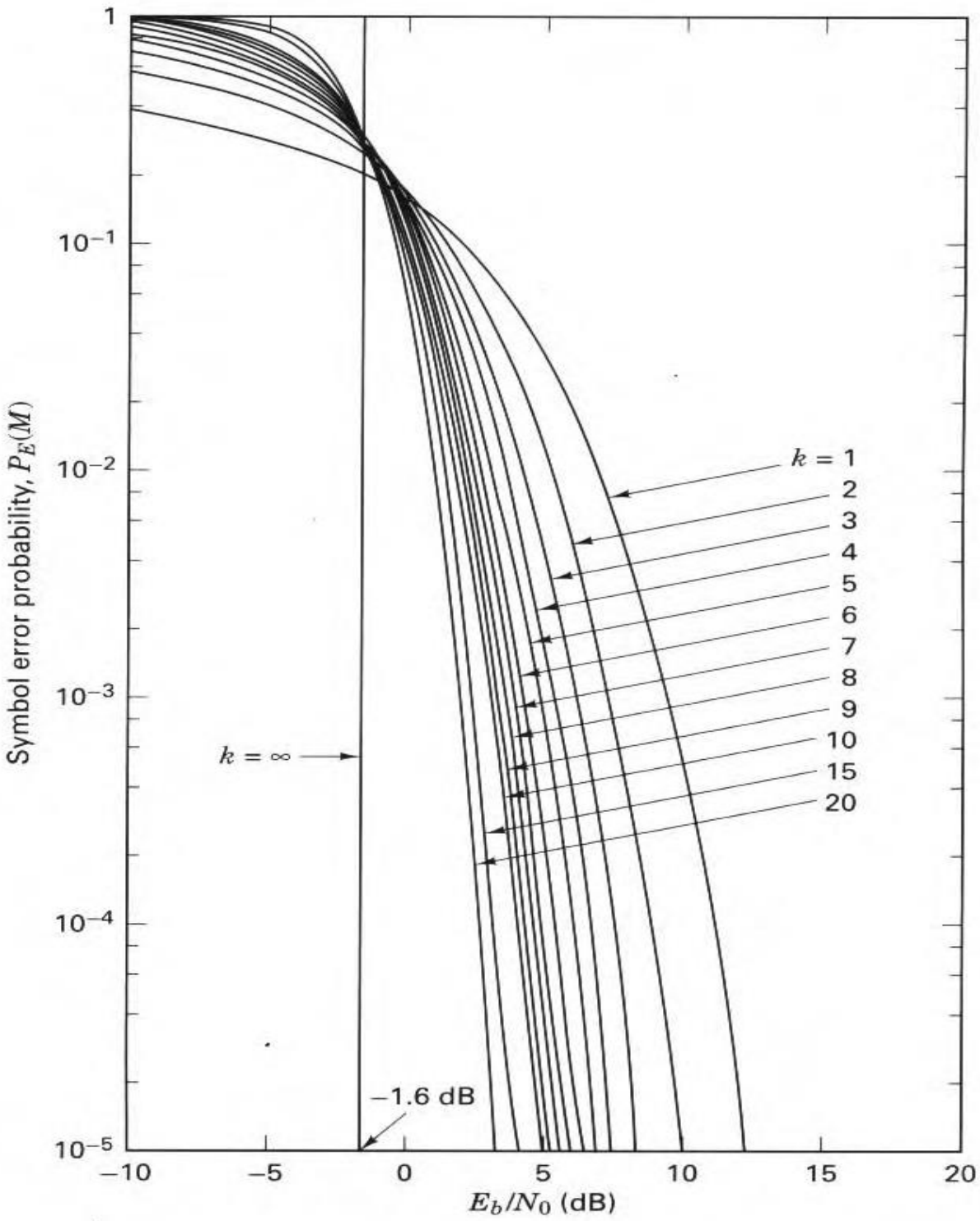
Probability of Symbol Error for MFSK

The symbol error performance $P_E(M)$, for equally likely, *coherently* detected M-ary orthogonal signaling can be upper bounded [5] as

$$P_E(M) \leq (M-1)\,Q\left(\sqrt{\frac{E_s}{N_0}}\right) \tag{4.107}$$

where $E_s = E_b(\log_2 M)$ is the energy per symbol and M is the size of the symbol set. The $P_E(M)$ performance curves for coherently detected M-ary orthogonal signaling are plotted versus E_b/N_0 in Figure 4.36.

The symbol error performance for equally likely, *noncoherently* detected M-ary orthogonal signaling is [9]



Symbol error probability for coherently detected M-ary orthogonal signaling.

$$P_E(M) = \frac{1}{M}\exp\left(-\frac{E_s}{N_0}\right)\sum_{j=2}^{M}(-1)^j\binom{M}{j}\exp\left(\frac{E_s}{jN_0}\right) \tag{4.108}$$

where

$$\binom{M}{j} = \frac{M!}{j!\,(M-j)!} \tag{4.109}$$

is the standard binomial coefficient yielding the number of ways in which j symbols out of M may be in error. Note that for binary case, Equation (4.108) reduces to

$$P_B = \frac{1}{2}\exp\left(-\frac{E_b}{2N_0}\right) \tag{4.110}$$

which is the same result as that described by Equation (4.96). The $P_E(M)$ performance curves for noncoherently detected M-ary orthogonal signaling are plotted versus E_b/N_0 in Figure 4.37. If we compare this noncoherent orthogonal $P_E(M)$ performance with the corresponding $P_E(M)$ results for the coherent detection of orthogonal signals in Figure 4.36, it can be seen that for $k > 7$, there is a negligible difference. An upper bound for coherent as well as noncoherent reception of orthogonal signals is [9]

$$P_E(M) < \frac{M-1}{2}\exp\left(-\frac{E_s}{2N_0}\right) \tag{4.111}$$

where E_s is the energy per symbol and M is the size of the symbol set.

Bit Error Probability versus Symbol Error Probability for Orthogonal Signals

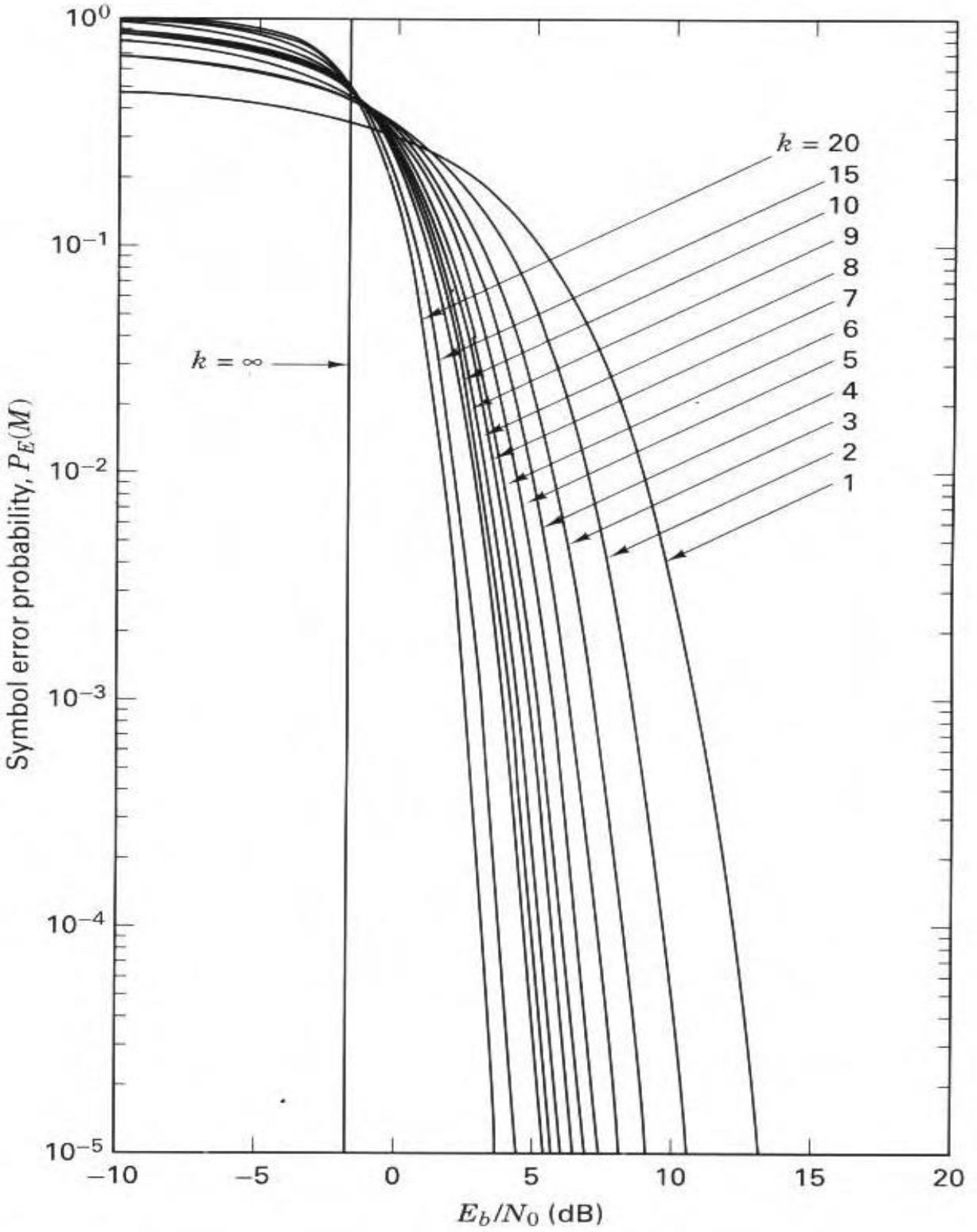
It can be shown [9] that the relationship between probability of bit error (P_B) and probability of symbol error (P_E) for an M-ary orthogonal signal set is

$$\frac{P_B}{P_E} = \frac{2^{k-1}}{2^k - 1} = \frac{M/2}{M-1} \tag{4.112}$$

In the limit as k increases, we get

$$\lim_{k\to\infty}\frac{P_B}{P_E} = \frac{1}{2}$$

A simple example will make Equation (4.112) intuitively acceptable. Figure 4.38 describes an octal message set. The message symbols (assumed equally likely) are to be transmitted on orthogonal waveforms such as FSK. With orthogonal signaling, a decision error will transform the correct signal into any one of the $(M-1)$ incorrect signals with equal probability. The example in Figure 4.38 indicates that the symbol comprising bits 0 1 1 was transmitted. An error might occur in any one of the other $2^k - 1 = 7$ symbols, with equal probability. Notice that just because a symbol error is made does not mean that all the bits within the symbol will be in error. In Figure 4.38, if the receiver decides that the transmitted symbol is the bot-



Symbol error probability for noncoherently detected *M*-ary orthogonal signaling.

			Bit position
	0	0	0
	0	0	1
	0	1	0
Transmitted symbol	0	1	1
	1	0	0
	1	0	1
	1	1	0
	1	1	1

Example of P_B versus P_E.

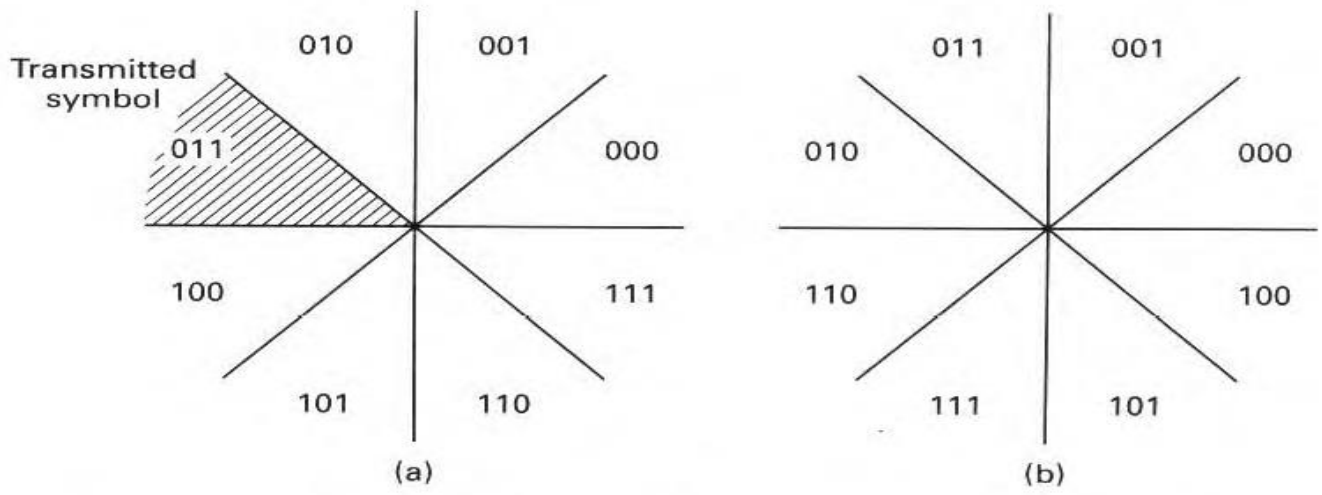
tom one listed, comprising bits 1 1 1, two of the three transmitted symbol bits will be correct; only one bit will be in error. It should be apparent that for nonbinary signaling, P_B will always be less than P_E (keep in mind that P_B and P_E reflect the frequency of making errors on the *average*.)

Consider any of the bit-position columns in Figure 4.38. For each bit position, the digit occupancy consists of 50% ones and 50% zeros. In the context of the first bit position (rightmost column) and the transmitted symbol, how many ways are there to cause an error to the binary one? There are $2^{k-1} = 4$ ways (four places where zeros appear in the column) that a bit error can be made; it is the same for each of the columns. The final relationship P_B/P_E, for orthogonal signaling, in Equation (4.112), is obtained by forming the following ratio: the number of ways that a bit error can be made (2^{k-1}) divided by the number of ways that a symbol error can be made ($2^k - 1$). For the Figure 4.38 example, $P_B/P_E = 4/7$.

Bit Error Probability Versus Symbol Error Probability for Multiple Phase Signaling

For the case of MPSK signaling, P_B is less than or equal to P_E, just as in the case of MFSK signaling. However, there is an important difference. For orthogonal signaling, selecting any one of the $(M-1)$ erroneous symbols is equally likely. In the case of MPSK signaling, each signal vector is not equidistant from all of the others. Figure 4.39a illustrates an 8-ary decision space with the pie-shaped decision regions denoted by the 8-ary symbols in binary notation. If symbol (0 1 1) is transmitted, it is clear that should an error occur, the transmitted signal will most likely be mistaken for one of its closest neighbors, (0 1 0) or (1 0 0). The likelihood that (0 1 1) would get mistaken for (1 1 1) is relatively remote. If the assignment of bits to symbols follows the binary sequence shown in the symbol decision regions of Figure 4.39a, some symbol errors will usually result in two or more bit errors, even with a large signal-to-noise ratio.

For nonorthogonal schemes, such as MPSK signaling, one often uses a binary-to-M-ary code such that binary sequences corresponding to adjacent sym-



Binary-coded versus Gray-coded decision regions in an MPSK signal space. (a) Binary coded. (b) Gray coded.

bols (phase shifts) differ in only one bit position; thus when an M-ary symbol error occurs, it is more likely that only one of the k input bits will be in error. A code that provides this desirable feature is the Gray code [7]; Figure 4.39b illustrates the bit-to-symbol assignment using a Gray code for 8-ary PSK. Here it can be seen that neighboring symbols differ from one another in only one bit position. Therefore, the occurrence of a multibit error, for a given symbol error, is much reduced compared to the uncoded binary assignment seen in Figure 4.39a. Implementing such a Gray code, represents one of the few cases in digital communications where a benefit can be achieved without incurring any cost. The Gray code is simply an assignment that requires no special or additional circuitry. Utilizing the Gray code assignment, it can be shown [5] that

$$P_B \approx \frac{P_E}{\log_2 M} \qquad (\text{for } P_E \ll 1) \tag{4.113}$$

Recall from Section 4.8.4 that BPSK and QPSK signaling have the same bit error probability. Here, in Equation (4.113), we verify that they do not have the same symbol error probability. For BPSK, $P_E = P_B$. However, for QPSK, $P_E \approx 2P_B$.

An exact closed-form expression for the bit-error probability P_B of 8-ary PSK, together with tight upper and lower bounds on P_B for M-ary PSK with larger M, may be found in Lee [10].

Effects of Intersymbol Interference

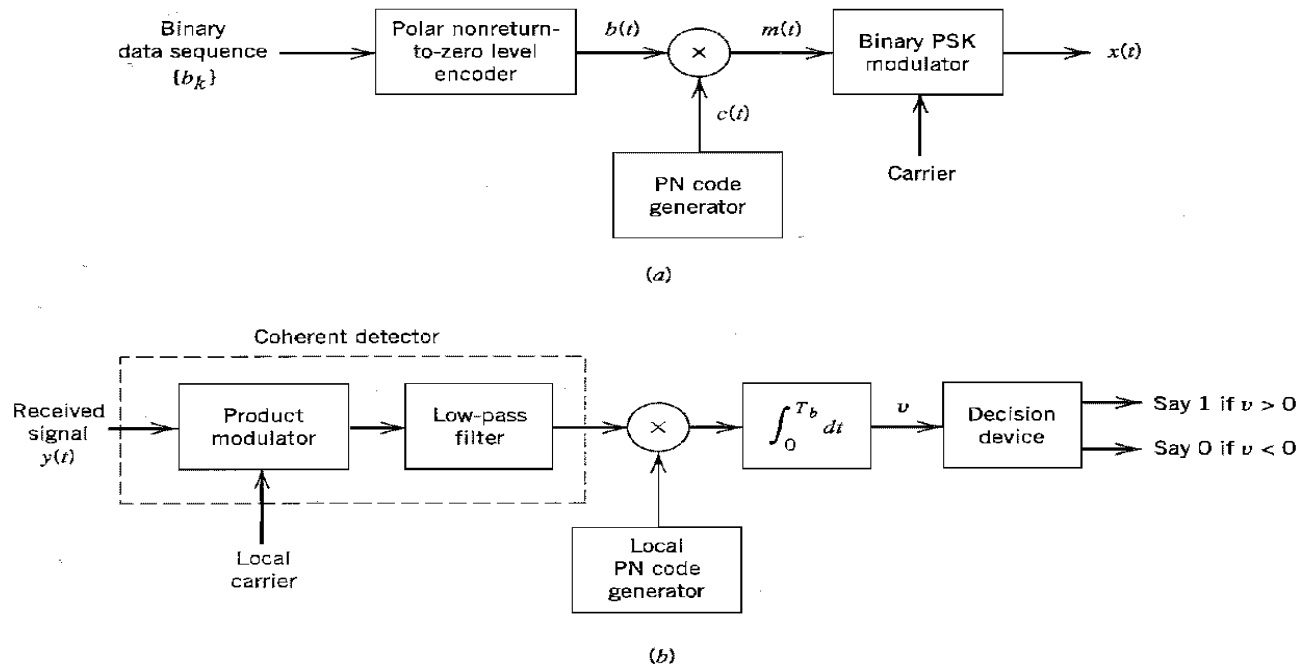
In the previous sections and in Chapter 3 we have treated the detection of signals in the presence of AWGN under the assumption that there is no intersymbol interference (ISI). Thus the analysis has been straightforward, since the zero-mean AWGN process is characterized by its variance alone. In practice we find that ISI is

often a second source of interference which must be accounted for. As explained in Section 3.3, ISI can be generated by the use of bandlimiting filters at the transmitter output, in the channel, or at the receiver input. The result of this additional interference is to degrade the error probabilities for coherent as well as for noncoherent reception. Calculating error performance in the presence of ISI in addition to AWGN is much more complicated since it involves the impulse response of the channel.

Direct-Sequence Spread Spectrum with Coherent Binary Phase-Shift Keying

The spread-spectrum technique described in the previous section is referred to as *direct-sequence spread spectrum*. The discussion presented there was in the context of baseband transmission. To provide for the use of this technique in passband transmission over a satellite channel, for example, we may incorporate *coherent binary phase-shift keying* (PSK) into the transmitter and receiver, as shown in Figure 7.7. The transmitter of Figure 7.7*a* first converts the incoming binary data sequence $\{b_k\}$ into a polar NRZ waveform $b(t)$, which is followed by two stages of modulation. The first stage consists of a product modulator or multiplier with the data signal $b(t)$ (representing a data sequence) and the PN signal $c(t)$ (representing the PN sequence) as inputs. The second stage consists of a binary PSK modulator. The transmitted signal $x(t)$ is thus a *direct-sequence spread binary phase-shift-keyed* (DS/BPSK) *signal*. The phase modulation $\theta(t)$ of $x(t)$ has one of two values, 0 and π, depending on the polarities of the message signal $b(t)$ and PN signal $c(t)$ at time t in accordance with the truth table of Table 7.3.

Figure 7.8 illustrates the waveforms for the second stage of modulation. Part of the modulated waveform shown in Figure 7.6*c* is reproduced in Figure 7.8*a*; the waveform shown here corresponds to one period of the PN sequence. Figure 7.8*b* shows the waveform of a sinusoidal carrier, and Figure 7.8*c* shows the DS/BPSK waveform that results from the second stage of modulation.



Direct-sequence spread coherent phase-shift keying. (*a*) Transmitter. (*b*) Receiver.

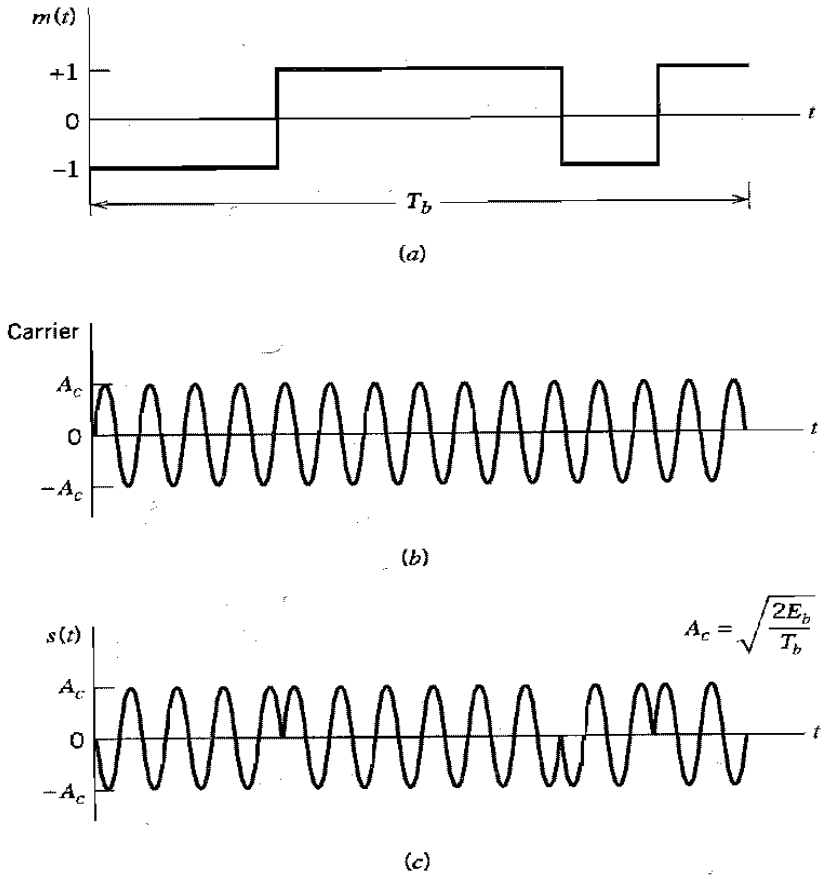
The receiver, shown in Figure 7.7*b*, consists of two stages of demodulation. In the first stage, the received signal $y(t)$ and a locally generated carrier are applied to a product modulator followed by a low-pass filter whose bandwidth is equal to that of the original message signal $m(t)$. This stage of the demodulation process reverses the phase-shift keying applied to the transmitted signal. The second stage of demodulation performs spectrum despreading by multiplying the low-pass filter output by a locally generated replica of the PN signal $c(t)$, followed by integration over a bit interval $0 \leq t \leq T_b$, and finally decision-making in the manner described in Section 7.3.

Model for Analysis

In the normal form of the transmitter, shown in Figure 7.7*a*, the spectrum spreading is performed prior to phase modulation. For the purpose of analysis, however, we find it more convenient to interchange the order of these operations, as shown in the model of

Truth table for phase modulation* $\theta(t)$, *radians

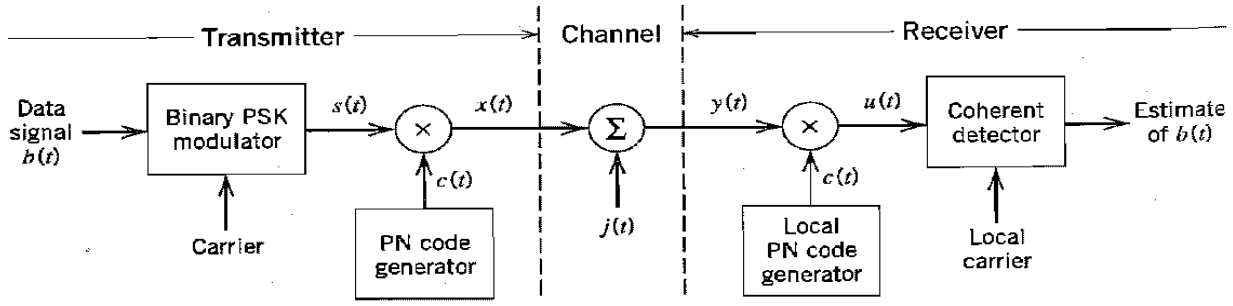
		Polarity of Data Sequence b(t) *at Time* t	
		+	−
Polarity of PN sequence c(t) *at time* t	+	0	π
	−	π	0



(*a*) Product signal $m(t) = c(t)b(t)$. (*b*) Sinusoidal carrier. (*c*) DS/BPSK signal.

We are permitted to do this because the spectrum spreading and the binary phase-shift keying are both linear operations; likewise for the phase demodulation and spectrum despreading. But for the interchange of operations to be feasible, it is important to synchronize the incoming data sequence and the PN sequence. The model of Figure 7.9 also includes representations of the channel and the receiver. In this model, it is assumed that the interference $j(t)$ limits performance, so that the effect of channel noise may be ignored. Accordingly, the channel output is given by

$$\begin{aligned} y(t) &= x(t) + j(t) \\ &= c(t)s(t) + j(t) \end{aligned} \tag{7.12}$$



Model of direct-sequence spread binary PSK system.

where $s(t)$ is the binary PSK signal, and $c(t)$ is the PN signal. In the channel model included in Figure 7.9, the interfering signal is denoted by $j(t)$. This notation is chosen purposely to be different from that used for the interference in Figure 7.5*b*. The channel model in Figure 7.9 is passband in spectral content, whereas that in Figure 7.5*b* is in baseband form.

In the receiver, the received signal $y(t)$ is first multiplied by the PN signal $c(t)$ yielding an output that equals the coherent detector input $u(t)$. Thus,

$$\begin{aligned} u(t) &= c(t)y(t) \\ &= c^2(t)s(t) + c(t)j(t) \\ &= s(t) + c(t)j(t) \end{aligned} \tag{7.13}$$

In the last line of Equation (7.13), we have noted that, by design, the PN signal $c(t)$ satisfies the property described in Equation (7.10), reproduced here for convenience:

$$c^2(t) = 1 \qquad \text{for all } t$$

Equation (7.13) shows that the coherent detector input $u(t)$ consists of a binary PSK signal $s(t)$ embedded in additive code-modulated interference denoted by $c(t)j(t)$. The modulated nature of the latter component forces the interference signal (jammer) to spread its spectrum such that the detection of information bits at the receiver output is afforded increased reliability.

SYNCHRONIZATION

For its proper operation, a spread-spectrum communication system requires that the locally generated PN sequence used in the receiver to despread the received signal be *synchronized* to the PN sequence used to spread the transmitted signal in the transmitter.[4] A solution to the synchronization problem consists of two parts: *acquisition* and *tracking*. In acquisition, or *coarse* synchronization, the two PN codes are aligned to within a fraction of the chip in as short a time as possible. Once the incoming PN code has been acquired, tracking, or *fine* synchronization, takes place. Typically, PN acquisition proceeds in two steps. First, the received signal is multiplied by a locally generated PN code to produce a measure of *correlation* between it and the PN code used in the transmitter. Next, an appropriate *decision-rule and search strategy* is used to process the measure of correlation so obtained to determine whether the two codes are in synchronism and what to do if they are not. As for tracking, it is accomplished using phase-lock techniques very similar to those used for the local generation of coherent carrier references. The principal difference between them lies in the way in which phase discrimination is implemented.

FREQUENCY HOPPING SYSTEMS

We now consider a spread-spectrum technique called frequency hopping (FH). The modulation most commonly used with this technique is M-ary frequency shift keying (MFSK), where $k = \log_2 M$ information bits are used to determine which

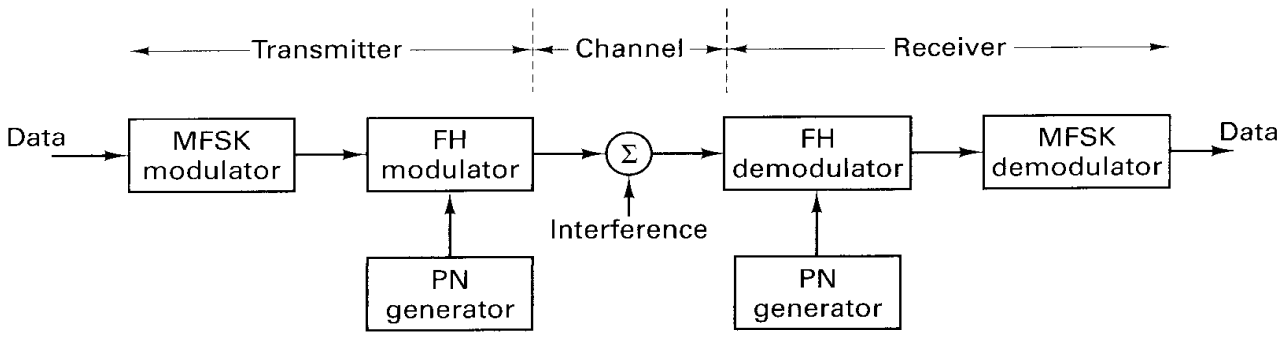
one of M frequencies is to be transmitted. The position of the M-ary signal set is shifted pseudorandomly by the frequency synthesizer over a hopping bandwidth W_{ss}. A typical FH/MFSK system block diagram is shown in Figure 12.11. In a conventional MFSK system, the data symbol modulates a *fixed frequency* carrier; in an FH/MFSK system, the data symbol modulates a carrier whose frequency is *pseudorandomly* determined. In either case, a single tone is transmitted. The FH system in Figure 12.11 can be thought of as a two-step modulation process—data modulation and frequency-hopping modulation—even though it can be implemented as a single step whereby the frequency synthesizer produces a transmission tone based on the simultaneous dictates of the PN code and the data. At each frequency hop time, a PN generator feeds the frequency synthesizer a frequency word (a sequence of ℓ chips), which dictates one of 2^{ℓ} symbol-set positions. The frequency-hopping bandwidth W_{ss}, and the minimum frequency spacing between consecutive hop positions Δf, dictate the minimum number of chips necessary in the frequency word.

For a given hop, the occupied transmission bandwidth is identical to the bandwidth of conventional MFSK, which is typically much smaller than W_{ss}. However, averaged over many hops, the FH/MFSK spectrum occupies the entire spread-spectrum bandwidth. Spread-spectrum technology permits FH bandwidths of the order of several gigahertz, which is an order of magnitude larger than implementable DS bandwidths [8], thus allowing for larger processing gains in FH compared to DS systems. Since frequency hopping techniques operate over such wide bandwidths, it is difficult to maintain phase coherence from hop to hop. Therefore, such schemes are usually configured using noncoherent demodulation. Nevertheless, consideration has been given to coherent FH in Reference [9].

In Figure 12.11 we see that the receiver reverses the signal processing steps of the transmitter. The received signal is first FH demodulated (dehopped) by mixing it with the same sequence of pseudorandomly selected frequency tones that was used for hopping. Then the dehopped signal is applied to a conventional bank of M noncoherent energy detectors to select the most likely symbol.

Frequency Word Size

A hopping bandwidth W_{ss} of 400 MHz and a frequency step size Δf of 100 Hz are specified. What is the minimum number of PN chips that are required for each frequency word?



FH/MFSK system.

Solution

$$\text{Number of tones contained in } W_{ss} = \frac{W_{ss}}{\Delta f} = \frac{400 \text{ MHz}}{100 \text{ Hz}}$$

$$= 4 \times 10^6$$

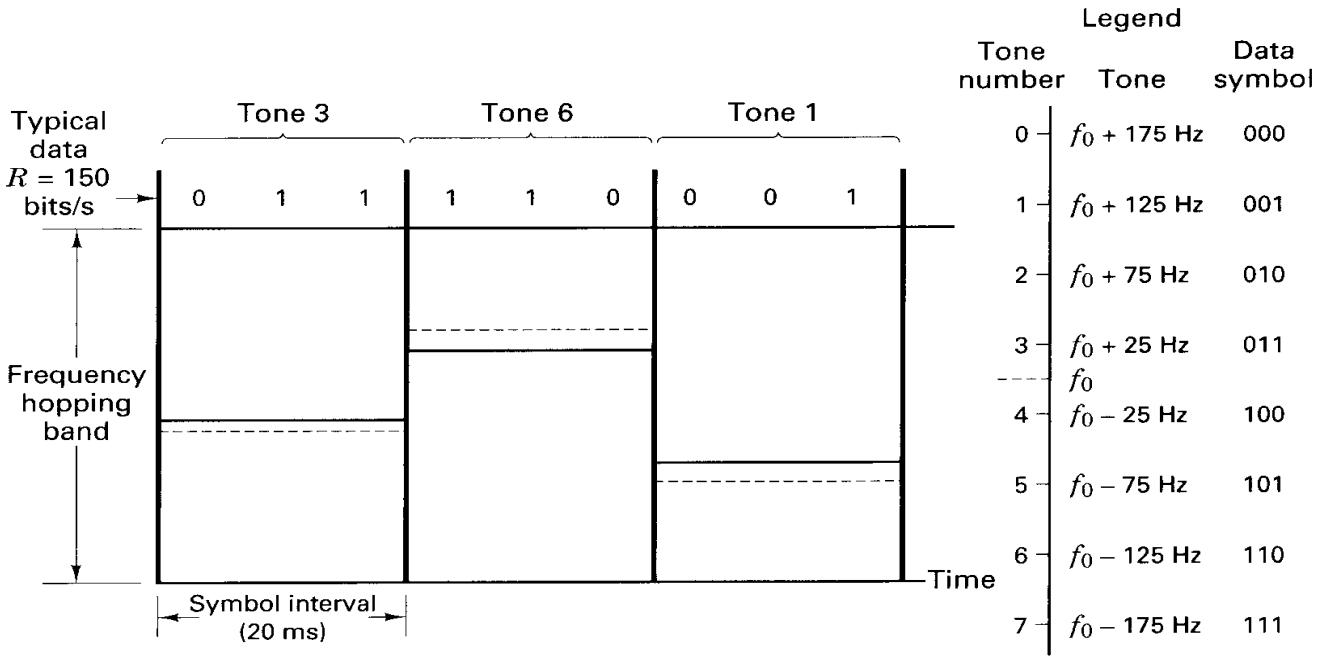
$$\text{Minimum number of chips} = \lceil \log_2 (4 \times 10^6) \rceil$$

$$= 22 \text{ chips}$$

where $\lceil x \rceil$ indicates the smallest integer value not less than x.

Frequency Hopping Example

Consider the frequency hopping example illustrated in Figure 12.12. The input data consist of a binary sequence with a data rate of $R = 150$ bits/s. The modulation is 8-ary FSK. Therefore, the symbol rate is $R_s = R/(\log_2 8) = 50$ symbols/s (the symbol duration $T = 1/50 = 20$ ms). The frequency is hopped once per symbol, and the hopping is time synchronous with the symbol boundaries. Thus, the hopping rate is 50 hops/s. Figure 12.12 depicts the time–bandwidth plane of the communication resource; the abscissa represents time, and the ordinate represents the hopping bandwidth, W_{ss}. The legend on the right side of the figure illustrates a set of 8-ary FSK symbol-to-tone assignments. Notice that the tone separation specified is $1/T = 50$ Hz, which corresponds to the minimum required tone spacing for the orthogonal signaling of this noncoherent FSK example (see Section 4.5.4).



Frequency-hopping example using 8-ary FSK modulation.

A typical binary data sequence is shown at the top of Figure 12.12. Since the modulation is 8-ary FSK, the bits are grouped three at a time to form symbols. In a *conventional* 8-ary FSK scheme, a single-sideband tone (offset from f_0, the *fixed* center frequency of the data band), would be transmitted according to an assignment like the one shown in the legend. The only difference in this FH/MFSK example is that the center frequency of the data band f_0 is *not fixed.* For each new symbol, f_0 hops to a new position in the hop bandwidth, and the entire data-band structure moves with it. In the example of Figure 12.12, the first symbol in the data sequence, 0 1 1, yields a tone 25 Hz above f_0. The diagram depicts f_0 with a dashed line and the symbol tone with a solid line. During the second symbol interval, f_0 has hopped to a new spectral location, as indicated by the dashed line. The second symbol, 1 1 0, dictates that a tone indicated by the solid line, 125 Hz below f_0, shall be transmitted. Similarly, the final symbol in this example, 0 0 1, calls for a tone 125 Hz above f_0. Again, the center frequency has moved, but the relative positions of the symbol tones remain fixed.

Robustness

A common dictionary definition describes the term *robustness* as the state of being strong and healthy; full of vigor; hardy. In the context of communications, the usage is not too different. Robustness characterizes a signal's ability to withstand impairments from the channel, such as noise, jamming, fading, and so on. A signal configured with multiple replicate copies, each transmitted on a different frequency, has a greater likelihood of survival than does a single such signal with equal total power. The greater the diversity (multiple transmissions, at different frequencies, spread in time), the more robust the signal against random interference.

The following example should clarify the concept. Consider a message consisting of four symbols: s_1, s_2, s_3, s_4. The introduction of diversity starts by repeating the message N times. Let us choose $N = 8$. Then, the repeated symbols, called *chips,* can be written.

$$s_1\,s_1\,s_1\,s_1\,s_1\,s_1\,s_1\,s_1\,s_2\,s_2\,s_2\,s_2\,s_2\,s_2\,s_2\,s_2\,s_3\,s_3\,s_3\,s_3\,s_3\,s_3\,s_3\,s_3\,s_4\,s_4\,s_4\,s_4\,s_4\,s_4\,s_4\,s_4$$

Each chip is transmitted at a different hopping frequency (the center of the data bandwidth is changed for each chip). The resulting transmissions at frequencies f_i, f_j, f_k, . . . yield a more robust signal than without such diversity. A target-shooting analogy is that a pellet from a barrage of shotgun pellets has a better chance of hitting a target, compared with the action of a single bullet.

Frequency Hopping with Diversity

In Figure 12.13 we extend the example illustrated in Figure 12.12, with the additional feature of a chip repeat factor of $N = 4$. During each 20-ms symbol interval, there are now four columns, corresponding to the four separate chips to be transmitted for each symbol. At the top of the figure we see the same data sequence,

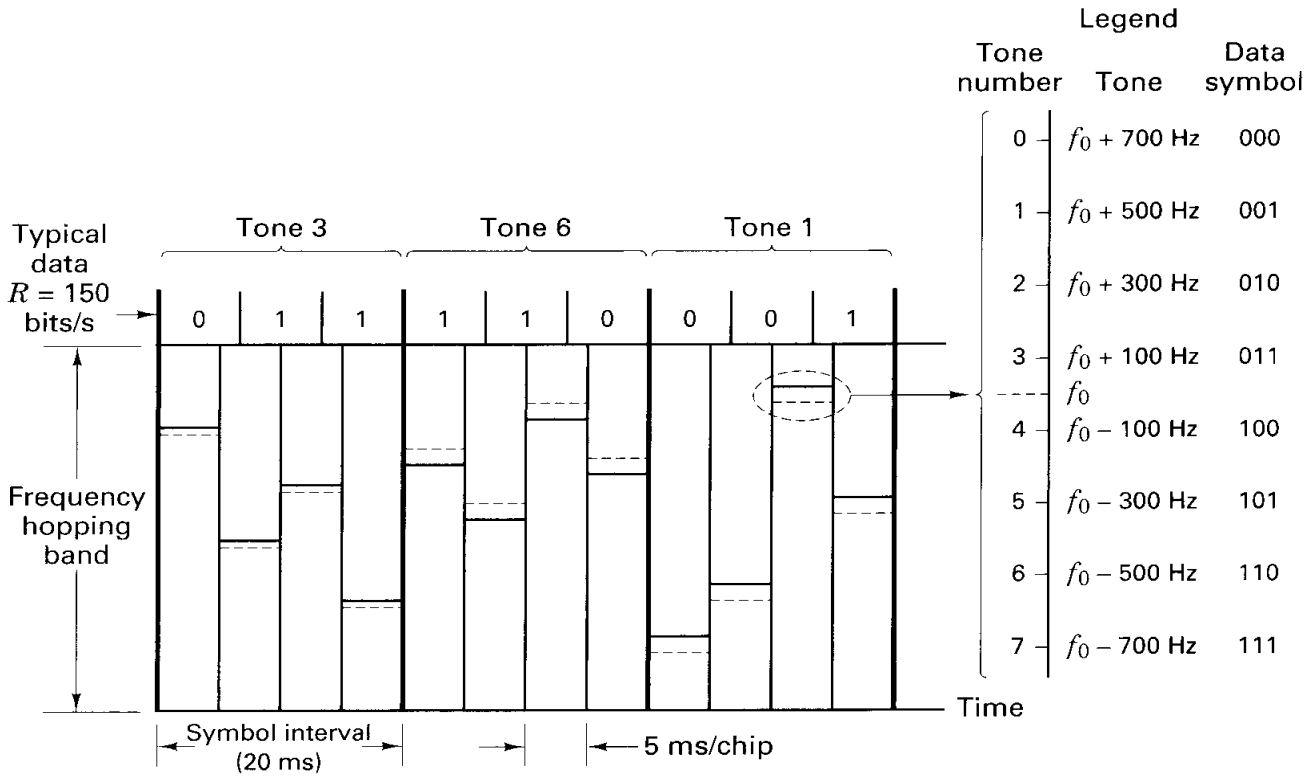
with $R = 150$ bps, as in the earlier example; and we see the same 3-bit partitioning to form the 8-ary symbols. Each symbol is transmitted four times, and for each transmission the center frequency of the data band is hopped to a new region of the hopping band, under the control of a PN code generator. Therefore, for this example, each chip interval, T_c, is equal to $T/N = 20$ ms/4 = 5 ms in duration, and the hopping rate is now

$$\frac{NR}{\log_2 8} = 200 \text{ hops/s}$$

Notice that the spacing between frequency tones must change to meet the changed requirement for orthogonality. Since the duration of each FSK tone is now equal to the chip duration, that is, $T_c = T/N$, the minimum separation between tones is $1/T_c = N/T = 200$ Hz. As in the earlier example, Figure 12.13 illustrates that the center of the data band (plus the modulation structure) is shifted at each new chip time. The position of the solid line (transmission frequency) has the same relationship to the dashed line (center of the data band) for each of the chips associated with a given symbol.

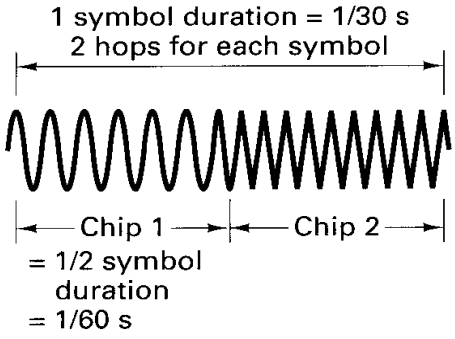
Fast Hopping versus Slow Hopping

In the case of direct-sequence spread-spectrum systems, the term "chip" refers to the PN code symbol (the symbol of shortest duration in a DS system). In a similar sense for frequency hopping systems, the term "chip" is used to characterize the

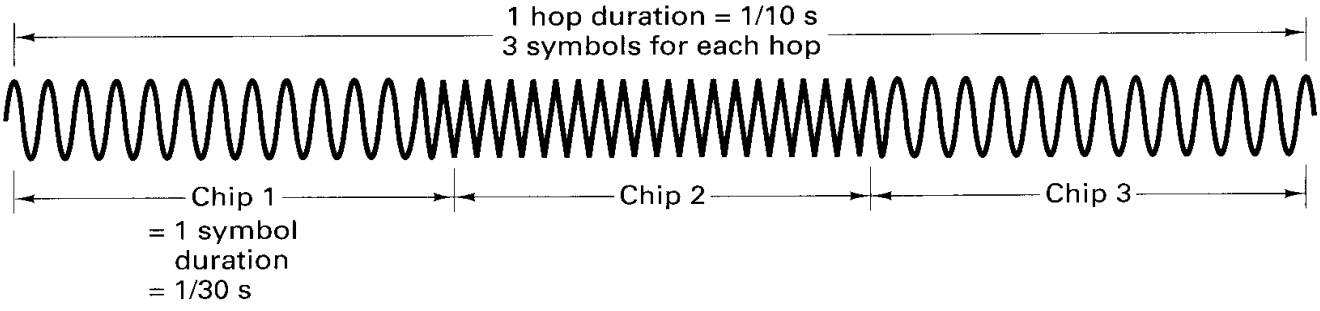


Frequency hopping example with diversity ($N = 4$).

shortest uninterrupted waveform in the system. Frequency hopping systems are classified as *slow-frequency hopping* (SFH), which means there are several modulation symbols per hop, or as *fast-frequency hopping* (FFH), which means that there are several frequency hops per modulation symbol. For SFH, the shortest uninterrupted waveform in the system is that of the data symbol; however, for FFH, the shortest uninterrupted waveform is that of the hop. Figure 12.14a illustrates an example of FFH; the data symbol rate is 30 symbols/s and the frequency hopping rate is 60 hops/s. The figure illustrates the waveform $s(t)$ over one symbol duration ($\frac{1}{30}$ s). The waveform change in (the middle of) $s(t)$ is due to a new frequency hop. In this example, a chip corresponds to a hop since the hop duration is shorter than the symbol duration. Each chip corresponds to half a symbol. Figure 14.14b illustrates an example of SFH; the data symbol rate is still 30 symbols/s, but the frequency hopping rate has been reduced to 10 hops/s. The waveform $s(t)$ is shown over a duration of three symbols ($\frac{1}{10}$ s). In this example, the hopping boundaries appear only at the beginning and end of the three-symbol duration. Here, the changes in the waveform are due to the modulation state changes; therefore, in this



(a)



(b)

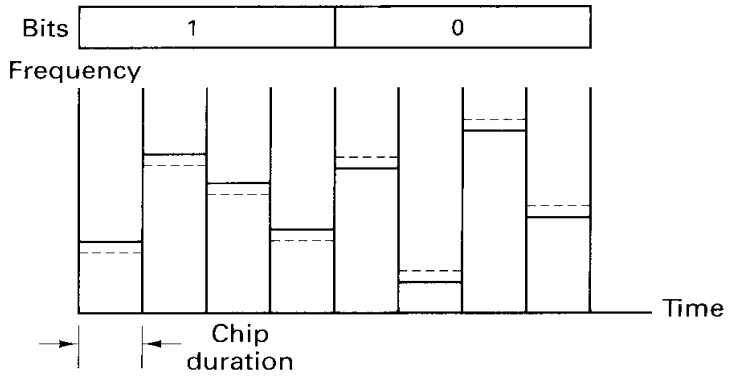
Chip—in the context of an FH/MFSK system. (a) Example 1: Frequency hopping MFSK system with symbol rate = 30 symbols/s and hopping rate = 60 hops/s. 1 chip = 1 hop. (b) Example 2: Same as part (a) except hopping rate = 10 hops/s. 1 chip = 1 symbol.

example a chip corresponds to a data symbol, since the data symbol is shorter than the hop duration.

FFH example of a binary FSK system. The diversity is $N = 4$. There are 4 chips transmitted per bit. As in Figure 12.13, the dashed line in each column corresponds to the center of the data band and the solid line corresponds to the symbol frequency. Here, for FFH, the chip duration is the hop duration. Figure 12.15b illustrates an example of an SFH binary FSK system. In this case, there are 3 bits transmitted during the time duration of a single hop. Here, for SFH, the chip duration is the bit duration. If this SFH example were changed from a binary system to an 8-ary system, what would the chip duration then correspond to? If the system were implemented as an 8-ary scheme, each 3 bits would be transmitted as a single data symbol. The symbol boundaries and the hop boundaries would then be the same, and the chip duration, the hop duration, and the symbol duration would all be the same.

FFH/MFSK Demodulator

Figure 12.16 illustrates the schematic for a typical fast frequency hopping MFSK (FFH/MFSK) demodulator. First, the signal is dehopped using a PN generator identical to the one used for hopping. Then, after filtering with a low-pass filter that



Fast hopping versus slow hopping in a binary system. (a) Fast-hopping example: 4 hops/bit. (b) Slow-hopping example: 3 bits/hop.

has a bandwidth equal to the data bandwidth, the signal is demodulated using a bank of M envelope or energy detectors. Each envelope detector is followed by a clipping circuit and an accumulator. The clipping circuit serves an important function in the presence of an intentional jammer or other strong unpredictable interference; it is treated in a later section. The demodulator does *not* make symbol decisions on a chip-by-chip basis. Instead, the energy from the N chips are accumulated, and after the energy from the Nth chip is added to the $N-1$ earlier ones, the demodulator makes a symbol decision by choosing the symbol that corresponds to the accumulator, z_i $(i = 1, 2, \ldots, M)$, with maximum energy.

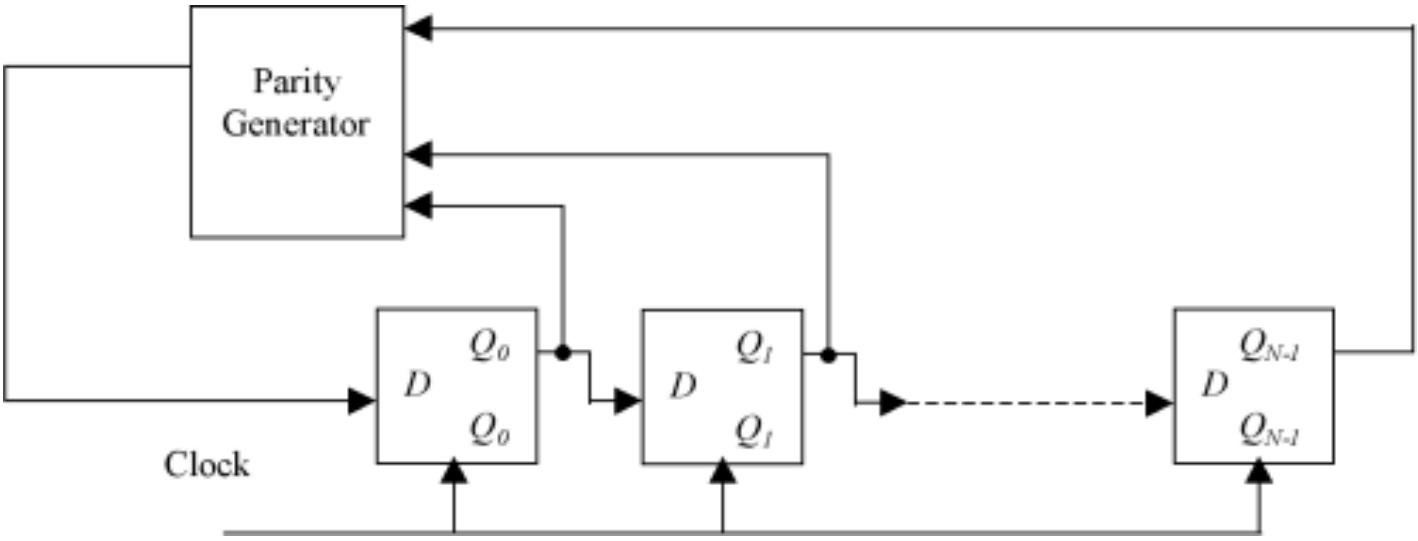
Processing Gain

Equation (12.27) shows the general expression for processing gain as $G_p = W_{ss}/R$. In the case of direct-sequence spread-spectrum, W_{ss} was set equal to the chip rate R_{ch}. In the case of frequency hopping, Equation (12.27) still expresses the processing gain, but here we set W_{ss} equal to the frequency band over which the system may hop. We designate this band as the *hopping band* $W_{hopping}$, and thus the processing gain for frequency hopping systems is written as

$$G_p = \frac{W_{\text{hopping}}}{R} \tag{12.29}$$

Generation and Characteristics of PN Sequences

A piece of hardware used to generate the PN sequences. It is mainly formed of a shift register made up of D flip-flops, with some selected outputs of the flip-flops connected to a parity generator. The parity generator is constructed of an array of exclusive-OR gates and it generates a logic 0 output when the input has an even number of 0's, and generates 1 when the input has an odd number of 1's.



The code generated depends on the number N of the flip-flops and on the selection of which flip-flop outputs are connected to the parity generator. At the beginning, all the Q's take specific pattern of logic values. Then, with each clock pulse these values are shifted and in general, the output of the parity generator will change. Since such hardware is fixed, the sequence is not truly random, i.e. it is deterministic. Also at some particular state, the sequence will start to repeat itself. In order to make that PN sequence almost random, a large number of flip-flops is used (today MOS large-scale integration may have 2000 flip-flops on a single chip), hence the sequence length will be large enough to consider the sequence as a random one. In fact, the maximum sequence length is L=2N-1. Note that the all-0's-state has been excluded because if the register shall ever arrive at such a state, it will remain in it permanently. The above maximal sequence length (L) can be attained for every (N) using certain flip-flop outputs as inputs for the parity generator. For example, at N=15, the sequence length can be made maximum, i.e. $L = 2^{15}-1 = 32767$, if the outputs of the flip-flops number 13 and 14 are used as inputs for the parity generator. It is worth noting that PN sequences are available at both outputs of any flip-flop. Yet, they are not independent, i.e. one can be derived from the other by a simple time shift or by complementing the bits or both. However, there are logic designs that produce sequences with small correlation to one another. The number of such independent sequences has the upper bound S given by S < (L-1)/N. Note that these independent sequences can be divided into two groups, one has sequences that are the mirror images of those in the other. Mirror image sequences have the same bit sequence when one is read forward and the other is read backward in time.

iii. Autocorrelation of a PN sequence

The autocorrelation function $R_d(t)$ of a truly random sequence d(t) with bit duration T_b is definedas

$R_d(t) = E\{d(t)\ d(T+t)\} = \text{Integral } d(t)d(T+t)\ dT$

And has the form shown in the figure below when d(t) = ± 1 volt

The autocorrelation function Rd(t) of a PN sequence is $R_{PN}(t) = E\{g(t)\ g(t + t)\}$ Where g(t) assumes the values ± 1 volt. As might be expected $R_{PN}(0) = 1$. Rpn(t) for t = nTc, where n is an integer and T_c is the chip duration.

$R_{PN}(t = nTc) = E\{g(t)g(t + nTc)\} = E\{ -g(t+kT_c)\}$

Now g(t + kTc) is a PN sequence and , in the course of L chips there is one more 1 than 0. Hence the average value of -g(t + kTc) is -1/L. Finally,since the sequence has a period

LT_c so too has $R_{PN}(t)$.

From the above, we deduce that an SS signal modulated with a PN sequence can only be demodulated by the exact PN sequence (t = exactly zero). For any other value of t, $R_d(t)$ will equal zero i.e. the energy of the received signal equals zero (no signal is received).

Use of SS with CDMA

In CDMA, each user is provided with a unique PN code which is almost uncorrelated with the other ones.To illustrate the principle of operation of CDMA, consider that at a given time, each of k users is transmitting data at the same carrier frequency fo, using DS-SS, and his particular code is gi(t). Then, each receiver is presented with the same input waveform, where each signal is assumed to present the same power P_s to the receiver, each pseudo-random sequence $g_i(t)$ has the same chip rate f_c and $d_i(t)$ is the data transmitted by the user i. The data rate for each user is the same, f_b.

$$v(t) = \sum_{i=1}^{k} \sqrt{2P_s}\, g_i(t) d_i(t) \cos(\omega_o t + \phi_1)$$

If the receiver is required to receive each of the k users it needs k correlators. At receiver 1, the signal v(t) will be multiplied by $g_1(t)$ and also by the corresponding carrier,

$$\sqrt{2}\cos(\omega_o t + \phi_1)$$

to generate the signal v_{o1} which is to be applied to the integrator to give the output

$$v'_{o1} = \sum_{i=1}^{k} \sqrt{P_s}\, g_1(t) g_i(t) d_i(t) \cos(\phi_i - \phi_1)$$

$$= \sqrt{P_s}\, d_1(t) + \sum_{i=2}^{k} \sqrt{P_s}\, g_1(t) g_i(t) d_i(t) \cos(\phi_i - \phi_1)$$

Note that at the output, the required signal is no more a SS signal, while all the other signals are still SS signals. This resulted from the fact that only $g_1^2(t)=1$, while the product $g_1(t)g_i(t)$ is still a random sequence itself having the same chip rate f_c as has any of the gi's individually.

The total power spectral density of k-1 independent interferers is the sum of the power spectral densities. The integrator which is essentially a low-pass filter with cutoff frequency $f_b(=1/T_b)$ allows the k-1 interferers to pass with total power spectral density

$$G_J(f) \approx (k-1)\frac{P_s}{4f_c} ; |f| \le f_b$$

The above approximation is valid since $f_b<c$, hence $G_J(f)$ can be considered constant over the given interval. This shows that, over that interval, the power of the interferers is much less than that of the required signal ($G_s(f)=P_s/f_b$).

Questions

A circuit that generates the upper and lower sidebands but no carrier is called a

A major benefit of SSB is

An AM signal without a carrier is called

An AM signal without a carrier is called a ---------- signal.

An amplitude modulator performs the mathematical operation of

Distortion of the modulating signal produces harmonics which cause an increase in the Signal

For a 100% modulation, the value of modulation index is

For a square law device,

For ideal AM,

In DSB-FC AM signal, greatest power is consumed by

In SSB, no signal is transmitted when

In SSB, which sideband is the best to use?

Indicate the false statement regarding the advantages of SSB over DSB-SC

Information in the AM signal is conveyed in the

Lower sideband (LSB) contain frequencies from

Modulation causes the information signal to be _____ to a higher frequency for more efficient transmission.

Over modulation occurs when

Recovering information from a carrier is known as

Square law modulator is used to generate ---------

The acronym SSB-SC means

The balanced modulator eliminates which of the following from its output?

The bandwidth required to transmit DSBSC wave is

The circuit that recovers the original modulating information from an AM signal is known as a

The degree or depth of modulation expressed as a percentage is computed using the Expression

The most widely used amplitude demodulator is called a

The new signals produced by modulation are called

The outline of the peaks of a carrier has the shape of the modulating signal and is called the

The purpose of a ---------- is to recover the original modulating signal from an AM wave.

The Vmax is a phasor addition of

To retain all the information,Two sidebands to be transmitted,Carrier to be transmitted

Upper sideband (USB) contain frequencies from

Vestigial Sideband modulation is normally used for

What is the minimum AM signal needed to transmit information?

Which of the following is not another name for modulation index?

Which of the following is not major communications medium?

Which of the following is not true about AM?

Which of the following is the most correct?

Which of the followings is not a source of noise?

Which one of the following cannot be used to remove the unwanted sideband in SSB.

Which one of the following is non linear device?

opt1	opt2	opt3	opt4
amplitude modulator	diode detector	class C amplifier	balanced modulator
higher power can be put into the sideband	greater power consumption	More carrier power	double the sideband power
SSB.	Vestigial sideband.	FM signal	DSB.
double sideband with carrier(DSB-FC)	single sideband (SS)	double sideband with no carrier(DSB-SC)	single sideband with no carrier(SSB)
Addition	Subtraction	Multiplication	Division
carrier power	bandwidth	sideband power	envelope voltage
1		0 <1	>1
output current is proportional to input voltage	output current is proportional to square of input voltage	output voltage is proportional to input current	output voltage is proportional to square of input current
m = 0	m = 1	m < 1	m > 1
carrier	upper side band	lower sideband	none of the above
the information signal is present	the information signal is absent	the carrier signal is present	the carrier signal is not present

upper	lower	neither	either upper or lower
more channel space is available.	more carrier power	signal is more noise resistant	less power is consumed
Carrier only.	sidebands only.	carrier and USB.	carrier and LSB.
fc-fm to fc	fc to fc +fm	fc to fc +2fm	fc-2fm to fc
Translated.	Multiplexed.	Multiplied.	Duplexed.
Vc > Vm	Vc < Vm	Vc = Vm	None of the above
Demultiplexing.	Modulation.	Detection.	Carrier recovery.
DSB-FC wave	VSB wave	SSB wave	none of the above
single sideband with carrier	single sideband with suppressed carrier	double sideband with no carrier	double sideband with carrier
upper sideband	carrier	lower sideband	both sidebands
equal to carrier frequency	twice the maximum modulating signal frequency	equal to the maximum modulating signal frequency	twice the carrier frequency

modulator	demodulator	mixer	crystal set
2 Vm.	100 / Vm.	m / 100.	100 % * m.
variable resistor	rectifier	diode detector	adder
Spurious emissions	Harmonics	Intermodulation products	Side bands.
Trace.	Wave shape.	Envelope.	Carrier variation.
Demodulator	Modulator	rectifier	adder
Vc+Vlsb+Vusb	Vc-Vlsb-Vusb	-Vc-Vlsb-Vusb	-Vc+Vlsb+Vusb
both A & B	only A	only B	neither A nor B
fc-fm to fc	fc to fc +fm	fc to fc +2fm	fc-2fm to fc
HF point-to-point communication	monaural broadcasting	TV broadcasting	stereo broadcasting
Carrier plus sidebands.	Carrier only.	One sideband.	Both sidebands.

Modulation reciprocal.	Modulation factor.	Degree of modulation.	Modulation Co-efficient.
Free space.	Water.	Wires.	Fiber-optic cable.
The carrier amplitude varies.	The carrier frequency remains constant.	The carrier frequency changes.	The information signal amplitude changes the carrier amplitude.
Vm should be greater than Vc.	Vc should be greater than Vm.	Vm must be equal to or less than Vc.	Vc must always be equal to Vm.
Another communications signal.	Atmospheric effects.	Manufactured electronic components.	Noise from industry.
filter system	phase shift method	third method	balanced modulator
multiplexer	modulator	diode	rectifier

opt5	opt6	Answer
		balanced modulator
		higher power can be put into the sideband
		DSB.
		double sideband with no carrier(DSB-SC)
		Multiplication
		bandwidth
		1
		output current is proportional to square of input voltage
		m = 1
		carrier
		the information signal is absent

either upper or lower

more carrier power

sidebands only.

fc-fm to fc

Translated.

Vc < Vm

Detection.

DSB-FC wave

single sideband with
suppressed carrier

carrier

twice the maximum
modulating signal
frequency

demodulator

100 % * m.

diode detector

Side bands.

Envelope.

Demodulator

Vc+Vlsb+Vusb

neither A nor B

fc to fc +fm

TV broadcasting

One sideband.

Modulation reciprocal.

Water.

The carrier frequency changes.

Vc should be greater than Vm.

Another communications signal.

balanced modulator

diode

Questions

A PLL frequency demodulator automatically compensate for changes in ---------

A PLL frequency demodulator requires

A varying phase shift produces a

An FM signal produces more -------- than an AM signal.

Assertion: An FM signal produces more sidebands than an AM signal.Reason:

Balanced slope detector consists of ---------- single ended slope detectors.

Both FM and PM are types of _____ modulation.

Carson's rule is used to calculate

For a narrowband FM, the modulation index is

For a wideband FM, the modulation index is

Foster Seeley discriminator can also be called as

If the amplitude of the modulating signal decreases the carrier deviation

In balanced slope detector, two single ended slope detectors are connected in

In direct method of FM generation, --------- is varied in accordance with baseband signal

In FM , the total transmitted power

In FM, as the modulating signal amplitude goes positive, the carrier frequency -------

In FM, the required bandwidth -------

In PM, carrier frequency deviation is not proportional to

Indicate which one of the following is not an advantage of FM over AM

Pre emphasis is used to boosting

The amount of frequency shift during modulation is called the

The amount of frequency shift in FM is directly proportional to the ---
----

The amplitudes of the sidebands in an FM signal are dependent upon a mathematical Process

The area of coverage of FM system is --------- than AM system.

The FM produced by PM is called

The general name given to both FM and PM is ---------- modulation.

The indirect method of FM generation needs

The low index FM are known as

The main advantage of FM over AM is its immunity from

The number of sidebands in FM is

The phenomenon of a strong FM signal dominating a weaker signal on a common frequency is referred to as the

The primary disadvantage of FM is its

The rate at which frequency variation takes place in FM is equal to the

The ratio detector is more immune to

The sidebands of FM signals are separated from the carrier by

Voltage versus frequency response curve of Foster Seeley discriminator is called as

What is the bandwidth of a FM signal for a 70 KHz carrier has a frequency deviation of 4 KHz with a 1000 Hz signal.

Which of the following is not a major benefit of FM over AM

Which one of the following has good noise immunity?

Which one of the following is not present in the PLL demodulator?

opt1	opt2	opt3
load	carrier frequency	carrier amplitude
no transformer	no tuned circuit	no coupling
amplitude deviation	frequency deviation	modulation index.
noise	attenuation	sidebands
The amplitude of the carrier signal is high.	The frequency of the carrier signal is high.	The frequency of modulated signal is high.
three	two	one
Amplitude.	Phase.	Analog.
Maximum deviation	Modulation Index	Deviation sensitivity
less than unity	exceed unity	equal to unity
less than unity	exceed unity	equal to unity

balanced slope detector	ratio detector	slope detector
Increases.	Decreases.	Remains constant.
series	cascade	parallel
carrier frequency	carrier amplitude	carrier phase
remains constant	increases with depth of modulation	decreases with depth of modulation
increases	decreases	remains constant
remains constant	increases with depth of modulation	decreases with depth of modulation
Modulating signal amplitude.	Carrier Amplitude and frequency.	Modulating signal frequency.
better noise immunity	lower bandwidth	better transmission efficiency
low frequencies	medium frequencies	high frequencies
amplitude	phase.	deviation.

amplitude of the carrier signal	amplitude of the modulating signal	frequency of the carrier signal
Taylor’s series	Fourier series	Bernaulis series
larger	smaller	same
Direct FM.	Indirect FM.	Direct AM.
frequency	amplitude	angle
amplitude modulator	frequency modulator	differentiator
Broad Band	Narrow band	Low band
noise	atmospheric condition	attenuation
Small	Large	Zero
3	infinite	finite
blot out	quieting factor	domination syndrome

Higher cost and complexity.	Excessive use of spectrum space.	Nose susceptibility.
carrier amplitude	carrier frequency	modulating signal amplitude
frequency variation in input signal	phase variation in input signal	amplitude variation in input signal
more linear than that of slope detector	more non-linear than that of slope detector	more linear than that of ratio detector
B-curve	S-curve	D-curve
At a maximum.	At a minimum.	Zero.
Greater Efficiency.	Noise Immunity.	Capture Effect.
AM	PM	FM

opt4	opt5	opt6	Answer
modulating frequency			carrier frequency
no buffer			no tuned circuit
capture effect.			frequency deviation
amplification			sidebands
The amplitude of the modulating signal is high.			The frequency of modulated signal is high.
more than two			two
Angle.			Angle.
Bandwidth			Bandwidth
zero			less than unity
zero			exceed unity

phase shift discriminator	phase shift discriminator
Goes to Zero.	Increases.
mixed configuration	parallel
modulating signal amplitude	carrier frequency
increases linearly with depth of modulation	remains constant
linearily increases	increases
increases linearly with depth of modulation	increases with depth of modulation
modulator phase shift.	Modulating signal frequency.
less modulating power	lower bandwidth
radio frequencies	high frequencies
angle.	deviation.

frequency of the modulating signal	amplitude of the modulating signal
Bessel’s function	Bessel’s function
much larger	smaller
Indirect PM.	Indirect FM.
phase	angle
phase modulator	phase modulator
Medium band	Narrow band
weather condition	noise
Infinite	Small
2	infinite
capture effect	capture effect

Lower Efficiency.

Excessive use of spectrum space.

modulating frequency

modulating frequency

atmospheric condition

amplitude variation in input signal

more non-linear than that of ratio detector

more linear than that of slope detector

U-curve

S-curve

Infinity.

At a maximum.

Lower complexity and cost.

Lower complexity and cost.

None of the above

FM

Questions	opt1
Noise from the sun and stars is called	Extraterrestrial noise
Atmospheric noise comes mainly from	star
The main source of internal noise is -----	fluorescent light
The S/N ratio is usually expressed in	watts
For best reception, the S/N ratio should be	low
Increasing the temperature of a component causes its noise power to	remains constant
Another name for thermal noise is A Johnson noiseB.White noise	both A & B
Narrowing the bandwidth of a circuit causes the noise level to	remains constant
The noise voltage produced across a 75 ohm input resistance at atemperature of 25oC with a banwidth of 1.5 MHz is --------	1 microvolt
Two types of noise produced by tubes or transistors are	shot noise & transit time noise

The receiver amplifies only signalB. only noise	both A & B
Noise temperature in degrees Kelvin is used to express the noise in a system at --------- frequencies.	low
The noise figure of an amplifier is 2.6. The noise temperature is	564 K
The noise is more of a problem at -------- frequencies.	audio
The stages of a receiver that contribute the most noise are the	RF amplifier & mixer
Which one of the following doesn’t come under external noise?	industrial noise
Which one of the following doesn’t come under internal noise?	shot noise
Which circuit contributes most to the noise in a receiver?	IF amplifier
Noise can be reduced by	widening the bandwidth
Atmospheric noise becomes less severe at frequencies	above 30 MHz.
One of the following types of noise becomes of great importance at high frequencies. It is the	shot noise

The value of a resistor creating thermal noise is doubled. The noise power generated is therefore	halved
Indicate the noise whose source is in a category different from that of the other three.	solar noise
Indicate the false statement. The square of the thermal noise voltage generated by a resistor is proportional to	its resistance
Indicate the false statement. The square of the shot noise current is proportional to	its resistance
The noise figure is defined as the ratio of	output S/N to input S/N
The ideal value of noise figure is	0
In a practical receiver, the the value of noise figure is	0
In a receiver, the output S/N will be --------- the input S/N	equal to
The equivalent noise temperature is given by	To(F+1)
A receiver connected to an antenna whose resistance is 50 ohm has an equivalent noise resistance of 30 ohm. The receiver's noise figure in db is	1.6
The available thermal noise power is	4RkTB

The noise voltage of the equivalent series resistance is square root of	En1+En2+En3+….
The noise current of the equivalent parallel resistance is square root of	I n1+In2+In3+….
The shot noise of a component, with a direct current of 1mA flowing across a semiconductor junction , given that the effective noise bandwidth of 1MHz is	18 nA
Flicker or 1/f noise appears	below the frequencies of few hertz
Noise figure is a ---------- expressed in decibels.	noise temperature
---------- formula is used for the calculation of noise temperature.	Friss's
Excess noise ratio(ENR) is defined as the ratio of	Tc-Th to Tc
The power ratio termed Y factor is	Y=Ph/Pc

opt2	opt3	opt4	opt5
Atmospheric noise	Thermal noise	Shot noise	
sun	lightning	space	
thermal agitation	automobiles	lightning	
HP	volts	decibels	
moderate	high	very low	
decreases	increases	exponentially increases	
only A	only B	neither A nor B	
decreases	increases	exponentially increases	
1.56 microvolt	1.46 microvolt	1.36 microvolt	
shot noise & thermal noise	thermal noise & transit time noise	internal & external noise	

only A	only B	neither A nor B
medium	microwave	audio
645 K	464 K	428K
low	medium	high
IF amplifier & mixer	RF amplifier & IF amplifier	detector & audio amplifier
atmospheric noise	space noise	thermal noise
transit time noise	space noise	thermal noise
demodulator	AF amplifier	mixer
narrowing the bandwidth	increasing the temperature	increasing the transistor current level
below 30MHz	above 1GHz	below 1GHz
random noise	impulse noise	transit time noise

quadrupled	doubled	unchanged
cosmic noise	atmospheric noise	galactic noise
its temperature	Boltzmann’s constant	the bandwidth over which it is measured
direct diode current temperature	charge of the electron	the bandwidth
input S/N to output S/N	signal power to noise power	noise power to signal power
greater than 1	1	less than 1
less than 1	1	greater than 1
less than	greater than	much less than
Te(F+1)	To(F-1)	To(1-F)
6.1	4.02	2.04
RkTB	4kTB	kTB

En12+En22+En32+….	En12 En22 En32….	En1 En2 En3….
I n12+I n22+I n32+….	I n1 n2I n3….	I n12 I n22 I n32….
9nA	36 nA	18mA
above the frequencies of few kilohertz	below the frequencies of few kilohertz	above the frequencies of few hertz
equivalent noise temperature	noise bandwidth	noise factor
Willie’s	Taylor’s	Legrange’s
Th-Tc to Tc	Th- Tc to Th	Tc -Th to Th
Y=Pc/Ph	Y= Th/Tc	Y=Tc/Th

opt6	Answer
	Extraterr estrial noise
	lightning
	thermal agitation
	decibels
	high
	increases
	both A & B
	decreases
	1.36 microvolt
	shot noise & transit time

both A & B

microwave

464 K

high

high

thermal noise

space noise

mixer

narrowing the bandwidth above 30 MHz.

random noise

doubled

atmosphe
ric noise

Boltzman
n’s
constant

its
resistanc
e

input
S/N to
output
S/N
1

greater
than 1

less than

To(F-1)

2.04

kTB

En12+En
22+En32
+….

I n12+I
n22+I
n32+….

18 nA

below
the
frequenci
es of
noise
factor

Friss’s

Th-Tc
to Tc

Y=Ph/Pc

Questions	opt1	
In a receiver, the output S/N will be --------- the input S/N	equal to	
The equivalent noise temperature is given by	To(F+1)	
A receiver connected to an antenna whose resistance is 50 ohm has an equivalent noise resistance of 30 ohm. The receiver's noise figure in db is		1.6
The available thermal noise power is	4RkTB	
The noise voltage of the equivalent series resistance is square root of	En1+En2+En3+….	
The noise current of the equivalent parallel resistance is square root of	I n1+In2+In3+….	
The shot noise of a component, with a direct current of 1mA flowing across a semiconductor junction , given that the effective noise bandwidth of 1MHz is	18 nA	
Flicker or 1/f noise appears	below the frequencies of few hertz	
Noise figure is a ---------- expressed in decibels.	noise temperature	
---------- formula is used for the calculation of noise temperature.	Friss's	

Excess noise ratio(ENR) is defined as the ratio of	Tc-Th to Tc
The power ratio termed Y factor is	Y=Ph/Pc
The relationship between ENR and noise factor is	F= Y/(ENR-1)
The overall noise figure of a three stage cascaded amplifier, each stage having a power gain of 10db and a noise figure of 6db is	3.44
DSB-SC system contains ---------- power in sidebands.	2/3
For AM, the detector noise output voltage is A Independent of the noise frequency .B. Same at all frequencies	both A & B
Noise power for AM follows ---------- distribution.	triangular
For FM system, the noise output power is proportional to	input noise voltage
In AM system, the post detection signal to noise power ratio is ---------- greater than the pre detection signal to noise power ratio.	2 times
In receiver system, input SNR and output SNR are measured at	mixer
For SSB system, the signal to noise power ratios at the input of the demodulator is--------- the output of the demodulator.	less than that in

The bandwidth of the filter in SSB-SC will be ------- equal to
-- that of the filter in DSB-SC.

From the point of noise improvement, the performance of SSB is --------- that of DSB. identical to

For a given signal power at the input, the ------- at the output is identical for DSB and SSB. noise to signal ratio

As the bandwidth increases in FM system, the input noise also increases and the point called -------- is reached. threshold

The threshold effect is more pronounced in ------- system. AM

The condition at which the noise is not distinguishable from signal is called capture effect

The pre emphasis circuit boosts --------- at the transmitter. low frequencies

The de emphasis compensate --------- at the receiver. low frequencies

The pre emphasis and de emphasis are applied in ---- AM
--------- system for improvement of SNR.

opt2 less than	opt3 greater than	opt4 much less than
Te(F+1)	To(F-1)	To(1-F)
6.1	4.02	2.04
RkTB	4kTB	kTB
En12+En22+En32+….	En12 En22 En32….	En1 En2 En3….
I n12+I n22+I n32+….	I n1 n2I n3….	I n12 I n22 I n32….
9nA	36 nA	18mA
above the frequencies of few kilohertz	below the frequencies of few kilohertz	above the frequencies of few hertz
equivalent noise temperature	noise bandwidth	noise factor
Willie’s	Taylor’s	Legrange’s

Th-Tc to Tc	Th- Tc to Th	Tc -Th to Th
Y=Pc/Ph	Y= Th/Tc	Y=Tc/Th
F= Y/(ENR+1)	ENR =(Y-1)/F	f= ENR/(Y-!)
4.33	3.43	3.23
1/3	entire	½
only A	only B	neither A nor B
square	circular	rectangular
frequency deviation	bandwidth	noise temperature
3 times	4 times	5 times
IF amplifier	demodulator	audio amplifier
greater than that in	identical to	twice that in

half	twice	thrice
better than	worst than	excellent than
threshold voltage	signal to noise ratio	capture effect
capture	breakdown	avalanche
DSB-SC	SSB-SC	FM
threshold effect	null effect	avalanche effect
medium frequencies	high frequencies	radio frequencies.
medium frequencies	high frequencies	radio frequencies.
DSB-SC	SSB-SC	FM

opt5	opt6	Answer
		less than
		To(F-1)
		2.04
		kTB
		En12+En22+En32+….
		I n12+I n22+I n32+….
		18 nA
		below the frequencies of few kilohertz
		noise factor
		Friss’s

Th-Tc to Tc

Y=Ph/Pc

f= ENR/(Y-!)

4.33

entire

both A & B

rectangular

frequency deviation

2 times

demodulator

identical to

half

identical to

signal to noise ratio

threshold

FM

threshold effect

high frequencies

high frequencies

FM

Questions	opt1
A sampled signal is	Analog signal
Attenuation is more in	analog communication
Band pass signal is	modulated signal
Bandwidth required for transmission of non sinusoidal signal is	greater than that of sinusoidal signal
Base band communication needs	unguided media
Base band communication suitable for	long distance communication
Base band communication suitable for	short distance communication
Base band signal is	modulated signal
Baud rate is defined as	number of symbols transmitted in one second
Bit rate is	directly proportional to baud rate

Bit rate is defined as	number of symbols transmitted in one sec
Communication channel is the medium	used for transfer of signal
Electrical signal is converted into voice signal by	loud speaker
For a good communication system, signal to noise ratio should be	low
Frequency band used for TV transmission is	VHF & SHF
In a communication system, noise is most likely to affect the signal	at the transmitter
Increasing the number of voltage levels for transmission	decrease the capacity
Indicate the false statement. Modulation is used to	reduce the bandwidth used
Indicate the noise whose source is in a category different from that of the other	Solar noise
Interference is more in analog communication because	the signal is in digital form
Maximum number of bits transmitted in one sec is	Capacity

Modem is used to	multiplex the signal
Nyquist formula relating bandwidth and capacity is	C = 2B log (2M)
Optical fiber carries	gamma ray
Pass band communication suitable for	long distance communication
Pass band communication suitable for	short distance communication
Radio communication use	coaxial cable
Shannon’s capacity theorem is	C= 2B log (1+SNR)
Signal to noise ratio is	ratio of noise power to signal power
Signals that travel through free space for long distance are	electromagnetic waves
The frequencies above 1 GHz are called	Radiowaves
The frequency range from 300-3400Hz is	voice frequency range

The radio frequency range starts at	1MHz
The wave used for medical application is	optical waves
Voice signal is converted into electrical signal by	loud speaker
Which one of the following is not an information source?	sound
Which one of the following is not in the communication system	transmitter
Which one of the following is not present in the analog modulation process	receiver

opt2	opt3	opt4	opt5
Digital signal	Discrete signal	Continuous signal	
digital communication	both analog and digital communication	none of the above	
encoded signal	decoded signal	original signal	
less than that of sinusoidal signal	same as that of sinusoidal signal	none of the above	
modulator	demodulator	guided media	
short distance communication	broad band communication	narrow band communication	
Long distance communication	both long & short distance communication	none of the above	
encoded signal	decoded signal	original signal	
number of bits transmitted in one second	number of symbols transmitted in one minute	number of bits transmitted in one minute	
inversely proportional to baud rate	directly proportional to bit interval	inversely proportional to bit interval	

number of bits transmitted in one sec	number of symbols transmitted in one minute	number of bits transmitted in one minute
used for encoding	used for receiving signal	used for transmitting signal
modulator	demodulator	microphone
very low	high	moderate
VHF & UHF	UHF& SHF	UHF & EHF
in the channel	in the information source	at the destination
decrease the signal to noise ratio	increase the capacity	increase the signal to noise ratio
separate differing transmissions	for long distance communication	allow the use of parabolic antenna
Cosmic noise	atmospherics noise	Galactic noise
the signal is in analog form	both signal and noise are in analog form	noise is in analog form
bit rate	baud rate	bit interval

modulate the signal	demodulate the signal	both modulate and demodulate the signal
C = B log (M)	C = B log (2M)	C = 2B log (M)
infrared wave	electromagnetic wave	light wave
short distance communication	broad band communication	narrow band communication
Long distance communication	both long & short distance communication	none of the above
free space	optical fiber	microwave links
C= B log (1+SNR)	C= B log (3+SNR)	C= B log (2+SNR)
ratio of noise power to modulated signal power	ratio of signal power to noise power	ratio of signal voltage to noise voltage
baseband signals	sinusoidal signals	non sinusoidal signals
microwaves	opticalwaves	gammarays
audio frequency range	radio frequency range	microwave frequency range

100MHz	200MHz	300MHz
radio waves	microwaves	x-rays
modulator	demodulator	microphone
video	samples	audio
receiver	channel	rectifier
channel encoder	modulator	demodulator

opt6	Answer Discrete signal
	digital communication
	modulated signal
	greater than that of sinusoidal signal
	guided media
	short distance communication
	short distance communication
	original signal
	number of symbols transmitted in one second
	inversely proportional to bit interval

number of bits
transmitted in one sec

used for transfer of signal

loud speaker

high

VHF & UHF

in the channel

increase the capacity

for long distance
communication

atmospherics noise

both signal and noise are
in analog form

Capacity

both modulate and
demodulate the signal

C = 2B log (M)

light wave

long distance
communication

Long distance
communication

free space

C= B log (1+SNR)

ratio of signal power to
noise power

electromagnetic waves

microwaves

voice frequency range

200MHz

x-rays

microphone

samples

rectifier

channel encoder