

REFERENCES

1. J. Singh, *Great Ideas of Modern Mathematics*, Dover, Boston, MA, 1959.
2. M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions*, National Bureau of Standards, Washington, DC, 1964, sec. 26.
3. The Chemical Rubber Co., *CRC Standard Mathematical Tables*, 26th ed., 1980.
4. J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*, Wiley, New York, 1965, p. 83.
5. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed., McGraw-Hill, New York, 1995.
6. N. S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice-Hall, Englewood Cliffs, NJ, 1984.

PROBLEMS

- 10.1-1** A card is drawn randomly from a regular deck of cards. Assign probability to the event that the card drawn is: (a) a red card; (b) a black queen; (c) a picture card (count an ace as a picture card); (d) a number card with number 7; (e) a number card with number ≤ 5 .
- 10.1-2** Three regular dice are thrown. Assign probabilities to the following events: the sum of the points appearing on the three dice is: (a) 4; (b) 9; (c) 15.
- 10.1-3** The probability that the number i appears on a throw of a certain loaded dice is k_i ($i = 1, 2, \dots, 6$). Assign probabilities to all six outcomes.
- 10.1-4** A bin contains three oscillator microchips, marked O_1 and O_2 , and O_3 , and two PLL microchips, marked P_1 and P_2 . Two chips are picked randomly in succession without replacement.
- (a) How many outcomes are possible, that is, how many points are in the sample space? List all the outcomes and assign probabilities to each of them.
 - (b) Express the following events as unions of the outcomes in part (a) (i) one chip drawn is oscillator and the other is PLL; (ii) both chips are PLL; (iii) both chips are oscillator; and (iv) both chips are of the same kind. Assign probabilities to each of these events.
- 10.1-5** Find the probabilities in Prob. 10.1-4, part (b), using Eq. (10.12).
- 10.1-6** In Prob. 10.1-4, determine the probability that:
- (a) The second pick is an oscillator chip given that the first pick is a PLL chip.
 - (b) The second pick is an oscillator chip given that the first pick is also an oscillator chip.
- 10.1-7** A binary source generates digits 1 and 0 randomly with equal probability. Assign probabilities to the following events. In ten digits generated by the source (a) there are exactly two 1's and eight 0's; (b) there are at least four 0's.
- 10.1-8** In the California lottery (Lotto), a player chooses any 6 numbers out of 49 numbers (1 through 49). Six balls are drawn randomly (without replacement) from the 49 balls numbered 1 through 49.
- (a) Find the probability of matching all 6 balls to the 6 numbers chosen by the player.
 - (b) Find the probability of matching exactly 5 balls.
 - (c) Find the probability of matching exactly 4 balls.
 - (d) Find the probability of matching exactly 3 balls.

- 10.1-9** A system consists of ten subsystems $s_1, s_2, \dots, s_{10}$ in cascade (Fig. P10.1-9). If any one of the subsystems fails, the entire system fails. The probability of failure of any one of the subsystems is 0.01.

- (a) What is the probability of failure of the system? *Hint:* Consider the probability that none of the subsystems fails.
- (b) The reliability of a system is the probability of not failing. If the system reliability is required to be 0.99, what must be the failure probability of each subsystem?

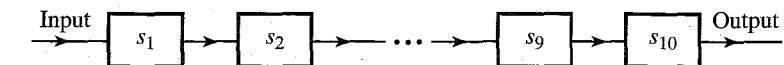


Figure P10.1-9

- 10.1-10** System reliability improves by using redundant systems. The reliability of the system in Prob. 10.1-9 (Fig. P10.1-9) can be improved by using two such systems in parallel (Fig. P10.1-10). Thus, if one system fails, the other one will still function.

- (a) Using the data in Prob. 10.1-9, determine the reliability of the system in Fig. P10.1-10.
- (b) If the reliability of this system is required to be 0.999, what must be the failure probability of each subsystem?

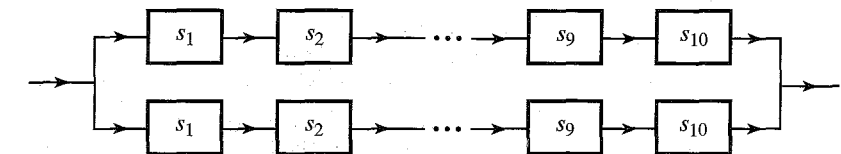


Figure P10.1-10

- 10.1-11** Compare the reliability of the two systems in Fig. P10.1-11, given that the probability of failure of subsystems s_1 and s_2 is p each.

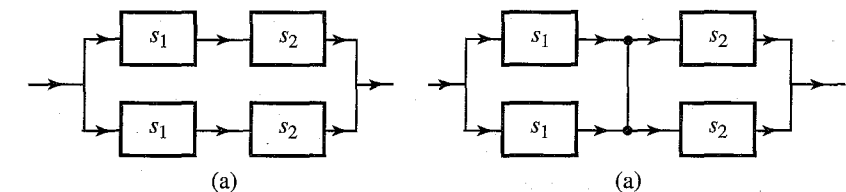


Figure P10.1-11

- 10.1-12** In a poker game each player is dealt five cards from a regular deck. What is the probability that a player will get a flush (all five cards of one suit)?

- 10.1-13** Two dice are tossed. One die is regular and the other is biased with probabilities:

$$P(1) = P(6) = \frac{1}{6}, \quad P(2) = P(4) = 0, \quad P(3) = P(5) = \frac{1}{3}$$

Determine the probabilities of obtaining a sum: (a) 4; (b) 5.

10.1-14 In Sec. 10.1, Example 10.5, determine:

- (a) $P(B)$, the probability of drawing a red ace in the second draw.
 (b) $P(A|B)$, the probability that the first draw was a red ace given that the second draw is a red ace. *Hint:* Event B can occur in two ways: the first draw is a red ace and the second draw is a red ace, or the first draw is not a red ace and the second draw is a red ace. This is $AB \cup A^cB$ (see Fig. 10.2).

10.1-15 A binary source generates digits 1 and 0 randomly with probabilities $P(1) = 0.8$ and $P(0) = 0.2$.

- (a) What is the probability that two 1's and three 0's will occur in a five-digit sequence?
 (b) What is the probability that at least three 1's will occur in a five-digit sequence?

10.1-16 In a binary communication channel, the receiver detects binary pulses with an error probability P_e . What is the probability that out of 100 received digits, no more than three digits are in error?

10.1-17 A PCM channel is made up of 10 links, with a regenerative repeater at the end of each link. If the detection error probabilities of the 10 detectors are $p_1, p_2, \dots, p_{10}$, determine the detection error probability of the entire channel if $p_i \ll 1$.

10.1-18 Example 10.8 in Sec. 10.1 considers the possibility of improving reliability by three repetitions of a digit. Repeat this analysis for five repetitions.

10.1-19 A bin contains nine bad microchips. A good microchip is thrown into the bin by mistake. Someone is trying to retrieve the good chip. He draws a chip randomly and tests it. If the chip is bad he throws it out and draws another chip randomly, repeating the procedure until he finds the good chip.

- (a) What is the probability that he will find the good chip in the first trial?
 (b) What is the probability that he will find the good chip in five trials?

10.1-20 One out of a group of 10 people is to be selected for a suicide mission by drawing straws. There are 10 straws—nine are of the same length and the tenth is shorter than the others. The straws are drawn one by one by all ten people. The person who draws the shortest straw is selected for the mission. Determine which position in the sequence favors the most and which favors the least drawing the short straw.

10.1-21 A binary source generates digits 1 and 0 randomly with probabilities $P(1) = 0.7$ and $P(0) = 0.3$.

- (a) What is the probability that no 1's will be generated in a sequence of 10 digits?
 (b) What is the probability that eight 1's and two 0's will be generated in a sequence of 10 digits?
 (c) What is the probability that at least five 0's will be generated in a sequence of 10 digits?

10.2-1 For a certain binary nonsymmetric channel it is given that

$$P_{y|x}(0|1) = 0.1 \quad \text{and} \quad P_{y|x}(1|0) = 0.2$$

where x is the transmitted digit and y is the received digit. If $P_x(0) = 0.4$, determine $P_y(0)$ and $P_y(1)$.

10.2-2 A binary symmetric channel (see Example 10.6) has an error probability P_e . The probability of transmitting 1 is Q , and the probability of transmitting 0 is $1 - Q$. If the receiver detects an incoming digit as 1, what is the probability that the corresponding transmitted digit was:

(a) 1; (b) 0? *Hint:* If x is the transmitted digit and y is the received digit, you are given $P_{y|x}(0|1) = P_{y|x}(1|0) = P_e$. Now using Bayes' rule, find $P_{x|y}(1|1)$ and $P_{x|y}(0|1)$.

10.2-3 The PDF of amplitude x of a certain signal $x(t)$ is given by $p_x(x) = 0.5|x|e^{-|x|}$.

- (a) Find the probability that $x \geq 1$.
 (b) Find the probability that $-1 < x \leq 2$.
 (c) Find the probability that $x \leq -2$.

10.2-4 The PDF of an amplitude x of a gaussian signal $x(t)$ is given by

$$p_x(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$$

This signal is applied to the input of a half-wave rectifier circuit (Fig. P10.2-4). Assuming an ideal diode, determine $F_y(y)$ and $p_y(y)$ of the output signal amplitude y .

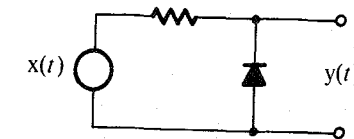


Figure P10.2-4

10.2-5 The PDF of a gaussian variable x is given by

$$p_x(x) = \frac{1}{3\sqrt{2\pi}} e^{-(x-4)^2/18}$$

Determine: (a) $P(x \geq 4)$; (b) $P(x \geq 0)$; (c) $P(x \geq -2)$.

10.2-6 For an RV x with PDF

$$p_x(x) = \frac{1}{2\sqrt{2\pi}} e^{-x^2/32} u(x)$$

- (a) Sketch $p_x(x)$ and state (with reasons) if this is a gaussian RV.
 (b) Determine: (i) $P(x \geq 1)$, (ii) $P(1 < x \leq 2)$.
 (c) Can you generate RV x from another gaussian RV? Explain.

10.2-7 The joint PDF of RVs x and y is shown in Fig. P10.2-7.

- (a) Determine: (i) A ; (ii) $p_x(x)$; (iii) $p_y(y)$; (iv) $P_{x|y}(x|y)$; (v) $P_{y|x}(y|x)$.
 (b) Are x and y independent? Explain.

10.2-8 The joint PDF $p_{xy}(x, y)$ of two continuous RVs is given by

$$p_{xy}(x, y) = xy e^{-(x^2+y^2)/2} u(x)u(y)$$

- (a) Find $p_x(x)$, $p_y(y)$, $p_{x|y}(x|y)$, and $p_{y|x}(y|x)$.
 (b) Are x and y independent?

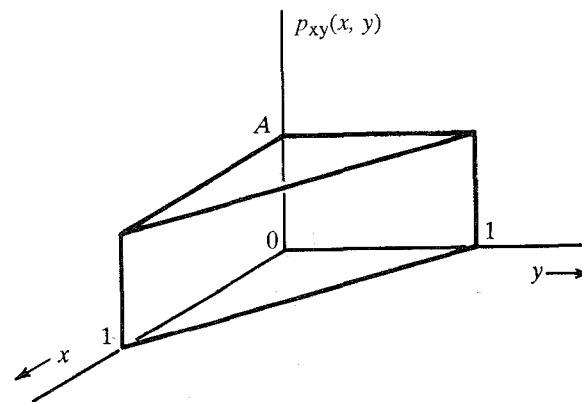


Figure P10.2-7

10.2-9 RVs x and y are said to be jointly gaussian if their joint PDF is given by

$$p_{xy}(x, y) = \frac{1}{2\pi\sqrt{M}} e^{-(ax^2 + by^2 - 2cxy)/2M}$$

where $M = ab - c^2$. Show that $p_x(x)$, $p_y(y)$, $p_{x|y}(x|y)$, and $p_{y|x}(y|x)$ are all gaussian and that $\overline{x^2} = b$, $\overline{y^2} = a$, and $\overline{xy} = c$. *Hint:* Use

$$\int_{-\infty}^{\infty} e^{-px^2 + qx} dx = \sqrt{\frac{\pi}{p}} e^{q^2/4p}$$

10.2-10 The joint PDF of RVs x and y is given by

$$p_{xy}(x, y) = ke^{-(x^2 + xy + y^2)}$$

Determine: (a) the constant k ; (b) $p_x(x)$; (c) $p_y(y)$; (d) $p_{x|y}(x, y)$; (e) $p_{y|x}(y|x)$. Are x and y independent?

10.2-11 In the example on threshold detection (Example 10.13), it was assumed that the digits 1 and 0 were transmitted with equal probability. If $P_x(1)$ and $P_x(0)$, the probabilities of transmitting 1 and 0, respectively, are not equal, show that the optimum threshold is not 0 but is a , where

$$a = \frac{\sigma_n^2}{2A_p} \ln \frac{P_x(0)}{P_x(1)}$$

Hint: Assume that the optimum threshold is a , and write P_e in terms of the Q functions. For the optimum case, $dP_e/da = 0$. Use the fact that

$$Q(x) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy$$

and

$$\frac{dQ(x)}{dx} = -\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

10.3-1 If an amplitude x of a gaussian signal $x(t)$ has a mean value of 2 and an rms value of 3, determine its PDF.

10.3-2 Determine the mean, the mean square, and the variance of the RV x in Prob. 10.2-3.

10.3-3 Determine the mean and the mean square value of y in Prob. 10.2-4.

10.3-4 Determine the mean and the mean square value of x in Prob. 10.2-6.

10.3-5 Find the mean, the mean square, and the variance of the RV x whose PDF is shown in Fig. P10.3-5.

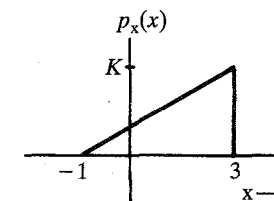


Figure P10.3-5

10.3-6 The sum of the points on two tossed dice is a discrete RV x , as analyzed in Example 10.9. Determine the mean, the mean square, and the variance of x .

10.3-7 For a gaussian PDF $p_x(x) = (1/\sigma_x\sqrt{2\pi})e^{-x^2/2\sigma_x^2}$, show that

$$\overline{x^n} = \begin{cases} (1)(3)(5)\cdots(n-1)\sigma_x^n & n \text{ even} \\ 0 & n \text{ odd} \end{cases}$$

Hint: See appropriate definite integrals in any standard mathematical table.

10.3-8 Ten regular dice are thrown. The sum of the numbers appearing on these ten dice is an RV x . Find $\overline{x}$, $\overline{x^2}$, and σ_x^2 . *Hint:* Remember that the outcome of each die is independent.

10.4-1 The random binary signal $x(t)$, shown in Fig. P10.4-1a, can take on only two values, 3 and 0, with equal probability. A gaussian channel noise $n(t)$ shown in Fig. P10.4-1b is added to this signal, giving the received signal $y(t)$. The PDF of the noise amplitude n is gaussian with a zero mean and an rms value of 2. Determine and sketch the PDF of the amplitude y . *Hint:* Use of Eq. (10.72) yields $p_y(y) = p_x(x) * p_n(n)$.

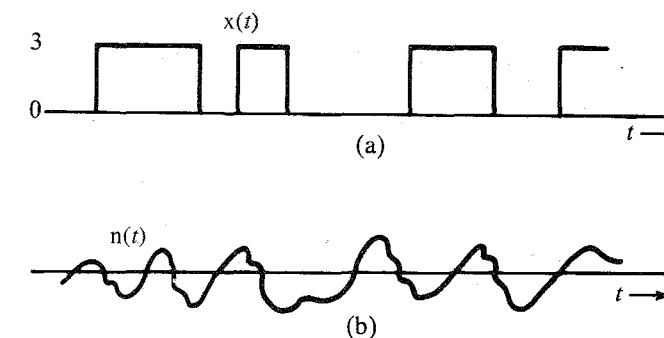


Figure P10.4-1

10.4-2 Repeat Prob. 10.4-1 if the amplitudes 3 and 0 of $x(t)$ are not equiprobable but $P_x(3) = 0.6$ and $P_x(0) = 0.4$.

10.3-4 If $x(t)$ and $y(t)$ are both independent binary signals each taking on values -1 and 1 only with

$$P_x(1) = Q = 1 - P_x(-1)$$

$$P_y(1) = P = 1 - P_y(-1)$$

determine $P_z(z_i)$ where $z = x + y$.

10.4-4 If $z = x + y$, where x and y are independent gaussian RVs with

$$p_x(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-(x-\bar{x})^2/2\sigma_x^2} \quad \text{and} \quad p_y(y) = \frac{1}{\sigma_y \sqrt{2\pi}} e^{-(y-\bar{y})^2/2\sigma_y^2}$$

then show that z is also Gaussian with

$$\bar{z} = \bar{x} + \bar{y} \quad \text{and} \quad \sigma_z^2 = \sigma_x^2 + \sigma_y^2$$

Hint: Use Eq. (3.43) to convolve $p_x(x)$ and $p_y(y)$. See pair 22 in Table 3.1.

10.5-1 Show that $|\rho_{xy}| \leq 1$, where ρ_{xy} is the correlation coefficient [Eq. (10.76)] of RVs x and y . *Hint:* For any real number a ,

$$[a(x - \bar{x}) - (y - \bar{y})]^2 \geq 0$$

The discriminant of this quadratic in a is nonpositive.

10.5-2 Show that if two RVs x and y are related by

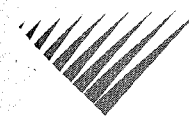
$$y = k_1 x + k_2$$

where k_1 and k_2 are arbitrary constants, the correlation coefficient $\rho_{xy} = 1$ if k_1 is positive, and $\rho_{xy} = -1$ if k_1 is negative.

10.5-3 Given $x = \cos \Theta$ and $y = \sin \Theta$, where Θ is an RV uniformly distributed in the range $(0, 2\pi)$, show that x and y are uncorrelated but are not independent.

10.6-1 In Example 10.21, design the optimum third-order predictor processor for speech signals and determine the SNR improvement. Values of various correlation coefficients for speech signals are given in Example 10.21.

11 RANDOM PROCESSES



The notion of a random process is an extension of the random variable (RV). Consider, for example, the temperature x of a certain city at noon. The temperature x is an RV and takes on different values every day. To get the complete statistics of x , we need to record values of x at noon over many days (a large number of trials). From this data, we can determine $p_x(x)$, the PDF of the RV x (the temperature at noon).

But the temperature is also a function of time. At 1 P.M., for example, the temperature may have an entirely different distribution from that of the temperature at noon. Thus, the RV x is a function of time and can be expressed as $x(t)$. An RV that is a function of time* is called a **random process**, or **stochastic process**, which is the subject of this chapter. Thus, a random process is a collection of an infinite number of RVs.

11.1 FROM RANDOM VARIABLE TO RANDOM PROCESS

To specify an RV x , we repeat the experiment a large number of times and from the outcomes determine $p_x(x)$. Similarly, to specify the random process $x(t)$, we do the same thing for each value of t . To continue with our example of the random process $x(t)$, the temperature of the city, we need to record daily temperatures for each value of t (for each time of the day). This can be done by recording temperatures at every instant of the day, which gives a waveform $x(t, \zeta_i)$, where ζ_i indicates the day for which the record was taken. We need to repeat this procedure every day for a large number of days. The collection of all possible waveforms is known as the **ensemble** (corresponding to the sample space) of the random process $x(t)$. A waveform in this collection is a **sample function** (rather than a sample point) of the random process (Fig. 11.1). Sample-function amplitudes at some instant $t = t_1$ are the values taken by the RV $x(t_1)$ in various trials.

We can view a random process in another way. In the case of an RV, the outcome of each trial of the random experiment is a number. We can view a random process also as the

* Actually, to qualify as a random process, x could be a function of any other variable, such as distance. A random process may also be a function of more than one variable.

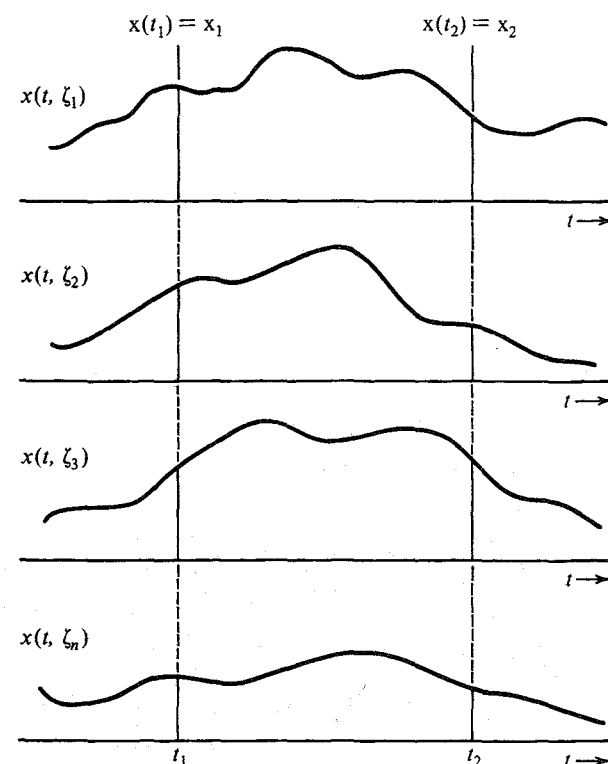


Figure 11.1 Random process to represent the temperature of a city.

outcome of a random experiment, where the outcome of each trial is a waveform (a sample function) that is a function of t . The number of waveforms in an ensemble may be finite or infinite. In the case of the random process $x(t)$ (the temperature of a city), the ensemble has infinite waveforms. On the other hand, if we consider the output of a binary signal generator (over the period 0 to $10T$), there are at most 2^{10} waveforms in this ensemble (Fig. 11.2).

One fine point that needs clarification is that the waveforms (sample functions) in the ensemble are not random. They are deterministic. Randomness in this situation is associated not with the waveform but with the uncertainty as to which waveform will occur in a given trial. This is completely analogous to the situation of an RV. For example, in the experiment of tossing a coin four times in succession (Example 10.4), 16 possible outcomes exist, all of which are known. The randomness in this situation is associated not with the outcomes but with the uncertainty as to which of the 16 outcomes will occur in a given trial.

Specification of a Random Process

The next important question is how to specify a random process. In some cases, we may be able to describe it analytically. Consider, for instance, a random process described by $x(t) = A \cos(\omega_c t + \Theta)$, where Θ is an RV uniformly distributed over the range $(0, 2\pi)$. This analytical expression completely describes a random process (and its ensemble). Each sample function is a sinusoid of amplitude A and frequency ω_c . But the phase is random (see Fig. 11.5). It is equally likely to take any value in the range $(0, 2\pi)$.

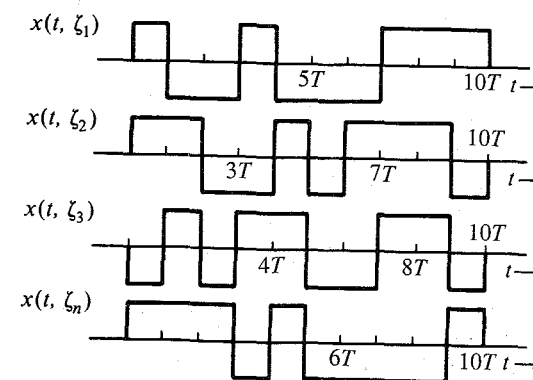


Figure 11.2 Ensemble with a finite number of sample functions.

Unfortunately, it is not always possible to be able to describe a random process analytically. We may have just an ensemble obtained experimentally. The ensemble has the complete information about the random process. From this ensemble, we must find some quantitative measure that will specify or characterize the random process. In this case, we consider the random process as an RV x that is a function of time. Thus, a random process is just a collection of an infinite number of RVs, which are generally dependent. We know that the complete information of several dependent RVs is provided by the joint PDF of those variables. Let x_i represent the RV $x(t_i)$ generated by the amplitudes of the random process at instant $t = t_i$. Thus, x_1 is the RV generated by the amplitudes at $t = t_1$, and x_2 is the RV generated by the amplitudes at $t = t_2$, and so on, as shown in Fig. 11.1. The n RVs $x_1, x_2, x_3, \dots, x_n$ generated by the amplitudes at $t = t_1, t_2, t_3, \dots, t_n$, respectively, are dependent in general. From this discussion, it follows that the random process is completely described by the joint PDF $p_{x_1 x_2 \dots x_n}(x_1, x_2, \dots, x_n)$ for all n (up to ∞) and for any choice of $t_1, t_2, t_3, \dots, t_n$. This n th-order joint PDF is also expressed as $p_{x_1 x_2 \dots x_n}(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n)$. Determining this PDF (of infinite order) is a formidable task. Fortunately, we shall soon see that when dealing with random processes in conjunction with linear systems, we can get by with only the first- and second-order statistics.

We can always derive a lower order PDF from a higher order PDF by simple integration. For instance,

$$p_{x_1}(x_1) = \int_{-\infty}^{\infty} p_{x_1 x_2}(x_1, x_2) dx_2$$

Hence, when the n th-order PDF is available, there is no need to specify PDFs of order lower than n .

The mean $\overline{x(t)}$ of a random process $x(t)$ can be determined from the first-order PDF as

$$\overline{x(t)} = \int_{-\infty}^{\infty} x p_x(x; t) dx \quad (11.1)$$

Why Do We Need Ensemble Statistics?

The preceding discussion shows that to specify a random process, we need ensemble statistics. For instance, to determine the PDF $p_{x_1}(x_1)$, we need to find the amplitudes of all the sample functions at $t = t_1$. This is *ensemble statistics*. In the same way, all possible statistics in the

specification of a random process necessitates some kind of ensemble statistics. In deterministic signals, we are used to studying the data of a waveform (or waveforms) as a function of time. Hence, the idea of investigating ensemble statistics makes us feel a bit uncomfortable at first. Theoretically, we may accept it, but does it have any practical significance? How is this concept useful in practice? We shall now answer this question.

To understand the necessity of ensemble statistics, consider the problem of threshold detection in Example 10.13. A **1** is transmitted by $p(t)$ and a **0** is transmitted by $-p(t)$ (polar signaling). The peak pulse amplitude is A_p . When **1** is transmitted, the received sample value is $A_p + n$, where n is the noise. We shall make a decision error if the noise amplitude at the sampling instant t_s is less than $-A_p$. To find this error probability, we repeat the experiment N times ($N \rightarrow \infty$), and see how many times the noise amplitude at $t = t_s$ is less than $-A_p$ (Fig. 11.3). This information is precisely one of ensemble statistics of the noise process $n(t)$ at instant t_s .

The importance of ensemble statistics is clear from this example. When we are dealing with a random process or processes, we do not know which sample function will occur in a given trial. Hence, for any specification, characterization, or optimization, we need to average over the entire ensemble. This is the basic physical reason for the appearance of ensemble statistics in random processes.

Autocorrelation Function of a Random Process

One of the most important characteristics of a random process is its **autocorrelation function**, which leads to the spectral information of the random process. The frequency content of a process depends on the rapidity of the amplitude change with time. This can be measured by correlating amplitudes at t_1 and $t_1 + \tau$. The random process $x(t)$ in Fig. 11.4a is a slowly varying process compared to the process $y(t)$ in Fig. 11.4b. For $x(t)$, the amplitudes at t_1 and $t_1 + \tau$ are similar (Fig. 11.4a), that is, have stronger correlation. On the other hand, for $y(t)$, the amplitudes at t_1 and $t_1 + \tau$ have little resemblance (Fig. 11.4b), that is, have weaker correlation. Recall that correlation is a measure of the similarity of two RVs. Hence, we can use correlation

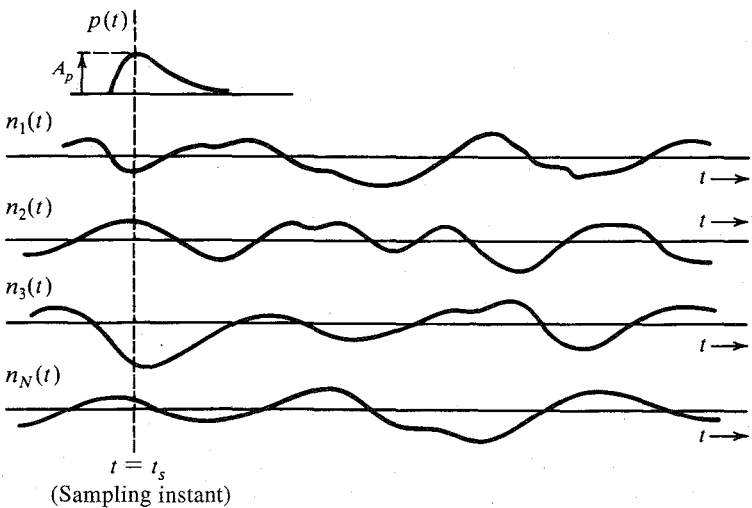


Figure 11.3 A random process to represent a channel noise.

to measure the similarity of amplitudes at t_1 and $t_2 = t_1 + \tau$. If the RVs $x(t_1)$ and $x(t_2)$ are denoted by x_1 and x_2 , respectively, then for a real random process,* the autocorrelation function $R_x(t_1, t_2)$ is defined as

$$R_x(t_1, t_2) = \overline{x(t_1)x(t_2)} = \overline{x_1x_2} \tag{11.2a}$$

This is the correlation of RVs $x(t_1)$ and $x(t_2)$. It is computed by multiplying amplitudes at t_1 and t_2 of a sample function and then averaging this product over the ensemble. It can be seen that for a small τ , the product x_1x_2 will be positive for most sample functions of $x(t)$, but the product y_1y_2 will be equally likely to be positive or negative. Hence, $\overline{x_1x_2}$ will be larger than $\overline{y_1y_2}$. Moreover, x_1 and x_2 will show correlation for considerably larger values of τ , whereas y_1 and y_2 will lose correlation quickly, even for small τ , as shown in Fig. 11.4c.

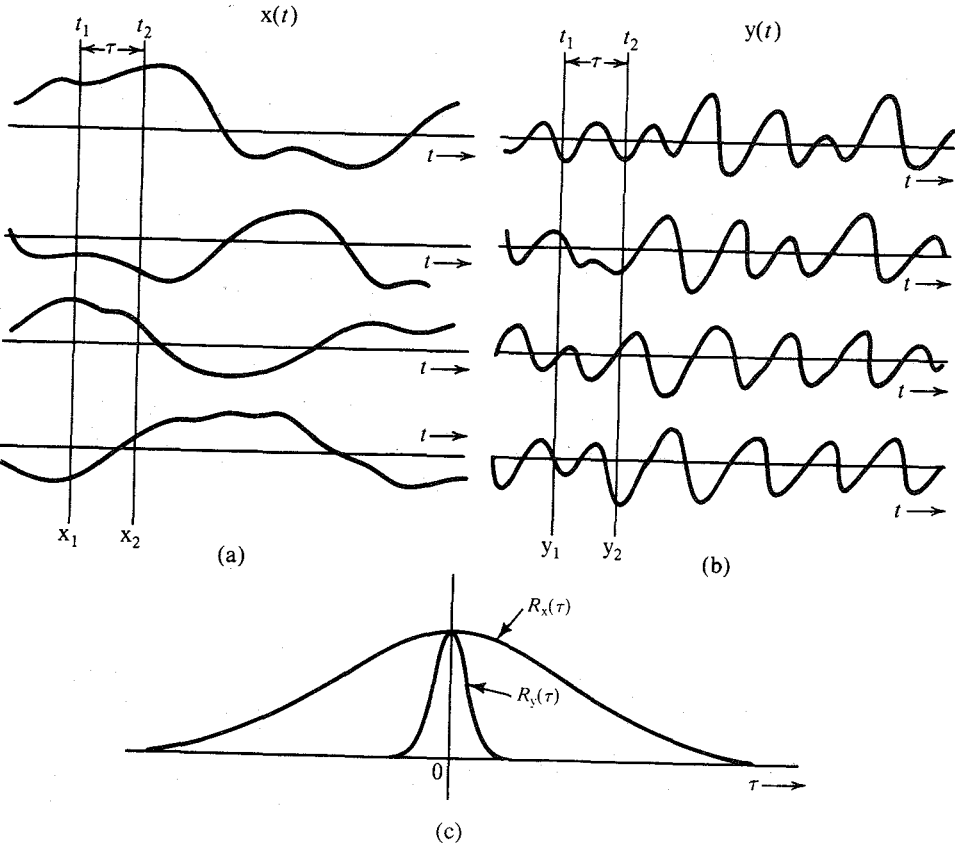


Figure 11.4 Autocorrelation functions for a slowly varying and a rapidly varying random process.

* For a complex random process $x(t)$, the autocorrelation function is defined as

$$R_x(t_1, t_2) = \overline{x^*(t_1)x(t_2)}$$

Thus, $R_x(t_1, t_2)$, the autocorrelation function of $x(t)$, provides valuable information about the frequency content of the process. In fact, we shall show that the PSD of $x(t)$ is the Fourier transform of its autocorrelation function, given by (for real processes)

$$R_x(t_1, t_2) = \overline{x_1 x_2} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 p_{x_1 x_2}(x_1, x_2) dx_1 dx_2 \quad (11.2b)$$

Hence, $R_x(t_1, t_2)$ can be derived from the joint PDF of x_1 and x_2 , which is the second-order PDF.

Classification of Random Processes

Random processes may be classified in the following broad categories.

Stationary and Nonstationary Random Processes: A random process whose statistical characteristics do not change with time is classified as a **stationary random process**. For a stationary process, we can say that a shift of time origin will be impossible to detect; the process will appear to be the same. Suppose we determine $p_x(x; t_1)$, then shift the origin by t_0 , and again determine $p_x(x; t_1)$. The instant t_1 in the new frame of reference is $t_2 = t_1 + t_0$ in the old frame of reference. Hence, the PDFs of x at t_1 and $t_2 = t_1 + t_0$ must be the same, that is, $p_x(x; t_1)$ and $p_x(x; t_2)$ must be identical for a stationary random process. This is possible only if $p_x(x; t)$ is independent of t . Thus, the first-order density of a stationary random process can be expressed as

$$p_x(x; t) = p_x(x)$$

Similarly, for a stationary random process the autocorrelation function $R_x(t_1, t_2)$ must depend on t_1 and t_2 only through the difference $t_2 - t_1$. If not, we could determine a unique time origin. Hence, for a real stationary process,

$$R_x(t_1, t_2) = R_x(t_2 - t_1)$$

Therefore,

$$R_x(\tau) = \overline{x(t)x(t+\tau)} \quad (11.3)$$

For a stationary process, the joint PDF for x_1 and x_2 must also depend only on $t_2 - t_1$. Similarly, higher order PDFs, such as $p_{x_1 x_2 \dots x_n}(x_1, x_2, \dots, x_n)$ where $x_i = x(t_i)$, are all independent of the choice of origin.

The random process $x(t)$ representing the temperature of a city is an example of a nonstationary random process, because the temperature statistics (mean value, for example) depend on the time of the day. On the other hand, the noise process in Fig. 11.3 is stationary, because its statistics (the mean and the mean square values, for example) do not change with time. In general, it is not easy to determine whether a process is stationary, because it involves investigation of the n th-order ($n = \infty$) statistics. In practice, we can ascertain stationarity if there is no change in the signal-generating mechanism. Such is the case for the noise process in Fig. 11.3.

Wide-Sense (or Weakly) Stationary Processes: A process may not be stationary in the *strict* sense, as discussed in the last subsection, yet it may have a mean value and an autocorrelation function that are independent of the shift of time origin. This means

$$\overline{x(t)} = \text{constant}$$

and

$$R_x(t_1, t_2) = R_x(\tau) \quad \tau = t_2 - t_1 \quad (11.4)$$

Such a process is known as a **wide-sense stationary**, or **weakly stationary**, process. Note that stationarity is a much stronger condition than wide-sense stationarity. All stationary processes are wide-sense stationary, but the converse is not necessarily true.

Just as no sinusoidal signal exists in actual practice, no truly stationary process can occur in real life. All processes in practice are nonstationary because they must begin at some finite time and must terminate at some finite time. A truly stationary process must start at $t = -\infty$ and go on forever. Many processes can be considered stationary for the time interval of interest, however, and the stationarity assumption allows a manageable mathematical model. The use of a stationary model is analogous to the use of a sinusoidal model in deterministic analysis.

EXAMPLE 11.1 Show that the random process

$$x(t) = A \cos(\omega_c t + \Theta)$$

where Θ is an RV uniformly distributed in the range $(0, 2\pi)$, is a wide-sense stationary process.

The ensemble (Fig. 11.5) consists of sinusoids of constant amplitude A and constant frequency ω_c , but the phase Θ is random. For any sample function, the phase is equally likely to have any value in the range $(0, 2\pi)$. Because Θ is an RV uniformly distributed over the range $(0, 2\pi)$, one can determine¹ $p_x(x, t)$ and, hence, $\overline{x(t)}$, as in Eq. (11.1). For this particular case, however, $\overline{x(t)}$ can be determined directly as follows:

$$\overline{x(t)} = \overline{A \cos(\omega_c t + \Theta)} = A \overline{\cos(\omega_c t + \Theta)}$$

Because $\cos(\omega_c t + \Theta)$ is a function of an RV Θ , we have [see Eq. (10.57b)]

$$\overline{\cos(\omega_c t + \Theta)} = \int_0^{2\pi} \cos(\omega_c t + \theta) p_{\Theta}(\theta) d\theta$$

Because $p_{\Theta}(\theta) = 1/2\pi$ over $(0, 2\pi)$ and 0 outside this range,

$$\overline{\cos(\omega_c t + \Theta)} = \frac{1}{2\pi} \int_0^{2\pi} \cos(\omega_c t + \theta) d\theta = 0$$

Hence,

$$\overline{x(t)} = 0 \quad (11.5a)$$

Thus, the ensemble mean of sample-function amplitudes at any instant t is zero.

The autocorrelation function $R_x(t_1, t_2)$ for this process also can be determined directly from Eq. (11.2a),

$$\begin{aligned} R_x(t_1, t_2) &= \overline{A^2 \cos(\omega_c t_1 + \Theta) \cos(\omega_c t_2 + \Theta)} \\ &= A^2 \overline{\cos(\omega_c t_1 + \Theta) \cos(\omega_c t_2 + \Theta)} \\ &= \frac{A^2}{2} \left\{ \overline{\cos[\omega_c(t_2 - t_1)]} + \overline{\cos[\omega_c(t_2 + t_1) + 2\Theta]} \right\} \end{aligned}$$

The first term on the right-hand side contains no RV. Hence, $\overline{\cos[\omega_c(t_2 - t_1)]}$ is $\cos[\omega_c(t_2 - t_1)]$ itself. The second term is a function of the RV Θ , and its mean is

$$\overline{\cos[\omega_c(t_2 + t_1) + 2\Theta]} = \frac{1}{2\pi} \int_0^{2\pi} \cos[\omega_c(t_2 + t_1) + 2\theta] d\theta = 0$$

Hence,

$$R_x(t_1, t_2) = \frac{A^2}{2} \cos[\omega_c(t_2 - t_1)] \tag{11.5b}$$

or

$$R_x(\tau) = \frac{A^2}{2} \cos \omega_c \tau \quad \tau = t_2 - t_1 \tag{11.5c}$$

From Eqs. (11.5a,b) it is clear that $x(t)$ is a wide-sense stationary process.

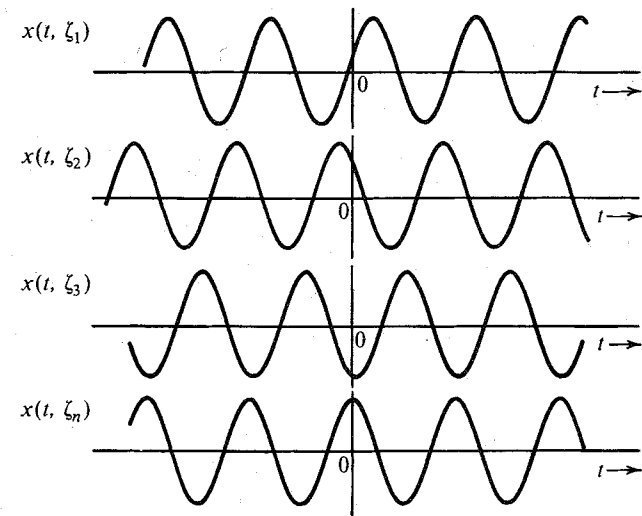


Figure 11.5 Ensemble for the random process $A \cos(\omega_c t + \Theta)$.

Ergodic Processes: We have studied the mean and the autocorrelation function of a random process. These are ensemble averages. For example, $\overline{x(t)}$ is the ensemble average of sample-function amplitudes at t , and $R_x(t_1, t_2) = \overline{x_1 x_2}$ is the ensemble average of the product of sample-function amplitudes $x(t_1)$ and $x(t_2)$.

We can also define time averages for each sample function. For example, a time mean $\widetilde{x(t)}$ of a sample function $x(t)$ is*

$$\widetilde{x(t)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t) dt \tag{11.6a}$$

Similarly, the time-autocorrelation function $R_x(\tau)$ defined in Eq. (3.82a) is

$$R_x(\tau) = \widetilde{x(t)x(t+\tau)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x(t+\tau) dt \tag{11.6b}$$

For **ergodic processes**, ensemble averages are equal to the time averages of any sample function. Thus, for an ergodic process $x(t)$,

$$\overline{x(t)} = \widetilde{x(t)} \tag{11.7a}$$

$$R_x(\tau) = R_x(\tau) \tag{11.7b}$$

These are just two of the many possible averages. For an ergodic process, all possible ensemble averages are equal to the corresponding time averages of one of its sample functions. Because a time average cannot be a function of time, it is evident that an ergodic process is necessarily a stationary process; but the converse is not true (see Fig. 11.6).

It is difficult to test whether a process is ergodic or not, because we must test all possible orders of time and ensemble averages. Nevertheless, in practice many of the stationary processes are ergodic with respect to at least second-order averages, such as the mean and the autocorrelation. For the process in Example 11.1 (Fig. 11.5), we can show that $\widetilde{x(t)} = 0$ and $R_x(\tau) = (A^2/2) \cos \omega_c \tau$ (see Problem 3.8-1). Therefore, this process is ergodic at least with respect to the first- and second-order averages.

The ergodicity concept can be explained by a simple example of traffic lights in a city. Suppose the city is well planned, with all its streets in E-W and N-S directions only and with traffic lights at each intersection. Assume that each light stays green for 0.75 second in the E-W direction and 0.25 second in the N-S direction and that switching of any light is independent of the other lights. For the sake of simplicity, we ignore the orange light.

If we consider a certain person driving a car arriving at any traffic light randomly in the E-W direction, the probability that the person will have a green light is 0.75; that is, on the average, 75% of the time the person will observe a green light. On the other hand, if we consider a large number of drivers arriving at a traffic light in the E-W direction at some instant

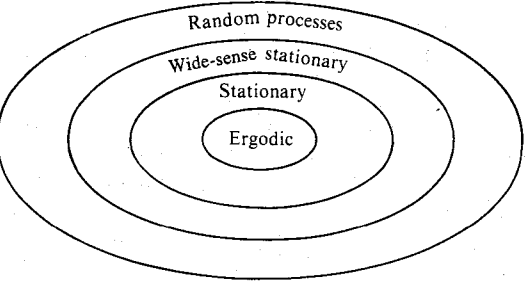


Figure 11.6 Classification of random processes.

* Here a sample function $x(t, \zeta_i)$ is represented by $x(t)$ for convenience.

t , then 75% of the drivers will have a green light, and the remaining 25% will have a red light. Thus, the experience of a single driver arriving randomly many times at a traffic light will contain the same statistical information (sample-function statistics) as that of a large number of drivers arriving simultaneously at various traffic lights (ensemble statistics) at one instant.

The ergodicity notion is extremely important, because we do not have a large number of sample functions available in practice from which to compute ensemble averages. If the process is known to be ergodic, then we need only one sample function to compute ensemble averages. As mentioned earlier, many of the stationary processes encountered in practice are ergodic with respect to at least second-order averages. As we shall see in dealing with stationary processes in conjunction with linear systems, we need only the first- and second-order averages. This means that in most cases we can get by with a single sample function.

11.2 POWER SPECTRAL DENSITY OF A RANDOM PROCESS

An electrical engineer instinctively thinks of signals and linear systems in terms of their frequency-domain descriptions. Linear systems are characterized by their frequency response (the transfer function), and signals are expressed in terms of the relative amplitudes and phases of their frequency components (the Fourier transform). From a knowledge of the input spectrum and transfer function, the response of a linear system to a given signal can be obtained in terms of the frequency content of that signal. This is an important procedure for deterministic signals. We may wonder if similar methods may be found for random processes. Ideally, all the sample functions of a random process are assumed to exist over the entire time interval $(-\infty, \infty)$ and thus, are power signals.* We therefore inquire about the existence of a power spectral density (PSD). Superficially, the concept of a PSD of a random process may appear ridiculous for the following reasons. In the first place, we may not be able to describe a sample function analytically. Second, for a given process, every sample function may be different from another one. Hence, even if a PSD does exist for each sample function, it may be different for different sample functions. Fortunately, both problems can be neatly resolved, and it is possible to define a meaningful PSD for a stationary (at least in the wide sense) random process. For nonstationary processes, the PSD does not exist.

Whenever randomness is involved, our inquiries can at best provide answers in terms of averages. When tossing a coin, for instance, the most we can say about the outcome is that on the average we will obtain heads in about half the trials and tails in the remaining half of the trials. For random signals or RVs, we do not have enough information to predict the outcome with certainty, and we must accept answers in terms of averages. It is not possible to transcend this limit of knowledge because of the fundamental ignorance of the process. It seems reasonable to define the PSD of a random process as a weighted mean of the PSDs of all sample functions. This is the only sensible solution, because we do not know exactly which of the sample functions may occur in a given trial. We must be prepared for any sample function. Consider, for example, the problem of filtering a certain random process. We would not want to design a filter with respect to any one particular sample function because any of the sample functions in the ensemble may be present at the input. A sensible approach is to design the filter with respect to the mean parameters of the input process. In designing a system to perform

* As we shall soon see, for the PSD to exist, the process must be stationary (at least in the wide sense). Stationary processes, because their statistics do not change with time, are power signals.

certain operations, one must design it with respect to the whole ensemble. We are therefore justified in defining the PSD $S_x(\omega)$ of a random process $x(t)$ as the ensemble average of the PSDs of all sample functions. Thus [see Eq. (3.79)],

$$S_x(\omega) = \lim_{T \rightarrow \infty} \overline{\left[\frac{|X_T(\omega)|^2}{T} \right]} \quad (11.8a)$$

where $X_T(\omega)$ is the Fourier transform of the truncated random process $x(t) \text{ rect}(t/T)$ and the bar represents ensemble average. Note that ensemble averaging is done before the limiting operation. We shall now show that the PSD as defined in Eq. (11.8a) is the Fourier transform of the autocorrelation function $R_x(\tau)$ of the process $x(t)$; that is,

$$R_x(\tau) \Longleftrightarrow S_x(\omega) \quad (11.8b)$$

This can be proved as follows:

$$X_T(\omega) = \int_{-\infty}^{\infty} x_T(t) e^{-j\omega t} dt = \int_{-T/2}^{T/2} x(t) e^{-j\omega t} dt \quad (11.9)$$

Thus, for real $x(t)$,

$$\begin{aligned} |X_T(\omega)|^2 &= X_T(-\omega) X_T(\omega) \\ &= \int_{-T/2}^{T/2} x(t_1) e^{j\omega t_1} dt_1 \int_{-T/2}^{T/2} x(t_2) e^{-j\omega t_2} dt_2 \\ &= \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} x(t_1) x(t_2) e^{-j\omega(t_2-t_1)} dt_1 dt_2 \end{aligned}$$

and

$$\begin{aligned} S_x(\omega) &= \lim_{T \rightarrow \infty} \overline{\left[\frac{|X_T(\omega)|^2}{T} \right]} \\ &= \lim_{T \rightarrow \infty} \overline{\left[\frac{1}{T} \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} x(t_1) x(t_2) e^{-j\omega(t_2-t_1)} dt_1 dt_2 \right]} \quad (11.10) \end{aligned}$$

Interchanging the operation of integration and ensemble averaging,* we get

$$\begin{aligned} S_x(\omega) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} \overline{x(t_1) x(t_2)} e^{-j\omega(t_2-t_1)} dt_1 dt_2 \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} R_x(t_2 - t_1) e^{-j\omega(t_2-t_1)} dt_1 dt_2 \end{aligned}$$

Here we are assuming that the process $x(t)$ is at least wide-sense stationary, so that $\overline{x(t_1) x(t_2)} = R_x(t_2 - t_1)$. For convenience, let

* The operation of ensemble averaging is also an operation of integration. Hence, interchanging integration with ensemble averaging is equivalent to interchanging the order of integration.

$$R_x(t_2 - t_1)e^{-j\omega(t_2 - t_1)} = \varphi(t_2 - t_1) \quad (11.11)$$

Then,

$$S_x(\omega) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} \varphi(t_2 - t_1) dt_1 dt_2 \quad (11.12)$$

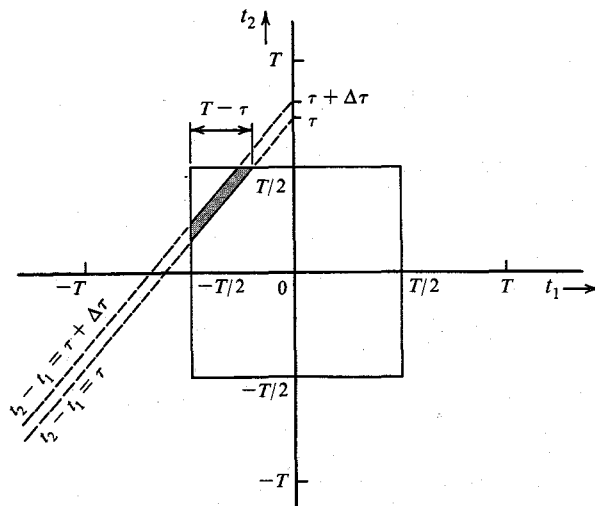
The integral on the right-hand side is a double integral over the range $(-T/2, T/2)$ for each of the variables t_1 and t_2 . The square region of integration in the t_1 - t_2 plane is shown in Fig. 11.7. The integral in Eq. (11.12) is a volume under the surface $\varphi(t_2 - t_1)$ over the square region in Fig. 11.7. The double integral in Eq. (11.12) can be converted to a single integral by observing that $\varphi(t_2 - t_1)$ is constant along any line $t_2 - t_1 = \tau$ (a constant) in the t_1 - t_2 plane (Fig. 11.7).

Let us consider two such lines, $t_2 - t_1 = \tau$ and $t_2 - t_1 = \tau + \Delta\tau$. If $\Delta\tau \rightarrow 0$, $\varphi(t_2 - t_1) \simeq \varphi(\tau)$ over the shaded region whose area is $(T - \tau) \Delta\tau$. Hence, the volume under the surface $\varphi(t_2 - t_1)$ over the shaded region is $\varphi(\tau)(T - \tau) \Delta\tau$. If τ were negative, the volume would be $\varphi(\tau)(T + \tau) \Delta\tau$. Hence, in general, the volume over the shaded region is $\varphi(\tau)(T - |\tau|) \Delta\tau$. The desired volume over the square region in Fig. 11.7 is the sum of the volumes over the shaded strips and is obtained by integrating $\varphi(\tau)(T - |\tau|)$ over the range of τ , which is $(-T, T)$ (see Fig. 11.7). Hence,

$$\begin{aligned} S_x(\omega) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T}^T \varphi(\tau)(T - |\tau|) d\tau \\ &= \lim_{T \rightarrow \infty} \int_{-T/2}^{T/2} \varphi(\tau) \left(1 - \frac{|\tau|}{T}\right) d\tau \\ &= \int_{-\infty}^{\infty} \varphi(\tau) d\tau \end{aligned}$$

provided $\int_{-\infty}^{\infty} |\tau| \varphi(\tau) d\tau$ is bounded. Substituting Eq. (11.11) into this equation, we have

Figure 11.7 Derivation of Wiener-Khinchine theorem.



$$S_x(\omega) = \int_{-\infty}^{\infty} R_x(\tau) e^{-j\omega\tau} d\tau \quad (11.13)$$

provided $\int_{-\infty}^{\infty} |\tau| R_x(\tau) e^{-j\omega\tau} d\tau$ is bounded. Thus, the PSD of a wide-sense stationary random process is the Fourier transform of its autocorrelation function,*

$$R_x(\tau) \Longleftrightarrow S_x(\omega) \quad (11.14)$$

This is the well-known Wiener-Khinchine theorem.

From the discussion thus far, the autocorrelation function emerges as one of the most significant entities in the spectral analysis of a random process. Earlier we showed heuristically how the autocorrelation function is connected with the frequency content of a random process.

The autocorrelation function $R_x(\tau)$ for real processes is an even function of τ . This can be proved in two ways. First, because $|X_T(\omega)|^2 = |X_T(\omega)X_T^*(\omega)| = |X_T(\omega)X_T(-\omega)|$ is an even function of ω , $S_x(\omega)$ is also an even function of ω , and $R_x(\tau)$, its inverse transform, is also an even function of τ (see Prob. 3.1-1). Alternately, we may argue that

$$R_x(\tau) = \overline{x(t)x(t+\tau)} \quad \text{and} \quad R_x(-\tau) = \overline{x(t)x(t-\tau)}$$

Letting $t - \tau = \sigma$, we have

$$R_x(-\tau) = \overline{x(\sigma)x(\sigma+\tau)} = R_x(\tau) \quad (11.15)$$

The PSD $S_x(\omega)$ is also a real and even function of ω .

The mean square value $\overline{x^2(t)}$ of the random process $x(t)$ is $R_x(0)$,

$$R_x(0) = \overline{x(t)x(t)} = \overline{x^2(t)} = \overline{x^2} \quad (11.16)$$

The mean square value $\overline{x^2}$ is not the time mean square of a sample function but the ensemble average of the amplitude squares of sample functions at any instant t .

The Power of a Random Process

The power P_x (average power) of a wide-sense random process $x(t)$ is its mean square value $\overline{x^2}$. From Eq. (11.14),

$$R_x(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_x(\omega) e^{j\omega\tau} d\omega$$

Hence, from Eq. (11.16),

$$P_x = \overline{x^2} = R_x(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_x(\omega) d\omega \quad (11.17a)$$

Because $S_x(\omega)$ is an even function of ω , we have

$$P_x = \overline{x^2} = 2 \int_0^{\infty} S_x(\omega) df \quad f = \frac{\omega}{2\pi} \quad (11.17b)$$

where f is the frequency in hertz. This is the same relationship as that derived for deterministic signals in Chapter 3 [Eq. (3.81)]. The power P_x is the area under the PSD. Also, $P_x = \overline{x^2}$ is the ensemble mean of the square amplitudes of the sample functions at any instant.

* It can be shown that Eq. (11.13) holds also for complex random processes, for which we define $R_x(\tau) = \overline{x^*(t)x(t+\tau)}$.

It is helpful to repeat here, once again, that the PSD does not exist for processes that are not wide-sense stationary. Hence, in our future discussion, random processes will be assumed to be at least wide-sense stationary, unless specifically stated otherwise.

EXAMPLE 11.2 Determine the autocorrelation function $R_x(\tau)$ and the power P_x of a low-pass random process with a white-noise PSD $S_x(\omega) = \mathcal{N}/2$ (Fig. 11.8a).

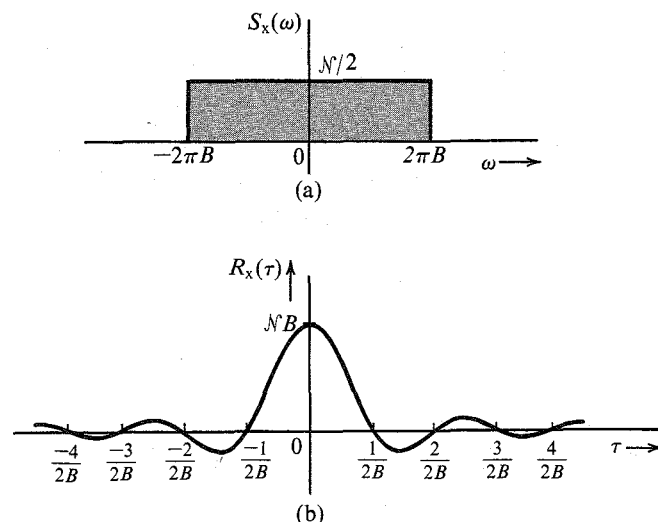


Figure 11.8 Bandpass white-noise PSD and its autocorrelation function.

We have

$$S_x(\omega) = \frac{\mathcal{N}}{2} \text{rect}\left(\frac{\omega}{4\pi B}\right) \quad (11.18a)$$

Hence, from Table 3.1 (pair 17),

$$R_x(\tau) = \mathcal{N}B \text{sinc}(2\pi B\tau) \quad (11.18b)$$

This is shown in Fig. 11.8b. Also,

$$P_x = \overline{x^2} = R_x(0) = \mathcal{N}B \quad (11.18c)$$

Alternately,

$$\begin{aligned} P_x &= 2 \int_0^\infty S_x(\omega) df \\ &= 2 \int_0^B \frac{\mathcal{N}}{2} df \\ &= \mathcal{N}B \end{aligned} \quad (11.18d)$$

EXAMPLE 11.3 Determine the PSD and the mean square value of a random process

$$x(t) = A \cos(\omega_c t + \Theta) \quad (11.19a)$$

where Θ is an RV uniformly distributed over $(0, 2\pi)$.

For this case $R_x(\tau)$ is already determined [Eq. (11.5c)],

$$R_x(\tau) = \frac{A^2}{2} \cos \omega_c \tau \quad (11.19b)$$

Hence,

$$S_x(\omega) = \frac{\pi A^2}{2} [\delta(\omega + \omega_c) + \delta(\omega - \omega_c)] \quad (11.19c)$$

$$P_x = \overline{x^2} = R_x(0) = \frac{A^2}{2} \quad (11.19d)$$

Thus, the power, or the mean square value, of the process $x(t) = A \cos(\omega_c t + \Theta)$ is $A^2/2$. The power P_x can also be obtained by integrating $S_x(\omega)$ with respect to ω (or f).

EXAMPLE 11.4 Amplitude Modulation

Determine the autocorrelation function and the PSD of the DSB-SC-modulated process $m(t) \cos(\omega_c t + \Theta)$, where $m(t)$ is a wide-sense stationary random process, and Θ is an RV uniformly distributed over $(0, 2\pi)$ and independent of $m(t)$.

Let

$$\varphi(t) = m(t) \cos(\omega_c t + \Theta)$$

Then

$$R_\varphi(\tau) = \overline{[m(t) \cos(\omega_c t + \Theta)][m(t + \tau) \cos(\omega_c(t + \tau) + \Theta)]}$$

Because $m(t)$ and Θ are independent, we can write [see Eqs. (10.60b) and (11.5c)]

$$\begin{aligned} R_\varphi(\tau) &= \overline{m(t)m(t + \tau) \cos(\omega_c t + \Theta) \cos[\omega_c(t + \tau) + \Theta]} \\ &= \frac{1}{2} R_m(\tau) \cos \omega_c \tau \end{aligned} \quad (11.20a)$$

Consequently,*

$$S_\varphi(\omega) = \frac{1}{4} [S_m(\omega + \omega_c) + S_m(\omega - \omega_c)] \quad (11.20b)$$

From Eq. (11.20a) it follows that

$$\overline{\varphi^2(t)} = R_\varphi(0) = \frac{1}{2} R_m(0) = \frac{1}{2} \overline{m^2(t)} \quad (11.20c)$$

Hence, the power of the DSB-SC-modulated signal is half the power of the modulating signal. We derived the same result earlier [Eq. (3.92)] for deterministic signals.

* We obtain the same result even if $\varphi(t) = m(t) \sin(\omega_c t + \Theta)$.

EXAMPLE 11.5 Random Binary Process

In this example we shall consider a random binary process for which a typical sample function is shown in Fig. 11.9a. The signal can assume only two states (values), 1 or -1, with equal probability. The transition from one state to another can take place only at node points, which occur every T_b seconds. The probability of a transition from one state to the other is 0.5. The first node is equally likely to be situated at any instant within the interval 0 to T_b from the origin.* The amplitudes at t represent RV x_1 , and those at $t + \tau$ represent RV x_2 . Note that x_1 and x_2 are discrete and each can assume only two values, -1 and 1. Hence,

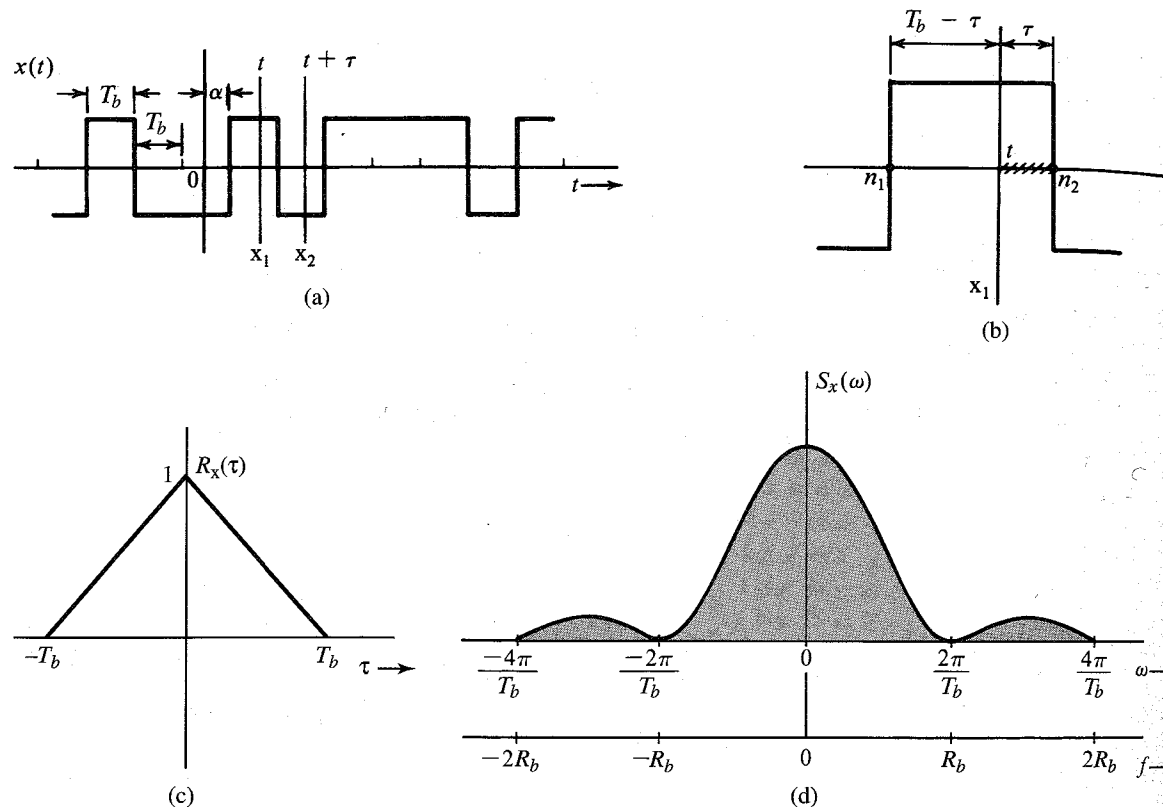


Figure 11.9 Derivation of autocorrelation function and PSD of a random binary process.

$$R_x(\tau) = \overline{x_1 x_2} = \sum_{x_1} \sum_{x_2} x_1 x_2 P_{x_1 x_2}(x_1, x_2) \\ = P_{x_1 x_2}(1, 1) + P_{x_1 x_2}(-1, -1) - P_{x_1 x_2}(-1, 1) - P_{x_1 x_2}(1, -1) \quad (11.21a)$$

* Analytically, we can represent $x(t)$ as

$$x(t) = \sum_n a_n p(t - nT_b - \alpha)$$

where α is an RV uniformly distributed over the range $(0, T_b)$ and $p(t)$ is the basic pulse (in this case $\text{rect}[(t - T_b/2)/T_b]$). Note that α is the distance of the first node from the origin, and it varies randomly from sample function to sample function. In addition, a_n is random, taking values 1 or -1 with equal probability.

By symmetry, the first two terms and the last two terms on the right-hand side are equal. Therefore,

$$R_x(\tau) = 2[P_{x_1 x_2}(1, 1) - P_{x_1 x_2}(1, -1)] \quad (11.21b)$$

Using Bayes' rule, we have

$$R_x(\tau) = 2P_{x_1}(1)[P_{x_2|x_1}(1|1) - P_{x_2|x_1}(-1|1)] \\ = P_{x_2|x_1}(1|1) - P_{x_2|x_1}(-1|1) \quad (11.21c)$$

Moreover,

$$P_{x_2|x_1}(1|1) = 1 - P_{x_2|x_1}(-1|1)$$

Hence,

$$R_x(\tau) = 1 - 2P_{x_2|x_1}(-1|1)$$

It is helpful to compute $R_x(\tau)$ for small values of τ first. Let us consider the case $\tau < T_b$, where, at most, one node is in the interval t to $t + \tau$. In this case, the event $x_2 = -1$ given $x_1 = 1$ is a joint event AB , where the event A is "a node in the interval $(t, t + \tau)$ " and B is "the state change at this node." Because A and B are independent events,

$$P_{x_2|x_1}(-1|1) = P(\text{a node lies in } t \text{ to } t + \tau)P(\text{state change}) \\ = \frac{1}{2}P(\text{a node lies in } t \text{ to } t + \tau)$$

Figure 11.9b shows adjacent nodes n_1 and n_2 , between which t lies. We mark off the interval τ from the node n_2 . If t lies anywhere in this interval (hatched area), the node n_2 lies within t and $t + \tau$. But because the instant t is chosen arbitrarily between nodes n_1 and n_2 , it is equally likely to be at any instant over the T_b seconds between n_1 and n_2 , and the probability that t lies in the shaded interval is simply τ/T_b . Therefore,

$$P_{x_2|x_1}(-1|1) = \frac{1}{2} \left(\frac{\tau}{T_b} \right) \quad (11.22)$$

and

$$R_x(\tau) = 1 - \frac{\tau}{T_b} \quad \tau < T_b \quad (11.23)$$

Because $R_x(\tau)$ is an even function of τ , we have

$$R_x(\tau) = 1 - \frac{|\tau|}{T_b} \quad |\tau| < T_b \quad (11.24)$$

Next, consider the range $\tau > T_b$. In this case at least one node lies in the interval t to $t + \tau$. Hence, x_1 and x_2 become independent, and

$$R_x(\tau) = \overline{x_1 x_2} = \overline{x_1} \overline{x_2} = 0 \quad \tau > T_b$$

where, by inspection, we observe that $\overline{x_1} = \overline{x_2} = 0$ (Fig. 11.9a). This result can also be obtained by observing that for $|\tau| > T_b$, x_1 and x_2 are independent, and it is equally likely that $x_2 = 1$ or -1 given that $x_1 = 1$ (or -1). Hence, all four probabilities in Eq. (11.21a) are equal to 1/4, and

$$R_x(\tau) = 0 \quad \tau > T_b$$

Therefore,

$$R_x(\tau) = \begin{cases} 1 - |\tau|/T_b & |\tau| < T_b \\ 0 & |\tau| > T_b \end{cases} \quad (11.25a)$$

and

$$S_x(\omega) = T_b \operatorname{sinc}^2\left(\frac{\omega T_b}{2}\right) \quad (11.25b)$$

The autocorrelation function and the PSD of this process are shown in Fig. 11.9c and d. Observe that $\overline{x^2} = R_x(0) = 1$, as expected.

Let us now consider a more general case of the pulse train $y(t)$, discussed in Sec. 7.2 (Fig. 7.3). From the knowledge of the PSD of this train, we can derive the PSD of on-off, polar, bipolar, duobinary, split-phase, and many more important digital signals.

EXAMPLE 11.6 Random PAM Pulse Train

Digital data is transmitted using a basic pulse $p(t)$, as shown in Fig. 11.10a. The successive pulses are separated by T_b seconds, and the k th pulse is $a_k p(t)$, where a_k is an RV. The distance α of the first pulse (corresponding to $k = 0$) from the origin is equally likely to be any value in the range $(0, T_b)$. Find the autocorrelation function and the PSD of such a random pulse train $y(t)$ whose sample function is shown in Fig. 11.10b. The random process $y(t)$ can be described as

$$y(t) = \sum_{k=-\infty}^{\infty} a_k p(t - kT_b - \alpha)$$

where α , is an RV uniformly distributed in the interval $(0, T_b)$. Thus, α is different for each sample function. Note that $p(\alpha) = 1/T_b$ over the interval $(0, T_b)$, and is zero everywhere else.* It can be shown that $\overline{y(t)} = (\overline{a_k}/T_b) \int_{-\infty}^{\infty} p(t) dt$ is a constant.†

We have the expression

$$\begin{aligned} R_y(\tau) &= \overline{y(t)y(t+\tau)} \\ &= \overline{\sum_{k=-\infty}^{\infty} a_k p(t - kT_b - \alpha) \sum_{m=-\infty}^{\infty} a_m p(t + \tau - mT_b - \alpha)} \\ &= \sum_{k=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \overline{a_k a_m p(t - kT_b - \alpha) p(t + \tau - mT_b - \alpha)} \end{aligned}$$

* If $\alpha = 0$, the process can be expressed as $y(t) = \sum_{k=-\infty}^{\infty} a_k p(t - kT_b)$. In this case $\overline{y(t)} = \overline{a_k} \sum_{k=-\infty}^{\infty} p(t - kT_b)$ is not constant, but is periodic with period T_b . Similarly, we can show that the autocorrelation function is periodic with the same period T_b . This is an example of a **cyclostationary**, or periodically stationary, process (a process whose statistics are invariant to a shift of the time origin by integral multiples of a constant T_b). Cyclostationary processes, as seen here, are clearly not wide-sense stationary. But they can be made wide-sense stationary with slight modification by adding the RV α in the expression of $y(t)$, as in this example.

† Using exactly the same approach, as used in the derivation of Eq. (11.26) we can show that $\overline{y(t)} = (\overline{a_k}/T_b) \int_{-\infty}^{\infty} p(t) dt$.

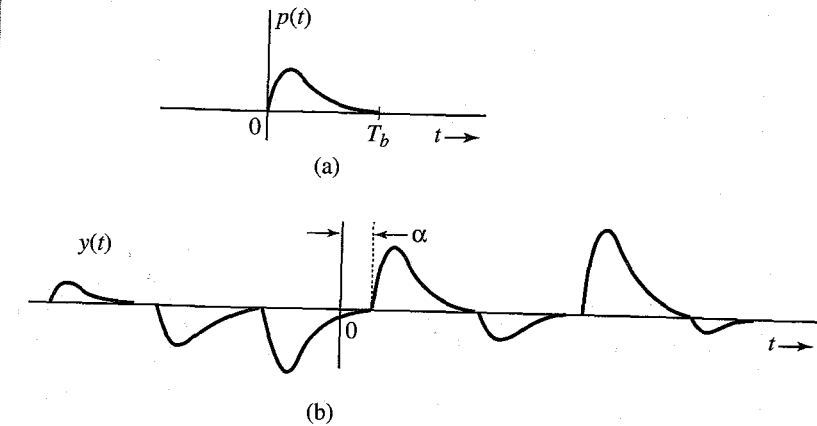


Figure 11.10 Random PAM process.

Because a_k and a_m are independent of α ,

$$R_y(\tau) = \sum_{k=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \overline{a_k a_m} \overline{p(t - kT_b - \alpha) p(t + \tau - mT_b - \alpha)}$$

Both k and m are integers. Letting $m = k + n$, this expression can be written

$$R_y(\tau) = \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \overline{a_k a_{k+n}} \overline{p(t - kT_b - \alpha) p(t + \tau - [k + n]T_b - \alpha)}$$

The first term under the double sum is the correlation of RVs a_k and a_{k+n} and will be denoted by $\Re_n$. The second term, being a mean with respect to the RV α , can be expressed as an integral. Thus,

$$R_y(\tau) = \sum_{n=-\infty}^{\infty} \Re_n \sum_{k=-\infty}^{\infty} \int_0^{T_b} p(t - kT_b - \alpha) p(t + \tau - [k + n]T_b - \alpha) p(\alpha) d\alpha$$

Recall that α is uniformly distributed over the interval 0 to T_b . Hence, $p(\alpha) = 1/T_b$ over the interval $(0, T_b)$, and is zero otherwise. Therefore,

$$\begin{aligned} R_y(\tau) &= \sum_{n=-\infty}^{\infty} \Re_n \sum_{k=-\infty}^{\infty} \frac{1}{T_b} \int_0^{T_b} p(t - kT_b - \alpha) p(t + \tau - [k + n]T_b - \alpha) d\alpha \\ &= \frac{1}{T_b} \sum_{n=-\infty}^{\infty} \Re_n \sum_{k=-\infty}^{\infty} \int_{t-(k+1)T_b}^{t-kT_b} p(\beta) p(\beta + \tau - nT_b) d\beta \\ &= \frac{1}{T_b} \sum_{n=-\infty}^{\infty} \Re_n \int_{-\infty}^{\infty} p(\beta) p(\beta + \tau - nT_b) d\beta \end{aligned}$$

The integral on the right-hand side is the time-autocorrelation function of the pulse $p(t)$ with the argument $\tau - nT_b$. Thus,

$$R_y(\tau) = \frac{1}{T_b} \sum_{n=-\infty}^{\infty} \Re_n \psi_p(\tau - nT_b) \quad (11.26)$$

where

$$\Re_n = \overline{a_k a_{k+n}} \quad (11.27)$$

and

$$\psi_p(\tau) = \int_{-\infty}^{\infty} p(t) p(t + \tau) dt \quad (11.28)$$

As seen in Eq. (3.73), if $p(t) \iff P(\omega)$, then $\psi_p(\tau) \iff |P(\omega)|^2$. Therefore, the PSD of $y(t)$, which is the Fourier transform of $R_y(\tau)$, is given by

$$\begin{aligned} S_y(\omega) &= \frac{1}{T_b} \sum_{n=-\infty}^{\infty} \Re_n |P(\omega)|^2 e^{-jn\omega T_b} \\ &= \frac{|P(\omega)|^2}{T_b} \sum_{n=-\infty}^{\infty} \Re_n e^{-jn\omega T_b} \end{aligned} \quad (11.29)$$

This result is similar to that found in Eq. (7.10b). The only difference is the use of the ensemble average in defining $\Re_n$ in this chapter, whereas R_n in Chapter 7 is the time average.

EXAMPLE 11.7 Find the PSD $S_y(\omega)$ for a polar binary random signal where **1** is transmitted by a pulse $p(t)$ (Fig. 11.11) whose Fourier transform is $P(\omega)$, and **0** is transmitted by $-p(t)$. The digits **1** and **0** are equally likely, and one digit is transmitted every T_b seconds. Each digit is independent of the other digits.

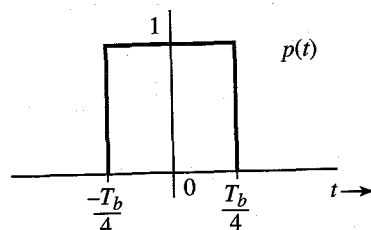


Figure 11.11 Basic pulse for a random binary polar signal.

In this case, a_k can take on values 1 and -1 with probability $1/2$ each. Hence,

$$\begin{aligned} \overline{a_k} &= \sum_{k=1, -1} a_k P(a_k) = (1)P_{a_k}(1) + (-1)P_{a_k}(-1) \\ &= \frac{1}{2} - \frac{1}{2} = 0 \\ \Re_0 = \overline{a_k^2} &= \sum_{k=1, -1} a_k^2 P(a_k) = (1)^2 P_{a_k}(1) + (-1)^2 P_{a_k}(-1) \\ &= \frac{1}{2}(1)^2 + \frac{1}{2}(-1)^2 = 1 \end{aligned}$$

and because each digit is independent of the remaining digits,

$$\Re_n = \overline{a_k a_{k+n}} = \overline{a_k} \overline{a_{k+n}} = 0 \quad n \geq 1$$

Hence, from Eq. (11.29),

$$S_y(\omega) = \frac{|P(\omega)|^2}{T_b}$$

We already found this result in Eq. (7.12), where we used time averaging instead of ensemble averaging. When a process is ergodic of second order (or higher), the ensemble and time averages yield the same result. Note that Example 11.5 is a special case of this result, where $p(t)$ is a full-width rectangular pulse $\text{rect}(t/T_b)$ with $P(\omega) = T_b \text{sinc}(\omega T_b/2)$, and

$$S_y(\omega) = \frac{|P(\omega)|^2}{T_b} = T_b \text{sinc}^2\left(\frac{\omega T_b}{2}\right)$$

EXAMPLE 11.8 Find the PSD $S_y(\omega)$ for on-off and bipolar random signals which use a basic pulse for $p(t)$, as shown in Fig. 11.11. The digits **1** and **0** are equally likely, and digits are transmitted every T_b seconds. Each digit is independent of the remaining digits. All these line codes are described in Sec. 7.2.

In each case we shall first determine $\Re_0, \Re_1, \Re_2, \dots, \Re_n$.

(a) *On-off signaling*: In this case, a_n can take on values 1 and 0 with probability $1/2$ each. Hence,

$$\begin{aligned} \overline{a_k} &= (1)P_{a_k}(1) + (0)P_{a_k}(0) = \frac{1}{2}(1) + \frac{1}{2}(0) = \frac{1}{2} \\ \Re_0 = \overline{a_k^2} &= (1)^2 P_{a_k}(1) + (0)^2 P_{a_k}(0) = \frac{1}{2}(1)^2 + \frac{1}{2}(0)^2 = \frac{1}{2} \end{aligned}$$

and because each digit is independent of the remaining digits,

$$\Re_n = \overline{a_k a_{k+n}} = \overline{a_k} \overline{a_{k+n}} = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4} \quad n \geq 1$$

Therefore, from Eq. (11.29),

$$S_y(\omega) = \frac{|P(\omega)|^2}{T_b} \left[\frac{1}{2} + \frac{1}{4} \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} e^{-jn\omega T_b} \right] \quad (11.30a)$$

$$= \frac{|P(\omega)|^2}{T_b} \left[\frac{1}{4} + \frac{1}{4} \sum_{n=-\infty}^{\infty} e^{-jn\omega T_b} \right] \quad (11.30b)$$

Equation (11.30b) is obtained from Eq. (11.30a) by splitting the term $1/2$ corresponding to $\Re_0$ into two: $1/4$ outside the summation and $1/4$ inside the summation (corresponding to $n = 0$). This result is identical to Eq. (7.17b) found earlier by using time averages.

We now use the formula (see the footnote* for a proof),

$$\sum_{n=-\infty}^{\infty} e^{-jn\omega T_b} = \frac{2\pi}{T_b} \sum_{n=-\infty}^{\infty} \delta\left(\omega - \frac{2\pi n}{T_b}\right)$$

Substitution of this result into Eq. (11.30b) yields

$$S_y(\omega) = \frac{|P(\omega)|^2}{4T_b} \left[1 + \frac{2\pi}{T_b} \sum_{n=-\infty}^{\infty} \delta\left(\omega - \frac{2\pi n}{T_b}\right) \right] \quad (11.30c)$$

Note that the spectrum $S_y(\omega)$ consists of both a discrete and a continuous part. A discrete component of clock frequency ($R_b = 1/T_b$) is present in the spectrum. The continuous component of the spectrum is $|P(\omega)|^2/4T_b$ is identical (except for a scaling factor 1/4) to the spectrum of the polar signal in Example 11.7. This is a logical result because as Fig. 7.2 shows, an on-off signal can be expressed as a sum of a polar and a periodic component. The polar component is exactly half the polar signal discussed earlier. Hence, the PSD of this component is one-fourth of the PSD of the polar signal. The periodic component is of clock frequency R_b , and consists of discrete components of frequency R_b and its harmonics.

(b) *Bipolar signaling*: in this case, a_k can take on values 0, 1, and -1 with probabilities 1/2, 1/4, and 1/4, respectively. Hence,

$$\begin{aligned} \bar{a}_k &= (0)P_{a_k}(0) + (1)P_{a_k}(1) + (-1)P_{a_k}(-1) \\ &= \frac{1}{2}(0) + \frac{1}{4}(1) + \frac{1}{4}(-1) = 0 \\ \mathfrak{R}_0 &= \bar{a}_k^2 = (0)^2P_{a_k}(0) + (1)^2P_{a_k}(1) + (-1)^2P_{a_k}(-1) \\ &= \frac{1}{2}(0)^2 + \frac{1}{4}(1)^2 + \frac{1}{4}(-1)^2 = \frac{1}{2} \end{aligned}$$

Also,

$$\mathfrak{R}_1 = \overline{a_k a_{k+1}} = \sum_k \sum_{k+1} a_k a_{k+1} P_{a_k a_{k+1}}(a_k a_{k+1})$$

Because a_k and a_{k+1} can take three values each, the sum on the right-hand side has nine terms, of which only four terms (corresponding to values ± 1 for a_k and a_{k+1}) are nonzero. Thus,

$$\begin{aligned} \mathfrak{R}_1 &= (1)(1)P_{a_k a_{k+1}}(1, 1) + (-1)(1)P_{a_k a_{k+1}}(-1, 1) \\ &\quad + (1)(-1)P_{a_k a_{k+1}}(1, -1) + (-1)(-1)P_{a_k a_{k+1}}(-1, -1) \end{aligned}$$

* The impulse train in Fig. 3.24a is $\delta_{T_b}(t)$ can be expressed as $\delta_{T_b}(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_b)$. Also $\delta(t - nT_b) \Leftrightarrow e^{-jn\omega T_b}$. Hence, the Fourier transform of this impulse train is $\sum_{n=-\infty}^{\infty} e^{-jn\omega T_b}$. But we found the alternate form of the Fourier transform of this train in Eq. (3.42) (Example 3.13). Hence,

$$\sum_{n=-\infty}^{\infty} e^{-jn\omega T_b} = \frac{2\pi}{T_b} \sum_{n=-\infty}^{\infty} \delta\left(\omega - \frac{2\pi n}{T_b}\right)$$

Because of the bipolar rule,

$$P_{a_k a_{k+1}}(1, 1) = P_{a_k a_{k+1}}(-1, -1) = 0$$

and

$$P_{a_k a_{k+1}}(-1, 1) = P_{a_k}(-1)P_{a_{k+1}|a_k}(1|-1) = \left(\frac{1}{4}\right)\left(\frac{1}{2}\right) = \frac{1}{8}$$

Similarly, we find $P_{a_k a_{k+1}}(1, -1) = 1/8$. Substitution of these values in $\mathfrak{R}_1$ yields

$$\mathfrak{R}_1 = -\frac{1}{4}$$

For $n \geq 2$, the pulse strengths a_k and a_{k+1} become independent. Hence,

$$\mathfrak{R}_n = \overline{a_k a_{k+n}} = \bar{a}_k \bar{a}_{k+n} = (0)(0) = 0 \quad n \geq 2$$

Substitution of these values in Eq. (11.29) and noting that $\mathfrak{R}_n$ is an even function of n , yields

$$S_y(\omega) = \frac{|P(\omega)|^2}{T_b} \sin^2\left(\frac{\omega T_b}{2}\right)$$

This result is identical to Eq. (7.20b) found earlier by using time averages.

11.3 MULTIPLE RANDOM PROCESSES

For two real random processes $x(t)$ and $y(t)$, we define the **crosscorrelation function*** $R_{xy}(t_1, t_2)$ as

$$R_{xy}(t_1, t_2) = \overline{x(t_1)y(t_2)} \quad (11.31a)$$

The two processes are said to be **jointly stationary** (in the wide sense) if each of the processes is individually wide-sense stationary and if

$$\begin{aligned} R_{xy}(t_1, t_2) &= R_{xy}(t_2 - t_1) \\ &= R_{xy}(\tau) \end{aligned} \quad (11.31b)$$

Uncorrelated, Orthogonal (Incoherent), and Independent Processes

Two processes $x(t)$ and $y(t)$ are said to be **uncorrelated** if their crosscorrelation function is equal to the product of their means; that is,

$$R_{xy}(\tau) = \overline{x(t)y(t+\tau)} = \bar{x}\bar{y} \quad (11.32)$$

This implies that RVs $x(t)$ and $y(t+\tau)$ are uncorrelated for all t and τ .

* For complex random processes, the crosscorrelation function is defined as

$$R_{xy}(t_1, t_2) = \overline{x^*(t_1)y(t_2)} \quad (11.31a)$$

Processes $x(t)$ and $y(t)$ are said to be **incoherent**, or **orthogonal**, if

$$R_{xy}(\tau) = 0 \quad (11.33)$$

Incoherent, or orthogonal, processes are uncorrelated processes with $\bar{x}$ and/or $\bar{y} = 0$.

Processes $x(t)$ and $y(t)$ are **independent** random processes if for any t_1 and t_2 , $x(t_1)$ and $y(t_2)$ are independent.

Cross-Power Spectral Density

We define the **cross-power spectral density** $S_{xy}(\omega)$ for two random processes $x(t)$ and $y(t)$ as

$$S_{xy}(\omega) = \lim_{T \rightarrow \infty} \frac{X_T^*(\omega) Y_T(\omega)}{T} \quad (11.34)$$

where $X_T(\omega)$ and $Y_T(\omega)$ are the Fourier transforms of the truncated processes $x(t) \text{ rect}(t/T)$ and $y(t) \text{ rect}(t/T)$, respectively. Proceeding along the lines of the derivation of Eq. (11.14), it can be shown that*

$$R_{xy}(\tau) \iff S_{xy}(\omega) \quad (11.35a)$$

It can be seen from Eqs. (11.31) that for real random processes $x(t)$ and $y(t)$,

$$R_{xy}(\tau) = R_{yx}(-\tau) \quad (11.35b)$$

Therefore,

$$S_{xy}(\omega) = S_{yx}(-\omega) \quad (11.35c)$$

11.4 TRANSMISSION OF RANDOM PROCESSES THROUGH LINEAR SYSTEMS

If a random process $x(t)$ is applied at the input of a linear time-invariant system (Fig. 11.12) with transfer function $H(\omega)$, we can determine the autocorrelation function and the PSD of the output process $y(t)$. We now show that

$$R_y(\tau) = h(\tau) * h(-\tau) * R_x(\tau) \quad (11.36)$$

and

$$S_y(\omega) = |H(\omega)|^2 S_x(\omega) \quad (11.37)$$

To prove this, we observe that

$$y(t) = \int_{-\infty}^{\infty} h(\alpha) x(t - \alpha) d\alpha$$

and

$$y(t + \tau) = \int_{-\infty}^{\infty} h(\alpha) x(t + \tau - \alpha) d\alpha$$

* Equation (11.35a) is valid for complex processes as well.

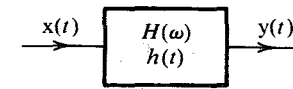


Figure 11.12 Transmission of a random process through a linear time-invariant system.

Hence,*

$$\begin{aligned} R_y(\tau) &= \overline{y(t)y(t+\tau)} = \overline{\int_{-\infty}^{\infty} h(\alpha)x(t-\alpha) d\alpha \int_{-\infty}^{\infty} h(\beta)x(t+\tau-\beta) d\beta} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\alpha)h(\beta) \overline{x(t-\alpha)x(t+\tau-\beta)} d\alpha d\beta \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\alpha)h(\beta) R_x(\tau + \alpha - \beta) d\alpha d\beta \end{aligned}$$

This double integral is precisely the double convolution $h(\tau) * h(-\tau) * R_x(\tau)$. Hence, Eqs. (11.36) and (11.37) follow.

EXAMPLE 11.9 Thermal Noise

Random thermal motion of electrons in a resistor R causes a random voltage across its terminals. This voltage $n(t)$ is known as the **thermal noise**. Its PSD $S_n(\omega)$ is practically flat over a very large band (up to 1000 GHz at room temperature) and is given by¹

$$S_n(\omega) = 2kTR \quad (11.38)$$

where k is the Boltzmann constant (1.38×10^{-23}) and T is the ambient temperature in kelvins. A resistor R at a temperature T K can be represented by a noiseless resistor R in series with a random white-noise voltage source (thermal noise) having a PSD of $2kTR$ (Fig. 11.13a). Observe that the thermal noise power over a band Δf is $(2kTR) 2\Delta f = 4kTR\Delta f$.

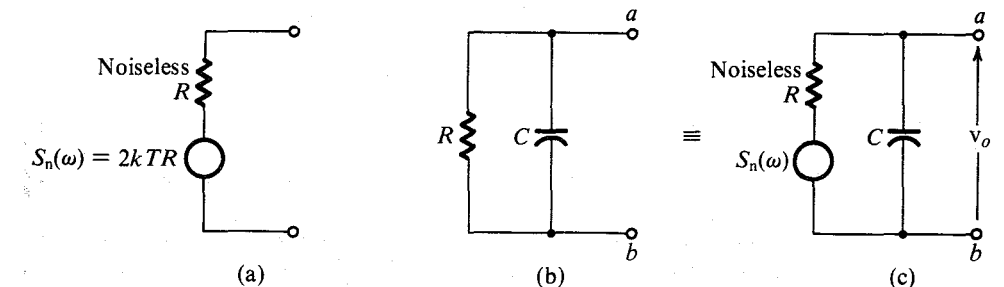


Figure 11.13 Thermal noise representation in a resistor.

Let us calculate the thermal noise voltage (rms value) across the simple RC circuit in Fig. 11.13b. The resistor R is replaced by an equivalent noiseless resistor in series with the thermal-noise voltage source. The transfer function $H(\omega)$ relating the voltage v_o at terminals ab to the thermal noise voltage is given by

* In this development, we interchange the operations of averaging and integrating. Because averaging is really an operation of integration, we are really changing the order of integration, and we assume that such a change is permissible.

$$H(\omega) = \frac{1/j\omega C}{R + 1/j\omega C} = \frac{1}{1 + j\omega RC}$$

If $S_0(\omega)$ is the PSD of the voltage v_o , then from Eq. (11.37) we have

$$\begin{aligned} S_0(\omega) &= \left| \frac{1}{1 + j\omega RC} \right|^2 2kTR \\ &= \frac{2kTR}{1 + \omega^2 R^2 C^2} \end{aligned}$$

The mean square value $\overline{v_o^2}$ is given by

$$\begin{aligned} \overline{v_o^2} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{2kTR}{1 + \omega^2 R^2 C^2} d\omega \\ &= \frac{kT}{C} \end{aligned} \quad (11.39)$$

Hence, the rms thermal noise voltage across the capacitor is $\sqrt{kT/C}$.

Sum of Random Processes

If two stationary processes (at least in the wide sense) $x(t)$ and $y(t)$ are added to form a process $z(t)$, the statistics of $z(t)$ can be determined in terms of those of $x(t)$ and $y(t)$. If

$$z(t) = x(t) + y(t) \quad (11.40a)$$

then

$$\begin{aligned} R_z(\tau) &= \overline{z(t)z(t+\tau)} = \overline{[x(t) + y(t)][x(t+\tau) + y(t+\tau)]} \\ &= R_x(\tau) + R_y(\tau) + R_{xy}(\tau) + R_{yx}(\tau) \end{aligned} \quad (11.40b)$$

If $x(t)$ and $y(t)$ are uncorrelated, then from Eq. (11.32),

$$R_{xy}(\tau) = R_{yx}(\tau) = \bar{x}\bar{y}$$

and

$$R_z(\tau) = R_x(\tau) + R_y(\tau) + 2\bar{x}\bar{y} \quad (11.41)$$

Most processes of interest in communication problems have zero means. If processes $x(t)$ and $y(t)$ are uncorrelated with either $\bar{x}$ or $\bar{y} = 0$ (that is, if $x(t)$ and $y(t)$ are incoherent), then

$$R_z(\tau) = R_x(\tau) + R_y(\tau) \quad (11.42a)$$

and

$$S_z(\omega) = S_x(\omega) + S_y(\omega) \quad (11.42b)$$

It also follows from Eqs. (11.42a) and (11.17) that

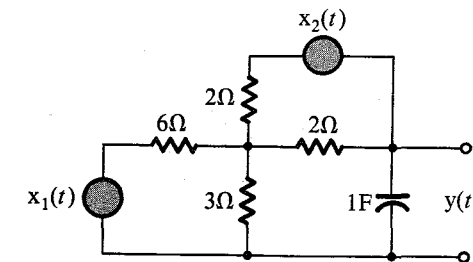
$$\overline{z^2} = \overline{x^2} + \overline{y^2} \quad (11.42c)$$

Hence, the mean square of a sum of incoherent (or orthogonal) processes is equal to the sum of the mean squares of these processes.

EXAMPLE 11.10 Two independent random voltage processes $x_1(t)$ and $x_2(t)$ are applied to an RC network, as shown in Fig. 11.14. It is given that

$$S_{x_1}(\omega) = K \quad S_{x_2}(\omega) = \frac{2\alpha}{\alpha^2 + \omega^2}$$

Figure 11.14 Noise calculations in a resistive circuit.



Determine the PSD and the power P_y of the output random process $y(t)$. Assume that the resistors in the circuit contribute negligible thermal noise (i.e., assume that they are noiseless).

Because the network is linear, the output voltage $y(t)$ can be expressed as

$$y(t) = y_1(t) + y_2(t)$$

where $y_1(t)$ is the output from input $x_1(t)$ [assuming $x_2(t) = 0$] and $y_2(t)$ is the output from input $x_2(t)$ [assuming $x_1(t) = 0$]. The transfer functions relating $y(t)$ to $x_1(t)$ and $x_2(t)$ are $H_1(\omega)$ and $H_2(\omega)$, respectively, given by

$$H_1(\omega) = \frac{1}{3(3j\omega + 1)}, \quad H_2(\omega) = \frac{1}{2(3j\omega + 1)}$$

Hence,

$$S_{y_1}(\omega) = |H_1(\omega)|^2 S_{x_1}(\omega) = \frac{K}{9(9\omega^2 + 1)}$$

and

$$S_{y_2}(\omega) = |H_2(\omega)|^2 S_{x_2}(\omega) = \frac{\alpha}{2(9\omega^2 + 1)(\alpha^2 + \omega^2)}$$

Because the input processes $x_1(t)$ and $x_2(t)$ are independent, the outputs $y_1(t)$ and $y_2(t)$ generated by them will also be independent. Also, the PSDs of $y_1(t)$ and $y_2(t)$ have no impulses at $\omega = 0$, implying that they have no dc components [i.e., $y_1(t) = y_2(t) = 0$]. Hence, $y_1(t)$ and $y_2(t)$ are incoherent, and

$$\begin{aligned} S_y(\omega) &= S_{y_1}(\omega) + S_{y_2}(\omega) \\ &= \frac{2K(\alpha^2 + \omega^2) + 9\alpha}{18(9\omega^2 + 1)(\alpha^2 + \omega^2)} \end{aligned}$$

The power P_y (or the mean square value $\overline{y^2}$) can be determined in two ways. We can find $R_y(\tau)$ by taking the inverse transforms of $S_{y_1}(\omega)$ and $S_{y_2}(\omega)$ as

$$R_y(\tau) = \underbrace{\frac{K}{54} e^{-|\tau|/3}}_{R_{y_1}(\tau)} + \underbrace{\frac{3\alpha - e^{-\alpha|\tau|}}{4(9\alpha^2 - 1)}}_{R_{y_2}(\tau)}$$

and

$$P_y = \overline{y^2} = R_y(0) = \frac{K}{54} + \frac{3\alpha - 1}{4(9\alpha^2 - 1)}$$

Alternatively, we can determine $\overline{y^2}$ by integrating $S_y(\omega)$ with respect to ω (or f) [see Eq. (11.17)].

11.5 BANDPASS RANDOM PROCESSES

If the PSD of a random process is confined to a certain passband (Fig. 11.15), the process is a **bandpass** random process. Just as a bandpass signal can be represented in terms of quadrature components [see Eq. (3.38)], we can express a bandpass random process $x(t)$ in terms of quadrature components as follows:

$$x(t) = x_c(t) \cos \omega_c t + x_s(t) \sin \omega_c t \quad (11.43)$$

This can be proved by considering the system in Fig. 11.16a, where $H_0(\omega)$ is an ideal low-pass filter (Fig. 11.16b) with unit impulse response $h_0(t)$. First we show that the system in Fig. 11.16a is an ideal bandpass filter with the transfer function $H(\omega)$ shown in Fig. 11.16c. This can be conveniently done by computing the response $h(t)$ to the unit impulse input $\delta(t)$. Because the system contains time-varying multipliers, however, we must also test whether it is a time-varying or a time-invariant system. It is therefore appropriate to consider the system response to an input $\delta(t - \alpha)$. This is an impulse at $t = \alpha$. Using the fact that [see Eq. (2.18b)] $f(t) \delta(t - \alpha) = f(\alpha) \delta(t - \alpha)$, we can express the signals at various points as follows:

$$\begin{aligned} \text{Signal at } a_1 & 2 \cos(\omega_c \alpha + \theta) \delta(t - \alpha) \\ a_2 & 2 \sin(\omega_c \alpha + \theta) \delta(t - \alpha) \\ b_1 & 2 \cos(\omega_c \alpha + \theta) h_0(t - \alpha) \\ b_2 & 2 \sin(\omega_c \alpha + \theta) h_0(t - \alpha) \\ c_1 & 2 \cos(\omega_c \alpha + \theta) \cos(\omega_c t + \theta) h_0(t - \alpha) \\ c_2 & 2 \sin(\omega_c \alpha + \theta) \sin(\omega_c t + \theta) h_0(t - \alpha) \\ d & 2h_0(t - \alpha) [\cos(\omega_c \alpha + \theta) \cos(\omega_c t + \theta) + \sin(\omega_c \alpha + \theta) \sin(\omega_c t + \theta)] \\ & = 2h_0(t - \alpha) \cos[\omega_c(t - \alpha)] \end{aligned}$$

Thus, the system response to the input $\delta(t - \alpha)$ is $2h_0(t - \alpha) \cos[\omega_c(t - \alpha)]$. Clearly, the system is linear time-invariant, with impulse response

$$h(t) = 2h_0(t) \cos \omega_c t$$

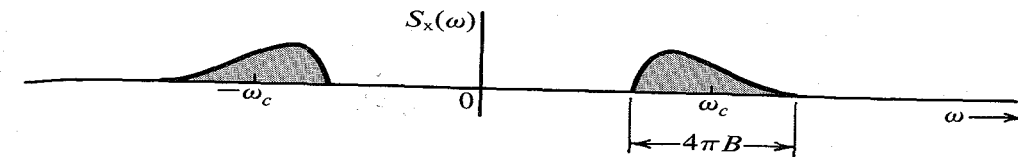


Figure 11.15 PSD of a bandpass random process.

and

$$H(\omega) = H_0(\omega + \omega_c) + H_0(\omega - \omega_c)$$

The transfer function $H(\omega)$ (Fig. 11.16c) represents an ideal bandpass filter.

If we apply the bandpass process $x(t)$ (Fig. 11.15) to the input of this system, the output will be $x(t)$. If the processes at points b_1 and b_2 (low-pass filter outputs) are denoted by $x_c(t)$ and $x_s(t)$, respectively, then the output $x(t)$ can be written as

$$x(t) = x_c(t) \cos(\omega_c t + \theta) + x_s(t) \sin(\omega_c t + \theta) \quad (11.44)$$

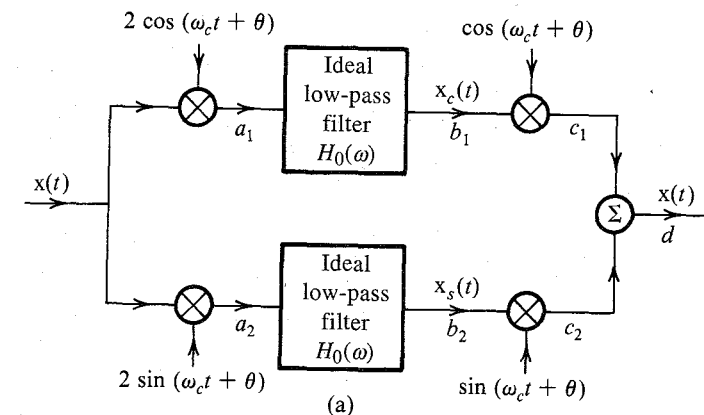
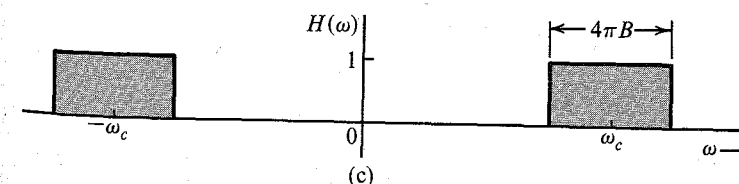
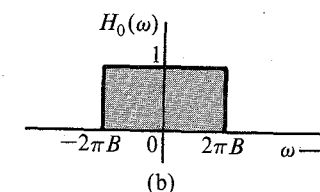


Figure 11.16 (a) Equivalent circuit of an ideal bandpass filter. (b) Ideal low-pass filter frequency response. (c) Ideal bandpass filter frequency response.



where $x_c(t)$ and $x_s(t)$ are low-pass random processes band-limited to B Hz (because they are the outputs of low-pass filters of bandwidth B). Because Eq. (11.44) is valid for any value of θ , by substituting $\theta = 0$, we get the desired representation in Eq. (11.43).

In order to characterize $x_c(t)$ and $x_s(t)$, consider once again Fig. 11.16a with the input $x(t)$. Let θ be an RV uniformly distributed over the range $(0, 2\pi)$, that is, for a sample function, θ is equally likely to take on any value in the range $(0, 2\pi)$. In this case $x(t)$ is represented as in Eq. (11.44). We observe that $x_c(t)$ is obtained by multiplying $x(t)$ by $2 \cos(\omega_c t + \theta)$, and then passing the result through a low-pass filter. The PSD of $2x(t) \cos(\omega_c t + \theta)$ is [see Eq. (11.20b)]

$$4 \times \frac{1}{4} [S_x(\omega + \omega_c) + S_x(\omega - \omega_c)]$$

This PSD is $S_x(\omega)$ shifted up and down by ω_c , as shown in Fig. 11.17a. When this is passed through a low-pass filter, the resulting PSD of $x_c(t)$ is as shown in Fig. 11.17b. It is clear that

$$S_{x_c}(\omega) = \begin{cases} S_x(\omega + \omega_c) + S_x(\omega - \omega_c) & |\omega| \leq 2\pi B \\ 0 & |\omega| > 2\pi B \end{cases} \quad (11.45a)$$

We can obtain $S_{x_s}(\omega)$ in the same way. As far as the PSD is concerned, multiplication by $\cos(\omega_c t + \theta)$ or $\sin(\omega_c t + \theta)$ makes no difference (see footnote following Eq. (11.20a), and we get

$$S_{x_c}(\omega) = S_{x_s}(\omega) = \begin{cases} S_x(\omega + \omega_c) + S_x(\omega - \omega_c) & |\omega| \leq 2\pi B \\ 0 & |\omega| > 2\pi B \end{cases} \quad (11.45b)$$

From Figs. 11.15 and 11.17b, we make the interesting observation that the areas under the PSDs $S_x(\omega)$, $S_{x_c}(\omega)$, and $S_{x_s}(\omega)$ are equal. Hence, it follows that

$$\overline{x_c^2(t)} = \overline{x_s^2(t)} = \overline{x^2(t)} \quad (11.45c)$$

Thus, the mean square values (or powers) of $x_c(t)$ and $x_s(t)$ are identical to that of $x(t)$.

These results are derived by assuming Θ to be an RV. For the representation in Eq. (11.43), $\Theta = 0$, and Eqs. (11.45b, c) may not be true. Fortunately, Eqs. (11.45b, c) hold even for the

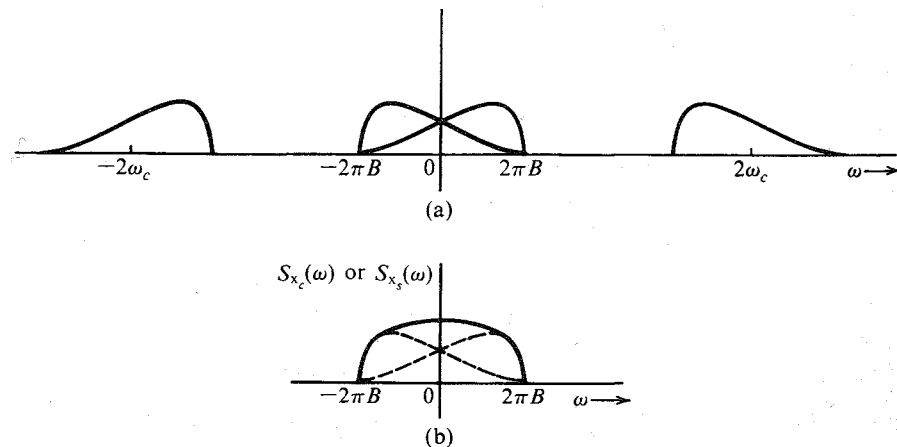


Figure 11.17 Derivation of PSDs of quadrature components of a bandpass random process.

case of $\Theta = 0$. The proof is rather long and cumbersome and will not be given here.^{1,2,3} It can also be shown that^{1,2,3}

$$\overline{x_c(t)x_s(t)} = R_{x_c x_s}(0) = 0 \quad (11.46)$$

That is, the amplitudes x_c and x_s at any given instant are uncorrelated. Moreover, if $S_x(\omega)$ is symmetrical about ω_c (as well as $-\omega_c$), then

$$R_{x_c x_s}(\tau) = 0 \quad (11.47)$$

EXAMPLE 11.11 The PSD of a bandpass white noise $n(t)$ is $\mathcal{N}/2$ (Fig. 11.18a). Represent this process in terms of quadrature components. Derive $S_{n_c}(\omega)$ and $S_{n_s}(\omega)$, and verify that $\overline{n_c^2} = \overline{n_s^2} = \overline{n^2}$.

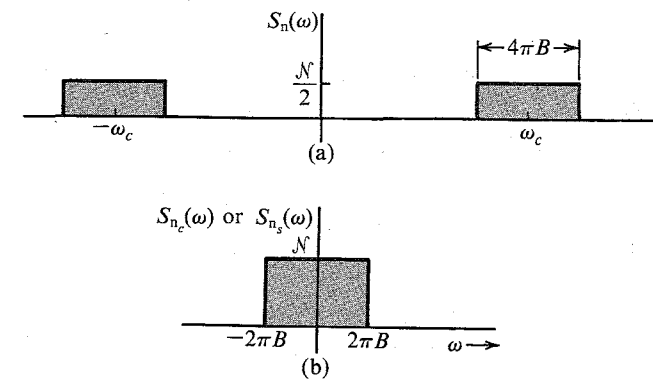


Figure 11.18 (a) PSD of a bandpass white-noise process. (b) PSD of its quadrature components.

We have the expression

$$n(t) = n_c(t) \cos \omega_c t + n_s(t) \sin \omega_c t \quad (11.48)$$

where

$$S_{n_c}(\omega) = S_{n_s}(\omega) = \begin{cases} S_n(\omega + \omega_c) + S_n(\omega - \omega_c) & |\omega| \leq 2\pi B \\ 0 & |\omega| > 2\pi B \end{cases}$$

It follows from this equation and from Fig. 11.18 that

$$S_{n_c}(\omega) = S_{n_s}(\omega) = \begin{cases} \mathcal{N} & |\omega| \leq 2\pi B \\ 0 & |\omega| > 2\pi B \end{cases} \quad (11.49)$$

Also,

$$\overline{n^2} = 2 \int_{f_c-B}^{f_c+B} \frac{\mathcal{N}}{2} df = 2\mathcal{N}B \quad (11.50a)$$

From Fig. 11.18b it follows that

$$\overline{n_c^2} = \overline{n_s^2} = 2 \int_0^B \mathcal{N} df = 2\mathcal{N}B \quad (11.50b)$$

Hence,

$$\overline{n_c^2} = \overline{n_s^2} = \overline{n^2} = 2\mathcal{N}B \quad (11.50c)$$

Nonuniqueness of the Quadrature Representation

No unique center frequency exists for a bandpass signal. For the spectrum in Fig. 11.19a, for example, we may consider the spectrum to have a bandwidth $2B$ centered at ω_c . The same spectrum can be considered to have a bandwidth $2B'$ centered at ω_1 , as also shown in Fig. 11.19a. The quadrature representation [Eq. (11.43)] is also possible for center frequency ω_1 :

$$x(t) = x_{c1}(t) \cos \omega_1 t + x_{s1}(t) \sin \omega_1(t)$$

where

$$S_{x_{c1}}(\omega) = S_{x_{s1}}(\omega) = \begin{cases} S_x(\omega + \omega_1) + S_x(\omega - \omega_1) & |\omega| \leq 2\pi B' \\ 0 & |\omega| > 2\pi B' \end{cases} \quad (11.51)$$

This is shown in Fig. 11.19b. Thus, the quadrature representation of a bandpass process is not unique. An infinite number of possible choices exist for the center frequency, and corresponding to each center frequency is a distinct quadrature representation.

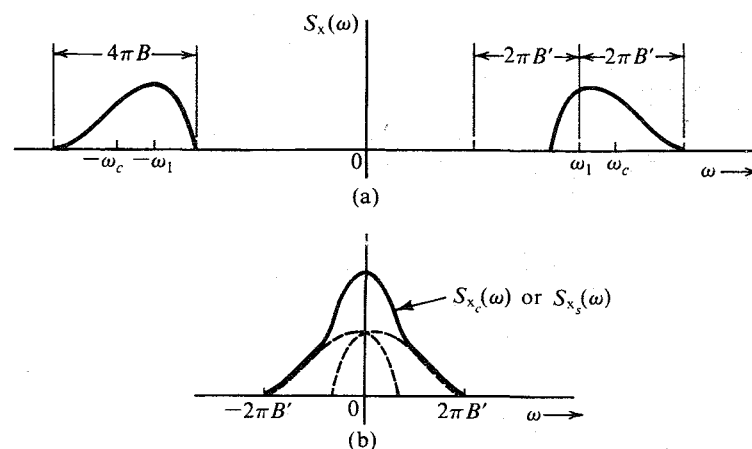


Figure 11.19 Nonunique nature of quadrature component representation of a bandpass process.

EXAMPLE 11.12 A bandpass white-noise PSD of an SSB channel (lower sideband) is shown in Fig. 11.20a. Represent this signal in terms of quadrature components with the carrier frequency ω_c .

The true center frequency of this PSD is not ω_c , but we can still use ω_c as the center frequency, as discussed earlier,

$$n(t) = n_c(t) \cos \omega_c t + n_s(t) \sin \omega_c t \quad (11.52)$$

The PSD $S_{n_c}(\omega)$ or $S_{n_s}(\omega)$ obtained by shifting $S_n(\omega)$ up and down by ω_c [see Eq. 11.51] is shown in Fig. 11.20b,

$$S_{n_c}(\omega) = S_{n_s}(\omega) = \begin{cases} \frac{\mathcal{N}}{2} & |\omega| \leq 2\pi B \\ 0 & |\omega| > 2\pi B \end{cases} \quad (11.53)$$

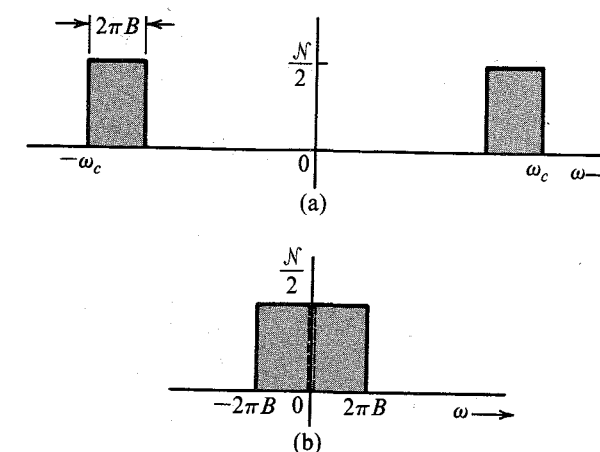


Figure 11.20 A possible form of quadrature component representation of noise in SSB.

From Fig. 11.20a it follows that

$$\overline{n^2} = \mathcal{N}B \quad (11.54a)$$

Similarly, from Fig. 11.20b we have

$$\overline{n_c^2} = \overline{n_s^2} = \mathcal{N}B \quad (11.54b)$$

Hence,

$$\overline{n_c^2} = \overline{n_s^2} = \overline{n^2} = \mathcal{N}B \quad (11.54c)$$

Bandpass White Gaussian Random Process

Thus far we have avoided defining a gaussian random process. The gaussian random process is perhaps the single most important random process in the area of communication. It requires a rather careful and unhurried discussion. Fortunately, we do not need to know much about the gaussian process at this point; therefore, its detailed discussion is postponed until Chapter 14 in order to avoid unnecessary digression. All we need to know here is that an RV $x(t)$ formed by sample-function amplitudes at instant t of a gaussian process is gaussian, with a PDF of the form of Eq. (10.38).

A gaussian random process with a uniform PSD is called a white gaussian random process. A bandpass white gaussian process $n(t)$ with PSD $\mathcal{N}/2$ centered at ω_c and with a bandwidth $2B$ (Fig. 11.18a) can be expressed in terms of quadrature components as

$$n(t) = n_c(t) \cos \omega_c t + n_s(t) \sin \omega_c t \quad (11.55)$$

where, from Eq. (11.49), we have

$$S_{n_c}(\omega) = S_{n_s}(\omega) = \begin{cases} \mathcal{N} & |\omega| \leq 2\pi B \\ 0 & |\omega| > 2\pi B \end{cases}$$

Also, from Eq. (11.50c),

$$\overline{n_c^2} = \overline{n_s^2} = \overline{n^2} = 2\mathcal{N}B \quad (11.56)$$

The bandpass signal can also be expressed in polar form [see Eq. (3.39)]:

$$n(t) = E(t) \cos(\omega_c t + \Theta) \quad (11.57a)$$

where

$$E(t) = \sqrt{n_c^2(t) + n_s^2(t)} \quad (11.57b)$$

$$\Theta(t) = -\tan^{-1} \frac{n_s(t)}{n_c(t)} \quad (11.57c)$$

The RVs $n_c(t)$ and $n_s(t)$ are uncorrelated [see Eq. (11.46)] gaussian RVs with zero means and variance $2\mathcal{N}B$ [Eq. (11.56)]. Hence, their PDFs are identical:

$$p_{n_c}(\alpha) = p_{n_s}(\alpha) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\alpha^2/2\sigma^2} \quad (11.58a)$$

where

$$\sigma^2 = 2\mathcal{N}B \quad (11.58b)$$

It will also be shown in Chapter 14 that if two gaussian RVs are uncorrelated, they are independent. In such a case, as shown in Example 10.14, $E(t)$ has a Rayleigh density

$$p_E(E) = \frac{E}{\sigma^2} e^{-E^2/2\sigma^2} \quad \sigma^2 = 2\mathcal{N}B \quad (11.59)$$

and Θ in Eq. (11.57a) is uniformly distributed over $(0, 2\pi)$.

Sinusoidal Signal in Noise

Another case of interest is a sinusoid plus a narrow-band gaussian noise. If $A \cos(\omega_c t + \varphi)$ is a sinusoid mixed with $n(t)$, a gaussian bandpass noise centered at ω_c , then the sum $y(t)$ is given by

$$y(t) = A \cos(\omega_c t + \varphi) + n(t)$$

Using Eq. (11.44) to represent the bandpass noise, we have

$$y(t) = [A + n_c(t)] \cos(\omega_c t + \varphi) + n_s(t) \sin(\omega_c t + \varphi) \quad (11.60a)$$

$$= E(t) \cos[\omega_c t + \Theta(t) + \varphi] \quad (11.60b)$$

where $E(t)$ is the envelope [$E(t) > 0$] and $\Theta(t)$ is the angle shown in Fig. 11.21,

$$E(t) = \sqrt{[A + n_c(t)]^2 + n_s^2(t)} \quad (11.61a)$$

$$\Theta(t) = -\tan^{-1} \frac{n_s(t)}{A + n_c(t)} \quad (11.61b)$$

Both $n_c(t)$ and $n_s(t)$ are gaussian, with variance σ^2 . For white gaussian noise, $\sigma^2 = 2\mathcal{N}B$ [Eq. (11.58b)]. Arguing in a manner analogous to that used in deriving Eq. (10.53), and observing that

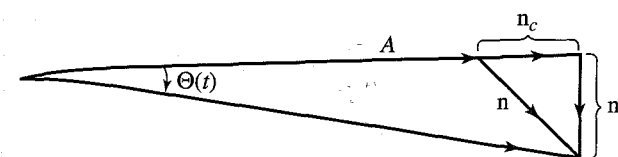


Figure 11.21 Phasor representation of a sinusoid and a narrow-band gaussian noise.

$$\begin{aligned} n_c^2 + n_s^2 &= E^2 - A^2 - 2An_c \\ &= E^2 - 2A(A + n_c) + A^2 \\ &= E^2 - 2AE \cos \Theta(t) + A^2 \end{aligned}$$

we have

$$p_{E\Theta}(E, \theta) = \frac{E}{2\pi\sigma^2} e^{-(E^2 - 2AE \cos \theta + A^2)/2\sigma^2} \quad (11.62)$$

where σ^2 is the variance of n_c (or n_s) and is equal to $2\mathcal{N}B$ for white noise. From Eq. (11.62) we have

$$\begin{aligned} p_E(E) &= \int_{-\pi}^{\pi} p_{E\Theta}(E, \theta) d\theta \\ &= \frac{E}{\sigma^2} e^{-(E^2 + A^2)/2\sigma^2} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{(AE/\sigma^2) \cos \theta} d\theta \right] \end{aligned} \quad (11.63)$$

The bracketed term on the right-hand side of Eq. (11.63) defines $I_0(AE/\sigma^2)$, where I_0 is the **modified zero-order Bessel function** of the first kind. Thus,

$$p_E(E) = \frac{E}{\sigma^2} e^{-(E^2 + A^2)/2\sigma^2} I_0\left(\frac{AE}{\sigma^2}\right) \quad (11.64a)$$

This is known as the **Rice density**, or **rician density**. For a large sinusoidal signal ($A \gg \sigma$), it can be shown that⁴

$$I_0\left(\frac{AE}{\sigma^2}\right) \simeq \sqrt{\frac{\sigma^2}{2\pi AE}} e^{AE/\sigma^2}$$

and

$$p_E(E) \simeq \sqrt{\frac{E}{2\pi A\sigma^2}} e^{-(E-A)^2/2\sigma^2} \quad (11.64b)$$

Because $A \gg \sigma$, $E \simeq A$, and $p_E(E)$ in Eq. (11.64b) is very nearly a gaussian density with mean A and variance σ ,

$$p_E(E) \simeq \frac{1}{\sigma\sqrt{2\pi}} e^{-(E-A)^2/2\sigma^2} \quad (11.64c)$$

Figure 11.22 shows the PDF of the normalized RV E/σ . Note that for $A/\sigma = 0$, we obtain the Rayleigh density.

From the joint PDF $p_{E\Theta}(E, \theta)$, we can also obtain $p_{\Theta}(\theta)$, the PDF of the phase Θ , by integrating the joint PDF with respect to E ,

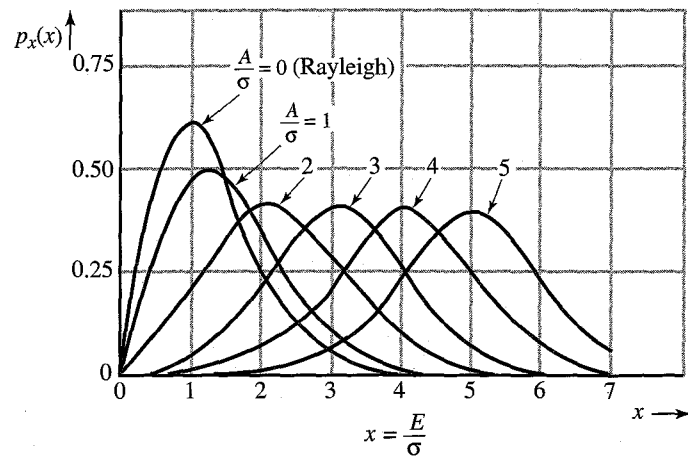


Figure 11.22 Rician PDF.

$$p_{\Theta}(\theta) = \int_0^{\infty} p_{E\Theta}(E, \theta) dE$$

Although the integration is straightforward, there are a number of involved steps, and for this reason it will not be repeated here. The final result is

$$p_{\Theta}(\theta) = \frac{1}{2\pi} e^{-A^2/2\sigma^2} \left\{ 1 + \frac{A}{\sigma} \sqrt{2\pi} \cos \theta e^{A^2 \cos^2 \theta / 2\sigma^2} \left[1 - Q\left(\frac{A \cos \theta}{\sigma}\right) \right] \right\} \quad (11.64d)$$

11.6 OPTIMUM FILTERING: WIENER-HOPF FILTER

When a desired signal is mixed with noise, the SNR can be improved by passing it through a filter that suppresses frequency components where the signal is weak but the noise is strong. The SNR improvement in this case can be explained qualitatively by considering a case of white noise mixed with a signal $m(t)$ whose PSD decreases at high frequencies. If the filter attenuates higher frequencies more, the signal will be reduced—in fact, distorted. The distortion component $m_e(t)$ may be considered an added noise. Thus, attenuation of higher frequencies will cause additional noise (from signal distortion), but, in compensation, it will reduce the channel noise, which is strong at high frequencies. Because at higher frequencies the signal has a small power content, the distortion component will be small compared to the reduction in channel noise, and the total noise may be smaller than before.

Let $H_{op}(\omega)$ be the optimum filter (Fig. 11.23a). This filter, not being ideal, will cause signal distortion. The distortion signal $m_e(t)$ can be found from Fig. 11.23b. The distortion signal power N_D appearing at the output is given by

$$N_D = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_m(\omega) |H_{op}(\omega) - 1|^2 d\omega$$

where $S_m(\omega)$ is the signal PSD at the input of the receiving filter. The channel noise power N_{ch} appearing at the filter output is given by

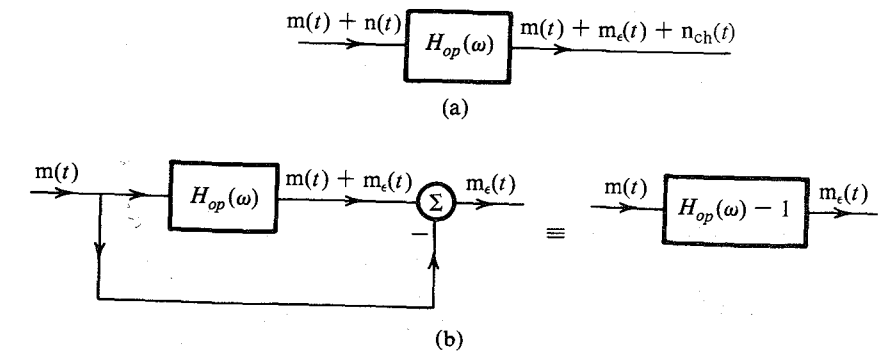


Figure 11.23 Wiener-Hopf filter calculations.

$$N_{ch} = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_n(\omega) |H_{op}(\omega)|^2 d\omega$$

where $S_n(\omega)$ is the noise PSD appearing at the input of the receiving filter. The distortion component acts as a noise. Because the signal and the channel noise are incoherent, the total noise N_o at the receiving filter output is the sum of the channel noise N_{ch} and the distortion noise N_D ,

$$N_o = N_{ch} + N_D = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[|H_{op}(\omega)|^2 S_n(\omega) + |H_{op}(\omega) - 1|^2 S_m(\omega) \right] d\omega \quad (11.65a)$$

Using the fact that $|A + B|^2 = (A + B)(A^* + B^*)$, and noting that both $S_m(\omega)$ and $S_n(\omega)$ are real, Eq. (11.65a) can be rearranged as

$$N_o = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\left| H_{op}(\omega) - \frac{S_m(\omega)}{S_r(\omega)} \right|^2 S_r(\omega) + \frac{S_m(\omega) S_n(\omega)}{S_r(\omega)} \right] d\omega \quad (11.65b)$$

where $S_r(\omega) = S_m(\omega) + S_n(\omega)$. The integrand on the right-hand side of Eq. (11.65b) is nonnegative. Moreover, it is a sum of two nonnegative terms. Hence, to minimize N_o , we must minimize each term. Because the second term $S_m(\omega) S_n(\omega) / S_r(\omega)$ is independent of $H_{op}(\omega)$, only the first term can be minimized. From Eq. (11.65b) it is obvious that this term is minimum when

$$H_{op}(\omega) = \frac{S_m(\omega)}{S_r(\omega)} = \frac{S_m(\omega)}{S_m(\omega) + S_n(\omega)} \quad (11.66a)$$

For this optimum choice, the output noise power N_o is given by

$$N_o = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{S_m(\omega) S_n(\omega)}{S_m(\omega) + S_n(\omega)} d\omega \quad (11.66b)$$

The optimum filter is known as the **Wiener-Hopf filter** in the literature. Equation (11.66a) shows that $H_{op}(\omega) \approx 1$ (no attenuation) when $S_m(\omega) \gg S_n(\omega)$. But when $S_m(\omega) \ll S_n(\omega)$, the filter has high attenuation. In other words, the optimum filter attenuates heavily the band where noise is relatively stronger. This causes some signal distortion, but at the same time it attenuates the noise more heavily so that the overall SNR is improved.

Comments on the Optimum Filter

If the SNR at the filter input is reasonably large—e.g., $S_m(\omega) > 100S_n(\omega)$ (SNR of 20 dB)—the optimum filter [Eq. (11.66a)] in this case is practically an ideal filter, and N_o [Eq. (11.66b)] is given by

$$N_o \simeq \frac{1}{2\pi} \int_{-\infty}^{\infty} S_n(\omega) d\omega$$

Hence for a large input SNR, optimization yields insignificant improvement. The Wiener-Hopf filter is therefore practical only when the input SNR is small (large-noise case).

Another issue is the realizability of the optimum filter in Eq. (11.66a). Because $S_m(\omega)$ and $S_n(\omega)$ are both even functions of ω , the optimum filter $H_{op}(\omega)$ is an even function of ω . Hence, the unit impulse response $h_{op}(t)$ is an even function of t (see Prob. 3.1-1). This makes $h_{op}(t)$ noncausal and the filter unrealizable. As noted earlier, such a filter can be realized approximately if we are willing to tolerate some delay in the output. If delay cannot be tolerated, the derivation of $H_{op}(\omega)$ must be repeated with a realizability constraint. Note that the realizable optimum filter can never be superior to the unrealizable optimum filter [Eq. (11.66a)]. Thus, the filter in Eq. (11.66a) gives the upper bound on performance (output SNR). Discussion of realizable optimum filters can be readily found in the literature.^{1,3}

EXAMPLE 11.13 A random process $m(t)$ (the signal) is mixed with a white channel noise $n(t)$. Given

$$S_m(\omega) = \frac{2\alpha}{\alpha^2 + \omega^2} \quad \text{and} \quad S_n(\omega) = \frac{\mathcal{N}}{2}$$

find the Wiener-Hopf filter to maximize the SNR. Find the resulting output noise power N_o .

From Eq. (11.66a),

$$\begin{aligned} H_{op}(\omega) &= \frac{4\alpha}{4\alpha + \mathcal{N}(\alpha^2 + \omega^2)} \\ &= \frac{4\alpha}{\mathcal{N}(\beta^2 + \omega^2)} \quad \beta^2 = \frac{4\alpha}{\mathcal{N}} + \alpha^2 \end{aligned} \quad (11.67a)$$

Hence,

$$h_{op}(t) = \frac{2\alpha}{\mathcal{N}\beta} e^{-\beta|t|} \quad (11.67b)$$

Figure 11.24a shows $h_{op}(t)$. It is evident that this is an unrealizable filter. However, a delayed version (Fig. 11.24b) of this filter, that is, $h_{op}(t - t_0)$, is closely realizable if we make $t_0 \geq 3/\beta$ and eliminate the tail for $t < 0$ (Fig. 11.24c).

The output noise power N_o is [Eq. (11.66b)]

$$N_o = \frac{1}{2\pi} \int_0^{\infty} \frac{2\alpha}{\beta^2 + \omega^2} d\omega = \frac{\alpha}{\beta} = \frac{\alpha}{\sqrt{\alpha^2 + (4\alpha/\mathcal{N})}} \quad (11.68)$$

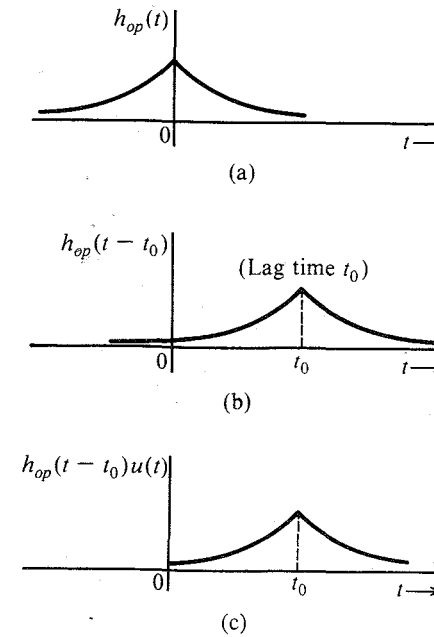


Figure 11.24 Close realization of an unrealizable filter using delay.

REFERENCES

1. B. P. Lathi, *An Introduction to Random Signals and Communication Theory*, International Textbook Co., Scranton, PA, 1968.
2. J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*, Wiley, New York, 1965.
3. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 2nd ed., McGraw-Hill, New York, 1984.
4. S. O. Rice, "Mathematical Analysis of Random Noise," *Bell. Syst. Tech. J.*, vol. 23, pp. 282–332, July 1944; vol. 24, pp. 46–156, Jan. 1945.

PROBLEMS

11.1-1 (a) Sketch the ensemble of the random process

$$x(t) = a \cos(\omega t + \Theta)$$

where ω and Θ are constants and a is an RV uniformly distributed in the range $(-A, A)$.

(b) Just by observing the ensemble, determine whether this is a stationary or a nonstationary process. Give your reasons.

11.1-2 Repeat Prob. 11.1-1 part (a) if a and Θ are constants but ω is an RV uniformly distributed in the range $(0, 100)$.

- 11.1-3 (a) Sketch the ensemble of the random process

$$x(t) = at + b$$

where b is a constant and a is an RV uniformly distributed in the range $(-2, 2)$.

- (b) Just by observing the ensemble, state whether this is a stationary or a nonstationary process.

- 11.1-4 Determine $\overline{x(t)}$ and $R_x(t_1, t_2)$ for the random process in Prob. 11.1-1, and determine whether this is a wide-sense stationary process.

- 11.1-5 Repeat Prob. 11.1-4 for the process $x(t)$ in Prob. 11.1-2.

- 11.1-6 Repeat Prob. 11.1-4 for the process $x(t)$ in Prob. 11.1-3.

- 11.1-7 Given a random process $x(t) = k$, where k is an RV uniformly distributed in the range $(-1, 1)$.

- (a) Sketch the ensemble of this process.
 (b) Determine $\overline{x(t)}$.
 (c) Determine $R_x(t_1, t_2)$.
 (d) Is the process wide-sense stationary?
 (e) Is the process ergodic?
 (f) If the process is wide-sense stationary, what is its power P_x [that is, its mean square value $\overline{x^2(t)}$]?]

- 11.1-8 Repeat Prob. 11.1-7 for the random process

$$x(t) = a \cos(\omega_c t + \Theta)$$

where ω_c is a constant and a and Θ are independent RVs uniformly distributed in the ranges $(-1, 1)$ and $(0, 2\pi)$, respectively.

- 11.2-1 For each of the following functions, state whether it can be a valid PSD of a real random process.

- (a) $\frac{\omega^2}{\omega^2 + 16}$ (b) $\frac{1}{\omega^2 - 16}$
 (c) $\frac{\omega}{\omega^2 + 16}$ (d) $\delta(\omega) + \frac{1}{\omega^2 + 16}$
 (e) $\delta(\omega + \omega_0) - \delta(\omega - \omega_0)$ (f) $j[\delta(\omega + \omega_0) + \delta(\omega - \omega_0)]$
 (g) $\frac{j\omega^2}{\omega^2 + 16}$

- 11.2-2 Show that for a wide-sense stationary process $x(t)$,

- (a) $R_x(0) \geq |R_x(\tau)|$ $\tau \neq 0$
 Hint: $(x_1 \pm x_2)^2 = x_1^2 + x_2^2 \pm 2x_1x_2 \geq 0$. Let $x_1 = x(t_1)$ and $x_2 = x(t_2)$.
 (b) $\lim_{\tau \rightarrow \infty} R_x(\tau) = \bar{x}^2$ Hint: As $\tau \rightarrow \infty$, x_1 and x_2 tend to become independent.

- 11.2-3 Show that if the PSD of a random process $x(t)$ is band-limited to B Hz, and if

$$R_x\left(\frac{n}{2B}\right) = \begin{cases} 1 & n = 0 \\ 0 & n = \pm 1, \pm 2, \pm 3, \dots \end{cases}$$

then $x(t)$ is a white bandlimited process; that is, $S_x(\omega) = k \text{rect}(\omega/4\pi B)$. Hint: Using the interpolation formula, reconstruct $R_x(\tau)$.

- 11.2-4 For the random binary process in Example 11.5 (Fig. 11.9a), determine $R_x(\tau)$ and $S_x(\omega)$ if the probability of transition (from 1 to -1 or vice versa) at each node is 0.6 instead of 0.5.

- 11.2-5 A wide-sense stationary white process $m(t)$ band-limited to B Hz is sampled at the Nyquist rate. Each sample is transmitted by a basic pulse $p(t)$ multiplied by the sample value. This is a PAM signal. Show that the PSD of the PAM signal is $2BR_m(0)|P(\omega)|^2$. Hint: Use Eq. (11.29). Show that Nyquist samples a_k and a_{k+n} ($n \geq 1$) are uncorrelated.

- 11.2-6 A duobinary line code proposed by Lender is a ternary scheme similar to bipolar, but requires only half the bandwidth of the latter. In this code, 0 is transmitted by no pulse, and 1 is transmitted by pulse $p(t)$ or $-p(t)$ using the following rule: A 1 is encoded by the same pulse as that used to encode the preceding 1 if the two 1's are separated by an even number of 0's. It is encoded by the negative of the pulse used to encode the preceding 1 if the two 1's are separated by an odd number of 0's. Random binary digits are transmitted every T_b seconds. Assuming $P(0) = P(1) = 0.5$, show that

$$S_y(\omega) = \frac{|P(\omega)|^2}{T_b} \cos^2\left(\frac{\omega T_b}{2}\right)$$

Find $S_y(\omega)$ if $p(t)$, the basic pulse used, is a half-width rectangular pulse $\text{rect}(2t/T_b)$.

- 11.2-7 Determine $S_y(\omega)$ for polar signaling if $P(1) = Q$ and $P(0) = 1 - Q$.

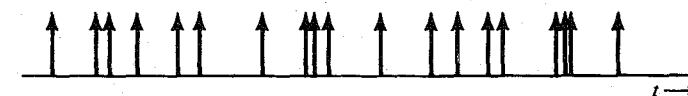


Figure P11.2-8

- 11.2-8 An impulse noise $x(t)$ can be modeled by a sequence of unit impulses located at random instants (Fig. P11.2-8). There are an average of α impulses per second, and the location of any impulse is independent of the locations of other impulses. Show that $R_x(\tau) = \alpha \delta(\tau) + \alpha^2$.

- 11.2-9 Repeat Prob. 11.2-8 if the impulses are equally likely to be positive and negative.

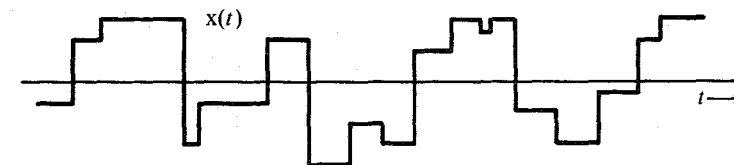


Figure P11.2-10

- 11.2-10 A sample function of a random process $x(t)$ is shown in Fig. P11.2-10. The signal $x(t)$ changes abruptly in amplitude at random instants. There are an average of β amplitude changes (or shifts) per second. The probability that there will be no amplitude shift in τ seconds is given

by $P_0(\tau) = e^{-\beta\tau}$. The amplitude after a shift is independent of the amplitude before the shift. The amplitudes are randomly distributed, with a PDF $p_x(x)$. Show that

$$R_x(\tau) = \overline{x^2} e^{-\beta|\tau|} \quad \text{and} \quad S_x(\omega) = \frac{2\beta\overline{x^2}}{\beta^2 + \omega^2}$$

This process represents a model for thermal noise.¹

11.3-1 Show that for jointly wide-sense stationary, real, random processes $x(t)$ and $y(t)$,

$$|R_{xy}(\tau)| \leq [R_x(0)R_y(0)]^{1/2}$$

Hint: For any real number a , $(ax - y)^2 \geq 0$.

11.3-2 If $x(t)$ and $y(t)$ are two incoherent random processes, and two new processes $u(t)$ and $v(t)$ are formed as follows:

$$u(t) = x(t) + y(t) \quad v(t) = 2x(t) + 3y(t)$$

find $R_u(\tau)$, $R_v(\tau)$, $R_{uv}(\tau)$, and $R_{vu}(\tau)$ in terms of $R_x(\tau)$ and $R_y(\tau)$.

11.3-3 Two random processes $x(t)$ and $y(t)$ are

$$x(t) = A \cos(\omega_0 t + \varphi) \quad \text{and} \quad y(t) = B \cos(n\omega_0 t + n\varphi)$$

where $n = \text{integer} \neq 1$ and A , B , and ω_0 are constants and φ is an RV uniformly distributed in the range $(0, 2\pi)$. Show that the two processes are incoherent.

11.3-4 A sample function of a periodic random process $x(t)$ is shown in Fig. P11.3-4. The interval b where the first pulse begins is an RV uniformly distributed in the range $(0, T_0)$. If a sample function with $b = 0$ is expressed by a compact trigonometric Fourier series,

$$x(t) = C_0 + \sum_{n=1}^{\infty} C_n \cos(n\omega_0 t + \theta_n)$$

then show that

$$R_x(\tau) = C_0^2 + \frac{1}{2} \sum_{n=1}^{\infty} C_n^2 \cos n\omega_0 \tau \quad \omega_0 = \frac{2\pi}{T_0}$$

Find the PSD $S_x(\omega)$. *Hint:*

$$x(t) = C_0 + \sum_{n=1}^{\infty} C_n \cos[n\omega_0(t - b) + \theta_n]$$

where b is an RV uniformly distributed in the range $(0, T_0)$. Now use the results in Prob. 11.3-3.

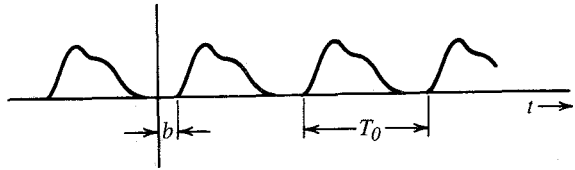


Figure P11.3-4

11.4-1 A simple RC circuit has two resistors R_1 and R_2 in parallel (Fig. P11.4-1a). Calculate the rms value of the thermal noise voltage v_o across the capacitor in two ways:

- (a) Consider resistors R_1 and R_2 as two separate resistors, with respective thermal noise voltages of PSD $2kTR_1$ and $2kTR_2$ (Fig. P11.4-1b). Note that the two sources are independent.
- (b) Consider the parallel combination of R_1 and R_2 as a single resistor of value $R_1 R_2 / (R_1 + R_2)$, with its thermal-noise voltage source of PSD $2kTR_1 R_2 / (R_1 + R_2)$ (Fig. P11.4-1c). Comment.

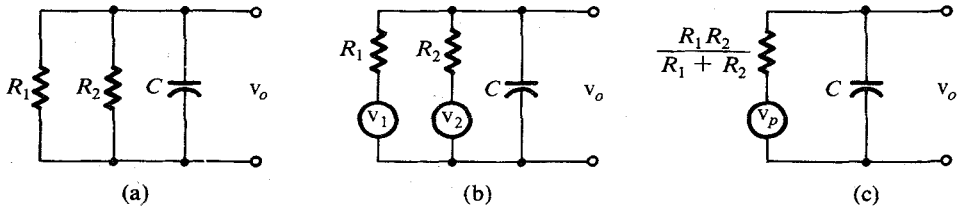


Figure P11.4-1

11.4-2 Show that $R_{xy}(\tau)$, the crosscorrelation function of the input process $x(t)$ and the output process $y(t)$ in Fig. 11.12, is

$$R_{xy}(\tau) = h(\tau) * R_x(\tau) \quad \text{and} \quad S_{xy}(\omega) = H(\omega)S_x(\omega)$$

Hence, show that for the thermal noise $n(t)$ and the output $v_o(t)$ in Fig. 11.13 (Example 11.9),

$$S_{nv_o}(\omega) = \frac{2kTR}{1 + j\omega RC} \quad \text{and} \quad R_{nv_o}(\tau) = \frac{2kT}{C} e^{-\tau/RC} u(\tau)$$

11.4-3 A shot noise is similar to impulse noise described in Prob. 11.2-8 except that instead of random impulses, we have pulses of finite width. If we replace each impulse in Fig. P11.2-8 by a pulse $h(t)$ whose width is large compared to $1/\alpha$, so that there is a considerable overlapping of pulses, we get shot noise. The result of pulse overlapping is that the signal looks like a continuous random signal, as shown in Fig. P11.4-3.

- (a) Derive the autocorrelation function and the PSD of such a random process. *Hint:* Shot noise results from passing impulse noise through a suitable filter. First derive the PSD of the shot noise and then obtain the autocorrelation function from the PSD. The answers will be in terms of α , $h(t)$, or $H(\omega)$.
- (b) The shot noise in transistors can be modeled by

$$h(t) = \frac{q}{T} e^{-t/T} u(t)$$

where q is the charge on an electron and T is the electron transit time. Determine and sketch the autocorrelation function and the PSD of the transistor shot noise.



Figure P11.4-3

- 11.5-1** A white process of PSD $\mathcal{N}/2$ is transmitted through a bandpass filter $H(\omega)$ (Fig. P11.5-1). Represent the filter output $n(t)$ in terms of quadrature components, and determine $S_{n_c}(\omega)$, $S_{n_s}(\omega)$, $\overline{n_c^2}$, $\overline{n_s^2}$, and $\overline{n^2}$ when the center frequency used in this representation is 100 kHz (that is, $\omega_c = 200\pi \times 10^3$).

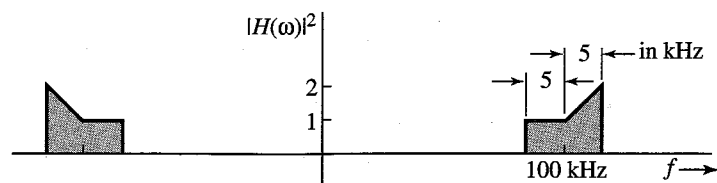


Figure P11.5-1

- 11.5-2** Repeat Prob. 11.5-1 if the center frequency ω_c used in the representation is not a true center frequency. Consider three cases: (a) $f_c = 105$ kHz; (b) $f_c = 95$ kHz; (c) $f_c = 90$ kHz.
- 11.5-3** A random process $x(t)$ with the PSD shown in Fig. P11.5-3a is passed through a bandpass filter (Fig. 11.5-3b). Determine the PSDs and mean square values of the quadrature components of the output process. Assume the center frequency in the representation to be 0.5 MHz.

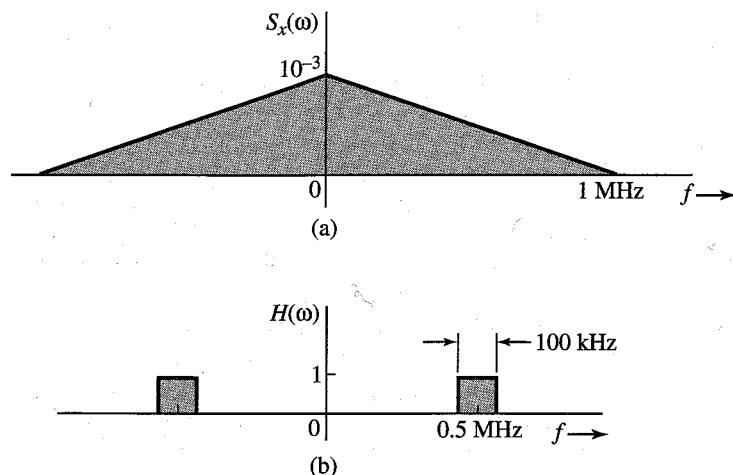


Figure P11.5-3

- 11.5-4** A signal process $m(t)$ is mixed with a channel noise $n(t)$. The respective PSDs are

$$S_m(\omega) = \frac{6}{9 + \omega^2} \quad \text{and} \quad S_n(\omega) = 6$$

- (a) Find the optimum Wiener-Hopf filter.
 (b) Sketch its unit impulse response.
 (c) Estimate the amount of delay necessary to make this filter closely realizable.
 (d) Compute the noise power at the input and the output of the filter.

- (e) What is the SNR improvement realized by the use of this filter? *Hint:* Note that because of the special way in which the problem of optimum filtering is formulated, the signal powers at the input and the output of the filter are identical (see Fig. 11.23).

- 11.5-5** Repeat Prob. 11.5-4 if

$$S_m(\omega) = \frac{4}{4 + \omega^2} \quad \text{and} \quad S_n(\omega) = \frac{32}{64 + \omega^2}$$

12 BEHAVIOR OF ANALOG SYSTEMS IN THE PRESENCE OF NOISE

In this chapter, we analyze the behavior of analog communication systems in the presence of noise to facilitate the comparison of various systems.

Figure 12.1 shows a schematic of a communication system. A certain signal power S_T is transmitted over a channel.* The transmitted signal is corrupted by channel noise during transmission. We shall assume channel noise to be additive. The channel will attenuate (and may also distort) the signal. At the receiver input, we have a signal mixed with noise. The signal and noise powers at the receiver input are S_i and N_i , respectively. The receiver processes (filters, demodulates, etc.) the signal to yield the desired signal plus noise. The signal and noise powers at the receiver output are S_o and N_o , respectively. In analog systems, the quality of the received signal is determined by S_o/N_o , the output SNR. Hence, we shall focus our attention on this parameter. But S_o/N_o can be increased as much as desired simply by increasing the transmitted power S_T . In practice, however, the maximum value of S_T is limited by other considerations, such as transmitter cost, channel capability, interference with other channels, and so on. Hence, the value of S_o/N_o for a given transmitted power is an appropriate figure of merit in an analog communication system. In practice, it is more convenient to deal with the received power S_i rather than the transmitted power S_T . From Fig. 12.1, it is apparent that S_i is proportional to S_T . Hence, the value of S_o/N_o for a given S_i will serve equally well as a figure of merit.

12.1 BASEBAND SYSTEMS

In baseband systems, the signal is transmitted directly without any modulation. This mode of communication is suitable over a pair of wires, optical fiber, or coaxial cables. It is mainly used in short-haul links. The study of baseband systems is important because many of the basic concepts and parameters encountered in baseband systems are carried over directly to modulated systems. Second, baseband systems serve as a basis against which other systems may be compared.

* Here the channel is used in the sense of a transmission medium.

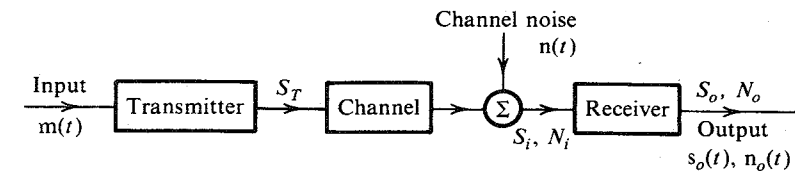


Figure 12.1 Communication system model.

For a baseband system, the transmitter and the receiver are ideal baseband filters (Fig. 12.2). The low-pass filter $H_p(\omega)$ at the transmitter limits the input signal spectrum to a given bandwidth. The low-pass filter $H_d(\omega)$ at the receiver eliminates the out-of-band noise and other channel interference. These filters can also serve an additional purpose, that of preemphasis and deemphasis, which optimizes the signal-to-noise ratio (SNR) at the receiver (or minimizes the channel noise interference).

The baseband signal $m(t)$ is assumed to be a zero mean, wide-sense stationary random process band-limited to B Hz. To begin with, we shall consider the case of ideal low-pass (or baseband) filters with bandwidth B at the transmitter and the receiver (Fig. 12.2). The channel is assumed to be distortionless. For this case,

$$S_o = S_i \quad (12.1a)$$

and

$$N_o = 2 \int_0^B S_n(\omega) df \quad (12.1b)$$

where $S_n(\omega)$ is the PSD of the channel noise. For the case of a white noise, $S_n(\omega) = \mathcal{N}/2$, and

$$N_o = 2 \int_0^B \frac{\mathcal{N}}{2} df = \mathcal{N}B \quad (12.1c)$$

and

$$\frac{S_o}{N_o} = \frac{S_i}{\mathcal{N}B} \quad (12.1d)$$

We define a parameter γ as

$$\gamma = \frac{S_i}{\mathcal{N}B} \quad (12.2)$$

From Eqs. (12.1d) and (12.2) we have

$$\frac{S_o}{N_o} = \gamma \quad (12.3)$$

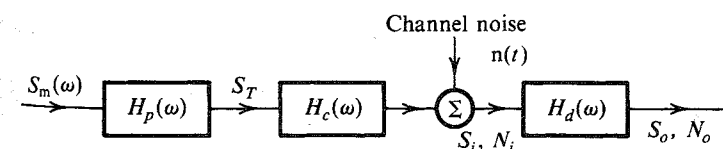


Figure 12.2 Baseband system.

The parameter γ is directly proportional to S_i and, therefore, directly proportional to S_T . Hence, a given S_T (or S_i) implies a given γ . Equation (12.3) is precisely the result we are looking for. It gives the receiver output SNR for a given S_T (or S_i).

The value of the SNR in Eq. (12.3) will serve as a standard against which the output SNR of other systems will be measured.

The power, or the mean square value, of $m(t)$ is $\overline{m^2}$, given by

$$\overline{m^2} = 2 \int_0^B S_m(\omega) df \quad (12.4)$$

In analog signals, the SNR is basic in specifying the signal quality. For voice signals, an SNR of 5 to 10 dB at the receiver implies a barely intelligible signal. Telephone quality signals have an SNR of 25 to 35 dB, whereas for television, an SNR of 45 to 55 dB is required.

12.2 AMPLITUDE-MODULATED SYSTEMS

We shall analyze DSB-SC, SSB-SC, and AM systems separately.

DSB-SC

A basic DSB-SC system is shown in Fig. 12.3.* The modulated signal is a bandpass signal centered at ω_c with a bandwidth $2B$. The channel noise is assumed to be additive. The channel and the filters in Fig. 12.3 are assumed to be ideal.

Let S_i and S_o represent the useful signal powers at the input and the output of the demodulator, and let N_o represent the noise power at the demodulator output. The signal at the demodulator input is $\sqrt{2}m(t)\cos\omega_c t + n(t)$, where $n(t)$ is the bandpass channel noise. Its spectrum is centered at ω_c and has a bandwidth $2B$. The input signal power S_i is the power of the modulated signal† $\sqrt{2}m(t)\cos\omega_c t$. From Eq. (11.20c),

$$S_i = [\sqrt{2}m(t)\cos\omega_c t]^2 = (\sqrt{2})^2 [\overline{m(t)\cos\omega_c t}]^2 = \overline{m^2(t)} = \overline{m^2} \quad (12.5)$$

The reader may now appreciate our use of $\sqrt{2}\cos\omega_c t$ (rather than $\cos\omega_c t$) in the modulator (Fig. 12.3). This was done to make the received power equal to that in the baseband system in order to facilitate comparison. We shall use a similar artifice in our analysis of the SSB system.

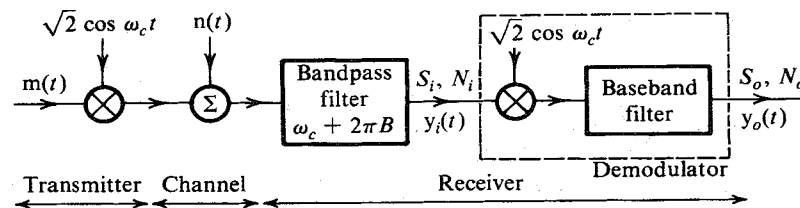


Figure 12.3 DSB-SC system.

* The use of an input bandpass filter in the receiver may appear redundant because the out-of-band noise components will be suppressed by the final baseband filter. In practice, an input filter is useful because by removing the out-of-band noise, it reduces the probability of nonlinear distortion from overload effects.

† The modulated signal also has a random phase Θ , which is uniformly distributed in the range $(0, 2\pi)$. This random phase [which is independent of $m(t)$] does not affect the final results and, hence, is ignored in this discussion.

To determine the output powers S_o and N_o , we note that the signal at the demodulator input is

$$y_i(t) = \sqrt{2}m(t)\cos\omega_c t + n(t)$$

Because $n(t)$ is a bandpass signal centered at ω_c , we can express it in terms of quadrature components, as in Eq. (11.48). This gives

$$y_i(t) = [\sqrt{2}m(t) + n_c(t)]\cos\omega_c t + n_s(t)\sin\omega_c t$$

When this signal is multiplied by $\sqrt{2}\cos\omega_c t$ (synchronous demodulation) and then low-pass filtered, the terms $m(t)\cos 2\omega_c t$ and $m(t)\sin 2\omega_c t$ are suppressed. The resulting demodulator output $y_o(t)$ is

$$y_o(t) = m(t) + \frac{1}{\sqrt{2}}n_c(t)$$

Hence,

$$S_o = \overline{m^2} = S_i \quad (12.6a)$$

$$N_o = \frac{1}{2}\overline{n_c^2(t)} \quad (12.6b)$$

For white noise with power density $\mathcal{N}/2$, we have [Eq. (11.50b)]

$$\overline{n_c^2(t)} = \overline{n^2(t)} = 2\mathcal{N}B$$

and

$$N_o = \mathcal{N}B \quad (12.7)$$

Hence, from Eqs. (12.6a) and (12.7) we have

$$\frac{S_o}{N_o} = \frac{S_i}{\mathcal{N}B} = \gamma \quad (12.8)$$

Comparison of Eqs. (12.8) and (12.3) shows that for a fixed transmitted power (which also implies a fixed signal power at the demodulator input), the SNR at the demodulator output is the same for the baseband and the DSB-SC systems. Moreover, quadrature multiplexing in DSB-SC can render its bandwidth requirement identical to that of baseband systems. Thus, theoretically, baseband and DSB-SC systems have identical capabilities.

SSB-SC

An SSB-SC system is shown in Fig. 12.4. The SSB signal* $\varphi_{SSB}(t)$ can be expressed as [see Eq. (4.17c)]

$$\varphi_{SSB}(t) = m(t)\cos\omega_c t + m_h(t)\sin\omega_c t \quad (12.9)$$

The spectrum of $\varphi_{SSB}(t)$ is shown in Fig. 4.15. This signal can be obtained (Fig. 12.4) by multiplying $m(t)$ by $2\cos\omega_c t$ and then suppressing the unwanted sideband. The power of the modulated signal $2m(t)\cos\omega_c t$ is $2\overline{m^2}$ [four times the power of $m(t)\cos\omega_c t$]. Suppression of one sideband halves the power. Hence S_i , the power of $\varphi_{SSB}(t)$, is

$$S_i = \overline{m^2} \quad (12.10)$$

* This is LSB. The discussion is valid for USB as well.

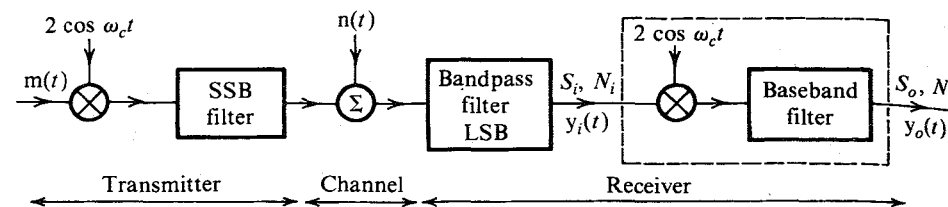


Figure 12.4 SSB-SC system.

Expressing the channel bandpass noise in terms of quadrature components as in Eq. (11.52), the signal at the detector input, $y_i(t)$, is

$$y_i(t) = [m(t) + n_c(t)] \cos \omega_c t + [m_h(t) + n_s(t)] \sin \omega_c t$$

This signal is multiplied by $2 \cos \omega_c t$ (synchronous demodulation) and then low-pass filtered to yield the demodulator output

$$y_o(t) = m(t) + n_c(t)$$

Hence,

$$\begin{aligned} S_o &= \overline{m^2} = S_i \\ N_o &= \overline{n_c^2} \end{aligned} \quad (12.11)$$

We have already found $\overline{n_c^2}$ for the SSB channel noise (lower sideband) in Eq. (11.54b) as

$$N_o = \overline{n_c^2} = \mathcal{N}B$$

Thus,

$$\frac{S_o}{N_o} = \frac{S_i}{\mathcal{N}B} = \gamma \quad (12.12)$$

This shows that baseband, DSB-SC, and SSB-SC systems perform identically in terms of resource utilization. All of them yield the same output SNR for a given transmitted power and transmission bandwidth.

EXAMPLE 12.1 In a DSB-SC system, the carrier frequency is $f_c = 500$ kHz, and the modulating signal $m(t)$ has a uniform PSD band-limited to 4 kHz. The modulated signal is transmitted over a distortionless channel with a noise PSD $S_n(\omega) = 1/(\omega^2 + a^2)$, where $a = 10^6 \pi$. The useful signal power at the receiver input is $1 \mu\text{W}$. The received signal is bandpass filtered, multiplied by $2 \cos \omega_c t$, and then low-pass filtered to obtain the output $s_o(t) + n_o(t)$. Determine the output SNR.

If the received signal is $km(t) \cos \omega_c t$, the demodulator input is $[km(t) + n_c(t)] \cos \omega_c t + n_s(t) \sin \omega_c t$. When this is multiplied by $2 \cos \omega_c t$ and low-pass filtered, the output is

$$s_o(t) + n_o(t) = km(t) + n_c(t) \quad (12.13)$$

Hence,

$$S_o = k^2 \overline{m^2} \quad \text{and} \quad N_o = \overline{n_c^2}$$

But the power of the received signal $km(t) \cos \omega_c t$ is $1 \mu\text{W}$. Hence,

$$\frac{k^2 \overline{m^2}}{2} = 10^{-6}$$

and

$$S_o = k^2 \overline{m^2} = 2 \times 10^{-6}$$

To compute $\overline{n_c^2}$, we use Eq. (11.45c):

$$\overline{n_c^2} = \overline{n^2}$$

where $\overline{n^2}$ is the power of the incoming bandpass noise of bandwidth 8 kHz centered at 500 kHz; that is,

$$\begin{aligned} \overline{n^2} &= 2 \int_{496,000}^{504,000} \frac{1}{\omega^2 + a^2} df \quad a = 10^6 \pi \\ &= \frac{1}{\pi} \int_{(2\pi)496,000}^{(2\pi)504,000} \frac{1}{\omega^2 + a^2} d\omega = \frac{1}{\pi a} \tan^{-1} \frac{\omega}{a} \Big|_{(2\pi)496,000}^{(2\pi)504,000} \\ &= 8.25 \times 10^{-10} = N_o \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{S_o}{N_o} &= \frac{2 \times 10^{-6}}{8.25 \times 10^{-10}} = 2.42 \times 10^3 \\ &= 33.83 \text{ dB} \end{aligned}$$

AM

AM signals can be demodulated synchronously or by envelope detection. The former is of theoretical interest only. It is useful, however, for comparing the noise performance of the envelope detector. For this reason, we shall consider both of these methods.

Coherent, or Synchronous, Demodulation of AM: Coherent AM is identical to DSB-SC in every respect except for the additional carrier. If the received signal $\sqrt{2}[A + m(t)] \cos \omega_c t$ is multiplied by $\sqrt{2} \cos \omega_c t$, the demodulator output is $m(t)$. Hence,

$$S_o = \overline{m^2}$$

The output noise will be exactly the same as that in DSB-SC [Eq. (12.7)]:

$$N_o = \overline{n_o^2} = \mathcal{N}B$$

The received signal is $\sqrt{2}[A + m(t)] \cos \omega_c t$. Hence,

$$\begin{aligned} S_i &= (\sqrt{2})^2 \frac{[A + m(t)]^2}{2} \\ &= [A + m(t)]^2 \\ &= A^2 + \overline{m^2(t)} + 2A \overline{m(t)} \end{aligned}$$

Because $m(t)$ is assumed to have zero mean,

$$S_i = A^2 + \overline{m^2(t)}$$

and

$$\begin{aligned} \frac{S_o}{N_o} &= \frac{\overline{m^2}}{\mathcal{N}B} \\ &= \frac{\overline{m^2}}{A^2 + \overline{m^2}} \frac{S_i}{\mathcal{N}B} \\ &= \frac{\overline{m^2}}{A^2 + \overline{m^2}} \gamma \end{aligned} \quad (12.14)$$

If $m(t)_{\max} = m_p$, then $A \geq m_p$. For the maximum SNR, $A = m_p$, and

$$\begin{aligned} \left(\frac{S_o}{N_o} \right)_{\max} &= \frac{\overline{m^2}}{m_p^2 + \overline{m^2}} \gamma \\ &= \frac{1}{(m_p^2/\overline{m^2} + 1)} \gamma \end{aligned} \quad (12.15a)$$

Because $(m_p^2/\overline{m^2}) \geq 1$,

$$\frac{S_o}{N_o} \leq \frac{\gamma}{2} \quad (12.15b)$$

It can be seen that the SNR in AM is at least 3 dB (and usually about 6 dB in practice) worse than that in DSB-SC and SSB-SC (depending on the modulation index and the signal waveform). For example, when $m(t)$ is sinusoidal, $m_p^2/\overline{m^2} = 2$, and AM requires three times as much power (4.77 dB) as that needed for DSB-SC or SSB-SC.

In many communication systems the transmitter is limited by peak power rather than average power transmitted. In such a case, AM fares even worse. It can be shown (see Prob. 12.2-6) that in tone modulation, for a fixed peak power transmitted, the output SNR of AM is 6 dB below that of DSB-SC and 9 dB below that of SSB-SC. These results are valid under conditions most favorable to AM, that is, with modulation index $\mu = 1$. For $\mu < 1$, AM would be worse than this. For this reason, volume compression and peak limiting are generally used in AM transmission in order to have full modulation most of the time.

AM Envelope Detection: Assuming the received signal to be $[A + m(t)] \cos \omega_c t$, the demodulator input is

$$y_i(t) = [A + m(t)] \cos \omega_c t + n(t)$$

Using the quadrature component representation for $n(t)$, we have

$$y_i(t) = [A + m(t) + n_c(t)] \cos \omega_c t + n_s(t) \sin \omega_c t \quad (12.16a)$$

The desired signal at the demodulator input is $[A + m(t)] \cos \omega_c t$. Hence, the signal power S_i

is [see Eq. (11.20c)]

$$S_i = \frac{[A + m(t)]^2}{2} = \frac{A^2 + \overline{m^2}}{2}$$

To find the envelope of $y_i(t)$, we rewrite Eq. (12.16a) in polar form as

$$y_i(t) = E_i(t) \cos [\omega_c t + \Theta_i(t)] \quad (12.16b)$$

where the envelope $E_i(t)$ is [see Eq. (3.40)]

$$E_i(t) = \sqrt{[A + m(t) + n_c(t)]^2 + n_s^2(t)} \quad (12.16c)$$

The envelope detector output is $E_i(t)$ [Eq. (12.16c)]. We shall consider two extreme cases: small noise and large noise.

1. Small-Noise Case: If $[A + m(t)] \gg n(t)$ for almost all t , then $[A + m(t)] \gg n_c(t)$ and $n_s(t)$ for almost all t .^{*} In this case $E_i(t)$ in Eq. (12.16c) can be approximated by $[A + m(t) + n_c(t)]$,

$$E_i(t) \simeq A + m(t) + n_c(t)$$

The dc component A of the envelope detector output E_i is blocked by a capacitor, yielding $m(t)$ as the useful signal and $n_c(t)$ as the noise. Hence,

$$S_o = \overline{m^2}$$

and from Eq. (11.50b),

$$N_o = \overline{n_c^2(t)} = 2\mathcal{N}B$$

and

$$\begin{aligned} \frac{S_o}{N_o} &= \frac{\overline{m^2}}{2\mathcal{N}B} \\ &= \frac{\overline{m^2}}{A^2 + \overline{m^2}} \frac{S_i}{\mathcal{N}B} \\ &= \frac{\overline{m^2}}{A^2 + \overline{m^2}} \gamma \end{aligned} \quad (12.17)$$

which is identical to the result for AM with synchronous demodulation [Eq. (12.14)]. Therefore for AM, when the noise is small compared to the signal, the performance of the envelope detector is identical to that of the synchronous detector.

2. Large-Noise Case: In this case $n(t) \gg [A + m(t)]$. Hence, $n_c(t)$ and $n_s(t) \gg [A + m(t)]$ for almost all t . Under this condition Eq. (12.16c) becomes

^{*} Here we use the term "almost all t " because $n_c(t)$ and $n_s(t)$ are both gaussian (amplitude range $-\infty$ to ∞), and in some instances $n_c(t)$ or $n_s(t)$ or both will exceed $A + m(t)$, no matter how large $A + m(t)$ is. For large signals, however, this occurs only over relatively short time intervals.

$$E_i(t) \simeq \sqrt{n_c^2(t) + n_s^2(t) + 2n_c(t)[A + m(t)]}$$

$$= E_n(t) \sqrt{1 + \frac{2[A + m(t)]}{E_n(t)} \cos \Theta_n(t)}$$

where $E_n(t)$ and $\Theta_n(t)$, the envelope and the phase of the noise $n(t)$, are [see Eqs. (3.40) and Fig. 12.5a]

$$E_n(t) = \sqrt{n_c^2(t) + n_s^2(t)} \quad (12.18a)$$

$$\Theta_n(t) = -\tan^{-1} \left[\frac{n_s(t)}{n_c(t)} \right] \quad (12.18b)$$

$$n_c(t) = E_n(t) \cos \Theta_n(t) \quad (12.18c)$$

$$n_s(t) = E_n(t) \sin \Theta_n(t) \quad (12.18d)$$

Because $E_n(t) \gg A + m(t)$, $E_i(t)$ may be further approximated as

$$E_i(t) \simeq E_n(t) \left[1 + \frac{A + m(t)}{E_n(t)} \cos \Theta_n(t) \right]$$

$$= E_n(t) + [A + m(t)] \cos \Theta_n(t) \quad (12.19)$$

A glance at Eq. (12.19) shows that the output contains no term proportional to $m(t)$. The signal $m(t) \cos \Theta_n(t)$ represents $m(t)$ multiplied by a time-varying function (actually a noise signal) $\cos \Theta_n(t)$ and, hence, is of no use in recovering $m(t)$. In all previous cases, the output signal contained a term of the form $am(t)$, where a was constant. Furthermore, the output noise was additive (even for envelope detection with small noise). In Eq. (12.19), the noise is multiplicative. In this situation the useful signal is badly mutilated. This is the threshold phenomenon, where the signal quality at the output undergoes disproportionately rapid deterioration when the input noise increases beyond a certain level (i.e., when γ drops below a certain value.)

Calculation of the SNR for the intermediate case (transition region) is quite complex.¹ Here we shall state the final results only:

$$\frac{S_o}{N_o} \simeq 0.916 A^2 \overline{m^2} \gamma^2$$

Figure 12.5b shows the plot of S_o/N_o as a function of γ for AM with synchronous detection and AM with envelope detection. The threshold effect is clearly seen from this figure. The threshold occurs when γ is on the order of 10 or less. For a reasonable-quality AM signal, γ should be on the order of 1000 (30 dB), and the threshold is rarely a limiting condition in practical cases.

EXAMPLE 12.2 Find γ_{thresh} , the value of γ at the threshold, in tone-modulated AM with $\mu = 1$ if the onset of the threshold is when $E_n > A$ with probability 0.01, E_n being the noise envelope.

Because $E_n = \sqrt{n_c^2 + n_s^2}$, where both n_c and n_s are gaussian with variance σ_n^2 , according to Eq. (10.54a) E_n has a Rayleigh PDF with variance σ_n^2 , and

$$P(E_n \geq A) = \int_A^\infty \frac{E_n}{\sigma_n^2} e^{-E_n^2/2\sigma_n^2} dE_n$$

$$= e^{-A^2/2\sigma_n^2} = 0.01$$

Hence, at the onset of the threshold,

$$\frac{A^2}{2\sigma_n^2} = 4.605$$

The variance σ_n^2 of the bandpass noise of PSD $\mathcal{N}/2$ and the bandwidth $2B$ is $2\mathcal{N}B$. Hence, at the onset of the threshold,

$$\frac{A^2}{4\mathcal{N}B} = 4.605$$

For tone modulation,

$$m(t) = \mu A \cos(\omega_m t + \Theta)$$

$$= A \cos(\omega_m t + \Theta) \quad \mu = 1$$

and

$$S_i = \frac{A^2 + \overline{m^2}}{2} = \frac{A^2 + 0.5A^2}{2} = \frac{3A^2}{4}$$

Hence,

$$\gamma_{\text{thresh}} = \frac{S_i}{\mathcal{N}B} = \frac{3A^2}{4\mathcal{N}B} = 3(4.605) = 13.8 \quad (\text{or } 12.4 \text{ dB})$$

12.3 ANGLE-MODULATED SYSTEMS

A block diagram of an angle-modulated system is shown in Fig. 12.6. The angle-modulated (or exponentially modulated) carrier $\varphi_{\text{EM}}(t)$ can be written as

$$\varphi_{\text{EM}}(t) = A \cos[\omega_c t + \psi(t)] \quad (12.20a)$$

where

$$\psi(t) = k_p m(t) \quad \text{for PM} \quad (12.20b)$$

$$= k_f \int_{-\infty}^t m(\alpha) d\alpha \quad \text{for FM} \quad (12.20c)$$

and $m(t)$ is the message signal. The channel noise $n(t)$ at the demodulator input is a bandpass noise with PSD $S_n(\omega)$ and bandwidth $2(\Delta f + B)$. The noise $n(t)$ can be expressed in terms of quadrature components as

$$n(t) = n_c(t) \cos \omega_c t + n_s(t) \sin \omega_c t \quad (12.21a)$$

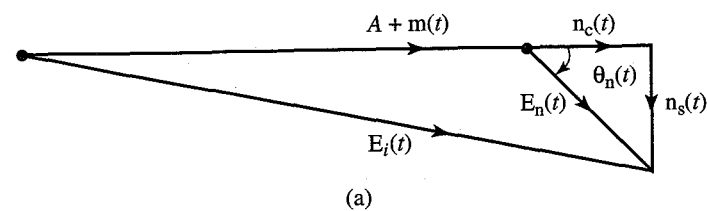
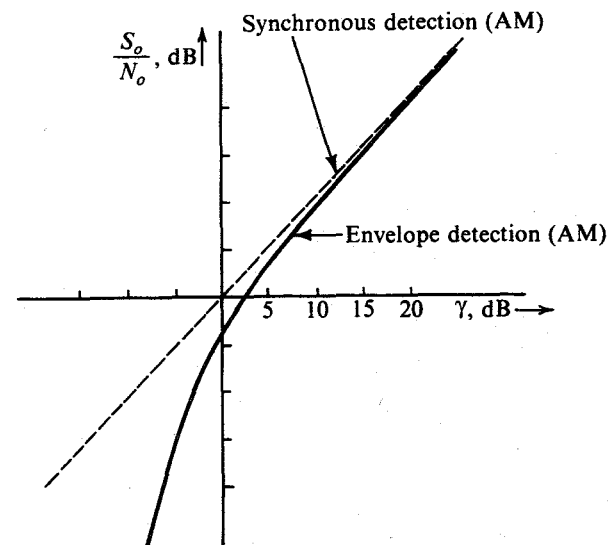


Figure 12.5 Performance of AM (synchronous detection and envelope detection).



(b)

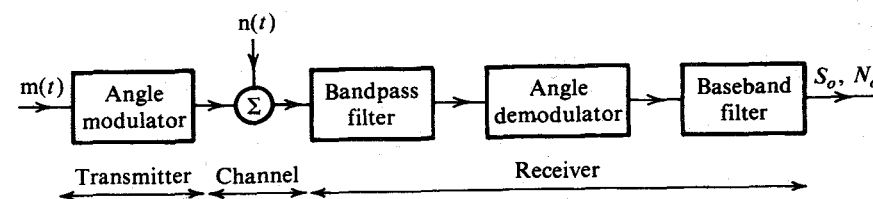


Figure 12.6 Angle-modulated system.

where $n_c(t)$ and $n_s(t)$ are low-pass signals of bandwidth $\Delta f + B$. The bandpass noise $n(t)$ may also be expressed in terms of the envelope $E_n(t)$ and phase $\Theta_n(t)$ as [see Fig. 12.5a and Eqs. (12.18)]

$$n(t) = E_n(t) \cos[\omega_c t + \Theta_n(t)] \quad (12.21b)$$

Angle modulation (and particularly wide-band angle modulation) is a nonlinear type of modulation. Hence, superposition does not apply. In AM, the signal output can be calculated

by assuming the channel noise to be zero, and the noise output can be calculated by assuming the modulating signal to be zero. This is a consequence of linearity. The signal and noise do not form intermodulation components. Unfortunately, exponential modulation is nonlinear, and we cannot use superposition to calculate the output, as can be done in AM. We shall show that because of special circumstances, however, even in angle modulation the noise output can be calculated by assuming the modulating signal to be zero. To prove this we shall first consider the case of PM and then extend those results to FM.

Phase Modulation

Because narrow-band modulation is approximately linear, we need to consider only wide-band modulation. The crux of the argument is that for wide-band modulation, the signal $m(t)$ changes very slowly relative to the noise $n(t)$. The modulating signal bandwidth is B , and the noise bandwidth is $2(\Delta f + B)$, with $\Delta f \gg B$. Hence, the phase and frequency variations of the modulated carrier are much slower than are the variations of $n(t)$. The modulated carrier appears to have constant frequency and phase over several cycles, and, hence, the carrier appears to be unmodulated. We may therefore calculate the output noise by assuming $m(t)$ to be zero (or a constant). This is a qualitative justification for the linearity argument. A quantitative justification is given in the following development.

To calculate the signal and noise powers at the output, we shall first construct a phasor diagram of the signal $y_i(t)$ at the demodulator input, as shown in Fig. 12.7,

$$\begin{aligned} y_i(t) &= A \cos[\omega_c t + \psi(t)] + n(t) \\ &= A \cos[\omega_c t + \psi(t)] + E_n(t) \cos[\omega_c t + \Theta_n(t)] \\ &= R(t) \cos[\omega_c t + \psi(t) + \Delta\psi(t)] \end{aligned} \quad (12.22)$$

where

$$\psi(t) = k_p m(t) \quad \text{for PM} \quad (12.23)$$

For the small-noise case, where $E_n(t) \ll A$ for "almost all t ," $\Delta\psi(t) \ll \pi/2$ for "almost all t ," and

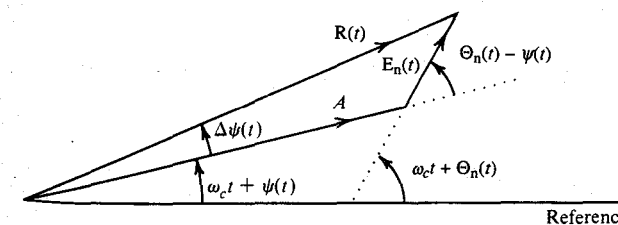
$$\Delta\psi(t) \simeq \frac{E_n(t)}{A} \sin[\Theta_n(t) - \psi(t)] \quad (12.24)$$

The demodulator detects the phase of the input $y_i(t)$. Hence, the demodulator output is

$$y_o(t) = \psi(t) + \Delta\psi(t) \quad (12.25a)$$

$$= k_p m(t) + \frac{E_n(t)}{A} \sin[\Theta_n(t) - \psi(t)] \quad (12.25b)$$

Figure 12.7 Phasor representation of signals in an angle-modulated system.



Note that the noise term $\Delta\psi(t)$ involves the signal $\psi(t)$ due to the nonlinear nature of angle modulation. Because $\psi(t)$ (baseband signal) varies much more slowly than $\Theta_n(t)$ (wide-band noise), we can approximate $\psi(t)$ by a constant ψ ,

$$\begin{aligned}\Delta\psi(t) &\simeq \frac{E_n(t)}{A} \sin[\Theta_n(t) - \psi] \\ &= \frac{E_n(t)}{A} \sin \Theta_n(t) \cos \psi - \frac{E_n(t)}{A} \cos \Theta_n(t) \sin \psi \\ &= \frac{n_s(t)}{A} \cos \psi - \frac{n_c(t)}{A} \sin \psi\end{aligned}$$

Also, because $n_c(t)$ and $n_s(t)$ are incoherent for white noise,

$$\begin{aligned}S_{\Delta\psi}(\omega) &= \frac{\cos^2 \psi}{A^2} S_{n_s}(\omega) + \frac{\sin^2 \psi}{A^2} S_{n_c}(\omega) \\ &= \frac{S_{n_s}(\omega)}{A^2} \quad [\text{because } S_{n_c}(\omega) = S_{n_s}(\omega)]\end{aligned}\quad (12.26)$$

For a white channel noise with PSD $\mathcal{N}/2$ [Eq. (11.49)],

$$S_{\Delta\psi}(\omega) = \begin{cases} \frac{\mathcal{N}}{A^2} & |f| \leq \Delta f + B \\ 0 & \text{otherwise} \end{cases} \quad (12.27a)$$

Note that if we had assumed $\psi(t) = 0$ (zero message signal) in Eq. (12.26), we would have obtained exactly the same result.*

The demodulated noise bandwidth is $\Delta f + B$. But because the useful signal bandwidth is only B , the demodulator output is passed through a low-pass filter of bandwidth B to remove the out-of-band noise. Hence, the PSD of the low-pass filter output noise is

$$S_{n_o}(\omega) = \begin{cases} \frac{\mathcal{N}}{A^2} & |\omega| \leq 2\pi B \\ 0 & |\omega| > 2\pi B \end{cases} \quad (12.27b)$$

and

$$N_o = 2B \left(\frac{\mathcal{N}}{A^2} \right) = \frac{2\mathcal{N}B}{A^2} \quad (12.28a)$$

From Eq. (12.25b) we have

$$S_o = k_p^2 \overline{m^2} \quad (12.28b)$$

Thus,

$$\frac{S_o}{N_o} = (Ak_p)^2 \frac{\overline{m^2}}{2\mathcal{N}B} \quad (12.28c)$$

These results are valid for small noise, and they apply to both WBPM and NBPM. We also have

$$\gamma = \frac{S_i}{\mathcal{N}B} = \frac{A^2/2}{\mathcal{N}B} = \frac{A^2}{2\mathcal{N}B}$$

* This follows from the fact that $E_n(t) \sin \Theta_n(t) = n_s(t)$. Hence, $\Delta\psi(t) = n_s(t)/A$, and $S_{\Delta\psi}(\omega) \simeq S_{n_s}(\omega)/A^2$.

and

$$\frac{S_o}{N_o} = k_p^2 \overline{m^2} \gamma \quad (12.29)$$

Also, for PM [Eq. (5.17a)],

$$\Delta\omega = k_p m'_p \quad m'_p = [\dot{m}(t)]_{\max}$$

Hence,

$$\frac{S_o}{N_o} = (\Delta\omega)^2 \left(\frac{\overline{m^2}}{m_p'^2} \right) \gamma \quad (12.30)$$

Note that the bandwidth of the angle-modulated waveform is about $2\Delta f$ (for the wide-band case). Thus, the output SNR increases with the square of the transmission bandwidth; that is, the output SNR increases by 6 dB for each doubling of the transmission bandwidth. Remember, however, that this result is valid only when the noise power is much smaller than the carrier power. Hence, the output SNR cannot be increased indefinitely by increasing the transmission bandwidth because this also increases the noise power, and at some stage the small-noise assumption is violated. When the noise power becomes comparable to the carrier power, the threshold appears as explained later, and a further increase in bandwidth actually reduces the output SNR instead of increasing it.

Let us apply Eq. (12.30) to tone modulation, where $m(t) = \alpha \cos \omega_m t$. For this case $\overline{m^2} = \alpha^2/2$, and $m'_p = |\dot{m}(t)|_{\max} = \alpha\omega_m$. Hence,

$$\frac{S_o}{N_o} = \frac{1}{2} \left(\frac{\Delta\omega}{\omega_m} \right)^2 \gamma = \frac{1}{2} \left(\frac{\Delta f}{f_m} \right)^2 \gamma \quad (12.31)$$

Note that Eqs. (12.28c), (12.29), (12.30), and (12.31) are valid for both NBPM and WBPM.

Frequency Modulation

Frequency modulation may be considered as a special case of phase modulation, where the modulating signal is $\int_{-\infty}^t m(\alpha) d\alpha$ (Fig. 12.8). At the receiver, we can demodulate FM with a PM demodulator followed by a differentiator, as shown in Fig. 12.8. The PM demodulator output is $k_f \int_{-\infty}^t m(\alpha) d\alpha$. The subsequent differentiator yields the output $k_f m(t)$, so that

$$S_o = k_f^2 \overline{m^2} \quad (12.32)$$

The phase demodulator output noise will be identical to that calculated earlier, with PSD $\mathcal{N}/A^2$ for white channel noise. This noise is passed through an ideal differentiator whose

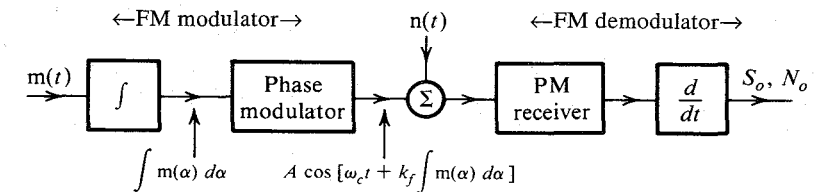


Figure 12.8 FM system as a special case of PM system.

transfer function is $j\omega$. Hence, the PSD $S_{n_o}(\omega)$ of the output noise is $|j\omega|^2$ times the PSD in Eq. (12.27b),

$$S_{n_o}(\omega) = \begin{cases} \frac{\mathcal{N}}{A^2} \omega^2 & |\omega| \leq 2\pi B \\ 0 & |\omega| > 2\pi B \end{cases} \quad (12.33)$$

The PSD of the output noise is parabolic (Fig. 12.9), and the output noise power is

$$\begin{aligned} N_o &= 2 \int_0^B \frac{\mathcal{N}}{A^2} (2\pi f)^2 df \\ &= \frac{8\pi^2 \mathcal{N} B^3}{3A^2} \end{aligned} \quad (12.34)$$

Hence, the output SNR is

$$\begin{aligned} \frac{S_o}{N_o} &= 3 \left(\frac{k_f^2 \overline{m^2}}{(2\pi B)^2} \right) \left(\frac{A^2/2}{\mathcal{N}B} \right) \\ &= 3 \left[\frac{k_f^2 \overline{m^2}}{(2\pi B)^2} \right] \gamma \end{aligned} \quad (12.35)$$

Because $\Delta\omega = k_f m_p$,

$$\frac{S_o}{N_o} = 3 \left(\frac{\Delta f}{B} \right)^2 \left(\frac{\overline{m^2}}{m_p^2} \right) \gamma \quad (12.36)$$

$$= 3\beta^2 \gamma \left(\frac{\overline{m^2}}{m_p^2} \right) \quad (12.37)$$

Recall that the transmission bandwidth is about $2\Delta f$. Hence, for each doubling of the bandwidth, the output SNR increases by 6 dB. Just as in the case of PM, the output SNR does not increase indefinitely because threshold appears as the increased bandwidth makes the channel noise power comparable to the carrier power.

For tone modulation, $\overline{m^2}/m_p^2 = 0.5$ and

$$\frac{S_o}{N_o} = \frac{3}{2} \beta^2 \gamma \quad (12.38)$$

The output SNR S_o/N_o (in decibel) in Eq. (12.38) is plotted in Fig. 12.10 as a function of γ (in decibels) for various values of β . The dotted portion of the curves indicates the threshold region (to be discussed later in this section). Although the curves in Fig. 12.10 are valid for tone modulation only ($\overline{m^2}/m_p^2 = 0.5$), they can be used for any other modulating signal $m(t)$ simply by shifting them vertically by a factor $(\overline{m^2}/m_p^2)/0.5 = 2\overline{m^2}/m_p^2$.

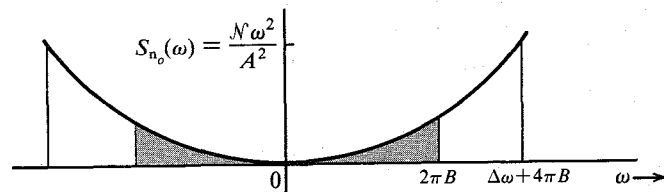


Figure 12.9 PSD of output noise in FM receiver.

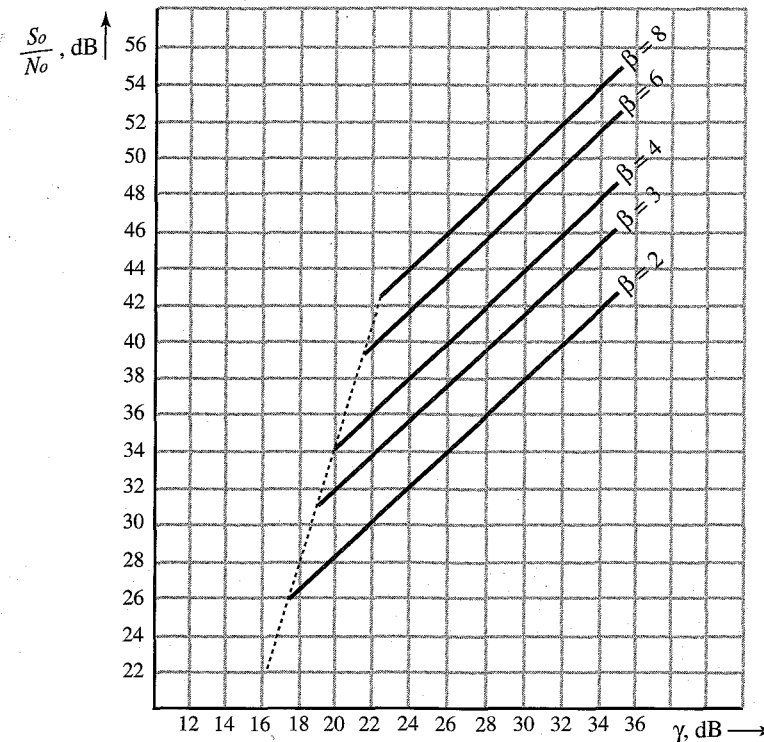


Figure 12.10 Performance of FM system.

From Eqs. (12.31) and (12.38), we observe that for tone modulation FM is superior to PM by a factor of 3. This does not mean that FM is superior to PM for other modulating signals as well. In fact, we shall see that PM is superior to FM for most of the practical signals. From Eqs. (12.30) and (12.36), it can be seen that

$$\frac{(S_o/N_o)_{PM}}{(S_o/N_o)_{FM}} = \frac{(2\pi B)^2 m_p^2}{3m_p'^2} \quad (12.39)$$

Hence, if $(2\pi B)^2 m_p^2 > 3m_p'^2$, PM is superior to FM. If the PSD of $m(t)$ is concentrated at lower frequencies, low-frequency components predominate in $m(t)$, and m_p' is small. This favors PM. Therefore, in general, PM is superior to FM when $S_m(\omega)$ is concentrated at lower frequencies and FM is superior to PM when $S_m(\omega)$ is concentrated at higher frequencies. This explains why FM is superior to PM for tone modulation, where all the signal power is concentrated at the highest frequency in the band. But for most of the practical signals, the signal power is concentrated at lower frequencies, and PM proves superior to FM.

EXAMPLE 12.3 For a zero mean gaussian random process $m(t)$ as the baseband signal, determine the output SNR for FM, assuming white gaussian channel noise.

For a gaussian $m(t)$, $m_p = \infty$. But because the probability that amplitude m lies beyond $3\sigma_m$ ($|m| \geq 3\sigma_m$) is about 0.0027, one may consider* m_p to be $3\sigma_m$. Hence,

* This is known as **3σ loading**. In the literature, 4σ loading [$m_p = 4\sigma_m$] is also used.

$$\overline{m^2} = \sigma_m^2 \quad \text{and} \quad m_p^2 = (3\sigma_m)^2$$

From Eq. (12.37) it follows that

$$\frac{S_o}{N_o} = \frac{1}{3} \beta^2 \gamma$$

Narrow-Band Modulation: The equations derived thus far are valid for both narrow-band and wide-band modulation. We observed in Chapter 5 that narrow-band exponential modulation (NBEM) is approximately linear and is very similar to AM. In fact, the output SNRs for NBEM and AM are similar. To see this, consider the cases of NBPM [Eq. (5.10)] and AM [Eq. (4.8a)],

$$\begin{aligned} \varphi_{AM}(t) &= A \cos \omega_c t + m(t) \cos \omega_c t \\ \varphi_{NBPM}(t) &= A \cos \omega_c t - A k_p m(t) \sin \omega_c t \\ &= A \cos \omega_c t - m_1(t) \sin \omega_c t \end{aligned}$$

where $m_1(t) = A k_p m(t)$. Both φ_{AM} and φ_{NBPM} contain a carrier and a DSB term. In φ_{NBPM} the carrier and the DSB component are out of phase by $\pi/2$ rad, whereas in φ_{AM} they are in phase. But the $\pi/2$ -rad phase difference has no effect on the power. Thus, $m(t)$ in φ_{AM} is analogous to $m_1(t)$ in φ_{NBPM} .

Now let us compare the output SNRs for AM and NBPM. For AM [Eq. (12.17)],

$$\left(\frac{S_o}{N_o} \right)_{AM} = \frac{\overline{m^2}}{A^2 + \overline{m^2}} \gamma$$

whereas for NBPM with $m_1(t) = A k_p m(t)$, Eq. (12.29) can be expressed as

$$\begin{aligned} \left(\frac{S_o}{N_o} \right)_{PM} &= k_p^2 \overline{m^2} \gamma \\ &= \frac{\overline{m_1^2}}{A^2} \gamma \end{aligned}$$

Note that for NBPM, we require that $|k_p m(t)| \ll 1$, that is, $m_1(t)/A \ll 1$. Hence,

$$A^2 \simeq A^2 + \overline{m_1^2}$$

and

$$\left(\frac{S_o}{N_o} \right)_{PM} \simeq \frac{\overline{m_1^2}}{A^2 + \overline{m_1^2}} \gamma$$

which is of the same form as $(S_o/N_o)_{AM}$. Hence, NBPM is very similar to AM. Under the best possible conditions, however, AM outperforms NBPM because for AM, we need only to satisfy the conditions $[A + m(t)] > 0$ [Eq. (4.9a)], which implies $[m(t)]_{\max} \leq A$. Thus for tone modulation, the modulation index for AM can be nearly equal to unity. For NBPM, however, the narrow-band condition would be equivalent to requiring $\mu \ll 1$. Hence, although AM and NBPM have identical performance for a given value of μ , AM has the edge over NBPM from the SNR viewpoint.

It is interesting to look for the line (in terms of Δf) that separates narrow-band and wide-band FM. We may consider the dividing line to be that value of Δf for which the output SNR for FM given in Eq. (12.37) is equal to the maximum output SNR for AM. The maximum SNR for AM occurs when $\mu = 1$, or when $A = m_p$. Hence, equating Eq. (12.37) with Eq. (12.17) [with $A = m_p$],

$$3\beta^2 \gamma \left(\frac{\overline{m^2}}{m_p^2} \right) = \frac{\overline{m^2}}{m_p^2 + \overline{m^2}} \gamma$$

or

$$\beta^2 = \frac{1}{3} \left[\frac{1}{1 + (\overline{m^2}/m_p^2)} \right] \quad (12.40)$$

Because $\overline{m^2}/m_p^2 < 1$ for practical signals, and

$$\beta^2 < \frac{1}{3}$$

or

$$\beta < 0.6 \quad (12.41a)$$

This gives

$$\Delta f = 0.6B \quad (12.41b)$$

Mean Square Bandwidth and Estimation of Angle-Modulation Bandwidth

The bandwidth of a signal is a measure of the width of the signal spectrum. Several definitions of bandwidth have appeared in the literature. All are meaningful and useful in different situations. A 3-dB bandwidth is commonly used in electronic circuits.

We now introduce one such useful measure of bandwidth. In Chapter 10, we noted that the standard deviation σ is a good measure of the width of a PDF. We can extend this idea to the PSD by normalizing the spectrum to a unit area. The variance of such a normalized spectrum is known as the **mean square bandwidth** of the spectrum. The mean square bandwidth is meaningful, extremely useful, and mathematically tractable for angle-modulated signals. A normalized PSD, by definition, has unit area. Hence, to normalize a PSD, we divide it by its area so that the resulting function has a unit area. For convenience, we shall use the variable f (in hertz) rather than the variable ω (in radians per second). This will give us the mean square bandwidth in hertz rather than in radians per second.

Using this concept, for a baseband signal $m(t)$, the normalized PSD is $S_m(\omega)/\int_{-\infty}^{\infty} S_m(\omega) d\omega$. Because the normalized PSD is symmetrical about the vertical axis ($f = 0$), it has a zero mean (in the sense of the PDF), and its variance B_m^2 is*

$$\overline{B_m^2} = \frac{\int_{-\infty}^{\infty} f^2 S_m(2\pi f) df}{\int_{-\infty}^{\infty} S_m(2\pi f) df} \text{ Hz} \quad (12.42a)$$

$$= \frac{1}{\overline{m^2}} \int_{-\infty}^{\infty} f^2 S_m(2\pi f) df \text{ Hz} \quad (12.42b)$$

* For $\overline{B_m^2}$ to exist, $S_m(\omega)$ must approach zero at a rate faster than $1/\omega^2$ for large values of ω .

EXAMPLE 12.4 For a low-pass signal with PSD $S_m(\omega) = \text{rect}(\omega/4\pi B)$, show that $\overline{B_m^2} = B^2/3$.

Because $S_m(\omega) = 1$ for $|f| < B$, and 0 for $|f| > B$, we have [Eqs. (12.42)]

$$\overline{B_m^2} = \frac{\int_{-B}^B f^2 df}{\int_{-B}^B df} = \frac{B^2}{3}$$

EXAMPLE 12.5 For a gaussian PSD $S_m(\omega) = ke^{-\omega^2/2\sigma^2}$ show that

$$\overline{B_m^2} = \left(\frac{\sigma}{2\pi}\right)^2$$

From Eqs. (12.42),

$$\overline{B_m^2} = \frac{k \int_{-\infty}^{\infty} f^2 e^{-4\pi^2 f^2/2\sigma^2} df}{k \int_{-\infty}^{\infty} e^{-4\pi^2 f^2/2\sigma^2} df} = \frac{\sigma^2}{4\pi^2}$$

We shall now investigate the dependence of the FM wave PSD on instantaneous frequency. In wideband modulation instantaneous frequency f_i in the range $(f, f + df)$ gives rise to power spectral components (components of the PSD) in the range $(f, f + df)$. The power contribution of these components is proportional to the relative time that f_i remains in this range. If $p_{fi}(f)$ is the PDF of instantaneous frequency f_i , then $p_{fi}(f) df$ is the probability of observing f_i in the range $(f, f + df)$. This probability is proportional to the time that f_i remains in the range $(f, f + df)$ and, hence, is proportional to the power contributed by spectral components in the range $(f, f + df)$. If $S_{FM}(\omega)$ is the PSD of an FM wave, then the preceding argument implies*

$$S_{FM}(2\pi f) = kp_{fi}(f) \quad f > 0 \quad (12.43)$$

where k is a constant of proportionality.†

To find the mean square bandwidth of $S_{FM}(2\pi f)$, we observe from Eq. (12.43) that $p_{fi}(f)$ is precisely $S_{FM}(2\pi f)$ normalized to unit area. Hence, the mean square bandwidth of $S_{FM}(2\pi f)$ is the variance σ_{fi}^2 of f_i , because

$$\begin{aligned} f_i &= \frac{\omega_i}{2\pi} = \frac{1}{2\pi}[\omega_c + k_f m(t)] \\ \overline{f_i} &= \frac{1}{2\pi}[\omega_c + k_f \overline{m(t)}] \\ &= f_c \quad [\text{because } \overline{m(t)} = 0] \end{aligned}$$

* Because $p_{fi}(f) = 0$ for $f < 0$, $S_{FM}(\omega)$ in Eq. (12.43) is the unilateral PSD. This means $S_{FM}(\omega) = 0$ for $\omega < 0$ and is twice the bilateral PSD for $\omega > 0$.

† By integrating both sides of Eq. (12.43) over $(0, \infty)$, it can be shown that $k = A^2/2$.

and

$$\begin{aligned} \sigma_{fi}^2 &= \overline{(f_i - f_c)^2} \\ &= \frac{1}{4\pi^2} \overline{[k_f m(t)]^2} \\ &= \frac{1}{4\pi^2} k_f^2 \overline{m^2} \end{aligned}$$

In other words,

$$\overline{B_{FM}^2} = \frac{1}{4\pi^2} k_f^2 \overline{m^2} \quad (12.44)$$

For phase modulation,

$$f_i = \frac{1}{2\pi} \omega_i = \frac{1}{2\pi} [\omega_c + k_p \dot{m}(t)]$$

and

$$\begin{aligned} \sigma_{fi}^2 &= \frac{1}{(2\pi)^2} \overline{(f_i - f_c)^2} \\ &= \frac{1}{4\pi^2} k_p^2 \overline{[\dot{m}(t)]^2} \end{aligned}$$

Because the PSD of $\dot{m}(t)$ is $\omega^2 S_m(\omega) = 4\pi^2 f^2 S_m(2\pi f)$,

$$\sigma_{fi}^2 = \frac{1}{4\pi^2} k_p^2 \int_{-\infty}^{\infty} 4\pi^2 f^2 S_m(2\pi f) df$$

Using Eq. (12.42),

$$\sigma_{fi}^2 = k_p^2 \overline{m^2} \overline{B_m^2}$$

that is,

$$\overline{B_{PM}^2} = k_p^2 \overline{m^2} \overline{B_m^2} \quad (12.45)$$

From Eqs. (12.44) and (12.45) we observe that the bandwidth of the FM wave is independent of the modulating signal spectrum, whereas the bandwidth of the PM wave is strongly influenced by the modulating signal spectrum.

The output SNRs in Eqs. (12.29) and (12.35) can now be expressed in terms of mean square bandwidths. For PM,

$$\frac{S_o}{N_o} = \frac{\overline{B_{PM}^2}}{\overline{B_m^2}} \gamma \quad (12.46)$$

and for FM,

$$\frac{S_o}{N_o} = 3 \frac{\overline{B_{FM}^2}}{\overline{B_m^2}} \gamma \quad (12.47)$$

Quantities $\overline{B_{PM}^2}$ and $\overline{B_{FM}^2}$ are the variances of the normalized PSDs of the PM and FM waves, respectively. As seen earlier, the actual transmission bandwidth will be several times the

standard deviation $\sqrt{B_{PM}^2}$ or $\sqrt{B_{FM}^2}$. For example, when the modulated signal PSD has a gaussian form,* 99.74% of the total power resides within 3σ of the carrier frequency. Hence, the bandwidth in this case may be taken as 6σ , or $6\sqrt{B_{PM}^2}$ for PM and $6\sqrt{B_{FM}^2}$ for FM. From Eqs. (12.46) and (12.47), it follows that in PM as well as in FM the output SNR improves by 6 dB for each doubling of the transmission bandwidth. It should be stressed that these results are valid only for small noise.

Which Is Superior: PM or FM? From Eqs. (12.46) and (12.47), we have

$$\frac{(S_o/N_o)_{PM}}{(S_o/N_o)_{FM}} = \left(\frac{B^2}{3B_m^2} \right) \left(\frac{B_{PM}^2}{B_{FM}^2} \right) \quad (12.48)$$

It will be instructive to compare the performance of PM and FM for the same transmission bandwidth (the same mean square bandwidth), that is, for $B_{PM}^2 = B_{FM}^2$. This gives

$$\frac{(S_o/N_o)_{PM}}{(S_o/N_o)_{FM}} = \frac{B^2}{3B_m^2} \quad (12.49a)$$

Thus, if

$$B^2 > 3B_m^2 \quad (12.49b)$$

PM is superior to FM. Otherwise, FM is superior to PM. We showed in Example 12.4 that when the PSD of $m(t)$ is constant (over a band B Hz), $3B_m^2 = B^2$. Hence, PM and FM perform equally well in that case. If the spectrum falls off with frequency, as it does for all real signals, B_m^2 is less than $B^2/3$, and PM is superior to FM. If, on the other hand, the spectrum is weighted heavily at higher frequencies, B_m^2 is greater than $B^2/3$, and FM is superior to PM. This is exactly what happens for tone modulation, where the spectrum is concentrated at the highest frequency B , with no power at lower frequencies, making $B_m^2 = B^2$. This is why for tone modulation FM is superior to PM by a factor of 3. Tone modulation proves to be grossly misleading in the SNR analysis of angle modulation. For practical signals, including audio, the spectrum falls off with frequency, and PM is superior to FM. Actually, so-called FM broadcast is not pure FM but is FM modified by preemphasis-deemphasis, as discussed in Sec. 5.5.

We have derived two different criteria [Eqs. (12.39) and (12.49a)] for comparing PM and FM performance. In Eq. (12.39), the output SNRs are compared for a given transmission bandwidth, whereas in Eq. (12.49a), the output SNRs are compared for a given mean square transmission bandwidth. Consequently, we may get slightly different answers by using these two criteria (see Probs. 12.3-4 and 12.3-8). Although the criterion in Eq. (12.39) is preferable to that in Eq. (12.49a), it is impossible in practice to determine the parameters required in Eq. (12.39), and the only way to compare may be through Eq. (12.49a).

EXAMPLE 12.6

If a baseband signal $m(t)$ has a gaussian PSD, show that PM is superior to FM by a factor of 3 (4.77 dB) when the bandwidth B is taken as 3σ , where σ is the standard deviation of the normalized PSD of $m(t)$.

* It can be shown² that for wide-band angle modulation, the PSD of the modulated carrier is gaussian when the modulating random process is gaussian.

$S_m(t)$ is gaussian of the form $ke^{-\omega^2/2\sigma^2}$. The radian bandwidth $W = 2\pi B = 3\sigma$, and $B = 3\sigma/2\pi$. Also, as seen in Example 12.5,

$$\overline{B_m^2} = \frac{\sigma^2}{4\pi^2} \quad \text{and} \quad B = \frac{3\sigma}{2\pi}$$

Hence, from Eqs. (12.49), it follows that PM is superior to FM by a factor of 3.

Threshold in Angle Modulation

When the noise power at the demodulator input is comparable to the carrier power, the threshold phenomenon (discussed earlier for AM systems) appears. In FM this effect is much more pronounced than in AM. Let us discuss qualitatively how the threshold effect appears. We refer back to Fig. 12.7 and observe that the phasor E_n rotates from the terminal of the phasor A . When $E_n \ll A$, the angle $\Delta\psi(t)$ is quite small, and because $\Theta_n(t)$ is random with uniform distribution in the range $(0, 2\pi)$, $\Delta\psi(t)$ assumes positive as well as negative values (Fig. 12.11a), which are usually much smaller than 2π . When E_n is large (on the order of A or greater), however, the resultant phasor is much more likely to rotate around the origin, and $\Delta\psi$ is more likely to go through changes of 2π (Fig. 12.11c) in a relatively short time, because the noise varies much faster than the modulating signal. The noise at the FM demodulator output is given by $\Delta\dot{\psi}(t)$.* This is shown in Fig. 12.11 for large and small noise. For large noise, we observe the appearance of spikes (of area 2π), which give rise to a crackling sound. When the noise is small ($E_n \ll A$), the PSD of the output noise $\Delta\dot{\psi}(t)$ is parabolic, and most of its power is in the frequencies greater than B and is therefore filtered out by the baseband filter at the output. For the large-noise case, on the other hand, we have the presence of spikes, which are like

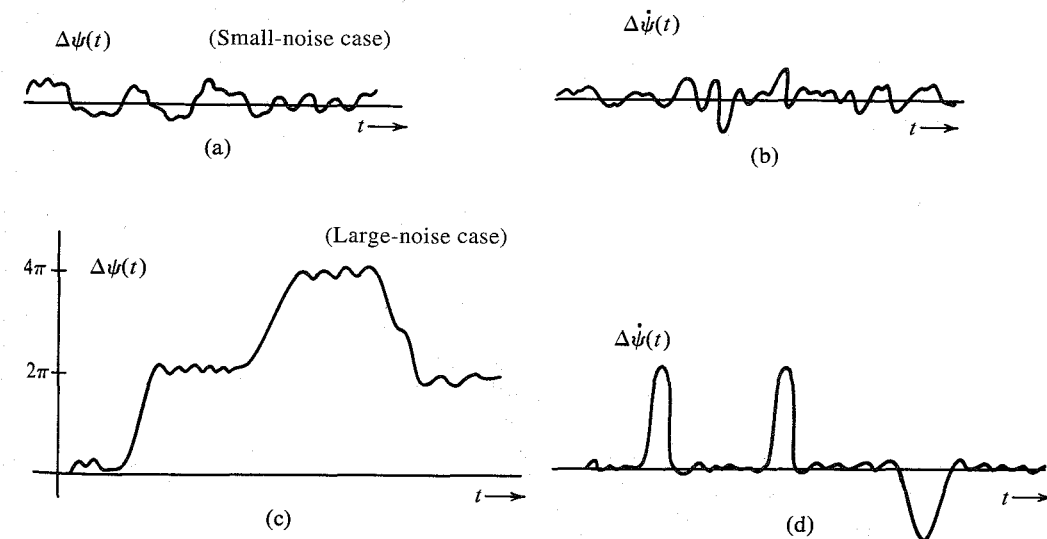


Figure 12.11 Nature of output noise in FM receiver for small and large channel noise.

* Recall that FM demodulation requires differentiation of the angle because the instantaneous frequency is proportional to the derivative of the angle.

impulses. Consequently, they have considerable power at lower frequencies. Hence, a spike will contribute much more noise at the output. For this reason, when E_n approaches the order of A , the output noise starts increasing disproportionately (Fig. 12.12). This is precisely the phenomenon of threshold.

It has been shown³ that the noise power caused by the spikes N_s , is

$$N_s = \frac{8\pi^2 B_{FM} B}{\sqrt{3}} Q \left(\sqrt{\frac{2B}{B_{FM}}} \gamma \right) \quad (12.50a)$$

and the total noise power N_T , is the sum of N_o in Eq. (12.34) and N_s in Eq. (12.50a). The output SNR is

$$\begin{aligned} \frac{S_o}{N_o + N_s} &= \frac{3\beta^2 \gamma (\overline{m^2}/m_p^2)}{1 + (2\sqrt{3} B_{FM}/B) \gamma Q[\sqrt{(2B/B_{FM})} \gamma]} \\ &= \frac{3\beta^2 \gamma (\overline{m^2}/m_p^2)}{1 + 4\sqrt{3}(\beta + 1) \gamma Q[\sqrt{\gamma/(\beta + 1)}]} \end{aligned} \quad (12.50b)$$

The onset of the threshold is when the carrier power is 10 times the channel noise power. The carrier power is $A^2/2$, and for white noise with a PSD $\mathcal{N}/2$, the noise power is $\mathcal{N}B_{FM}$, where B_{FM} , the bandwidth of an FM carrier, is

$$B_{FM} = 2(\Delta f + B) = 2B(\beta + 1)$$

Hence, the threshold occurs when

$$2\mathcal{N}B(\beta + 1) = \frac{1}{10} \frac{A^2}{2}$$

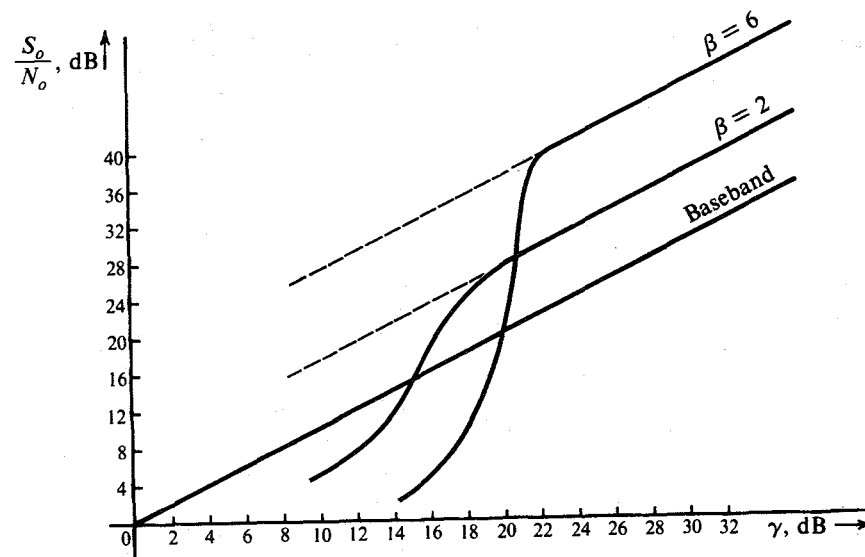


Figure 12.12 Threshold phenomenon in FM.

or

$$20(\beta + 1) = \frac{A^2}{2\mathcal{N}B}$$

By definition, the right-hand side of this equation is γ at threshold. Thus, γ_{thresh} , the value of γ at the onset of threshold, is

$$\gamma_{\text{thresh}} = 20(\beta + 1) \quad (12.51)$$

EXAMPLE 12.7 A gaussian $m(t)$ with 4σ loading (that is, $m_p = 4\sigma_m$) frequency-modulates a carrier using $\beta = 4$. The output SNR is found to be 20.5 dB. Determine whether the system is in threshold.

For $\beta = 4$, γ_{thresh} is

$$\gamma_{\text{thresh}} = 20(\beta + 1) = 20(5) = 100$$

For this value of γ [Eq. (12.37)],

$$\left(\frac{S_o}{N_o} \right) = 3(16)(100) \left(\frac{1}{16} \right) = 300$$

Because 20.5 dB is a ratio of 160, the SNR is below 300, and the system is in threshold.

Threshold Extension in Angle Modulation

The problem of threshold is rather serious in angle modulation. The ability to communicate even at low power levels simply by increasing the transmission bandwidth is the very raison d'être of angle-modulated systems, and threshold deprives angle modulation of its very essence. For this reason, much attention has been paid to the problem of pushing the threshold level further down and extending the useful operating range. The two principal methods used for this purpose utilize frequency compression feedback (FCF) and phase-lock loop (PLL). We shall look briefly into these techniques.

Frequency Compression Feedback (FCF): The basic circuit is shown in Fig. 12.13. Let us consider a frequency-modulated wave $A \cos[\omega_c t + \psi(t)]$ at the input. The quiescent frequency of the VCO is $\omega_c - \omega_o$. The instantaneous frequency ω_i of the VCO is given by

$$\omega_i = (\omega_c - \omega_o) + \alpha e(t)$$

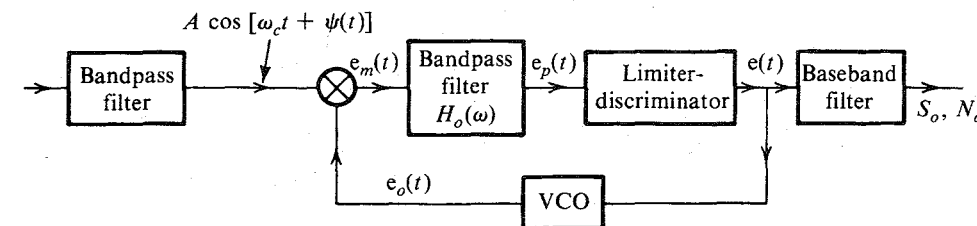


Figure 12.13 Frequency compression feedback (FCF).

and the VCO output $e_o(t)$ is given by* (assuming initial instant $t = 0$),

$$e_o(t) = 2 \cos \left[(\omega_c - \omega_o)t + \alpha \int_0^t e(\tau) d\tau \right] \quad (12.52)$$

The bandpass filter $H_o(\omega)$ is centered at ω_o . The multiplier output $e_m(t)$ has a spectrum centered at the sum and difference frequencies. The sum frequencies are centered at $2\omega_c - \omega_o$ and are suppressed by the bandpass filter $H_o(\omega)$, which allows only the difference frequencies centered at ω_o to pass. The output $e_p(t)$ of the bandpass filter is

$$e_p(t) = A \cos \left[\omega_o t + \psi(t) - \alpha \int_0^t e(\tau) d\tau \right] \quad (12.53)$$

The limiter-discriminator frequency demodulates the signal $e_p(t)$ by differentiating its angle. Therefore, the output $e(t)$ is given by

$$\begin{aligned} e(t) &= \frac{d}{dt} \left[\psi(t) - \alpha \int_0^t e(\tau) d\tau \right] \\ &= \dot{\psi}(t) - \alpha \dot{e}(t) \end{aligned}$$

Hence,

$$e(t) = \frac{\dot{\psi}(t)}{1 + \alpha}$$

Substitution of this value of $e(t)$ into Eq. (12.53) yields

$$e_p(t) = A \cos \left[\omega_o t + \frac{\psi(t)}{1 + \alpha} \right] \quad (12.54)$$

This interesting result shows that the signal $e_p(t)$ is another frequency-modulated signal with carrier frequency ω_o and is similar to the incoming carrier $A \cos [\omega_c t + \psi(t)]$. The difference is that the angle is $\psi(t)/(1 + \alpha)$ instead of $\psi(t)$. This implies a reduction in the frequency deviation and, consequently, a reduction in the bandwidth of the modulated signal by a factor† of $(1 + \alpha)$. Hence, for a wide-band case, the bandwidth of $H_o(\omega)$ need only be about $2\Delta f/(1 + \alpha)$. The second conclusion is that when $e_p(t)$ is applied to a limiter-discriminator, the output $e(t)$ is $\dot{\psi}(t)/(1 + \alpha)$. Hence, the FCF demodulator indeed frequency-demodulates the incoming frequency-modulated carrier.

Let us now see what happens when the input signal is a frequency-modulated carrier plus bandpass channel noise. The signal plus noise can be expressed as $R(t) \cos [\omega_c t + \psi(t) + \Delta\psi(t)]$ [Eq. (12.22)]. The amplitude variations $R(t)$ are eventually eliminated in the limiter-discriminator (see Sec. 5.4), and

$$e_p(t) = A \cos \left[\omega_c t + \frac{\psi(t) + \Delta\psi(t)}{1 + \alpha} \right] \quad (12.55)$$

Similarly,

$$e(t) = \frac{1}{1 + \alpha} [\dot{\psi}(t) + \Delta\dot{\psi}(t)]$$

* The amplitude of $e_o(t)$ is immaterial in our discussion. For convenience, it is considered to be 2.

† In order to avoid distortion of the modulating signal, the bandwidth $(\Delta f + B)/(1 + \alpha)$ must be greater than B .

where $\dot{\psi}$ is the useful signal and $\Delta\dot{\psi}$ is the noise. The output of the FCF demodulator is identical to that of the conventional demodulator except for the multiplicative factor $1/(1 + \alpha)$. Hence, the output SNR of the FCF demodulator is identical to that of the conventional demodulator. The advantage is gained, however, in extending the threshold. This can be seen from the fact that the FM wave passes unmolested through the filter $H_o(\omega)$ [Eq. (12.55)]. The channel noise, however, which had hitherto a bandwidth of about $2\Delta f$, has to pass through $H_o(\omega)$ having $1/(1 + \alpha)$ times its former bandwidth. Hence, the carrier-power-to-noise ratio at the input of the limiter-discriminator is enhanced by a factor of $1 + \alpha$. As seen earlier, this does not affect the SNR, but it does extend the threshold because of the relative enhancement of the carrier power in relation to noise power. It can be seen that the carrier power can now afford to be reduced by a factor of $1 + \alpha$ before the onset of the threshold. In other words, the threshold is extended roughly by $10 \log \alpha$ dB. Because of practical problems, however, this benefit is not completely realized. In practical FCF demodulators, threshold extension of about 5 to 7 dB can be realized.

Phase-Locked Loop (PLL): The functioning of the PLL was discussed in Chapters 4 and 5. For the small-noise case, the PLL performance is identical to that of a conventional demodulator (just as in FCF). For large noise, however, the PLL extends the threshold just as the FCF does, by reducing the filter bandwidth. The detailed analysis is beyond our scope.⁴ A practical PLL extends the threshold region by about 3 to 6 dB.

12.4 PULSE-MODULATED SYSTEMS

Among pulse-modulated systems (PAM, PWM, PPM, and PCM), only PCM is of practical importance. The other systems are rarely used in practice. For this reason we shall discuss only PCM in detail. It can be shown that the performance of PAM is similar to that of AM-SC systems (i.e., the output SNR is equal to γ). The PWM and PPM systems are capable of exchanging the transmission bandwidth with output SNR, as in angle-modulated systems. In PWM, the output SNR is proportional to the transmission bandwidth B_T . This performance is clearly inferior to that of angle-modulated systems, where the output SNR increases as B_T^2 . In PPM systems under optimum conditions, the output SNR increases as B_T^2 but is still inferior to FM by a factor of 6. For in-depth treatment of PAM, PWM, and PPM, the reader is referred to Panter¹ or Rowe.⁵

Another PM system that deserves mention is the delta-modulation (DM) system discussed in Chapter 6. For speech signals, this system's performance is comparable to that of PCM for a bandwidth expansion ratio B_T/B of 7 to 8. For $B_T/B > 8$, PCM is superior to DM, and for $B_T/B < 8$, DM is superior to PCM.

Pulse-Code Modulation

In PCM, a baseband signal $m(t)$ band-limited to B Hz and with amplitudes in the range of $-m_p$ to m_p is sampled at a rate of $2B$ samples per second. The sample amplitudes are quantized into L levels, each separated by $2m_p/L$. Each quantized sample is encoded into n binary digits ($2^n = L$). The binary signal is transmitted over a channel. The receiver detects the binary signal and reconstructs quantized samples (decoding). The quantized samples are then passed through a low-pass filter to obtain the desired signal $m(t)$.

There are two sources of error in PCM: (1) quantization or "rounding off" error, and (2) detection error. The latter is caused by error in the detection of the binary signal at the receiver.

As usual, $m(t)$ is assumed to be a wide-sense stationary random process. The random variable $m(kT_s)$, formed by sample-function amplitudes at $t = kT_s$, will be denoted by m_k . The k th sample m_k is rounded off, or quantized, to a value $\hat{m}_k$, which is encoded and transmitted as binary digits. Because of the channel noise, some of the digits may be detected erroneously at the receiver, and the reconstructed sample will be $\tilde{m}_k$ instead of $\hat{m}_k$. If q_k and ϵ_k are the quantization and detection errors, respectively, then

$$\begin{aligned} q_k &= m_k - \hat{m}_k \\ \epsilon_k &= \hat{m}_k - \tilde{m}_k \end{aligned} \quad (12.56)$$

and

$$m_k - \tilde{m}_k = q_k + \epsilon_k \quad (12.57)$$

Hence, the total error $m_k - \tilde{m}_k$ at the receiver is $q_k + \epsilon_k$. The receiver reconstructs the signal $\tilde{m}(t)$ from samples $\tilde{m}_k$ according to the interpolation formula in Eq. (6.10),

$$\begin{aligned} \tilde{m}(t) &= \sum_k \tilde{m}_k \text{sinc}(2\pi Bt - k\pi) \\ &= \sum_k [m_k - (q_k + \epsilon_k)] \text{sinc}(2\pi Bt - k\pi) \\ &= \sum_k m_k \text{sinc}(2\pi Bt - k\pi) - \sum_k (q_k + \epsilon_k) \text{sinc}(2\pi Bt - k\pi) \\ &= m(t) - e(t) \end{aligned} \quad (12.58a)$$

where

$$e(t) = \sum_k (q_k + \epsilon_k) \text{sinc}(2\pi Bt - k\pi) \quad (12.58b)$$

The receiver therefore receives the signal $m(t) - e(t)$ instead of $m(t)$. The error signal $e(t)$ is a random process with the k th sample $q_k + \epsilon_k$. Because the process is wide-sense stationary, the mean square value of the process is the same as the mean square value at any instant. Because $q_k + \epsilon_k$ is the value of $e(t)$ at $t = kT_s$,

$$\overline{e^2(t)} = \overline{(q_k + \epsilon_k)^2}$$

Because q_k and ϵ_k are independent with zero mean RVs (see Examples 10.17, 10.18, and 10.19),

$$\overline{e^2(t)} = \overline{q_k^2} + \overline{\epsilon_k^2}$$

We have already derived $\overline{q_k^2}$ and $\overline{\epsilon_k^2}$ in Examples 10.17 and 10.18 [Eqs. (10.64b) and (10.66b)]. Hence,

$$\overline{e^2(t)} = \frac{1}{3} \left(\frac{m_p}{L} \right)^2 + \frac{4m_p^2 P_e (L^2 - 1)}{3L^2} \quad (12.59a)$$

where P_e is the detection error probability. For binary coding, each sample is encoded into n binary digits. Hence, $2^n = L$, and

$$\overline{e^2(t)} = \frac{m_p^2}{3(2^{2n})} [1 + 4P_e(2^{2n} - 1)] \quad (12.59b)$$

As seen from Eq. (12.58a), the output $m(t) - e(t)$ contains the signal $m(t)$ and noise $e(t)$. Hence,

$$S_o = \overline{m^2} \quad N_o = \overline{e^2(t)}$$

and

$$\frac{S_o}{N_o} = \frac{3(2^{2n})}{1 + 4P_e(2^{2n} - 1)} \left(\frac{\overline{m^2}}{m_p^2} \right) \quad (12.60)$$

The error probability P_e depends on A_p , the peak pulse amplitude, and the channel noise power σ_n^2 [Eq. (10.41)],*

$$P_e = Q \left(\frac{A_p}{\sigma_n} \right)$$

It will be shown in Sec. 13.1 that A_p/σ_n can be maximized (that is, P_e can be minimized) by passing the incoming digital signal through an optimum filter (known as the **matched filter**). It will be shown that for polar signaling [Eqs. (13.9c) and (13.20a)],

$$\left(\frac{A_p}{\sigma_n} \right)_{\max} = \sqrt{\frac{2E_p}{\mathcal{N}}}$$

and

$$(P_e)_{\min} = Q \left(\sqrt{\frac{2E_p}{\mathcal{N}}} \right) \quad (12.61)$$

where E_p is the energy of the received binary pulse and the channel noise is assumed to be white with PSD $\mathcal{N}/2$. Because there are n binary pulses per sample and there are $2B$ samples per second, there are a total of $2Bn$ pulses per second. Hence, the received signal power $S_i = 2BnE_p$, and

$$\begin{aligned} P_e &= Q \left(\sqrt{\frac{S_i}{n\mathcal{N}B}} \right) \\ &= Q \left(\sqrt{\frac{\gamma}{n}} \right) \end{aligned} \quad (12.62)$$

and

$$\frac{S_o}{N_o} = \frac{3(2^{2n})}{1 + 4(2^{2n} - 1)Q(\sqrt{\gamma/n})} \left(\frac{\overline{m^2}}{m_p^2} \right) \quad (12.63)$$

* This assumes polar signaling. Bipolar (or pseudoternary) signaling requires about 3 dB more power than polar to achieve the same P_e (see Sec. 7.2). In practice, bipolar rather than polar signaling is used for PCM.

Figure 12.14 shows a plot of the output SNR as a function of γ for tone modulation ($\overline{m^2}/m_p^2 = 0.5$). It can also be used for a general case, with $\overline{m^2}/m_p^2 = \lambda$, where the SNR is $\lambda/0.5 = 2\lambda$ times that in Fig. 12.14. Hence, we simply shift the curves in Fig. 12.14 upward by $10 \log(2\lambda)$ dB.

Figure 12.14 shows two interesting features: the threshold and the saturation. When γ is too small, a large pulse-detection error results, and the decoded pulse sequence yields a sample value that has no relation to the actual sample transmitted. The received signal is thus meaningless, and we have the phenomenon of threshold. To explain saturation, we observe that when γ is sufficiently large (implying sufficiently large pulse amplitude), the detection error $P_e \rightarrow 0$, and Eq. (12.60) becomes

$$\frac{S_o}{N_o} = 3L^2 \left(\frac{\overline{m^2}}{m_p^2} \right) \quad (12.64a)$$

$$= 3(2^{2n}) \left(\frac{\overline{m^2}}{m_p^2} \right) \quad (12.64b)$$

The SNR in this case is practically independent of γ . Because the detection error approaches zero, the output noise now consists entirely of the quantization noise, which depends only on

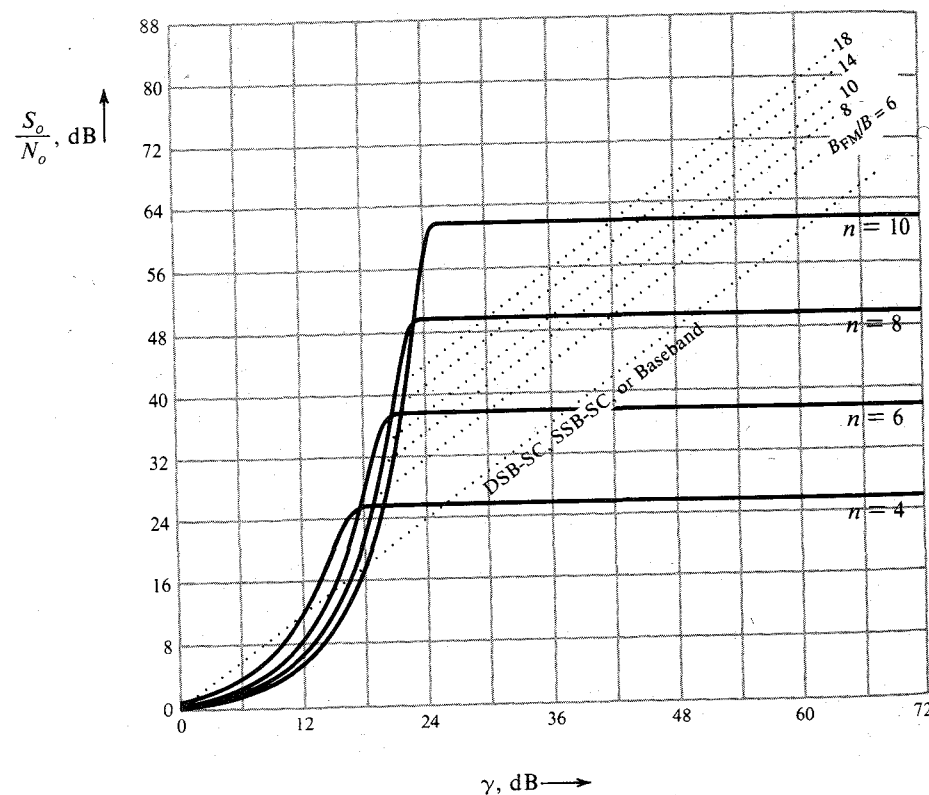


Figure 12.14 Performance of PCM.

L . Because the pulse amplitude is large enough so that there is very little probability of making a detection error, a further increase in γ by increasing the pulse amplitude buys no advantage, and we have the saturation effect.

In the saturation region,

$$\left(\frac{S_o}{N_o} \right)_{\text{dB}} = 10 \left[\log 3 + 2n \log 2 + \log \left(\frac{\overline{m^2}}{m_p^2} \right) \right] = \alpha + 6n \quad (12.64c)$$

where $\alpha = 4.77 + 10 \log_{10}(\overline{m^2}/m_p^2)$.

EXAMPLE 12.8 For PCM with $n = 8$, determine the output SNR for a gaussian $m(t)$. Assume the saturation region of operation.

For a gaussian signal, $m_p = \infty$. In practice, however, we may clip amplitudes $> 3\sigma_m$ or $4\sigma_m$, depending on the accuracy desired. For example, in the case of 3σ loading,

$$P(|m| > 3\sigma_m) = 2Q(3) = 0.0026$$

and for 4σ loading,

$$P(|m| > 4\sigma_m) = 2Q(4) = 6 \times 10^{-5}$$

If we take the case of 3σ loading,

$$\frac{\overline{m^2}}{m_p^2} = \frac{\sigma_m^2}{(3\sigma_m)^2} = \frac{1}{9}$$

and

$$\frac{S_o}{N_o} = 3(2)^{16} \left(\frac{1}{9} \right) = 21,845 = 43.4 \text{ dB}$$

For 4σ loading,

$$\frac{S_o}{N_o} = 3(2)^{16} \left(\frac{1}{16} \right) = 12,288 = 40.9 \text{ dB}$$

To facilitate the comparison of PCM with other types of modulation, the SNRs of FM and DSB-SC are superimposed on the SNR of PCM in Fig. 12.14. The theoretical bandwidth expansion ratio for PCM is $B_{\text{PCM}}/B = n$. In practice, this can be achieved by using duobinary signaling. Today's PCM systems use bipolar signaling, however, requiring $B_{\text{PCM}}/B = 2n$. Moreover, P_e in Eq. (12.62) is valid only for polar signaling. Bipolar signaling requires twice as much power. Hence, the plot in Fig. 12.14 is valid for bipolar signaling if 3 dB is added to each value of γ .

From Eq. (12.64b) we have

$$\frac{S_o}{N_o} = 3 \left(\frac{\overline{m^2}}{m_p^2} \right) 2^{2B_{\text{PCM}}/kB} \quad (12.65)$$

where $1 \leq k \leq 2$. For duobinary $k = 1$, and for bipolar $k = 2$.

It is clear from Eq. (12.65) that in PCM, the output SNR increases exponentially with the transmission bandwidth. Compare this with angle modulation, where the SNR increases as a square of the transmission bandwidth. In angle modulation, doubling the transmission bandwidth quadruples the output SNR. From Eq. (12.64c) we see that in PCM, increasing n by 1 quadruples the SNR. But increasing n by 1 increases the bandwidth only by the fraction $1/n$. For $n = 8$, a mere 12.5% increase in the transmission bandwidth quadruples the SNR. Therefore in PCM, the exchange of SNR for bandwidth is much more efficient than in angle modulation. This is particularly evident for large values of n , as can be seen from Fig. 12.14. For smaller values of n , the difference between FM and PCM is not as impressive as for large values of n .*

Actually, to do full justice to PCM, we must consider the use of regenerative repeaters, which cannot be used for angle modulation. The results for PCM derived thus far apply for a single link. If k regenerative repeaters are used along the path, the noise power over each link is reduced by the factor k , and hence P_e , the error probability over each link, is

$$P_e = Q\left(\sqrt{\frac{k\gamma}{n}}\right)$$

But because k links are now in tandem, the overall error probability is kP_e (see Example 10.7), and

$$\frac{S_o}{N_o} = \frac{3(2^{2n})}{1 + 4(2^{2n} - 1)kQ(\sqrt{k\gamma/n})} \left(\frac{\overline{m^2}}{m_p^2}\right) \quad (12.66)$$

To maintain a given S_o/N_o , the value of γ decreases as the number of repeaters k , increases. Figure 12.15 shows the γ vs. k needed to maintain $S_o/N_o = 49.5$ dB for $n = 8$.

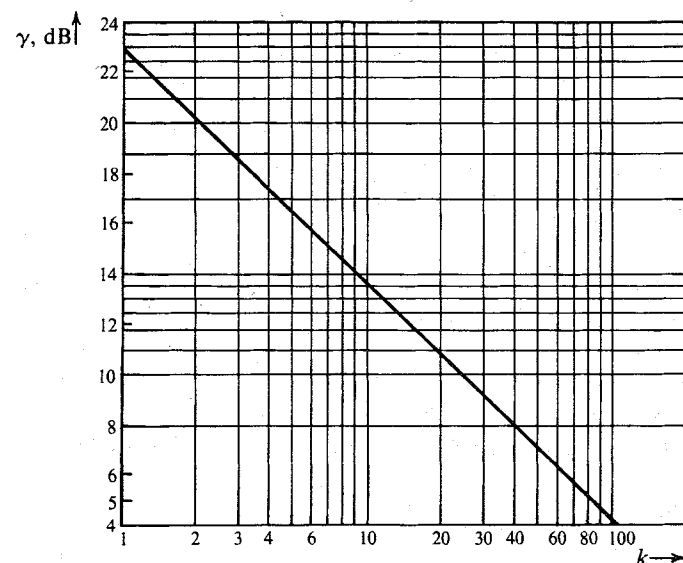


Figure 12.15 Power reduction as a function of the number of repeaters in PCM.

* The FM plots in Fig. 12.14 are without preemphasis and deemphasis.

Companded PCM

The output SNR of PCM is proportional to $\overline{m^2}/m_p^2$, where $\overline{m^2}$ is the power of the baseband signal $m(t)$ and m_p is the peak value of $m(t)$. It may appear that $\overline{m^2}/m_p^2$ will remain more or less constant regardless of the speech level, because $\overline{m^2}$ is proportional to m_p^2 . Unfortunately, m_p^2 is a constant of the quantizer with a quantization range $(-m_p, m_p)$. Once a quantizer is designed, m_p is fixed, and $\overline{m^2}/m_p^2$ is proportional to the speech signal power $\overline{m^2}$ only. This can vary from talker to talker (or even for the same talker) by as much as 40 dB, causing the output SNR to vary widely. This problem can be mitigated, and a relatively constant SNR over a large dynamic range of $\overline{m^2}$ can be obtained, either by nonuniform quantization or by signal companding. Both methods are equivalent, but the latter is simpler to implement. In this method, the signal amplitudes are nonlinearly compressed.

Figure 6.12 shows the input-output characteristics of the two most commonly used compressors (the μ -law and the A-law). For convenience, let us denote

$$x = \frac{m}{m_p}$$

Clearly, the peak value of x is 1 (when $m = m_p$). Moreover, the peak value of the compressor output y is also 1 (occurring when $m = m_p$). Thus, x and y are the normalized input and output of the compressor, each with unit peak value (Fig. 12.16). The input-output characteristics have an odd symmetry about $x = 0$. For convenience, we have only shown the region $x \geq 0$. The output signal samples in the range $(-1, 1)$ are uniformly quantized into L levels, with a quantization interval of $2/L$. Figure 12.16 shows the j th quantization interval for the output y as well as the input x . All input sample amplitudes that lie in the range Δ_j are mapped into y_j . For the input sample value x in the range Δ_j , the quantization error is $q = (x - x_j)$, and

$$2 \int_{x_j - (\Delta_j/2)}^{x_j + (\Delta_j/2)} (x - x_j)^2 p_x(x) dx$$

is the part of $\overline{q^2}$ (the mean square quantizing error) contributed by x in the region Δ_j . The factor 2 appears because there is an equal contribution from negative amplitudes of x centered at $-x_j$. Thus,

$$\overline{q^2} = 2 \sum_j \int_{x_j - (\Delta_j/2)}^{x_j + (\Delta_j/2)} (x - x_j)^2 p_x(x) dx$$

Because $L \gg 1$, the quantizing interval $(2/L)$ and Δ_j are very small, and $p_x(x)$ can be assumed to be constant over each interval. Hence,

$$\begin{aligned} \overline{q^2} &= 2 \sum_j p_x(x_j) \int_{x_j - (\Delta_j/2)}^{x_j + (\Delta_j/2)} (x - x_j)^2 dx \\ &= 2 \sum_j \frac{p_x(x_j) \Delta_j^3}{12} \end{aligned} \quad (12.67)$$

Because $2/L$ and Δ_j are very small, the compression characteristics can be assumed to be linear over each Δ_j , and

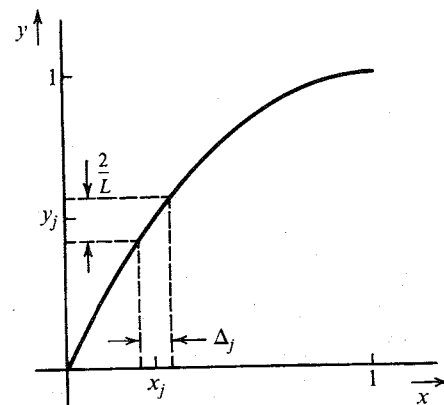


Figure 12.16 Input-output characteristic of a PCM compressor.

$$\dot{y}(x_j) \simeq \frac{2/L}{\Delta_j}$$

Substituting this in Eq. (12.67), we have

$$\overline{q^2} \simeq \frac{2}{3L^2} \sum_j \frac{p_x(x_j)}{[\dot{y}(x_j)]^2} \Delta_j$$

For L large enough, the preceding sum can be approximated by an integral

$$\overline{q^2} \simeq \frac{2}{3L^2} \int_0^1 \frac{p_x(x)}{[\dot{y}(x)]^2} dx \quad (12.68)$$

For the μ -law [Eq. (6.17a)],

$$y = \frac{\ln(1 + \mu x)}{\ln(1 + \mu)} \quad 0 \leq x \leq 1$$

and

$$\dot{y}(x) = \frac{\mu}{\ln(1 + \mu)} \left(\frac{1}{1 + \mu x} \right)$$

and

$$\overline{q^2} = \left(\frac{2}{3L^2} \right) \left[\frac{\ln(1 + \mu)}{\mu} \right]^2 \int_0^1 (1 + \mu x)^2 p_x(x) dx \quad (12.69)$$

If $p_x(x)$ is symmetrical about $x = 0$,

$$\sigma_x^2 = 2 \int_0^1 x^2 p_x(x) dx \quad (12.70a)$$

and $\overline{|x|}$, the mean of the rectified x , is

$$\overline{|x|} = 2 \int_0^1 x p_x(x) dx \quad (12.70b)$$

We can express $\overline{q^2}$ as

$$\overline{q^2} = \left[\frac{\ln(1 + \mu)}{\mu} \right]^2 \left[\frac{1 + \mu^2 \sigma_x^2 + 2\mu \overline{|x|}}{3L^2} \right] \quad (12.71a)$$

$$= \frac{[\ln(1 + \mu)]^2}{3L^2} \left(\sigma_x^2 + \frac{2\overline{|x|}}{\mu} + \frac{1}{\mu^2} \right) \quad (12.71b)$$

Recall that $\overline{q^2}$ in Eqs. (12.71) is the normalized quantization error. The actual error is $m_p^2 \overline{q^2}$. The normalized output signal is $x(t)$, and, hence, the normalized output power $S_o = \sigma_x^2 = \overline{m^2}/m_p^2$. The actual S_o will be $m_p^2 \sigma_x^2$. Hence,

$$\frac{S_o}{N_o} = \frac{\sigma_x^2}{\overline{q^2}} = \frac{3L^2}{[\ln(1 + \mu)]^2} \frac{\sigma_x^2}{(\sigma_x^2 + 2\overline{|x|}/\mu + 1/\mu^2)} \quad (12.72a)$$

$$= \frac{3L^2}{[\ln(1 + \mu)]^2} \frac{1}{(1 + 2\overline{|x|}/\mu \sigma_x^2 + 1/\mu^2 \sigma_x^2)} \quad (12.72b)$$

To get an idea of the relative importance of the various terms in the parentheses in Eq. (12.72b), we note that x is an RV distributed in the range $(-1, 1)$. Hence, σ_x^2 and $\overline{|x|}$ are both less than 1, and $\overline{|x|}/\sigma_x$ is typically in the range of 0.7 to 0.9. The values of μ used in practice are greater than 100. For example, the D2 channel bank used in conjunction with the T1 carrier system has $\mu = 255$. Thus, the second and third terms in the parentheses in Eq. (12.72b) are small compared to 1 if σ_x^2 is not too small, and so

$$\frac{S_o}{N_o} \simeq \frac{3L^2}{[\ln(1 + \mu)]^2} \quad (12.72c)$$

which is independent of σ_x^2 . The exact expression in Eq. (12.72b) has a weak dependence on σ_x^2 over a broad range of σ_x . Note that the SNR in Eq. (12.72b) also depends on the signal statistics $\overline{|x|}$ and σ_x^2 . But for most of the practical PDFs, $\overline{|x|}/\sigma_x$ is practically the same (in the range of 0.7 to 0.9). Hence, S_o/N_o depends only on $\overline{q^2}$. This means the plot of S_o/N_o vs. σ_x^2 will be practically independent of the PDF of x . Figure 12.17 shows the plot of S_o/N_o vs. σ_x^2 for two different PDFs; Laplacian and gaussian (see the next example). It can be seen that there is hardly any difference between the two curves.

Because $x = m/m_p$, $\sigma_x^2 = \sigma_m^2/m_p^2$, and $\overline{|x|} = \overline{|m|}/m_p$, Eq. (12.72a) becomes

$$\frac{S_o}{N_o} = \frac{3L^2}{[\ln(1 + \mu)]^2} \left[\frac{\sigma_m^2/m_p^2}{\sigma_m^2/m_p^2 + 2\overline{|m|}/\mu m_p + 1/\mu^2} \right] \quad (12.73)$$

One should be careful in interpreting m_p in Eq. (12.73). Once the system is designed for some $m(t)$, m_p is fixed. Hence, m_p is a constant of the system, not of the signal $m(t)$ that may be subsequently transmitted.

EXAMPLE 12.9 A voice signal amplitude PDF can be closely modeled by the Laplace density*

* A better but more complex model for speech signal amplitude m is the gamma density⁶

$$p_m(m) = \sqrt{\frac{k}{4\pi|m|}} e^{-k|m|}$$

$$p_m(m) = \frac{1}{\sigma_m \sqrt{2}} e^{-\sqrt{2}|m|/\sigma_m}$$

For a voice PCM system with $n = 8$ and $\mu = 255$, find and sketch the output SNR as a function of the normalized voice power σ_m^2/m_p^2 .

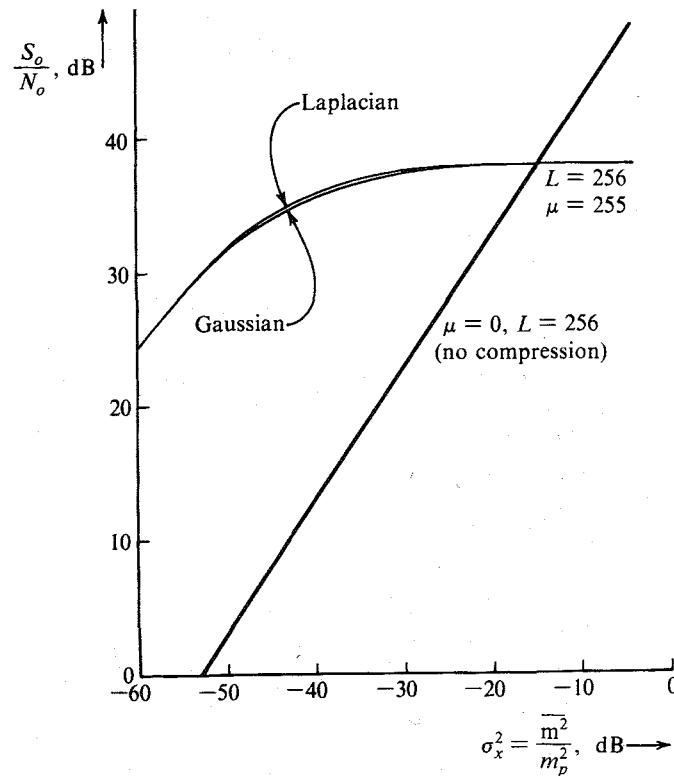


Figure 12.17 PCM performance with and without companding.

It is straightforward to show that the variance of this Laplace PDF is σ_m^2 . In practice, the speech amplitude will be limited by either 3σ or 4σ loading. In either case, the probability of observing m beyond this limit will be negligible, and in computing $\overline{|m|}$ (etc.), we may use the limits 0 to ∞ ,

$$\overline{|m|} = 2 \int_0^\infty \frac{m}{\sigma_m \sqrt{2}} e^{-\sqrt{2}m/\sigma_m} dm = 0.707\sigma_m$$

Hence, from Eq. (12.73),

$$\frac{S_o}{N_o} = \frac{6394(\sigma_m^2/m_p^2)}{(\sigma_m^2/m_p^2) + 0.00555(\sigma_m/m_p) + 1.53 \times 10^{-5}} \quad (12.74)$$

This is plotted as a function of (σ_m^2/m_p^2) in Fig. 12.17.

EXAMPLE 12.10 Repeat Example 12.9 for the gaussian $m(t)$.

In this case,

$$\overline{|m|} = 2 \int_0^\infty \frac{m}{\sigma_m \sqrt{2\pi}} e^{-m^2/2\sigma_m^2} dm = 0.798\sigma_m$$

and

$$\frac{S_o}{N_o} = \frac{6394(\sigma_m^2/m_p^2)}{(\sigma_m^2/m_p^2) + 0.0063(\sigma_m/m_p) + 1.53 \times 10^{-5}} \quad (12.75)$$

The SNR here is nearly the same as that in Eq. (12.74). The plot of SNR vs. σ_m^2/m_p^2 (Fig. 12.17) is practically indistinguishable from that in Example 12.9.

12.5 OPTIMUM PREEMPHASIS-DEEMPHASIS SYSTEMS

It is possible to increase the output SNR by deliberate distortion of the transmitted signal (preemphasis) and the corresponding compensation (deemphasis) at the receiver. For intuitive understanding of this process, consider a case of white channel noise and a signal $m(t)$ whose PSD decreases with frequency. In this case, we can boost the high-frequency components of $m(t)$ at the transmitter (preemphasis). Because the signal has relatively less power at high frequencies, this preemphasis will require only a small increase in transmitted power.* At the receiver, the high-frequency components are attenuated (or deemphasized) in order to undo the preemphasis at the transmitter. This will restore the useful signal to its original form. It is an entirely different story with the channel noise. Because the noise is added after the transmitter, it does not undergo preemphasis. At the receiver, however, it does undergo deemphasis (i.e., attenuation of high-frequency components). Thus, at the receiver output, the signal power is restored but the noise power is reduced. The output SNR is therefore increased.

We shall first consider a baseband system and then extend the discussion to modulated systems. A baseband system with a preemphasis filter $H_p(\omega)$ at the transmitter and the corresponding complementary deemphasis filter $H_d(\omega)$ at the receiver is shown in Fig. 12.18. The channel transfer function is $H_c(\omega)$, and the PSD of the input signal $m(t)$ is $S_m(\omega)$. We shall determine the optimum preemphasis-deemphasis (PDE) filters $H_p(\omega)$ and $H_d(\omega)$ required for distortionless transmission of the signal $m(t)$.

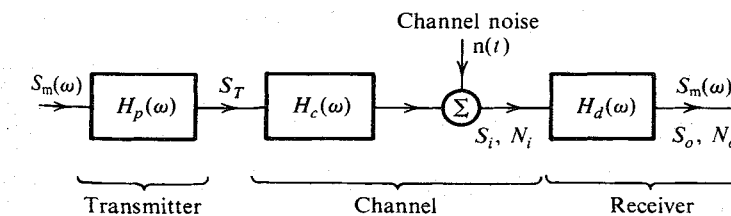


Figure 12.18 Optimum PDE in a baseband system.

* Actually, the transmitted power is maintained constant by attenuating the preemphasized signal slightly.

For distortionless transmission,

$$|H_p(\omega)H_c(\omega)H_d(\omega)| = G \quad (\text{a constant}) \quad (12.76a)$$

and

$$\theta_p(\omega) + \theta_c(\omega) + \theta_d(\omega) = -\omega t_d \quad (12.76b)$$

We want to maximize the output SNR, S_o/N_o , for a given transmitted power S_T .

Referring to Fig. 12.18, we have

$$S_T = \int_{-\infty}^{\infty} S_m(\omega) |H_p(\omega)|^2 df \quad (12.77a)$$

Because $|H_p(\omega)H_c(\omega)H_d(\omega)| = G$, the signal power S_o at the receiver output is

$$S_o = G^2 \int_{-\infty}^{\infty} S_m(\omega) df \quad (12.77b)$$

The noise power N_o at the receiver output is

$$N_o = \int_{-\infty}^{\infty} S_n(\omega) |H_d(\omega)|^2 df \quad (12.77c)$$

Thus,

$$\frac{S_o}{N_o} = \frac{G^2 \int_{-\infty}^{\infty} S_m(\omega) df}{\int_{-\infty}^{\infty} S_n(\omega) |H_d(\omega)|^2 df} \quad (12.78)$$

We wish to maximize this ratio subject to the condition in Eq. (12.77a) with S_T as a given constant. We can include this constraint by multiplying the numerator and the denominator of the right-hand side of Eq. (12.78) by the left-hand side and the right-hand side, respectively, of Eq. (12.77a). This gives

$$\frac{S_o}{N_o} = \frac{G^2 S_T \int_{-\infty}^{\infty} S_m(\omega) df}{\int_{-\infty}^{\infty} S_n(\omega) |H_d(\omega)|^2 df \int_{-\infty}^{\infty} S_m(\omega) |H_p(\omega)|^2 df} \quad (12.79)$$

The numerator of the right-hand side of Eq. (12.79) is fixed. Hence, to maximize S_o/N_o , we need only minimize the denominator of the right-hand side of Eq. (12.79). To do this, we use the Schwarz inequality (Appendix B),

$$\begin{aligned} \int_{-\infty}^{\infty} S_m(\omega) |H_p(\omega)|^2 df \int_{-\infty}^{\infty} S_n(\omega) |H_d(\omega)|^2 df \\ \geq \left| \int_{-\infty}^{\infty} [S_m(\omega)S_n(\omega)]^{1/2} |H_p(\omega)H_d(\omega)| df \right|^2 \end{aligned} \quad (12.80)$$

The equality holds only if

$$S_m(\omega) |H_p(\omega)|^2 = K^2 S_n(\omega) |H_d(\omega)|^2 \quad (12.81)$$

where K is an arbitrary constant. Thus to maximize S_o/N_o , Eq. (12.81) must be satisfied. Substitution of Eq. (12.76a) into Eq. (12.81) yields

$$|H_p(\omega)|_{\text{opt}}^2 = GK \frac{\sqrt{S_n(\omega)/S_m(\omega)}}{|H_c(\omega)|} \quad (12.82a)$$

$$|H_d(\omega)|_{\text{opt}}^2 = \frac{G \sqrt{S_m(\omega)/S_n(\omega)}}{K |H_c(\omega)|} \quad (12.82b)$$

The constant K is found by substituting Eq. (12.82a) into Eq. (12.77a) as

$$K = \frac{S_T}{G \int_{-\infty}^{\infty} [\sqrt{S_m(\omega)S_n(\omega)}/|H_c(\omega)|] df} \quad (12.82c)$$

Substitution of this value of K into Eqs. (12.82a, b) yields

$$|H_p(\omega)|_{\text{opt}}^2 = \frac{S_T \sqrt{S_n(\omega)/S_m(\omega)}}{|H_c(\omega)| \int_{-\infty}^{\infty} [\sqrt{S_m(\omega)S_n(\omega)}/|H_c(\omega)|] df} \quad (12.83a)$$

$$|H_d(\omega)|_{\text{opt}}^2 = \frac{G^2 \int_{-\infty}^{\infty} [\sqrt{S_m(\omega)S_n(\omega)}/|H_c(\omega)|] df}{S_T |H_c(\omega)| \sqrt{S_n(\omega)/S_m(\omega)}} \quad (12.83b)$$

The output SNR under optimum conditions is given by Eq. (12.79) with its denominator replaced with the right-hand side of Eq. (12.80). Now substituting $|H_p(\omega)H_d(\omega)| = G/|H_c(\omega)|$ yields

$$\left(\frac{S_o}{N_o} \right)_{\text{opt}} = \frac{S_T \int_{-\infty}^{\infty} S_m(\omega) df}{\left(\int_{-\infty}^{\infty} [\sqrt{S_m(\omega)S_n(\omega)}/|H_c(\omega)|] df \right)^2} \quad (12.83c)$$

Equations (12.83a) and (12.83b) give the magnitudes of the optimum filters $H_p(\omega)$ and $H_d(\omega)$. The phase functions must be chosen to satisfy the condition of distortionless transmission [Eq. (12.76b)].

Observe that the preemphasis filter in Eq. (12.82a) boosts frequency components where the signal is weak and suppresses frequency components where the signal is strong. The deemphasis filter in Eq. (12.82b) does exactly the opposite. Thus, the signal is unchanged but the noise is reduced.

EXAMPLE 12.11 Consider the case with $\alpha = 1400\pi$,

$$S_m(\omega) = \begin{cases} \frac{C}{\omega^2 + \alpha^2} & |\omega| \leq 8000\pi \\ 0 & |\omega| \geq 8000\pi \end{cases} \quad (12.84a)$$

The channel noise is white with PSD

$$S_n(\omega) = \frac{\mathcal{N}}{2} \quad (12.84b)$$

The channel is assumed to be ideal [$H_c(\omega) = 1$ and $G = 1$] over the band of interest (0 to 4000 Hz).

Without preemphasis-deemphasis, we have

$$\begin{aligned} S_o &= \frac{1}{2\pi} \int_{-8000\pi}^{8000\pi} S_m(\omega) d\omega \\ &= \frac{1}{\pi} \int_0^{8000\pi} \frac{C}{\omega^2 + \alpha^2} d\omega \quad \alpha = 1400\pi \\ &= 10^{-4} C \end{aligned}$$

Also, because $G = 1$, the transmitted power $S_T = S_o$,

$$S_o = S_T = 10^{-4}C$$

and the noise power without preemphasis-deemphasis is

$$N_o = \mathcal{N}B = 4000\mathcal{N}$$

Therefore,

$$\frac{S_o}{N_o} = 2.5 \times 10^{-8} \frac{C}{\mathcal{N}} \quad (12.85)$$

The optimum transmitting and receiving filters are given by [Eqs. (12.83a and b)]

$$|H_p(\omega)|^2 = \frac{10^{-4}\sqrt{\omega^2 + \alpha^2}}{\int_{-\infty}^{\infty} \left(1/\sqrt{\omega^2 + \alpha^2}\right) df} = \frac{1.286\sqrt{\omega^2 + \alpha^2}}{10^4} \quad |\omega| \leq 8000\pi \quad (12.86a)$$

$$|H_d(\omega)|^2 = \frac{10^4 \int_{-\infty}^{\infty} \left(1/\sqrt{\omega^2 + \alpha^2}\right) df}{\sqrt{\omega^2 + \alpha^2}} = \frac{0.778 \times 10^4}{\sqrt{\omega^2 + \alpha^2}} \quad |\omega| \leq 8000\pi \quad (12.86b)$$

The output SNR using optimum preemphasis and deemphasis is found from Eq. (12.83c) as

$$\begin{aligned} \left(\frac{S_o}{N_o}\right)_{\text{opt}} &= \frac{(10^{-4}C)^2}{(\mathcal{N}C/2) \left[\int_{-4000}^{4000} \left[1/\sqrt{4\pi^2 f^2 + (1400\pi)^2}\right] df \right]^2} \\ &= 3.3 \times 10^{-8} \frac{C}{\mathcal{N}} \end{aligned} \quad (12.87)$$

Comparison of Eq. (12.85) with Eq. (12.87) shows that preemphasis-deemphasis has increased the output SNR by a factor of 1.32.

Optimum Preemphasis-Deemphasis in AM Systems

Because the channel noise in a DSB system is in the band $\omega_c \pm 2\pi B$, the PDE should be carried out in this band. This means PDE filters $H_p(\omega)$ and $H_d(\omega)$ are bandpass filters located as shown in Fig. 12.19. The optimization is localized to the subsystem shown in the dashed box. This subsystem is identical to that in Fig. 12.18. Hence, $H_p(\omega)$ and $H_d(\omega)$ can be obtained from Eqs. (12.83a, b). However, because the signal PSD at the input of the subsystem in Fig. 12.19 is $\frac{1}{2}[S_m(\omega + \omega_c) + S_m(\omega - \omega_c)]$, we should use this PSD in place of $S_m(\omega)$ in Eqs. (12.83a, b). The same argument applies to SSB and VSB systems. In these cases, $H_p(\omega)$ and $H_d(\omega)$ can be computed from Eqs. (12.83a, b) by replacing $S_m(\omega)$ with the appropriate SSB or VSB PSD.

Optimum Preemphasis-Deemphasis in Angle Modulation

We shall consider the case of PDE in the baseband, that is, preemphasis before modulation and deemphasis after demodulation (Fig. 12.20a). The channel is assumed to be ideal, that is, $H_c(\omega) = 1$.

The problem here is very different from the baseband communication case. In the baseband case, the preemphasis filter changes the transmitted power but not its bandwidth,

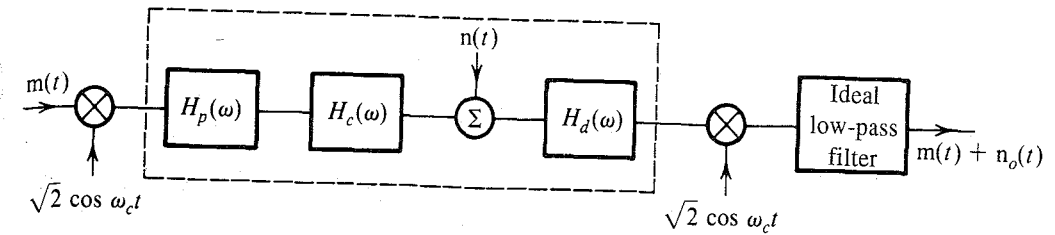


Figure 12.19 Optimum PDE in an amplitude-modulated system.

and optimization was performed under the constraint of a fixed transmitted power (because bandwidth remained fixed, it was of no concern). In FM, on the other hand, the transmitted power is always fixed ($A^2/2$), but the transmission bandwidth is modified by the preemphasis filter. But optimization must be performed under the constraint of a fixed transmission bandwidth. Which bandwidth should be used? The conventional ($2\Delta f$) or the mean square? A look at Eq. (12.36) shows that the SNR in terms of conventional bandwidth involves the constant m_p . When $m(t)$ is passed through $H_p(\omega)$, m_p will change, and no simple relationship exists between $m(t)$, $H_p(\omega)$, and the new m_p . The problem is mathematically intractable. No such difficulty appears when we use Eq. (12.47), which uses the mean square bandwidth. Hence, we shall optimize the ratio S_o/N_o with the constraint of a fixed mean square bandwidth of transmission. Fortunately, this constraint turns out to be similar to that used in the baseband system. This can be seen from Eq. (12.44):

$$\overline{B_{\text{FM}}^2} = \frac{1}{4\pi^2} k_f^2 \overline{m^2}$$

Thus to maintain $\overline{B_{\text{FM}}^2}$ fixed, $\overline{m^2}$, the power of the modulating signal, must be the same with or without preemphasis. This is exactly the constraint of the baseband system [Eq. (12.77a)]. If the input to the system in the dotted box (Fig. 12.20a) is $x(t)$, its output is $k_f x(t)$ plus the parabolic noise $n'(t)$ with the PSD in Eq. (12.33). Hence, the system in Fig. 12.20b is the equivalent of the system in Fig. 12.20a. Our problem is the optimization of the output SNR with the constraint that the output of $H_p(\omega)$ has a given mean square value. This problem is

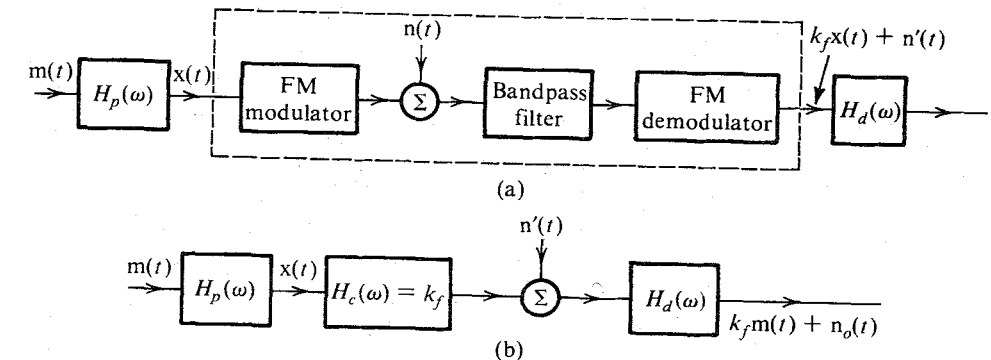


Figure 12.20 (a) Optimum PDE in FM system. (b) Equivalent of system in part (a).

identical to the optimization of the baseband system with $G = H_c(\omega) = k_f$. Hence, from Eqs. (12.83) with $S_n(\omega) = \mathcal{N}\omega^2/A^2$ and $S_T = \overline{m^2}$, we have

$$|H_p(\omega)|_{\text{opt}}^2 = \frac{\overline{m^2}\omega}{\sqrt{S_m(\omega)} \int_{-\infty}^{\infty} \sqrt{\omega'^2 S_m(\omega')} df} \quad (12.88a)$$

$$|H_d(\omega)|_{\text{opt}}^2 = \frac{\sqrt{S_m(\omega)} \int_{-\infty}^{\infty} \sqrt{\omega'^2 S_m(\omega')} df}{\overline{m^2}\omega} \quad (12.88b)$$

$$\left(\frac{S_o}{N_o}\right)_{\text{opt}} = \frac{\left(\frac{A^2 k_f^2}{4\pi^2 \mathcal{N}}\right) \overline{m^2} \int_0^B S_m(\omega) df}{2 \left[\int_0^B f \sqrt{S_m(\omega)} df \right]^2} \quad (12.88c)$$

Because $\gamma = A^2/2\mathcal{N}B$,

$$\left(\frac{S_o}{N_o}\right)_{\text{opt}} = \left[\frac{B^3 \overline{m^2}}{6 \left[\int_0^B f \sqrt{S_m(\omega)} df \right]^2} \right] \left(\frac{3B_{\text{FM}}^2}{B^2} \gamma \right) \quad (12.88d)$$

Comparison of Eq. (12.88d) with Eq. (12.47) shows that PDE in FM improves the SNR by the factor inside the brackets on the right-hand side of Eq. (12.88d). Optimum preemphasis and deemphasis is not used in commercial FM broadcasting for historical and practical reasons. The relatively simple suboptimum scheme, discussed in Sec. 5.5, is used instead.

REFERENCES

1. P. F. Panter, *Modulation, Noise, and Spectral Analysis*, McGraw-Hill, New York, 1965.
2. N. Abramson, "Bandwidth and Spectra of Phase- and Frequency-Modulated Waves," *IEEE Trans. Commun. Syst.*, vol. CS-11, pp. 407-414, Dec. 1963.
3. S. O. Rice, "Mathematical Analysis of Random Noise," *Bell Syst. Tech. J.*, vol. 23, pp. 282-332, July 1944; vol. 24, pp. 46-156, Jan. 1945.
4. A. J. Viterbi, *Principles of Coherent Communication*, McGraw-Hill, New York, 1966.
5. H. E. Rowe, *Signals and Noise in Communication Systems*, Van Nostrand, Princeton, NJ, 1965.
6. M. D. Paez and T. H. Glissom, "Minimum Mean Square Error Quantization in Speech, PCM, and DPCM Systems," *IEEE Trans. Commun. Technol.*, vol. COM-20, pp. 225-230, April 1972.

PROBLEMS

- 12.1-1** A certain telephone channel has $H_c(\omega) \simeq 10^{-3}$ over the signal band. The message signal PSD is $S_m(\omega) = \beta \text{rect}(\omega/2\alpha)$, with $\alpha = 8000\pi$. The channel noise PSD is $S_n(\omega) = 10^{-8}$. If the output SNR at the receiver is required to be at least 30 dB, what is the minimum transmitted power required? Calculate the value of β corresponding to this power.
- 12.1-2** A signal $m(t)$ with PSD $S_m(\omega) = \beta \text{rect}(\omega/2\alpha)$ and $\alpha = 8000\pi$ is transmitted over a telephone channel with transfer function $H_c(\omega) = 10^{-3}/(j\omega + \alpha)$. The channel noise PSD is $S_n(\omega) = 10^{-10}$. To compensate for the channel distortion, the receiver filter transfer function is chosen to be

$$H_d(\omega) = \left(\frac{j\omega + \alpha}{\alpha} \right) \text{rect} \left(\frac{\omega}{2\alpha} \right)$$

The receiver output SNR is required to be at least 35 dB. Determine the minimum required value of β and the corresponding transmitted power S_T and the power S_i received at the receiver input.

- 12.2-1** For a DSB-SC system with a channel noise PSD of $S_n(\omega) = 10^{-10}$ and a baseband signal of bandwidth 4 kHz, the receiver output SNR is required to be at least 30 dB. The receiver is as shown in Fig. 12.3.

- (a) What must be the signal power S_i received at the receiver input?
- (b) What is the receiver output noise power N_o ?
- (c) What is the minimum transmitted power S_T if the channel transfer function is $H_c(\omega) = 10^{-4}$ over the transmission band?

- 12.2-2** Repeat Prob. 12.2-1 for SSB-SC.

- 12.2-3** Determine the output SNR of each of the two quadrature multiplexed channels and compare the results with those of DSB-SC and SSB-SC.

- 12.2-4** Assume $[m(t)]_{\text{max}} = -[m(t)]_{\text{min}} = m_p$.

- (a) Show that for AM,

$$m_p = \mu A$$

- (b) Show that the output SNR for AM [Eq. (12.14)] can be expressed as

$$\frac{S_o}{N_o} = \frac{\mu^2}{k^2 + \mu^2} \gamma$$

where $k^2 = m_p^2/\overline{m^2}$.

- (c) Using the result in part (b), show that for tone modulation with $\mu = 1$, $S_o/N_o = \gamma/3$.
- (d) Show that if S_T and S'_T are the AM and DSB-SC transmitted powers, respectively, required to attain a given output SNR, then

$$S_T \simeq k^2 S'_T \quad \text{for } \mu = 1 \quad \text{and} \quad k^2 \gg 1.$$

- 12.2-5** A gaussian baseband random process $m(t)$ is transmitted by AM.

- (a) Show that for 3σ loading (that is, $m_p = 3\sigma$), the output SNR is $\gamma/10$ when $\mu = 1$.
- (b) Show that for 3σ loading and $\mu = 0.5$, the output SNR $\simeq \gamma/36$.

- 12.2-6** In many communication systems, the transmitted signal is limited by peak power rather than by average power. Under such a limitation, AM fares much worse than DSB-SC or SSB-SC. Show that for tone modulation for a fixed peak power transmitted, the output SNR of AM is 6 dB below that of DSB-SC and 9 dB below that of SSB-SC.

- 12.2-7** Determine γ_{thresh} in AM with $\mu = 1$ if the onset of the threshold is when $E_n > A$ with probability 0.01, where E_n is the noise envelope. Assume the modulating signal $m(t)$ to be gaussian and use 4σ loading.

- 12.3-1** For an FM communication system with $\beta = 2$ and white channel noise with PSD $S_n(\omega) = 10^{-10}$, the output SNR is found to be 28 dB. The baseband signal $m(t)$ is gaussian and band-limited

to 15 kHz, and 3σ loading is used. The demodulator constant $\alpha = 10^{-4}$. This means that the FM demodulator output is $\alpha\dot{\psi}(t)$ when the input is $A \cos[\omega_c t + \psi(t)]$. In the present case, the signal at the demodulator output is $\alpha k_f m(t)$. The output noise is also multiplied by α .

- Determine the received signal power S_i .
- Determine the output signal power S_o .
- Determine the output noise power N_o .

12.3-2 For the modulating signal $m(t)$ shown in Fig. P12.3-2, show that PM is superior to FM by a factor $3\pi^2/4$ from the SNR point of view. *Hint:* Assume the bandwidth of $m(t)$ to be the frequency of its third harmonic.

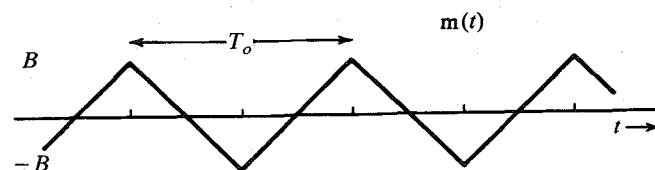


Figure P12.3-2

- 12.3-3** For a modulating signal $m(t) = \cos^3 \omega_o t$, show that PM is superior to FM by a factor of 2.25 from the SNR point of view. *Hint:* Determine m'_p by setting $\dot{m}(t) = 0$.
- 12.3-4** For $m(t) = a_1 \cos \omega_1 t + a_2 \cos \omega_2 t$, show that PM is superior to FM from the SNR point of view when $(1 + xy)^2 < (1 + x)^2/3$, where $x = a_1/a_2$ and $y = \omega_1/\omega_2$. *Hint:* Use Eq. (12.39).
- 12.3-5** Show that $\overline{B_m^2}$ in Eq. (12.42a) can also be expressed directly in the time domain as

$$\overline{B_m^2} = \frac{\int_{-\infty}^{\infty} [\dot{m}(t)]^2 dt}{\int_{-\infty}^{\infty} m^2(t) dt}$$

12.3-6 Show that when the modulating signal $m(t)$ has a Butterworth PSD, that is,

$$S_m(\omega) = \frac{1}{1 + (\omega/2\pi f_o)^{2k}}$$

then the mean square bandwidth $\overline{B_m^2}$ is given by

$$\overline{B_m^2} = f_o^2 \frac{\sin(\pi/2k)}{\sin(3\pi/2k)} \quad k \geq 2$$

and as $k \rightarrow \infty$, $\overline{B_m^2} \rightarrow \frac{1}{3} f_o^2$.

12.3-7 A modulating signal PSD is given by

$$S_m(\omega) = \frac{|\omega|}{\sigma^2} e^{-\omega^2/2\sigma^2}$$

The effective bandwidth B of $m(t)$ is that bandwidth which contains χ percent of the total power of $m(t)$. Determine which of the angle modulations is superior if: (a) $\chi = 99$; (b) $\chi = 95$; (c) $\chi = 90$.

12.3-8 In Prob. 12.3-4, FM and PM were compared for a given transmission bandwidth. If we compare them for a given mean square transmission bandwidth, we get a slightly different result. Show

that PM is superior to FM if $(1 + x^2 y^2) < (1 + x^2)/3$ when they are compared for a given mean square transmission bandwidth. This example shows that two signal spectra with the same mean square bandwidth can have different conventional bandwidths. *Hint:* In this case $S_m(\omega)$ is discrete. Normalize it, and then take the second moment about the origin. Show that

$$\overline{B_m^2} = \frac{a_1^2 f_1^2 + a_2^2 f_2^2}{a_1^2 + a_2^2}$$

12.3-9 In a certain FM system used in space communication, the output SNR is found to be 23.4 dB with $\beta = 2$. The modulating signal $m(t)$ is gaussian with a bandwidth of 10 kHz, and 3σ loading is used. The system with $\beta = 2$ is in the nonthreshold region of operation. The output SNR is required to be at least 40 dB. Because power is at premium in space communication, it is decided to increase the output SNR by increasing β (i.e., increasing the transmission bandwidth) as much as is possible.

- What are the maximum value of β and the corresponding transmission bandwidth that can be used without running into the threshold? What is the corresponding output SNR? *Hint:* Use Eqs. (12.51) and (12.37) to determine the minimum usable γ .
- What must be the minimum increase in the transmitted power required to attain an output SNR of 40 dB? What are the corresponding value of β and the transmission bandwidth?

12.3-10 Show that

- For tone modulation, the dividing line between narrow-band and wide-band modulation is $\beta = 0.47$.
- For a gaussian modulating signal with 3σ loading, the dividing line is at $\beta = 0.55$.
- For 4σ loading it is at $\beta = 0.56$.

12.3-11 For FM stereophonic broadcasting (Fig. 5.19), show that the $(L - R)$ channel is about 22 dB noisier than the $L + R$ channel. *Hint:* $L + R$ is the baseband signal. It is preemphasized to obtain $(L + R)'$. The signal $(L - R)$ is preemphasized to obtain $(L - R)'$, which is used to obtain $(L - R)' \cos \omega_c t$. The sum $(L + R)' + (L - R)' \cos \omega_c t$ now frequency-modulates a carrier. At the receiver, after frequency demodulation, $(L + R)'$ and $(L - R)' \cos \omega_c t$ are separated. $(L - R)'$ is deemphasized. $(L + R)' \cos \omega_c t$ is multiplied by $2 \cos \omega_c t$ to obtain $(L - R)'$, which is then deemphasized.

12.4-1 In M -ary PCM, pulses can take M distinct amplitudes (in contrast to two for binary PCM). Show that the signal-to-quantization-noise ratio for M -ary PCM is

$$\frac{S_o}{N_o} = 3M^{2n} \left(\frac{\overline{m^2}}{m_p^2} \right)$$

12.4-2 A TV signal band-limited to 4.5 MHz is to be transmitted by binary PCM. The receiver output signal-to-quantization-noise ratio is required to be at least 55 dB.

- If all brightness levels are assumed to be equally likely, that is, amplitudes of $m(t)$ are uniformly distributed in the range $(-m_p, m_p)$, find the minimum number of quantization levels L required. Select the nearest value of L to satisfy $L = 2^n$.
- For this value of L , compute the receiver output SNR and the transmission bandwidth, assuming the nonthreshold region of operation.
- If the output SNR is required to be increased by 6 dB (four times), what are the new value of L and the corresponding transmission bandwidth?

12.4-3 A modulating signal $m(t)$ band-limited to 4 kHz is sampled at a rate of 8000 samples per second. The samples are quantized into 256 levels, binary coded, and transmitted over a channel with $S_n(\omega) = 6.25 \times 10^{-7}$. Each received pulse has energy $E_p = 2 \times 10^{-5}$. Given that $m_p = 1$ and $\overline{m^2} = 1/9$,

- Find the output SNR assuming polar signal, and the error probability given in Eq. (12.62).
- If the transmitted power is reduced by 10 dB, find the new SNR.
- At the reduced power level in part (b), is it possible to increase the output SNR by changing the value of L ? Determine the maximum output SNR achievable and the corresponding value of L .

12.4-4 In a PCM channel using k identical regenerative links, we have shown that the error probability P_E of the overall channel is kP_e , where P_e is the error probability of an individual link (see Example 10.7). This shows that P_e is cumulative.

- Show that if $k - 1$ links are identical with error probability P_e and the remaining one link has an error probability P'_e , then

$$P_E = (k - 1)P_e + P'_e$$

- For a certain chain of repeaters with $k = 100$ (100 repeaters), it is found that γ over each of the 99 links is 25 dB, and over the remaining link γ is 23 dB. Calculate P_e and P'_e using Eq. (12.62) (with $n = 8$). Now compute P_E and show that P_E is primarily determined by the single weakest link in the chain.

12.4-5 For companded PCM with $n = 8$, $\mu = 255$, and amplitude m uniformly distributed in the range $(-A, A)$ ($A \leq m_p$), show that

$$\frac{S_o}{N_o} = \frac{6394(\sigma_m^2/m_p^2)}{(\sigma_m^2/m_p^2) + 0.0068(\sigma_m/m_p) + 1.53 \times 10^{-5}}$$

Note that m_p is a constant of the system, not of the signal. The peak signal A can vary from talker to talker, whereas m_p is fixed for a given system.

12.5-1 A message signal $m(t)$ with

$$S_m(\omega) = \frac{\alpha^2}{\omega^2 + \alpha^2} \quad (\alpha = 3000\pi)$$

DSB-SC modulates a carrier of 100 kHz. Assume an ideal channel with $H_c(\omega) = 10^{-3}$ and the channel noise PSD $S_n(\omega) = 2 \times 10^{-9}$. The transmitted power is required to be 1 kW, and $G = 10^{-2}$.

- Determine transfer functions of optimum preemphasis and deemphasis filters.
- Determine the output signal power, the noise power, and the output SNR.
- Determine γ at the demodulator input.

12.5-2 Repeat Prob. 12.5-1 for the SSB (USB) case.

12.5-3 It was shown in the text that when the baseband $m(t)$ is band-limited with a uniform PSD, PM and FM have identical performance from the SNR point of view. For such $m(t)$, show that optimum PDE filters in angle modulation can improve the output SNR by a factor of 4/3 (or 1.3 dB) only. Find the optimum PDE filter transfer functions.

13 BEHAVIOR OF DIGITAL COMMUNICATION SYSTEMS IN THE PRESENCE OF NOISE

In analog systems, the chief objective is the fidelity of reproduction of waveforms, and, hence, the suitable performance criterion is the output signal-to-noise ratio. The choice of this criterion stems from the fact that the signal-to-noise ratio is related to the ability of the listener to interpret a message. In digital communication systems, the transmitter input is chosen from a finite set of possible symbols, or messages. The objective at the receiver is not to reproduce the waveform with fidelity, because the possible waveforms are already known exactly. Our goal is to decide, from the noisy received signal, which of the waveforms has been transmitted. Logically, the appropriate figure of merit in a digital communication system is the probability of error in making such a decision at the receiver.

13.1 OPTIMUM THRESHOLD DETECTION

In the threshold detection method discussed in Example 10.13, the received pulse is sampled at its peak amplitude A_p . Because of channel noise, the sampled value is not A_p but $A_p + n$. The decision is made from the value $A_p + n$. It was shown in Example 10.13 that for the polar binary case the error probability P_e is

$$P_e = Q(\rho) \quad (13.1)$$

where

$$\rho = \frac{A_p}{\sigma_n} \quad (13.2a)$$

σ_n^2 being the variance of the received noise. To minimize P_e , we need to maximize ρ because $Q(\rho)$ decreases monotonically with ρ . Note that A_p is the signal amplitude and σ_n is the rms noise so that ρ is the signal amplitude to rms noise ratio.

Let the received pulse $p(t)$ be time limited to T_o (Fig. 13.1). Note that here we use a general symbol T_o rather than T_b because the pulse may not be binary. We shall keep the discussion as general as possible at this point. There is a possibility of increasing $\rho = A_p/\sigma_n$

by passing the received pulse through a filter that enhances the pulse amplitude at some instant t_m and simultaneously reduces the noise power σ_n^2 (Fig. 13.1). We thus seek a filter with a transfer function $H(\omega)$ that maximizes ρ , where

$$\rho^2 = \frac{P_o^2(t_m)}{\sigma_n^2} \quad (13.2b)$$

Because

$$\begin{aligned} p_o(t) &= \mathcal{F}^{-1}[P(\omega)H(\omega)] \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} P(\omega)H(\omega)e^{j\omega t} d\omega \end{aligned}$$

we have

$$p_o(t_m) = \frac{1}{2\pi} \int_{-\infty}^{\infty} P(\omega)H(\omega)e^{j\omega t_m} d\omega \quad (13.3)$$

Also,

$$\sigma_n^2 = \overline{n_o^2(t)} = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_n(\omega)|H(\omega)|^2 d\omega \quad (13.4)$$

Hence,

$$\rho^2 = \frac{[\int_{-\infty}^{\infty} H(\omega)P(\omega)e^{j\omega t_m} d\omega]^2}{2\pi \int_{-\infty}^{\infty} S_n(\omega)|H(\omega)|^2 d\omega} \quad (13.5)$$

In the Schwarz inequality (Appendix B), if we identify $X(\omega) = H(\omega)\sqrt{S_n(\omega)}$ and $Y(\omega) = P(\omega)e^{j\omega t_m}/\sqrt{S_n(\omega)}$, then it follows from Eq. (13.5) that

$$\rho^2 \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|P(\omega)|^2}{S_n(\omega)} d\omega \quad (13.6a)$$

with equality only if

$$H(\omega)\sqrt{S_n(\omega)} = k \left[\frac{P(\omega)e^{j\omega t_m}}{\sqrt{S_n(\omega)}} \right]^* = \frac{kP(-\omega)e^{-j\omega t_m}}{\sqrt{S_n(\omega)}}$$

Hence

$$H(\omega) = k \frac{P(-\omega)e^{-j\omega t_m}}{S_n(\omega)} \quad (13.6b)$$

where k is an arbitrary constant.

For white channel noise $S_n(\omega) = \mathcal{N}/2$, and Eqs. (13.6) become

$$\rho_{\max}^2 = \frac{1}{\pi \mathcal{N}} \int_{-\infty}^{\infty} |P(\omega)|^2 d\omega = \frac{2E_p}{\mathcal{N}} \quad (13.7a)$$

where E_p is the energy of $p(t)$, and

$$H(\omega) = k' P(-\omega)e^{-j\omega t_m} \quad (13.7b)$$

where $k' = 2k/\mathcal{N}$ is an arbitrary constant.

The unit impulse response $h(t)$ of the optimum filter is given by

$$h(t) = \mathcal{F}^{-1}[k' P(-\omega)e^{-j\omega t_m}]$$

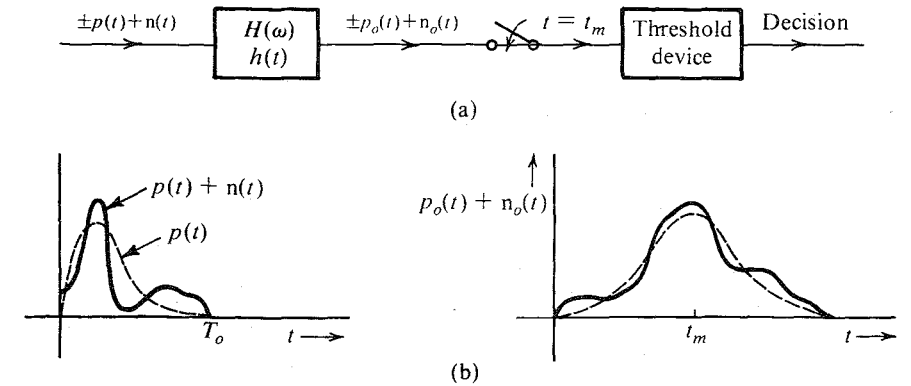


Figure 13.1 Scheme to minimize the error probability in threshold detection.

Note that $p(-t) \iff P(-\omega)$ and $e^{-j\omega t_m}$ represents the time delay of t_m seconds. Hence,

$$h(t) = k' p(t_m - t) \quad (13.7c)$$

The signal $p(t_m - t)$ is the signal $p(-t)$ delayed by t_m . Three cases, $t_m < T_o$, $t_m = T_o$, and $t_m > T_o$, are shown in Fig. 13.2. The first case, $t_m < T_o$, yields a noncausal impulse response, which is unrealizable.* Although the other two cases yield physically realizable filters, the last case, $t_m > T_o$, delays the decision-making instant t_m an unnecessary length of time. The case $t_m = T_o$ gives the minimum delay for decision making using a realizable filter. In our future discussion, we shall assume $t_m = T_o$, unless otherwise mentioned.

Observe that both $p(t)$ and $h(t)$ have a width of T_o seconds. Hence $p_o(t)$, which is a convolution of $p(t)$ and $h(t)$, has a width of $2T_o$ seconds, with its peak occurring at $t = T_o$. Also, because $P_o(\omega) = P(\omega)H(\omega) = k'|P(\omega)|^2 e^{-j\omega T_o}$, $p_o(t)$ is symmetrical about $t = T_o$ (Fig. 13.1).†

The arbitrary constant k' in Eq. (13.7) multiplies both the signal and the noise by the same factor and does not affect the ratio ρ . Hence, the error probability, or the system performance, is independent of the value of k' . For convenience, we choose $k' = 1$. This gives

$$h(t) = p(T_o - t) \quad (13.8a)$$

and

$$H(\omega) = P(-\omega)e^{-j\omega T_o} \quad (13.8b)$$

* The filter unrealizability can be readily understood intuitively when the decision-making instant is $t_m < T_o$. In this case, we are forced to make a decision even before the complete pulse is fed to the filter ($t_m < T_o$). This calls for a prophetic filter, which can respond to inputs even before they are applied. As we know, only unrealizable (noncausal) filters can do this job.

† This follows from the fact that because $|P(\omega)|^2$ is an even function of ω , its inverse transform is symmetrical about $t = 0$ (see Prob. 3.1-1). The output from the previous input pulse terminates and has a zero value at $t = T_o$. Similarly, the output from the following pulse starts and has a zero value at $t = T_o$. Hence, at the decision-making instant T_o , no intersymbol interference occurs.

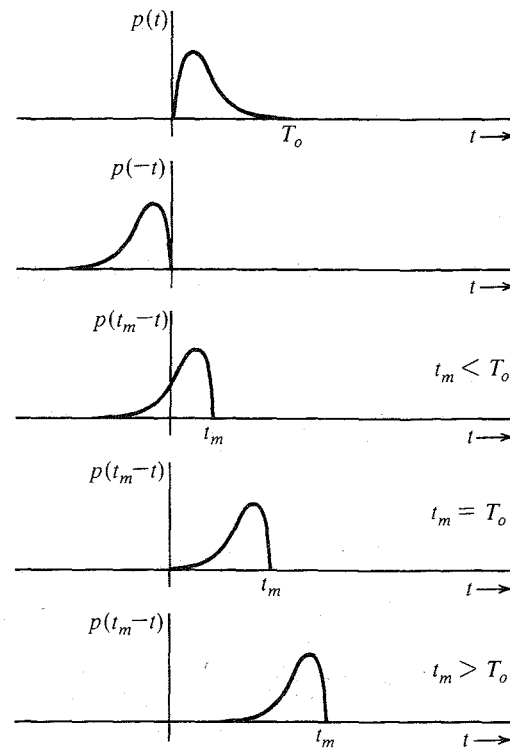


Figure 13.2 Optimum choice for sampling instant.

The optimum filter in Eqs. (13.8) is known as the **matched filter**.^{*} At the output of this filter, the signal to rms noise amplitude ratio is maximum at the decision-making instant $t = T_o$.

The matched filter is optimum in the sense that it maximizes the signal amplitude to rms noise ratio at the decision-making instant. Although it is reasonable to assume that maximization of this particular signal to noise ratio will minimize the detection error probability, we have not proved that threshold detection (sample and decide) is the optimum method from the detection error point of view. It will be shown in Chapter 14 that when the channel noise is white gaussian, the matched-filter receiver is indeed the optimum receiver that minimizes the detection error probability. The maximum value of this signal to rms noise ratio attained by the matched filter is given in Eq. (13.7a). The peak amplitude $p_o(t_m) = A_p$ is found by substituting Eq. (13.8b) (with $k' = 1$) into Eq. (13.3),

$$A_p = \frac{1}{2\pi} \int_{-\infty}^{\infty} |P(\omega)|^2 d\omega = E_p \quad (13.9a)$$

The noise power σ_n^2 is obtained by substituting Eq. (13.8b) (with $k' = 1$) into Eq. (13.4),

$$\sigma_n^2 = \frac{\mathcal{N}}{4\pi} \int_{-\infty}^{\infty} |P(\omega)|^2 d\omega = \frac{\mathcal{N}E_p}{2} \quad (13.9b)$$

^{*} It is important to remember that the optimum filter is the matched filter only when the channel noise is white. For a general case, the optimum filter is given in Eq. (13.6b).

Hence,

$$\rho_{\max}^2 = \frac{A_p^2}{\sigma_n^2} = \frac{2E_p}{\mathcal{N}} \quad (13.9c)$$

and

$$P_e = Q(\rho_{\max}) = Q\left(\sqrt{\frac{2E_p}{\mathcal{N}}}\right) \quad (13.9d)$$

Equations (13.9) are truly remarkable. They show that at the decision-making instant, the signal amplitude, and the rms noise amplitude depend on the waveform $p(t)$ only through its energy E_p . As far as the system performance is concerned, when the matched-filter receiver is used, all the waveforms used for $p(t)$ are equivalent as long as they have the same energy.

The matched filter may also be realized by the alternative arrangement shown in Fig. 13.3. If the input to the matched filter is $r(t)$, then the output $y(t)$ is given by

$$y(t) = \int_{-\infty}^{\infty} r(x)h(t-x)dx$$

where $h(t) = p(T_o - t)$ and

$$h(t-x) = p[T_o - (t-x)] = p(x + T_o - t)$$

Hence,

$$y(t) = \int_{-\infty}^{\infty} r(x)p(x + T_o - t)dx \quad (13.10a)$$

At the decision-making instant $t = T_o$, we have

$$y(T_o) = \int_{-\infty}^{\infty} r(x)p(x)dx \quad (13.10b)$$

Because the input $r(x)$ is assumed to start at $x = 0$ and $p(x) = 0$ for $x > T_o$,

$$y(T_o) = \int_0^{T_o} r(x)p(x)dx \quad (13.10c)$$

We can implement Eqs. (13.10) as shown in Fig. 13.3. This type of arrangement is known as the **correlation receiver** and is equivalent to the matched-filter receiver.

The right-hand side of Eq. (13.10a) is $\psi_{rp}(T_o - t)$, where $\psi_{rp}(\tau)$ is the crosscorrelation of the received pulse with $p(t)$. Recall that correlation basically measures the similarity of signals (see Sec. 2.6). Thus, the optimum detector measures the similarity of the received signal with the pulse $p(t)$. Based on this similarity measure, it decides whether $p(t)$ was transmitted or not.

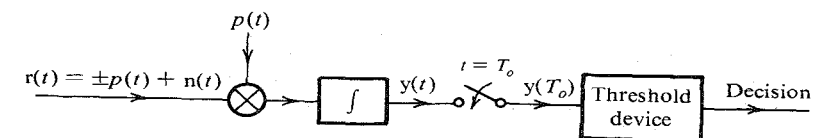


Figure 13.3 Correlation detector.

Thus far we have discussed polar signaling where only one basic pulse $p(t)$ is used. Generally, in binary communication, we use two distinct pulses $p(t)$ and $q(t)$ to represent the two symbols. The optimum receiver for such a case will now be discussed.

13.2 GENERAL ANALYSIS: OPTIMUM BINARY RECEIVER

In a binary scheme where symbols are transmitted every T_b seconds, let $p(t)$ and $q(t)$ be the two pulses used to transmit 1 and 0. The optimum receiver structure considered here is shown in Fig. 13.4a. The incoming pulse is transmitted through a filter $H(\omega)$, and the output $r(t)$ is sampled at T_b . The decision as to whether 0 or 1 was present at the input depends on whether $r(T_b) < \text{or} > a_o$, where a_o is the optimum threshold.

Let $p_o(t)$ and $q_o(t)$ be the response of $H(\omega)$ to inputs $p(t)$ and $q(t)$, respectively. From Eq. (13.3) it follows that

$$p_o(T_b) = \frac{1}{2\pi} \int_{-\infty}^{\infty} P(\omega) H(\omega) e^{j\omega T_b} d\omega \quad (13.11a)$$

$$q_o(T_b) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Q(\omega) H(\omega) e^{j\omega T_b} d\omega \quad (13.11b)$$

and σ_n^2 , the variance, or power, of the noise at the filter output, is

$$\sigma_n^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_n(\omega) |H(\omega)|^2 d\omega \quad (13.11c)$$

If n is the noise output at T_b , then the sampler output $r(T_b) = q_o(T_b) + n$ or $p_o(T_b) + n$, depending on whether $m = 0$ or $m = 1$, is received. Hence, r is a gaussian RV of variance σ_n^2 and mean $q_o(T_b)$ or $p_o(T_b)$, depending on whether $m = 0$ or 1. Thus, the conditional PDFs of the sampled output $r(T_b)$ are

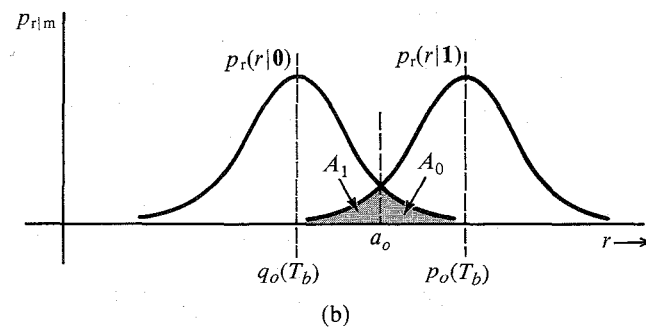
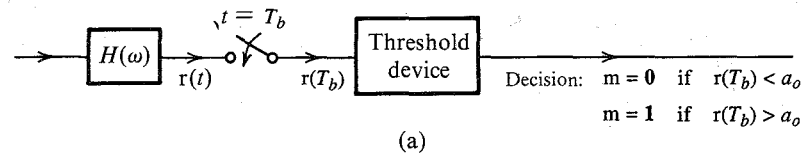


Figure 13.4 Optimum binary threshold detection.

$$p_{r|m}(r|0) = \frac{1}{\sigma_n \sqrt{2\pi}} e^{-[r - q_o(T_b)]^2 / 2\sigma_n^2}$$

$$p_{r|m}(r|1) = \frac{1}{\sigma_n \sqrt{2\pi}} e^{-[r - p_o(T_b)]^2 / 2\sigma_n^2}$$

The two PDFs are shown in Fig. 13.4b. If a_o is the optimum threshold of detection, then the decision is $m = 0$ if $r < a_o$ and $m = 1$ if $r > a_o$. The conditional error probability $P(\epsilon | m = 0)$ is the probability of making a wrong decision when $m = 0$. This is simply the area A_0 under $p_{r|m}(r|0)$ from a_o to ∞ . Similarly, $P(\epsilon | m = 1)$ is the area A_1 under $p_{r|m}(r|1)$ from $-\infty$ to a_o (Fig. 13.4b), and

$$P_e = \sum_i P(\epsilon | m_i) P(m_i) = \frac{1}{2} (A_0 + A_1)$$

assuming $P_m(0) = P_m(1) = 0.5$. From Fig. 13.4b it can be seen that the sum $A_0 + A_1$ of the shaded areas is minimized by choosing a_o at the intersection of the two PDFs. Thus,

$$a_o = \frac{p_o(T_b) + q_o(T_b)}{2} \quad (13.12a)$$

and the corresponding P_e is

$$\begin{aligned} P_e &= P(\epsilon|0) = P(\epsilon|1) \\ &= \frac{1}{\sigma_n \sqrt{2\pi}} \int_{a_o}^{\infty} e^{-[r - q_o(T_b)]^2 / 2\sigma_n^2} dr \\ &= Q \left[\frac{a_o - q_o(T_b)}{\sigma_n} \right] \\ &= Q \left[\frac{p_o(T_b) - q_o(T_b)}{2\sigma_n} \right] \end{aligned} \quad (13.12b)$$

$$= Q \left(\frac{\beta}{2} \right) \quad (13.12c)$$

where we define

$$\beta = \frac{p_o(T_b) - q_o(T_b)}{\sigma_n} \quad (13.13)$$

Substituting Eqs. (13.11) into Eq. (13.13), we get

$$\beta^2 = \frac{[\int_{-\infty}^{\infty} [P(\omega) - Q(\omega)] H(\omega) e^{j\omega T_b} d\omega]^2}{2\pi \int_{-\infty}^{\infty} S_n(\omega) |H(\omega)|^2 d\omega}$$

This equation is of the same form as Eq. (13.5) with $P(\omega)$ replaced by $P(\omega) - Q(\omega)$. Hence,

$$\beta_{\max}^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|P(\omega) - Q(\omega)|^2}{S_n(\omega)} d\omega \quad (13.14a)$$

and the optimum filter $H(\omega)$ is given by

$$H(\omega) = k \frac{[P(-\omega) - Q(-\omega)]e^{-j\omega T_b}}{S_n(\omega)} \quad (13.14b)$$

where k is an arbitrary constant.

For white noise $S_n(\omega) = \mathcal{N}/2$, and the optimum filter $H(\omega)$ is given by*

$$H(\omega) = [P(-\omega) - Q(-\omega)]e^{-j\omega T_b} \quad (13.15a)$$

and

$$h(t) = p(T_b - t) - q(T_b - t) \quad (13.15b)$$

This is a filter matched to the pulse $p(t) - q(t)$. The corresponding β is [Eq. (13.14a)]

$$\beta_{\max}^2 = \frac{1}{\pi \mathcal{N}} \int_{-\infty}^{\infty} |P(\omega) - Q(\omega)|^2 d\omega \quad (13.16a)$$

$$= \frac{2}{\mathcal{N}} \int_0^{T_b} [p(t) - q(t)]^2 dt \quad (13.16b)$$

$$= \frac{E_p + E_q - 2E_{pq}}{\mathcal{N}/2} \quad (13.16c)$$

where E_p and E_q are the energies of $p(t)$ and $q(t)$, respectively, and

$$E_{pq} = \int_0^{T_b} p(t)q(t) dt \quad (13.17)$$

So far, we have been using the notation P_e to denote error probability. In the binary case, this error probability is the **bit error probability** or **bit error rate** (BER), and will be denoted by P_b (rather than P_e). Thus, from Eqs. (13.12c) and (13.16c),

$$P_b = Q\left(\frac{\beta_{\max}}{2}\right) \quad (13.18a)$$

$$= Q\left(\sqrt{\frac{E_p + E_q - 2E_{pq}}{2\mathcal{N}}}\right) \quad (13.18b)$$

The optimum threshold a_o is obtained by substituting Eqs. (13.11a, b) and (13.15a) into Eq. (13.12a) and recognizing that (see Prob. 3.7-3)

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} P(\omega)Q(-\omega) d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} P(-\omega)Q(\omega) d\omega = E_{pq}$$

This gives

$$a_o = \frac{1}{2}(E_p - E_q) \quad (13.19)$$

In deriving the optimum binary receiver, we assumed a certain receiver structure (the threshold-detection receiver in Fig. 13.4). It is not clear yet whether there exists another structure that may have better performance than that in Fig. 13.4. It will be shown in Chapter 14 that for a gaussian noise, the receiver derived here is the absolute optimum. Equation (13.18b) gives P_b

* Because k in Eq. (13.14b) is arbitrary, we choose $k = \mathcal{N}/2$ for convenience.

for the optimum receiver when the channel noise is white. For the case of nonwhite noise, P_b is obtained by substituting $\beta_{\max}$ from Eq. (13.14a) into Eq. (13.18a).

13.2.1 Equivalent Optimum Binary Receivers

For the optimum receiver in Fig. 13.4a,

$$H(\omega) = P(-\omega)e^{-j\omega T_b} - Q(-\omega)e^{-j\omega T_b}$$

This filter can be realized as a parallel combination of two filters matched to $p(t)$ and $q(t)$, respectively, as shown in Fig. 13.5a. Yet another equivalent form is shown in Fig. 13.5b. Because the threshold is $(E_p - E_q)/2$, we subtract $E_p/2$ and $E_q/2$, respectively, from the two matched filter outputs. This is equivalent to shifting the threshold to 0. In the case where $E_p = E_q$, we need not subtract $E_p/2$ and $E_q/2$ from the two outputs, and the receiver simplifies to that shown in Fig. 13.5c.

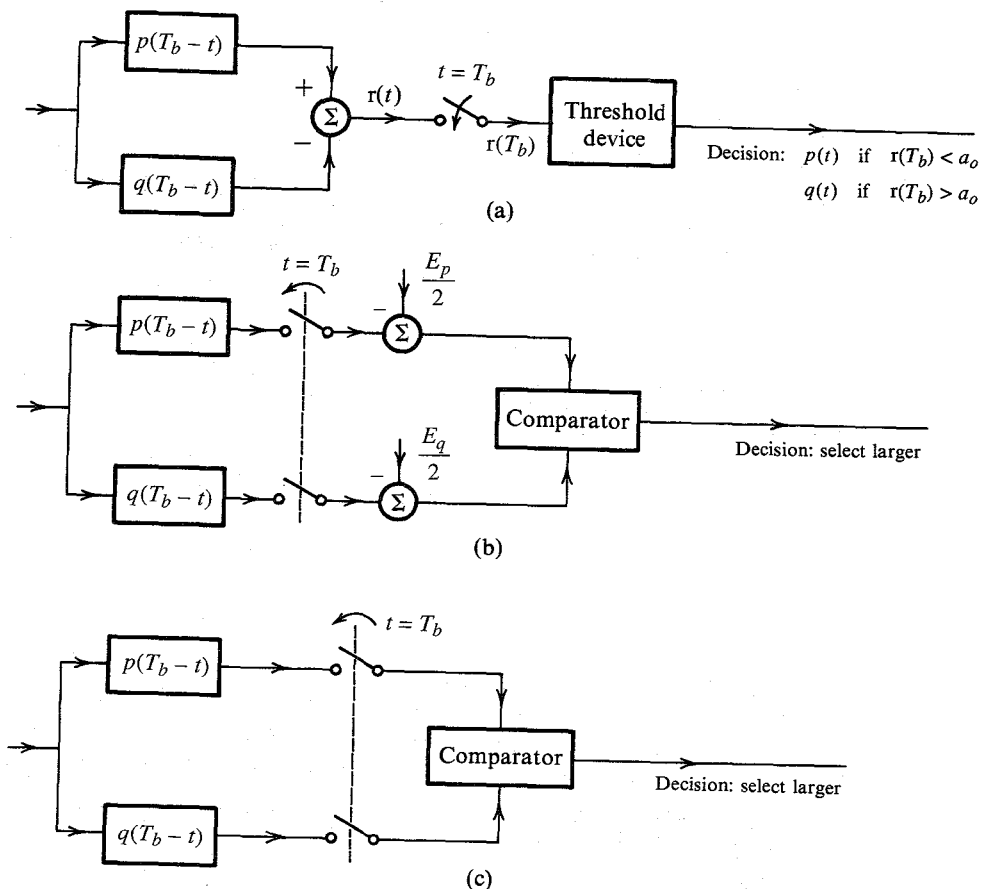


Figure 13.5 Realization of the optimum binary threshold detector.

Polar Signaling

In this case $q(t) = -p(t)$. Hence,

$$E_p = E_q \quad \text{and} \quad E_{pq} = - \int_{-\infty}^{\infty} p^2(t) dt = -E_p$$

Substituting these results into Eq. (13.18b) yields

$$P_b = Q \left(\sqrt{\frac{2E_p}{\mathcal{N}}} \right) \quad (13.20a)$$

Also from Eq. (13.15b),

$$h(t) = 2p(T_b - t)$$

Recall that the multiplication of $h(t)$ by any constant multiplies both the signal and the noise by the same factor, and hence does not affect the system performance. For convenience, we shall multiply $h(t)$ by 0.5 to obtain

$$h(t) = p(T_b - t) \quad (13.20b)$$

From Eq. (13.19), the threshold a_o is

$$a_o = 0 \quad (13.20c)$$

Therefore, for the polar case, the receiver in Fig. 13.5a reduces to that shown in Fig. 13.6a with threshold 0. This filter is equivalent to that in Fig. 13.3.

The error probability can be expressed in terms of a more basic parameter E_b , the energy per bit:

$$E_b = \text{Energy per bit}$$

In the polar case assuming 1 and 0 are equally likely, the bit energy E_b is the mean of E_p and E_q . Hence,

$$E_b = \frac{E_p + E_q}{2} = E_p \quad (13.21a)$$

and from Eq. (13.20a),

$$P_b = Q \left(\sqrt{\frac{2E_b}{\mathcal{N}}} \right) \quad (13.21b)$$

The parameter $E_b/\mathcal{N}$ is the normalized energy per bit, which will be seen in future discussions as a fundamental parameter serving as a figure of merit in digital communication.* Because the signal power is equal to E_b times the bit rate, a given E_b is the same as a given signal power (for a given bit rate). Hence, when we compare a system for a given value of E_b , we are comparing it for a given signal power. Using an asymptotic approximation [Eq. (10.36a)] for $Q(\sqrt{2E_b/\mathcal{N}})$, we obtain

* If the transmission rate is R_b pulses per second, the signal power S_i is $S_i = E_b R_b$, and $E_b/\mathcal{N} = S_i/\mathcal{N}R_b$. Observe that $S_i/\mathcal{N}R_b$ is similar to the parameter γ (signal-to-noise ratio $S_i/\mathcal{N}B$) used in analog systems.

$$P_b \simeq \frac{1}{2\sqrt{\pi E_b/\mathcal{N}}} e^{-E_b/\mathcal{N}} \quad E_b/\mathcal{N} \gg 1 \quad (13.21c)$$

Figure 13.6b shows the plot of P_b as a function of $E_b/\mathcal{N}$ (in decibels). Equation (13.21b) indicates that, for optimum threshold detection, the system performance does not depend on the pulse shape, but on the pulse energy.

On-Off Signaling

In this case, $q(t) = 0$, and Fig. 13.5a reduces to Fig. 13.6a (the same as that for the polar case) except that the threshold, as seen from Eq. (13.19), is $E_p/2$. Because $q(t) = 0$, From Eqs. (13.17) and (13.18),

$$E_q = 0, \quad E_{pq} = 0, \quad \text{and} \quad P_b = Q \left(\sqrt{\frac{E_p}{2\mathcal{N}}} \right)$$

Also

$$E_b = \frac{E_p + E_q}{2} = \frac{E_p}{2}$$

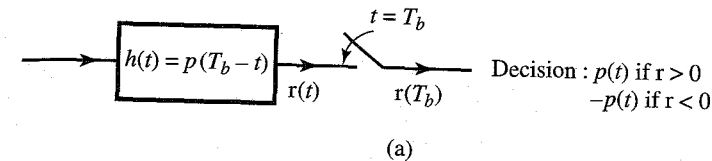
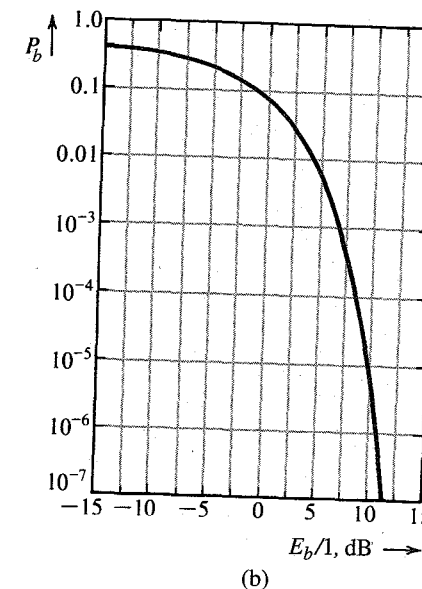


Figure 13.6 Optimum threshold detector and its error probability for polar signaling.



Therefore,

$$P_b = Q \left(\sqrt{\frac{E_b}{\mathcal{N}}} \right) \quad (13.22a)$$

$$\simeq \frac{1}{\sqrt{2\pi E_b/\mathcal{N}}} e^{-E_b/2\mathcal{N}} \quad E_b/\mathcal{N} \gg 1 \quad (13.22b)$$

A comparison of this with Eq. (13.21b) shows that on-off signaling requires twice as much energy per bit (3 dB more power) to achieve the same performance (i.e., the same P_b) as polar signaling.

Orthogonal Signaling

In orthogonal signaling, $p(t)$ and $q(t)$ are selected to be orthogonal over the interval $(0, T_b)$. This gives

$$E_{pq} = \int_0^{T_b} p(t)q(t) dt = 0$$

Two examples of binary orthogonal pulses are shown in Fig. 13.7. From Eq. (13.18),

$$P_b = Q \left(\sqrt{\frac{E_p + E_q}{2\mathcal{N}}} \right) \quad (13.23)$$

Assuming **1** and **0** to be equiprobable,

$$E_b = \frac{E_p + E_q}{2}$$

and

$$P_b = Q \left(\sqrt{\frac{E_b}{\mathcal{N}}} \right) \quad (13.24a)$$

$$\simeq \frac{1}{\sqrt{2\pi E_b/\mathcal{N}}} e^{-E_b/2\mathcal{N}} \quad E_b/\mathcal{N} \gg 1 \quad (13.24b)$$

This shows that the performance of orthogonal signaling is inferior to that of polar signaling by 3 dB, but it is identical to that of on-off signaling. This is not surprising in view of the fact that on-off is a special case of orthogonal signaling because $E_{pq} = 0$.

Bipolar Signaling

Bipolar signaling, although a binary scheme, uses three symbols: $p(t)$, $-p(t)$, and 0. Hence, the preceding results cannot be used directly without some modifications. In this case, we seek to distinguish between $p(t)$ and 0 or $-p(t)$ and 0 so that it is basically similar to the on-off case. Hence, we can use the receiver in Fig. 13.6a with threshold $\pm a_o = \pm E_p/2$. Thus, if $|r(T_b)| < E_p/2$, the decision is **0**, and if $|r(T_b)| > E_p/2$, the decision is **1**. This receiver is a matched filter in Fig. 13.1a. To compute P_b , we compute the conditional error probabilities. When **0** is transmitted by no pulse, the receiver output is just the noise n with variance $\sigma_n^2 = \mathcal{N}E_p/2$ [see Eq. (13.9b)]. Hence,

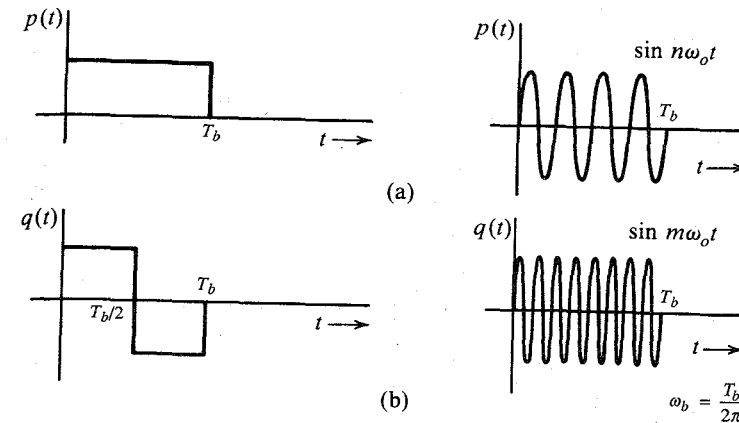


Figure 13.7 Examples of orthogonal signals.

$$\begin{aligned} P(\epsilon|0) &= \text{probability} \left(|n| > \frac{E_p}{2} \right) \\ &= 2Q \left(\frac{E_p}{2\sigma_n} \right) \\ &= 2Q \left(\sqrt{\frac{E_p}{2\mathcal{N}}} \right) \end{aligned}$$

When **1** is transmitted, the filter output at T_b is $A_p + n$, where $A_p = E_p$ when $p(t)$ is transmitted [Eq. (13.9a)]. Consequently, $A_p = -E_p$ when $-p(t)$ is transmitted. Thus,

$$\begin{aligned} P(\epsilon|1) &= \text{probability} \left(n < -\frac{E_p}{2} \right) \text{ when } p(t) \text{ is used, or} \\ &\quad \text{probability} \left(n > \frac{E_p}{2} \right) \text{ when } -p(t) \text{ is used} \\ &= Q \left(\frac{E_p}{2\sigma_n} \right) \\ &= Q \left(\sqrt{\frac{E_p}{2\mathcal{N}}} \right) \end{aligned}$$

Therefore, the average error probability is (assuming **1** and **0** equally likely)

$$P_b = \frac{1}{2} [P(\epsilon|0) + P(\epsilon|1)] = 1.5Q \left(\sqrt{\frac{E_p}{2\mathcal{N}}} \right)$$

This shows that P_b is 50% higher for bipolar signaling than for on-off signaling. But this difference can be compensated by a very small increase in E_p because the Q function, which varies exponentially with E_p , is extremely sensitive to variations in E_p . For example, to realize $P_b = 0.286 \times 10^{-6}$, we need $(E_p)_{\text{bipolar}} = 1.016(E_p)_{\text{on-off}}$.

13.3 CARRIER SYSTEMS: ASK, FSK, PSK, AND DPSK

In digital carrier systems, baseband pulses modulate a high-frequency carrier. We have briefly discussed amplitude-shift keying (ASK), frequency-shift keying (FSK), and phase-shift keying (PSK) in Chapter 7.

Figure 13.8 shows the three schemes, using a rectangular baseband pulse. The baseband pulse may be specifically shaped (e.g., a raised cosine) to eliminate intersymbol interference and to have a finite bandwidth.

The error probability of the optimum detector depends only on the pulse energy, not on the pulse shape. Hence, as far as the error probability is concerned, the performance of a modulated scheme will be identical to that of the baseband scheme of the same energy.

The incoming modulated pulses can be demodulated either coherently (synchronously) or noncoherently (by envelope detection). The former method is the optimum and requires much more sophisticated equipment. Naturally, it has a superior performance in comparison to the latter method.

13.3.1 Coherent Detection

Let the RF pulse $p(t) = \sqrt{2} p'(t) \cos \omega_c t$, where $p'(t)$ is a baseband pulse. The RF pulse can be detected by a filter matched to the RF pulse $p(t)$ followed by a sampler (Fig. 13.9a). In this case [Eq. (13.7a)],

$$\rho^2 = \frac{2E_p}{\mathcal{N}}$$

where E_p is the energy of $p(t)$.

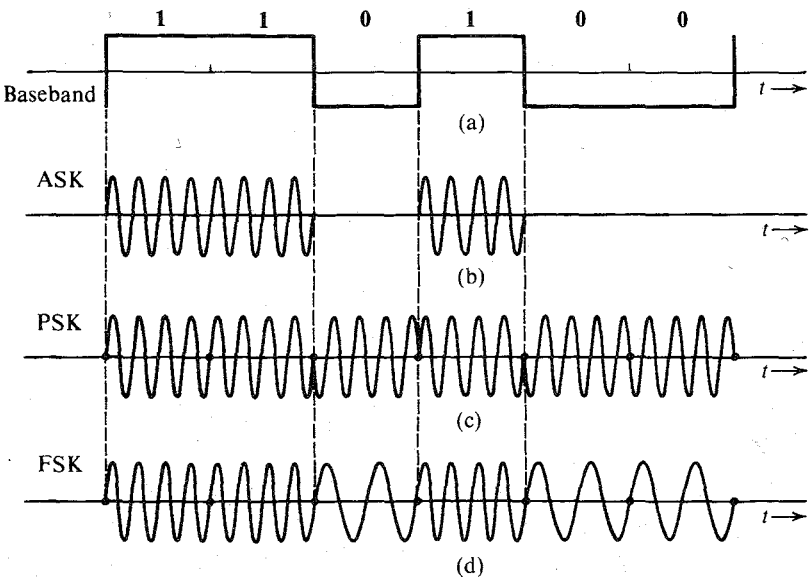


Figure 13.8 Digital modulated waveforms.

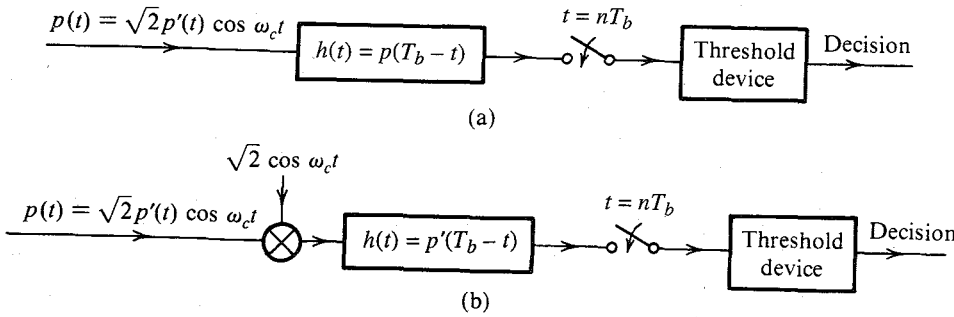


Figure 13.9 Coherent detection of digital modulated signals.

We may also detect the RF pulse by first demodulating it coherently by multiplying it by $\sqrt{2} \cos \omega_c t$. The product is the baseband pulse* $p'(t)$ plus a baseband noise with PSD $\mathcal{N}/2$ (see Example 11.11 and Sec. 12.2), and this is applied to a filter matched to the baseband pulse $p'(t)$. Because $p(t) = \sqrt{2} p'(t) \cos \omega_c t$, $E_p = E_{p'}$, and in this case also

$$\rho^2 = \frac{2E_p}{\mathcal{N}}$$

Hence, the two schemes are equivalent.

Let us consider the cases of PSK, ASK, and FSK individually.

Phase-Shift Keying

This is a case of polar signaling, and the results derived earlier [Eqs. (13.21b, c)] apply. The optimum detector is shown in Fig. 13.9 with threshold 0. From Eq. (13.21b),

$$P_b = Q\left(\sqrt{\frac{2E_b}{\mathcal{N}}}\right) \tag{13.25a}$$

$$\simeq \frac{1}{2\sqrt{\pi E_b/\mathcal{N}}} e^{-E_b/\mathcal{N}} \quad E_b/\mathcal{N} \gg 1 \tag{13.25b}$$

Amplitude-Shift Keying

This is a case of on-off signaling, and Eqs. (13.22) apply. Recall that the optimum detector for on-off signaling is the same as that for polar. Hence, the optimum detector for ASK is the same as that for PSK with threshold $E_p/2$ (Fig. 13.9). From Eqs. (13.22) we have

$$P_b = Q\left(\sqrt{\frac{E_b}{\mathcal{N}}}\right) \tag{13.26a}$$

$$\simeq \frac{1}{\sqrt{2\pi E_b/\mathcal{N}}} e^{-E_b/\mathcal{N}/2} \quad E_b/\mathcal{N} \gg 1 \tag{13.26b}$$

Comparison of Eqs. (13.26) with Eqs. (13.25) shows that for the same performance, the pulse energy in ASK must be twice that in PSK. Hence, ASK requires 3 dB more power than PSK. Thus, in coherent detection, PSK is always preferable to ASK. For this reason, ASK

* There is also a spectrum of $p'(t)$ centered at $2\omega_c$, which is eventually eliminated by the filter matched to $p'(t)$.

is of no practical importance in coherent detection. But ASK can be useful in noncoherent (envelope) detection. In PSK, the information lies in the phase, and, hence, it cannot be detected noncoherently.

The baseband pulses used in carrier systems should be shaped to minimize the ISI. The bandwidth of the PSK or ASK signal is twice that of the corresponding baseband signal because of modulation.*

Frequency-Shift Keying

In FSK, binary 0 and 1 are transmitted by RF pulses $\sqrt{2}p'(t) \cos [\omega_c - (\Delta\omega/2)t]$ and $\sqrt{2}p'(t) \cos [\omega_c + (\Delta\omega/2)t]$, respectively. Such a waveform may be considered to be two interleaved ASK waves. Hence, the PSD will consist of two PSDs, centered at $[f_c - (\Delta f/2)]$ and $[f_c + (\Delta f/2)]$. For a large $\Delta f/f_c$, the PSD will consist of two nonoverlapping PSDs. For a small $\Delta f/f_c$, the two spectra merge, and the bandwidth decreases. But in no case is the bandwidth less than that of ASK or PSK.

The receiver in Fig. 13.5a or b can serve as the optimum receiver. But because the pulses have equal energy, the simplified form of the optimum receiver in Fig. 13.5c is the most convenient. The filters $p(T_b - t)$ and $q(T_b - t)$ are matched to the two RF pulses and can be replaced by respective synchronous demodulators followed by filters matched to the baseband pulse $p'(t)$.

Consider the case:

$$q(t) = \sqrt{2} A \cos \left(\omega_c - \frac{\Delta\omega}{2} \right) t$$

$$p(t) = \sqrt{2} A \cos \left(\omega_c + \frac{\Delta\omega}{2} \right) t$$

To compute P_b from Eq. (13.18b), we need E_{pq} ,

$$\begin{aligned} E_{pq} &= \int_0^{T_b} p(t)q(t) dt \\ &= 2A^2 \int_0^{T_b} \cos \left(\omega_c - \frac{\Delta\omega}{2} \right) t \cos \left(\omega_c + \frac{\Delta\omega}{2} \right) t dt \\ &= A^2 \left[\int_0^{T_b} \cos (\Delta\omega)t dt + \int_0^{T_b} \cos 2\omega_c t dt \right] \\ &= A^2 T_b \left[\frac{\sin (\Delta\omega)T_b}{(\Delta\omega)T_b} + \frac{\sin 2\omega_c T_b}{2\omega_c T_b} \right] \end{aligned}$$

In practice $\omega_c T_b \gg 1$, and the second term on the right-hand side can be ignored. Therefore,

$$E_{pq} = A^2 T_b \operatorname{sinc} (\Delta\omega)T_b$$

Figure 13.10a shows $\operatorname{sinc} (\Delta\omega)T_b$ as a function of $(\Delta\omega)T_b$.

To minimize P_b [Eq. (13.18b)], E_{pq} must be minimized. From Fig. 13.10a, the minimum value of E_{pq} is $-0.217A^2T_b$ and occurs at $(\Delta\omega)T_b = 1.43\pi$ or when

$$\Delta f = \frac{0.715}{T_b} = 0.715R_b$$

* We can also use QAM (quadrature multiplexing) to double the data rate.

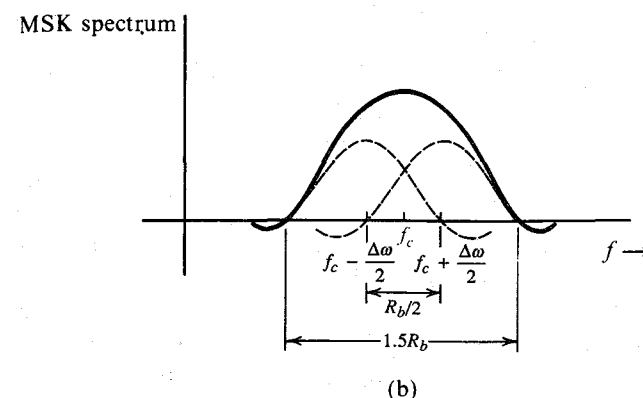
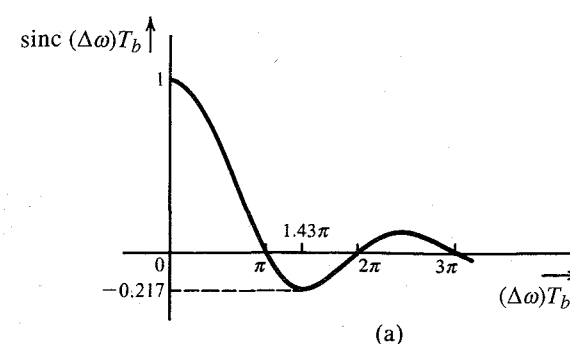


Figure 13.10 FSK and MSK spectra.

Assuming 1 and 0 equiprobable, $E_b = E_p = E_q = A^2 T_b$ and $E_{pq} = -0.217A^2 T_b$,

$$\begin{aligned} P_b &= Q \left(\sqrt{\frac{1.217 A^2 T_b}{\mathcal{N}}} \right) \\ &= Q \left(\sqrt{\frac{1.217 E_b}{\mathcal{N}}} \right) \end{aligned} \quad (13.27a)$$

When $E_{pq} = 0$, we have the case of orthogonal signaling. From Fig. 13.10a, it is clear that $E_{pq} = 0$ for $\Delta f = n/2T_b$, where n is any integer. Larger Δf means wider separation between signaling frequencies $\omega_c - (\Delta\omega/2)$ and $\omega_c + (\Delta\omega/2)$, and consequently larger transmission bandwidth. To minimize the bandwidth, Δf should be as small as possible. The minimum value of Δf that can be used for orthogonal signaling is $1/2T_b$. FSK using this value of Δf is known as **minimum-shift keying (MSK)** or **fast-frequency-shift keying**.

In MSK, abrupt phase changes at the bit-transition instants, characteristic of other FSK implementations, are avoided. FSK schemes where phase continuity is maintained are known as **continuous-phase FSK (CP-FSK)**, of which MSK is one example. These schemes have rapid spectral roll-off and improved efficiency.

To maintain phase continuity in CP-FSK (or MSK), the phase at every bit transition is made dependent on the past data sequence. Consider, for example, the data sequence **1001**... starting at $t = 0$. The first pulse corresponding to the first bit 1 is $\cos [\omega_c + (\Delta\omega/2)t]$ over

the interval 0 to T_b seconds. At $t = T_b$, this pulse ends with a phase $[\omega_c + (\Delta\omega/2)]T_b$. The next pulse, corresponding to the second data bit 0, is $\cos [\omega_c - (\Delta\omega/2)]t$. This pulse is given additional phase $[\omega_c + (\Delta\omega/2)]T_b$ in order to maintain phase continuity at the transition instant. We continue this way at each transition.

MSK being an orthogonal scheme, its error probability is given by

$$P_b = Q\left(\sqrt{\frac{E_b}{N}}\right) \quad (13.27b)$$

This performance appears inferior to that of the optimum case in Eq. (13.27a). Closer examination shows that MSK is actually superior to the so-called optimum case. If MSK were coherently detected as ordinary FSK using an observation interval of T_b , MSK would have $P_b = Q(\sqrt{E_b/N})$. But recall that MSK is CP-FSK, where the phase of each pulse is dependent on the past data sequence. Hence, better performance may be obtained by observing the received waveform over a period longer than T_b . It can be shown that if MSK is detected using an observation interval of $2T_b$, the performance of MSK is identical to that of PSK, that is,

$$P_b = Q\left(\sqrt{\frac{2E_b}{N}}\right) \quad (13.27c)$$

MSK also has other useful properties. It has self-synchronization capabilities and its bandwidth is only $1.5R_b$, as shown in Fig. 13.10b. This is only 50% higher than for duobinary signaling. Moreover, the MSK spectrum decays much more rapidly as $1/f^4$, in contrast to the PSK (or bipolar) spectrum, which decays only as $1/f^2$ [see Eqs. (7.14) and (7.21)]. Because of these properties, MSK has received a great deal of attention recently. More discussion of MSK can be found in the references.^{1,2}

13.3.2 Noncoherent Detection

If the phase Θ in the received RF pulse $\sqrt{2}p'(t) \cos(\omega_c t + \Theta)$ is unknown, we cannot use coherent detection techniques but must rely on noncoherent techniques, such as envelope detection. It can be shown^{3,4} that when the phase Θ of the received pulse is random and uniformly distributed over $(0, 2\pi)$, the optimum detector is a filter matched to the RF pulse $\sqrt{2}p'(t) \cos \omega_c t$ followed by an envelope detector, a sampler (to sample at $t = T_b$), and a comparator to make the decision (Fig. 13.11).

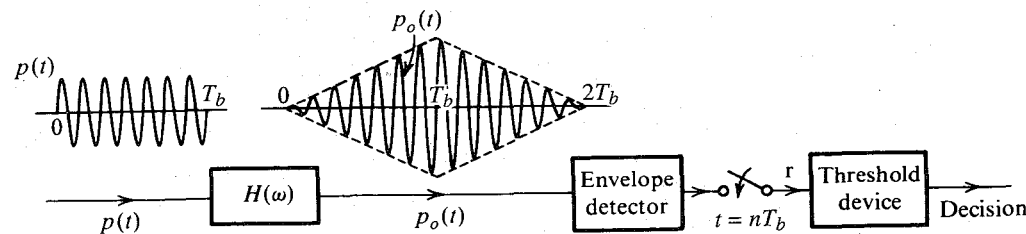


Figure 13.11 Noncoherent detection of digital modulated signals.

Amplitude-Shift Keying

The noncoherent detector for ASK is shown in Fig. 13.11. The filter $H(\omega)$ is a filter matched to the RF pulse, ignoring the phase. This means the filter output amplitude A_p will not necessarily be maximum at the sampling instant. But the envelope will be close to maximum at the sampling instant (Fig. 13.11). The matched filter output is now detected by an envelope detector. The envelope is sampled at $t = T_b$ for making the decision.

When a 1 is transmitted, the output of the envelope detector at $t = T_b$ is an envelope of a sine wave of amplitude A_p in a gaussian noise of variance σ_n^2 . In this case, the envelope r has a rician density, given by [Eq. (11.64a)]

$$p_r(r|m=1) = \frac{r}{\sigma_n^2} e^{-(r^2 + A_p^2)/2\sigma_n^2} I_0\left(\frac{rA_p}{\sigma_n^2}\right) \quad (13.28a)$$

Also, when $A_p \gg \sigma_n$ (small-noise case) from Eq. (11.64c), we have

$$p_r(r|m=1) \simeq \sqrt{\frac{r}{2\pi A_p \sigma_n^2}} e^{-(r-A_p)^2/2\sigma_n^2} \quad (13.28b)$$

$$\simeq \frac{1}{\sigma_n \sqrt{2\pi}} e^{-(r-A_p)^2/2\sigma_n^2} \quad (13.28c)$$

Observe that for small noise, the PDF of r is practically gaussian, with mean A_p and variance σ_n^2 . When 0 is transmitted, the output of the envelope detector is an envelope of a gaussian noise of variance σ_n^2 . The envelope in this case has a Rayleigh density, given by [Eq. (11.59)]

$$p_r(r|m=0) = \frac{r}{\sigma_n^2} e^{-r^2/2\sigma_n^2}$$

Both $p_r(r|m=1)$ and $p_r(r|m=0)$ are shown in Fig. 13.12. Using the argument used earlier (see Fig. 13.4), the optimum threshold is found to be the point where the two densities intersect. Hence, the optimum threshold a_o is

$$\frac{a_o}{\sigma_n^2} e^{-(a_o^2 + A_p^2)/2\sigma_n^2} I_0\left(\frac{A_p a_o}{\sigma_n^2}\right) = \frac{a_o}{\sigma_n^2} e^{-a_o^2/2\sigma_n^2}$$

or

$$e^{-A_p^2/2\sigma_n^2} I_0\left(\frac{A_p a_o}{\sigma_n^2}\right) = 1$$

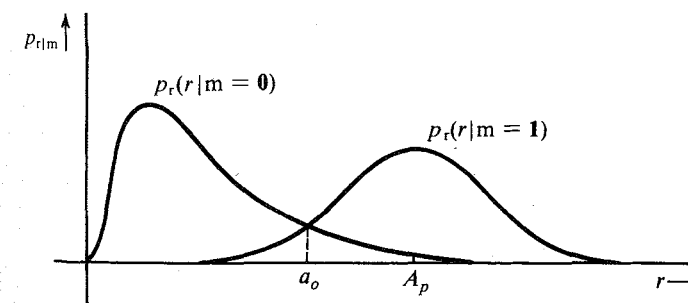


Figure 13.12 Conditional PDFs in noncoherent detection of ASK signals.

This equation is satisfied to a close approximation for

$$a_o = \frac{A_p}{2} \sqrt{1 + \frac{8\sigma_n^2}{A_p^2}}$$

Because the matched filter is used, $A_p = E_p$ and $\sigma_n^2 = \mathcal{N}E_p/2$ [see Eqs. (13.9a, b)]. Moreover, for ASK there are, on the average, only $R_b/2$ pulses per second. Thus, $E_b = E_p/2$. Hence,

$$\left(\frac{A_p}{\sigma_n}\right)^2 = \frac{2E_p}{\mathcal{N}} = 4\frac{E_b}{\mathcal{N}}$$

and

$$a_o = E_b \sqrt{1 + \frac{2}{E_b/\mathcal{N}}} \quad (13.29a)$$

Observe that the optimum threshold is not constant but depends on $E_b/\mathcal{N}$. This is a serious drawback in a fading channel. For a strong signal, $E_b/\mathcal{N} \gg 1$,

$$a_o \simeq E_b \quad (13.29b)$$

and

$$\begin{aligned} P(\epsilon|m=0) &= \int_{A_p/2}^{\infty} p_r(r|m=0) dr \\ &= \int_{A_p/2}^{\infty} \frac{r}{\sigma_n^2} e^{-r^2/2\sigma_n^2} dr \\ &= e^{-A_p^2/8\sigma_n^2} \\ &= e^{-\frac{1}{2}E_b/\mathcal{N}} \end{aligned} \quad (13.30)$$

Also,

$$P(\epsilon|m=1) = \int_{-\infty}^{A_p/2} p_r(r|m=1) dr$$

Evaluation of this integral is somewhat cumbersome.⁵ For a strong signal (that is, for $E_b/\mathcal{N} \gg 1$), the rician PDF can be approximated by the gaussian PDF [Eq. (11.64c)], and

$$\begin{aligned} P(\epsilon|m=1) &= \frac{1}{\sigma_n \sqrt{2\pi}} \int_{-\infty}^{A_p/2} e^{-(r-A_p)^2/2\sigma_n^2} dr \\ &= Q\left(\frac{A_p}{2\sigma_n}\right) \\ &= Q\left(\sqrt{\frac{E_b}{\mathcal{N}}}\right) \end{aligned} \quad (13.31)$$

Hence,

$$P_b = P_m(0)P(\epsilon|m=0) + P_m(1)P(\epsilon|m=1)$$

Assuming $P_m(1) = P_m(0) = 0.5$,

$$P_b = \frac{1}{2} \left[e^{-\frac{1}{2}E_b/\mathcal{N}} + Q\left(\sqrt{\frac{E_b}{\mathcal{N}}}\right) \right] \quad (13.32a)$$

Using the approximation in Eq. (10.36a),

$$P_b \simeq \frac{1}{2} \left(1 + \frac{1}{\sqrt{2\pi E_b/\mathcal{N}}} \right) e^{-\frac{1}{2}E_b/\mathcal{N}} \quad E_b/\mathcal{N} \gg 1 \quad (13.32b)$$

$$\simeq \frac{1}{2} e^{-\frac{1}{2}E_b/\mathcal{N}} \quad (13.32c)$$

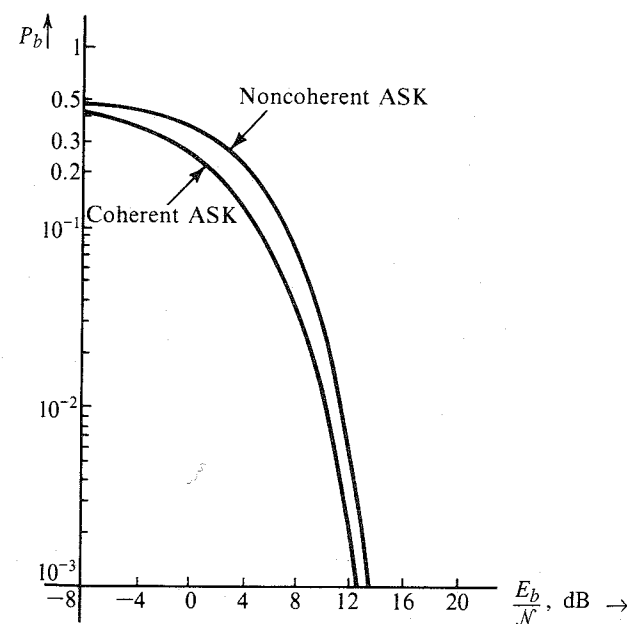
Note that in an optimum receiver, for $E_b/\mathcal{N} \gg 1$, $P(\epsilon|m=1)$ is much smaller than $P(\epsilon|m=0)$. For example, at $E_b/\mathcal{N} = 10$, $P(\epsilon|m=0) \simeq 8.7 P(\epsilon|m=1)$. Hence, mistaking 0 for 1 is the type of error that predominates. The timing information in noncoherent detection is extracted from the envelope of the received signal by methods discussed in Sec. 7.5.

For a coherent detector,

$$\begin{aligned} P_b &= Q\left(\sqrt{\frac{E_b}{\mathcal{N}}}\right) \\ &\simeq \frac{1}{\sqrt{2\pi E_b/\mathcal{N}}} e^{-\frac{1}{2}E_b/\mathcal{N}} \quad E_b/\mathcal{N} \gg 1 \end{aligned} \quad (13.33)$$

This appears similar to Eq. (13.32c) (the noncoherent case). Thus for a large $E_b/\mathcal{N}$, the performances of the coherent detector and the envelope detector are similar (Fig. 13.13). This is similar to the behavior observed in the case of analog signals.

Figure 13.13 Error probability of ASK.



Frequency-Shift Keying

The noncoherent receiver for FSK is shown in Fig. 13.14. The filters $H_0(\omega)$ and $H_1(\omega)$ are matched to the two RF pulses corresponding to 0 and 1, respectively. The outputs of the envelope detectors at $t = T_b$ are r_0 and r_1 , respectively. The noise components of outputs of filters $H_0(\omega)$ and $H_1(\omega)$ are the gaussian RVs n_0 and n_1 , respectively, with $\sigma_{n_0} = \sigma_{n_1} = \sigma_n$.

If 1 is transmitted ($m = 1$), then at the sampling instant, the envelope r_1 has the rician PDF*

$$p_{r_1}(r_1) = \frac{r_1}{\sigma_n^2} e^{-(r_1^2 + A_p^2)/2\sigma_n^2} I_0\left(\frac{r_1 A_p}{\sigma_n^2}\right)$$

and r_0 is the noise envelope with Rayleigh density

$$P_{r_0}(r_0) = \frac{r_0}{\sigma_n^2} e^{-r_0^2/2\sigma_n^2}$$

The decision is $m = 1$ if $r_1 > r_0$ and $m = 0$ if $r_1 < r_0$. Hence, when binary 1 is transmitted, an error is made if $r_0 > r_1$,

$$P(\epsilon|m = 1) = P(r_0 > r_1)$$

The event $r_0 > r_1$ is the same as the joint event " r_1 has any positive value† and r_0 has a value greater than r_1 ." This is simply the joint event ($0 < r_1 < \infty, r_0 > r_1$). Hence,

$$\begin{aligned} P(\epsilon|m = 1) &= P(0 < r_1 < \infty, r_0 > r_1) \\ &= \int_0^\infty \int_{r_1}^\infty p_{r_1 r_0}(r_1, r_0) dr_1 dr_0 \end{aligned}$$

Because r_1 and r_0 are independent, $p_{r_1 r_0} = p_{r_1} p_{r_0}$. Hence,

$$\begin{aligned} P(\epsilon|m = 1) &= \int_0^\infty \frac{r_1}{\sigma_n^2} e^{-(r_1^2 + A_p^2)/2\sigma_n^2} I_0\left(\frac{r_1 A_p}{\sigma_n^2}\right) \int_{r_1}^\infty \frac{r_0}{\sigma_n^2} e^{-r_0^2/2\sigma_n^2} dr_1 dr_0 \\ &= \int_0^\infty \frac{r_1}{\sigma_n^2} e^{-(2r_1^2 + A_p^2)/2\sigma_n^2} I_0\left(\frac{r_1 A_p}{\sigma_n^2}\right) dr_1 \end{aligned}$$

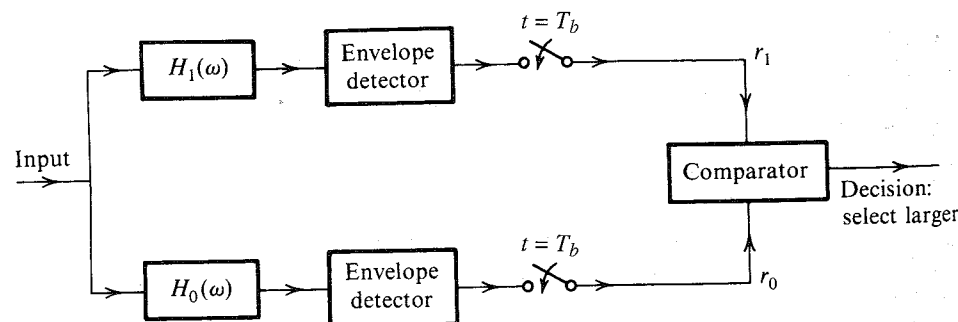


Figure 13.14 Noncoherent detection of binary FSK.

* An orthogonal FSK is assumed. This ensures that r_0 and r_1 have Rayleigh and Rice densities, respectively, when 1 is transmitted.

† r_1 is the envelope detector and can take only positive values.

Letting $x = \sqrt{2} r_1$ and $\alpha = A_p/\sqrt{2}$, we have

$$P(\epsilon|m = 1) = \frac{1}{2} e^{-A_p^2/4\sigma_n^2} \int_0^\infty \frac{x}{\sigma_n^2} e^{-(x^2 + \alpha^2)/2\sigma_n^2} I_0\left(\frac{x\alpha}{\sigma_n^2}\right) dx$$

Observe that the integrand is a rician density, and, hence, its integral is unity. Therefore,

$$P(\epsilon|m = 1) = \frac{1}{2} e^{-A_p^2/4\sigma_n^2} \quad (13.34a)$$

Note that for a matched filter,

$$\rho_{\max}^2 = \frac{A_p^2}{\sigma_n^2} = \frac{2E_p}{\mathcal{N}}$$

For FSK, $E_b = E_p$, and Eq. (13.34a) becomes

$$P(\epsilon|m = 1) = \frac{1}{2} e^{-\frac{1}{2} E_b/\mathcal{N}} \quad (13.34b)$$

Similarly,

$$P(\epsilon|m = 0) = \frac{1}{2} e^{-\frac{1}{2} E_b/\mathcal{N}} \quad (13.34c)$$

and

$$P_b = \frac{1}{2} e^{-\frac{1}{2} E_b/\mathcal{N}} \quad (13.35)$$

This behavior is similar to that of noncoherent ASK [Eq. (13.32c)]. Again we observe that for $E_b/\mathcal{N} \gg 1$, the performance of coherent and noncoherent FSK are essentially similar.

From the practical point of view, FSK is to be preferred over ASK because FSK has a fixed optimum threshold, whereas the optimum threshold of ASK depends on $E_b/\mathcal{N}$ (the signal level). Hence, ASK is particularly susceptible to signal fading. Because the decision requires a comparison between r_0 and r_1 , this problem does not arise in FSK. This is the outstanding advantage of noncoherent FSK over noncoherent ASK. In addition, unlike noncoherent ASK, probabilities $P(\epsilon|m = 1)$ and $P(\epsilon|m = 0)$ are equal in noncoherent FSK. The disadvantage of FSK is that it requires a larger bandwidth than ASK.

Differentially Coherent PSK

Just as it is impossible to demodulate a DSB-SC signal with an envelope detector, it is also impossible to demodulate PSK (which is really DSB-SC) noncoherently. We can, however, demodulate PSK without the synchronous, or coherent, local carrier by using what is known as **differentially coherent PSK (DPSK)**.

The optimum receiver is shown in Fig. 13.15. This receiver is very much like a correlation detector (Fig. 13.3), which is equivalent to a matched-filter detector. In a correlation detector, we multiply pulse $p(t)$ by a locally generated pulse $p(t)$. In the case of DPSK, we take advantage of the fact that the two RF pulses used in transmission are identical except for the sign. In the detector in Fig. 13.15, we multiply the incoming pulse by the preceding pulse. Hence, the preceding pulse serves as a substitute for the locally generated pulse. The only difference is that the preceding pulse is noisy because of channel noise, and this tends to degrade the performance in comparison to coherent PSK. When the output r is positive, the present pulse is identical to the previous one, and when r is negative, the present pulse is the negative of the previous pulse. Hence, from the knowledge of the first reference digit, it is

possible to detect all the received digits. Detection is facilitated by using so-called *differential coding*, as discussed in Sec. 7.3.

In order to derive the DPSK error probability, we observe that DPSK using differential coding is essentially an orthogonal signaling scheme. A binary 1 is transmitted by a sequence of two pulses (p, p) or ($-p, -p$) over $2T_b$ seconds (no transition). Similarly, a binary 0 is transmitted by a sequence of two pulses ($p, -p$) or ($-p, p$) over $2T_b$ seconds (transition). Either of the pulse sequences used for binary 1 is orthogonal to either of the pulse sequences used for binary 0. Because no local carrier is generated for demodulation, the detection is noncoherent, with an effective pulse energy equal to $2E_p$ (twice the energy of pulse p). The actual energy transmitted per digit is only E_p , however, the same as in noncoherent FSK. Consequently, the performance of DPSK is 3 dB superior to that of noncoherent FSK. Hence from Eq. (13.35), we can write P_b for DPSK as

$$P_b = \frac{1}{2} e^{-E_b/\mathcal{N}} \quad (13.36)$$

This error probability (Fig. 13.16) is superior to that of noncoherent FSK by 3 dB and is essentially similar to coherent PSK for $E_b/\mathcal{N} \gg 1$ [Eq. (13.25b)]. This is as expected, because we saw earlier that DPSK appears similar to PSK. Rigorous derivation of Eq. (13.36) can be found in the literature.⁶

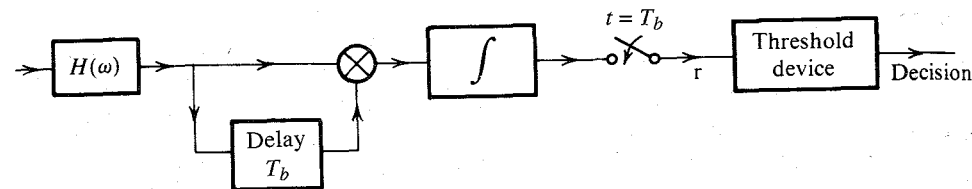


Figure 13.15 Differential PSK detection.

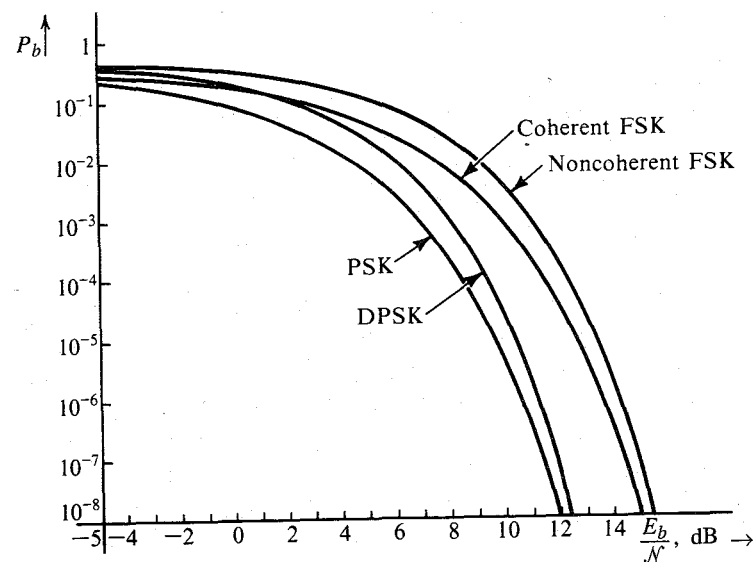


Figure 13.16 Error probability of PSK, DPSK, and coherent and noncoherent FSK.

13.4 PERFORMANCE OF SPREAD SPECTRUM SYSTEMS

In Sec. 9.2, we explained how spread spectrum systems work. Spread spectrum is basically a broad-band modulation scheme where the baseband spectrum is spread by using a pseudo-random (PN) code. At the receiver, a replica of the same PN code is used to “despread” the received signal to recover the original data. The key concept in spread spectrum systems is the PN code sequence, which is generally periodic and consists of periodic coded sequence of 1's and 0's with interesting autocorrelation property.

13.4.1 PN Sequence Generation

A good PN sequence is characterized by an autocorrelation that is similar to that of a white noise. This means the autocorrelation function of a PN sequence should be high near $\tau = 0$ and low for all $\tau \neq 0$, as shown in Fig. 13.17a. Moreover, in CDMA applications several users share the same band using different PN sequences. Hence, it is necessary that the crosscorrelation among different pairs of PN sequences be small to reduce mutual interference.

A PN code is periodic. A shift-register network with output feedback can generate a sequence with long periods and low susceptibility to structural identification by an outsider. The most widely known binary PN sequences are the **maximum-length** shift-register sequences (*m*-sequences). Such a sequence can be generated by an *m*-stage shift register with suitable feedback connections and has a length $n = 2^m - 1$ bits, which is a the maximum period for such a finite state machine. Figure 13.17b shows a shift-register encoder for $m = 6$ and $n = 63$.

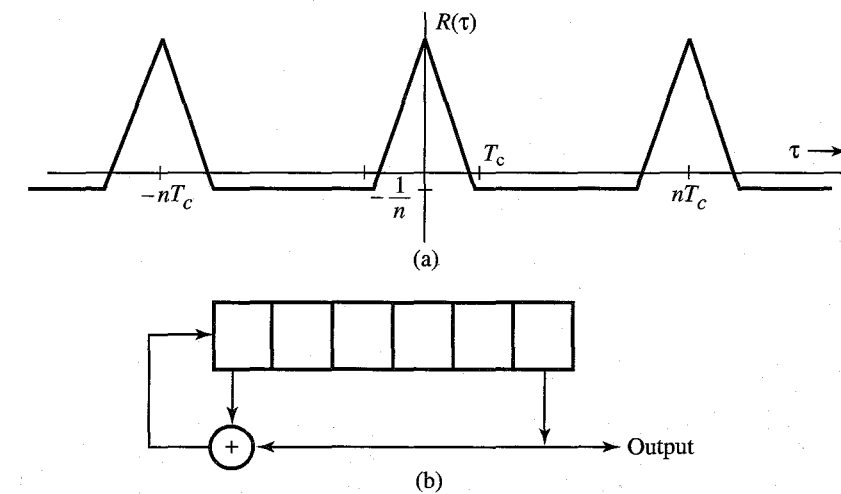


Figure 13.17 (a) PN sequence autocorrelation function. (b) Six-stage generator of a maximum-length PN sequence.

13.4.2 Optimum Detection of DS/SS PSK

The signal $m(t)$ is a polar signal consisting of a sequence of NRZ pulses of duration T_b and constant amplitude 1 or -1 .^{*} This signal is multiplied by the PN code $c(t)$ and then DSB-modulated with a carrier $\sqrt{2} \cos(\omega_c t + \theta)$.[†] Although the message signal here is $m(t)$ with binary values ± 1 , it is helpful to view the situation in a different way. During each bit interval T_b , we are transmitting either $\sqrt{2} c(t) \cos(\omega_c t + \theta)$ or $-\sqrt{2} c(t) \cos(\omega_c t + \theta)$. Hence, the basic binary pulse may be taken as $\sqrt{2} c(t) \cos(\omega_c t + \theta)$ [instead of $m(t)$], and the optimum correlation receiver would be of the form shown in Fig. 13.18 (see Fig. 13.3). The integrator integrates over the interval 0 to T_b , and its output is sampled at $t = T_b$ to decide whether the transmitted pulse was $\sqrt{2} c(t) \cos(\omega_c t + \theta)$ or $-\sqrt{2} c(t) \cos(\omega_c t + \theta)$; that is, whether $m(t)$ was 1 or -1 . The low-pass filter shown in the receiver is not necessary because the subsequent integrator itself acts as a low-pass filter. We have shown the filter explicitly to simplify understanding of the receiver.

In Eq. (13.21b) we showed that for a channel with (white) noise of PSD $\mathcal{N}/2$, the polar optimum receiver error probability is given by

$$P_b = Q\left(\sqrt{\frac{2E_b}{\mathcal{N}}}\right) \quad (13.37)$$

where E_b is the energy per bit (energy of one pulse). This result appears rather surprising in view of the fact that the error probability of an optimum receiver is unchanged whether we use spread spectrum or not. We have seen in Sec. 9.2 how a spread spectrum system can reduce the power of an interfering (jamming) signal by N , the processing gain. Is noise different from an intentional interference? A heuristic explanation of this apparent anomaly lies in the fact that the spread spectrum reduces the interfering signal power by spreading its spectrum

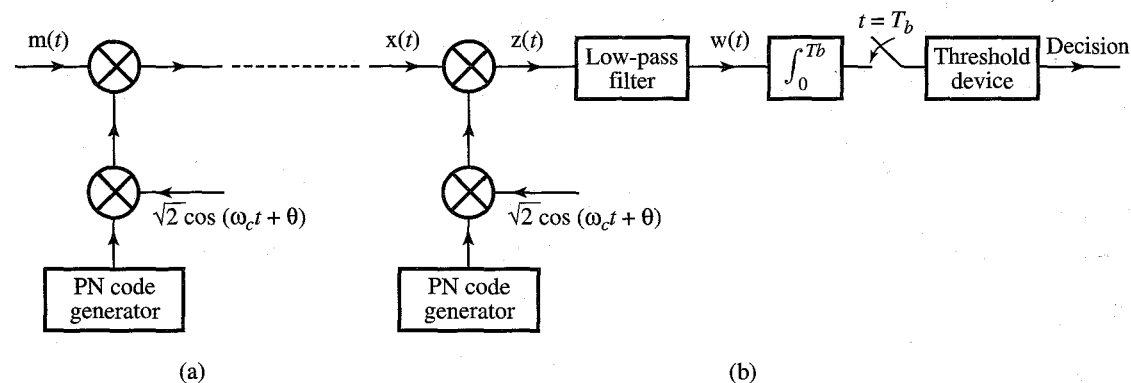


Figure 13.18 DS/SS PSK system. (a) Transmitter. (b) Receiver.

^{*} In practice, the amplitude may be $\pm A$. This makes no difference for the optimum receiver.

[†] The multipliers $\sqrt{2}$ in the carriers of the transmitter and the receiver in Fig. 13.18 are for convenience only. The final results (for error probability) are independent of the values of these multipliers. The choice of $\sqrt{2}$ is dictated by the fact that it leaves the power of the modulated (or the demodulated) signal identical to the signal before modulation (or demodulation). In the literature, the multiplier at the transmitter is often chosen as $\sqrt{2P}$ to ensure the transmitted signal power to be P .

over a very wide band. But the noise is a broad-band signal to begin with, and it cannot be effectively spread any more. The spectral spreading can be realized only if the interference signal bandwidth is on the order of the baseband signal.

In Sec. 9.2 we showed that a spread spectrum system reduces the power of the interfering signal by $N = T_b/T_c$. Hence, the effective noise power is the channel noise power plus the jamming signal power divided by N .

13.4.3 Multiple-Access Performance of DS/SS*

To analyze a DS/SS system with K multiple-access users, we compute the interference at the output of a certain receiver caused by the remaining $K - 1$ users. Let $m_1(t)$ and $m_k(t)$ be the data signals and $c_1(t)$ and $c_k(t)$ the spreading signals (PN codes) for the first user and the k th user, respectively. Let their respective carriers be $\sqrt{2} \cos(\omega_c t + \theta_1)$ and $\sqrt{2} \cos(\omega_c t + \theta_k)$. Moreover the two received signals will have different delays. Let the relative delay of the k th signal with respect to the first signal be τ_k . Hence, the signal $x_k(t)$ from the k th user at the input of the receiver of the first user is

$$\begin{aligned} x_k(t) &= \sqrt{2} m_k(t - \tau_k) c_k(t - \tau_k) \cos[\omega_c(t - \tau_k) + \theta_k] \\ &= \sqrt{2} m_k(t - \tau_k) c_k(t - \tau_k) \cos(\omega_c t + \phi_k) \quad \phi_k = \theta_k - \omega_c \tau_k \end{aligned}$$

The receiver multiplies $x_k(t)$ with $\sqrt{2} c_1(t) \cos(\omega_c t + \theta_1)$ to yield

$$z_k(t) = 2 m_k(t - \tau_k) c_k(t - \tau_k) c_1(t) \cos(\omega_c t + \phi_k) \cos(\omega_c t + \theta_1) \quad (13.38)$$

The high-frequency component is suppressed by the low-pass filter to yield

$$w_k(t) = m_k(t - \tau_k) c_k(t - \tau_k) c_1(t) \cos(\phi_k - \theta_1) \quad (13.39)$$

The output (interference) of the integrator sampler is

$$I_k = \cos(\phi_k - \theta_1) \int_0^{T_b} m_k(t - \tau_k) c_k(t - \tau_k) c_1(t) dt \quad (13.40a)$$

Recall that $m_k(t)$ is a polar NRZ binary signal with amplitudes ± 1 . The bit duration is T_b , and there is a possible change of amplitude of $m_k(t - \tau_k)$ at $t = \tau_k$. Let this amplitude be b_{-1} before $t = \tau_k$ and b_0 after $t = \tau_k$. Then

$$I_k = \cos(\phi_k - \theta_1) \left[b_{-1} \int_0^{\tau_k} c_k(t - \tau_k) c_1(t) dt + b_0 \int_{\tau_k}^{T_b} c_k(t - \tau_k) c_1(t) dt \right] \quad (13.40b)$$

This is the interference from the k th user at the output of the first user. This expression makes explicit the dependence of the interference on the crosscorrelation of the spreading codes $c_1(t)$ and $c_k(t)$. Ideally, we would like $\int c_k(t - \tau_k) c_1(t) dt = 0$ for any value of τ_k . This is not possible in practice, but it is possible to find many sets of sequences with good crosscorrelation properties.⁷

The Gaussian Approximation

The multiple-access interference (MAI) can be modeled as the sum of independent Bernoulli trials. One frequent approximation is to apply the central limit theorem, which implies that

^{*} Rest of the Section 13.4 is based on the material contributed by Prof. B. D. Woerner and R. M. Buehrer.

the sum of these tiny effects (interferences) tends toward a gaussian distribution. Since there are $K - 1$ independent identically distributed interferers, the total MAI, $I = \sum_{k=2}^K I_k$, may be approximated by a gaussian random variable. It can be shown that the resulting bit error rate P_b is given by^{8,9}

$$P_b = Q \left(\frac{1}{\sqrt{(K-1)/3N + \mathcal{N}/2E_b}} \right) \quad (13.41)$$

where N is the processing gain. We can derive this result heuristically as follows. The power of the interferer's signal $m_k(t)$ is $E_b T_b$. But because this signal is despread by the first receiver, the power of the interfering signal $w_k(t)$ is only $E_b T_b / N$ at the first receiver. Moreover, the bandwidth of a polar NRZ binary signal is T_b Hz. Hence, the PSD of the interfering signal $w_k(t)$ is E_b / N . Also there are $K - 1$ identically distributed interfering signals at the first receiver whose PSDs would add to yield the PSD $(K - 1)E_b / N$. However, each interfering signal has a different relative time delay τ_k , which effectively results in power reduction by a factor of 3 [recall the factor $\cos(\phi_k - \theta_1)$ in Eqs. (13.40)] to yield the interference PSD as $(K - 1)E_b / 3N$. Consequently, the noise PSD at the first receiver is not $\mathcal{N}/2$, but $(K - 1)E_b / 3N + \mathcal{N}/2$. When we substitute this PSD in place of $\mathcal{N}/2$ in Eq. (13.37), we obtain Eq. (13.41).

Observe that when a single user is present ($K = 1$), Eq. (13.41) reduces to Eq. (13.37), as expected. When the signal-to-noise ratio is very high ($E_b / \mathcal{N} \rightarrow \infty$), we obtain

$$\lim_{E_b / \mathcal{N} \rightarrow \infty} P_b = Q \left(\sqrt{\frac{3N}{K-1}} \right)$$

This shows the presence of an irreducible error floor for the MAI limited case.*

13.4.4 The Near/Far Problem

Equation (13.41) has been derived under the assumption that signals from all users are received with the same signal power. Within a mobile system, such an assumption may be reasonable for the forward link in which all transmissions from a centrally located base station originate with the same signal power and follow the same transmission path. However, this assumption is not realistic on the reverse link from the mobile to the base station because signals originate from diverse and moving locations. This effect can be partially compensated for by power control, which takes two forms. Under open loop power control, a mobile station adjusts its power based on the strength of the signal it receives from the base station. This presumes that a reciprocal relationship exists between forward and reverse links, an assumption that may not hold if the links operate in different frequency bands. As a result, closed loop power control is often required in which the base station orders changes in the mobile station's transmitted power.

* The gaussian approximation has limitations when used to predict system performance. While the central limit theorem implies that I will tend toward a gaussian distribution near the center of its distribution, convergence may be slow at its tails. Unfortunately, it is the extreme values of I that tend to produce errors. As a result, the gaussian distribution is optimistic for low BER.¹⁰ The problem is worsened in situations where there is a single strong interferer or a single strong multipath component. This is of concern because low BERs are difficult to simulate, so we particularly desire analytical techniques that are accurate at low BERs. Several improvements to the gaussian approximation have been proposed which represent MAI more accurately under a wide range of conditions.¹⁰⁻¹²

The effects of the near-far problem and power control may be investigated using the analytic techniques described earlier with nonequal signal powers. Although the accuracy of the gaussian approximation is reduced in the case of the near-far problem, quantitative results may still be obtained because the central limit theorem may be applied to the chips from each interferer as well as the number of interferers K .

By extending the qualitative interpretation of Eq. (13.41), the effects of the near-far problem can be found as follows. If users have unequal powers, each will contribute $E_b^{(k)} / 3N$ to the noise level, where $E_b^{(k)}$ is the energy per bit for user k . Adding these as before to the noise PSD, we can find a simple equation for the error probability of user 1, similar to Eq. (13.41):

$$P_b = Q \left[\frac{1}{\sqrt{\sum_{k=2}^K E_b^{(k)} / 3E_b^{(1)} N + \mathcal{N}/2E_b^{(1)}}} \right] \quad (13.42)$$

Equation (13.42) illustrates the deleterious effect of the near-far problem. If just a single user has a significantly higher power level than the desired user, that interferer will dominate the performance. Substituting $E_b^{(k)} = E_b^{(1)}$ results in Eq. (13.41), as expected.

As an example, the effect of imperfect power control is investigated by Woerner and Cameron.¹³ Field trial results have indicated that even after the operation of open and closed loop power control, the received signal level in the IS-95 digital cellular CDMA system reverse channel varies according to a log-normal distribution with a variance of 1 to 2 dB. The BER for this model is calculated using the method of Lehnert.¹⁰ For a wireless voice transmission system, BERs of 10^{-2} to 10^{-3} are required for satisfactory performance. Thus the capacity may be considered to be the maximum number of users for which the BER does not exceed a fixed threshold. The results of Fig. 13.19 indicate that even after power control is applied, the variance of received power levels can result in capacity losses of 10 to 30% in a CDMA system.

13.4.5 Multipath Propagation

Thus far we have assumed that each user's signal is transmitted via a single path. However, the wireless channel is characterized by multipath propagation as signals reflect and scatter off objects in the transmission path. The constructive and destructive interference between arriving multipath components can result in severe fading in narrow-band systems. The ability of spread spectrum signals to reject and even exploit multipath propagation is a key advantage in the wireless environment. A spread spectrum signal can resolve multipath components if the chip duration is of the same order as the multipath delay. In narrow-band systems, delays are normally much smaller than a bit period, making them indistinguishable.

Since the desired user's spreading code will have low autocorrelation, we can treat these multipath components as additional interferers provided that they are at least one chip apart. Practical experience indicates that at low microwave frequencies, 1 MHz of spread bandwidth is required to resolve a significant amount of the multipath components, and 10 MHz of bandwidth will resolve nearly all.¹⁴ Returning to the case of equal power levels, we can model each individual path as an interferer contributing $E_b / 3N$ to the noise level. Thus, there are $K - 1$ users with L paths along with $L - 1$ interfering versions of the desired signal contributing

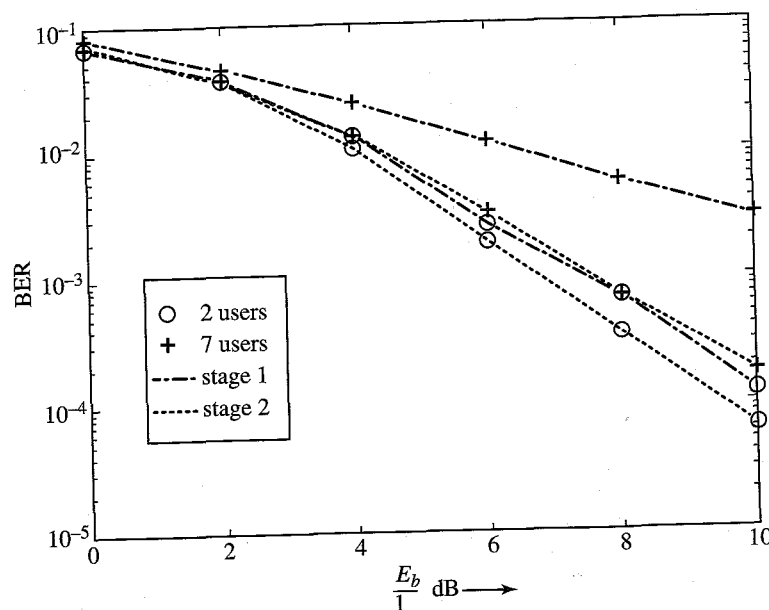


Figure 13.19 Error probability for number of users under imperfect power control.

to the noise level. This results in a total noise level of $(KL - 1)E_b/3N$ being added to the overall noise PSD. Utilizing this in Eq. (13.41) results in

$$P_b = Q \left[\frac{1}{\sqrt{(KL - 1)/3N + N/2E_b}} \right] \quad (13.43)$$

Though the numerical accuracy of this equation is limited, it qualitatively shows how the performance degrades as the number of arriving multipath components L increases. Also we can see that for $L = 1$, Eq. (13.43) reduces to Eq. (13.41). If there is only a single user ($K = 1$), this equation shows that performance is still below optimum due to the multipath. Equation (13.43) assumes that the multipath and interferers have power levels of the same magnitude as the desired user's primary path. Although this is rarely the case, Eq. (13.43) can be extended similarly to Eq. (13.42) for unequal power levels.

In addition to mitigating the multipath propagation, a spread spectrum may actually exploit this effect by using a Rake receiver.^{15,16} A Rake receiver is a device for coherently combining the energy from two or more received multipath components, thereby increasing the received signal power and providing a form of diversity reception. The Rake consists of a bank of correlation receivers, with each individual receiver correlating with a different arriving multipath component. By adjusting for the delays, the individual multipath components can contribute correlated energy to the decision statistic, rather than harmful interference. Analogous to a garden rake, the Rake receiver gathers as much signal energy as possible.

13.4.6 Performance of Frequency Hopped Systems

Although DS/SS is currently more commercially important, some commercial applications for FH/SS are emerging. FH/SS is used in some spread spectrum devices which provide low rate

data communication in the unlicensed ISM frequency band from 902 to 928 MHz, and both the cellular digital packet data (CDPD) and GSM standards have options for incorporating frequency hopping.

Normally noncoherent detection is used for FH/SS because of the difficulty in maintaining an accurate carrier phase reference under changing frequency conditions. For noncoherent detection, we have shown that [see Eq. (13.32c)]

$$P_b = \frac{1}{2} e^{-E_b/2N} \quad (13.44)$$

However, if two users transmit simultaneously in the same frequency band, a collision or "hit" occurs. In this case we will assume that the probability of error is $\frac{1}{2}$. (This is actually pessimistic since studies have shown that this value can be lower.) Thus the overall probability of bit error can be modeled as

$$P_b = \frac{1}{2} e^{-E_b/2N} (1 - P_h) + \frac{1}{2} P_h \quad (13.45)$$

where P_h is the probability of a hit, which we must determine. If there are J frequency slots, there is a $1/J$ probability that a given interferer will be present in the desired user's slot. If there are $K - 1$ interferers or other users, the probability that at least one is present in the desired frequency slot is

$$P_h = 1 - \left(1 - \frac{1}{J}\right)^{K-1} \approx \frac{K-1}{J} \quad (13.46)$$

assuming J is large. Substituting this into Eq. (13.45) gives

$$P_b = \frac{1}{2} e^{-E_b/2N} \left(1 - \frac{K-1}{J}\right) + \frac{1}{2} \frac{K-1}{J} \quad (13.47)$$

If $K = 1$, the probability of error reduces to Eq. (13.44), the standard probability of error for BFSK. Also, if we allow E_b/N to approach infinity, we see that

$$\lim_{E_b/N \rightarrow \infty} P_b = \frac{1}{2} \frac{K-1}{J} \quad (13.48)$$

which illustrates the irreducible error rate due to MAI.

The previous analysis assumes that all users hop their carrier frequencies synchronously. This is called **slotted** frequency hopping. This may not be a realistic scenario for many FH/SS systems. Even when synchronization can be achieved between individual user clocks, individual paths will not arrive synchronously to each other due to the various propagation delays. A simple development for asynchronous performance can be shown following the approach of Geronoitis and Pursley,¹⁷ which shows that the probability of a hit in the asynchronous case is

$$P_h = 1 - \left[1 - \frac{1}{J} \left(1 + \frac{1}{N_b}\right)\right]^{K-1} \quad (13.49)$$

where N_b is the number of bits per hop. Comparing Eqs. (13.49) and (13.46) we see that for the asynchronous case, the probability of a hit is increased, as expected. Using Eq. (13.49) in Eq. (13.45) we obtain the probability of error for the asynchronous case as

$$P_b = \frac{1}{2} e^{-E_b/2N} \left[1 - \frac{1}{J} \left(1 + \frac{1}{N_b}\right)\right]^{K-1} + \frac{1}{2} \left\{1 - \left[1 - \frac{1}{J} \left(1 + \frac{1}{N_b}\right)\right]^{K-1}\right\} \quad (13.50)$$

From the preceding analysis we can draw two inferences. First, FH/SS has an advantage over DS/SS in that it is not as susceptible to the near-far problem. Because signals are generally not utilizing the same frequency simultaneously, the relative power levels of signals are not as critical as in DS/SS. The near-far problem is not totally avoided, however, since there will be some interference caused by stronger signals bleeding into weaker signals due to imperfect filtering. The second inference drawn is that to achieve reasonable throughputs, error correction coding is required. By applying strong Reed-Solomon or other burst error correcting codes, the performance can be increased dramatically.

13.5 M-ARY COMMUNICATION

Thus far we have stressed binary communication, which happens to be perhaps the single most important mode of communication in practice today. In the binary case, only two symbols are used, whereas in the M -ary case, the total number of symbols used is M . Each M -ary symbol carries as much information as $\log_2 M$ binary digits [see Eq. (7.55)]. In compensation, we may have to increase the transmitted power (as in the multiamplitude or multiphase case) or transmission bandwidth (as in the multitone case) in order to maintain a given performance level.

Multiamplitude Signaling

In the binary case, we transmit two symbols, consisting of the pulses $p(t)$ and $-p(t)$, where $p(t)$ may be either a baseband pulse or a carrier modulated by a baseband pulse. In the multiamplitude (MASK) case, the M symbols are transmitted by M pulses $\pm p(t)$, $\pm 3p(t)$, $\pm 5p(t)$, ..., $\pm(M-1)p(t)$. Thus, to transmit R_M M -ary digits per second, we are required to transmit R_M pulses per second of the form $kp(t)$. Pulses are transmitted every T_M seconds, so that $T_M = 1/R_M$. If E_p is the energy of pulse $p(t)$, then assuming that pulses $\pm p(t)$, $\pm 3p(t)$, $\pm 5p(t)$, ..., $\pm(M-1)p(t)$ are equally likely, the average pulse energy E_{pM} is given by

$$\begin{aligned} E_{pM} &= \frac{2}{M} [E_p + 9E_p + 25E_p + \dots + (M-1)^2 E_p] \\ &= \frac{2E_p}{M} \sum_{k=0}^{M-2} (2k+1)^2 \\ &= \frac{M^2 - 1}{3} E_p \end{aligned} \quad (13.51a)$$

$$\simeq \frac{M^2}{3} E_p \quad M \gg 1 \quad (13.51b)$$

Recall that an M -ary symbol carries an information of $\log_2 M$ bits. Hence, the bit energy E_b is

$$E_b = \frac{E_{pM}}{\log_2 M} = \frac{M^2 - 1}{3 \log_2 M} E_p \quad (13.51c)$$

Because the transmission bandwidth is independent of the pulse amplitude, the M -ary bandwidth is the same as in the binary case for the given rate of pulses, yet it carries more information.

This means that for a given information rate, the MASK bandwidth is less than that of the binary case by a factor of $\log_2 M$.

To calculate the error probability, we observe that because we are dealing with the same basic pulse $p(t)$, the optimum M -ary receiver is a filter matched to $p(t)$. When the input pulse is $kp(t)$, the output $r(T_p)$ at the sampling instant will be $kA_p + n_o(T_M)$. Note that $A_p = E_p$, the energy of $p(t)$, and that σ_n^2 , the variance of $n_o(t)$, is $\mathcal{N}E_p/2$. Thus, the optimum receiver for the multiamplitude M -ary signaling case is identical to that of the polar binary case (see Fig. 13.3 or 13.6a). The sampler has M possible outputs $\pm kA_p + n_o(T_M)$ ($k = 1, 3, 5, \dots, M-1$) that we wish to detect. The conditional PDFs $p(r|m)$ are gaussian with mean kA_p and variance σ_n^2 , as shown in Fig. 13.20a. Let P_{eM} be the error probability detecting a symbol and $P(\epsilon|m)$ be the error probability given that the symbol m is transmitted.

To calculate P_{eM} , we observe that the case of the two extreme symbols [represented by $\pm(M-1)p(t)$] is similar to the binary case because they have to guard against only one neighbor. As for the remaining symbols, they must guard against neighbors on both sides, and, hence, $P(\epsilon|m)$ in this case is twice that of the extreme symbol. From Fig. 13.20a it is evident that $P(\epsilon|m_i)$ is $Q(A_p/\sigma_n)$ for the two extreme signals and is $2Q(A_p/\sigma_n)$ for the remaining $(M-2)$ symbols. Hence,

$$\begin{aligned} P_{eM} &= \sum_{i=1}^M P(m_i) P(\epsilon|m_i) \\ &= \frac{1}{M} \sum_{i=1}^M P(\epsilon|m_i) \\ &= \frac{1}{M} \left[Q\left(\frac{A_p}{\sigma_n}\right) + Q\left(\frac{A_p}{\sigma_n}\right) + (M-2)2Q\left(\frac{A_p}{\sigma_n}\right) \right] \\ &= \frac{2(M-1)}{M} Q\left(\frac{A_p}{\sigma_n}\right) \end{aligned} \quad (13.52a)$$

For a matched-filter receiver, $(A_p/\sigma_n)^2 = 2E_p/\mathcal{N}$, and

$$P_{eM} = 2\left(\frac{M-1}{M}\right) Q\left(\sqrt{\frac{2E_p}{\mathcal{N}}}\right) \quad (13.52b)$$

$$= 2\left(\frac{M-1}{M}\right) Q\left[\sqrt{\frac{6 \log_2 M}{M^2 - 1} \left(\frac{E_b}{\mathcal{N}}\right)}\right] \quad (13.52c)$$

$$\approx 2Q\left[\sqrt{\frac{6 \log_2 M}{M^2} \left(\frac{E_b}{\mathcal{N}}\right)}\right] \quad M \gg 1 \quad (13.52d)$$

Bit Error Rate (BER): It is somewhat unfair to compare M -ary signaling on the basis of P_{eM} , the error probability of an M -ary symbol, which conveys the information of $k = \log_2 M$ bits. This weighs unfairly against higher values of M . For a fair comparison, we should compare various schemes for P_b , the probability of bit error, rather than P_{eM} , the probability of symbol error (symbol error rate). We now show that for multiamplitude signaling $P_b = P_{eM}/\log_2 M$.

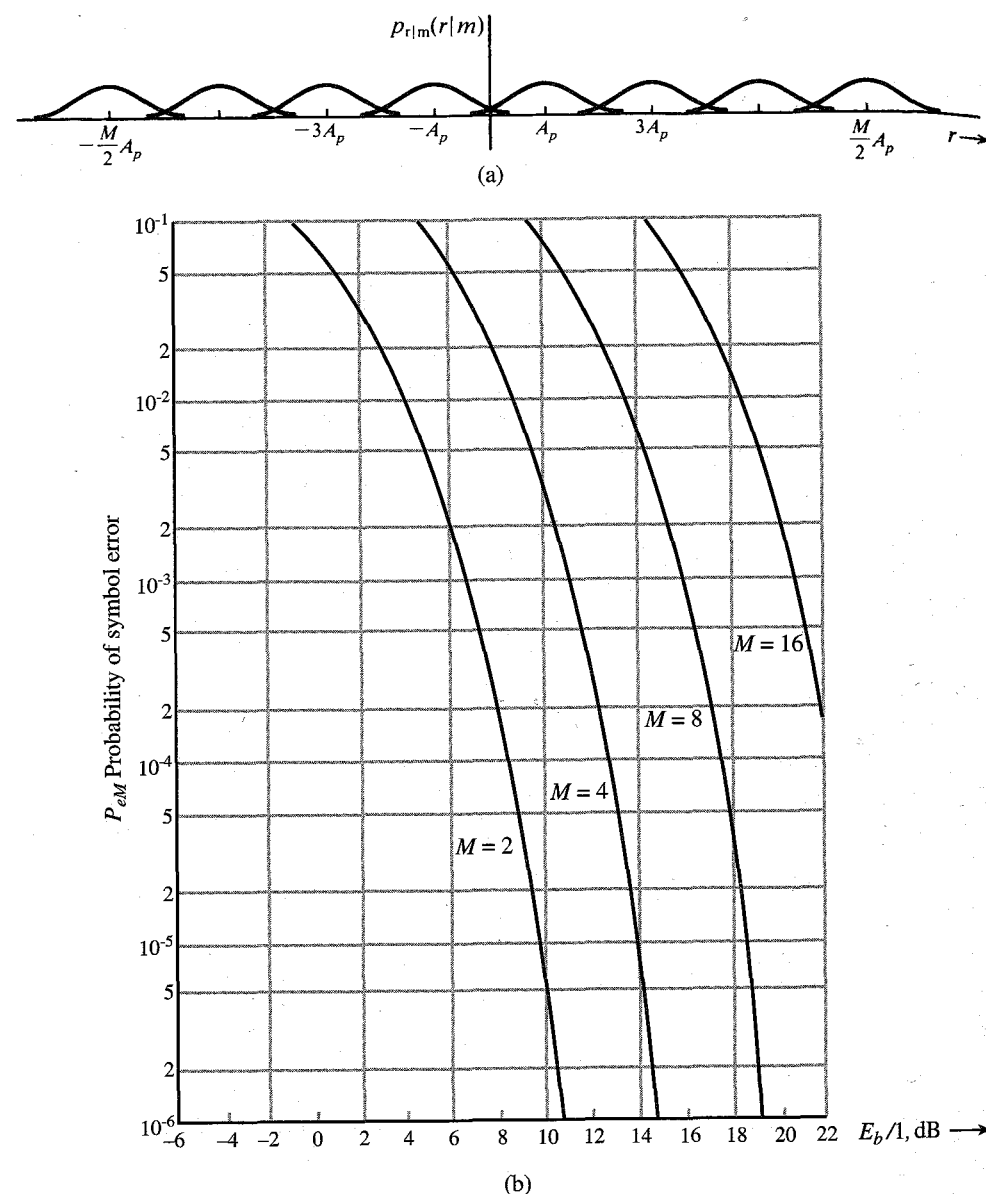


Figure 13.20 (a) Conditional PDFs in MASK. (b) Error probability in MASK.

Because the type of errors that predominate are those where a symbol is mistaken for its immediate neighbors (see Fig. 13.20a), it would be logical to assign neighboring M -ary symbols, binary code words that differ in the least possible digits. The Gray code* is suitable

* The Gray code can be constructed as follows: Construct an n -digit natural binary code (NBC) corresponding to 2^n decimal numbers. If $b_1b_2 \dots b_n$ is a code word in this code, then the corresponding Gray code word $g_1g_2 \dots g_n$ is obtained by the rule

for this because adjacent binary combinations in this code differ only by one digit. Hence, an error in one M -ary symbol detection will cause only one error in a group of $\log_2 M$ binary digits transmitted by the M -ary symbol. Hence, the bit error rate $P_b = P_{eM}/\log_2 M$. However, the factor $\log_2 M$ is negligible compared to other factors that vary exponentially with M in the expression for P_{eM} [see Eqs. (13.52)], and, hence, it is not as influential in determining the behavior of P_{eM} . Figure 13.20b shows P_{eM} as a function of $E_b/\mathcal{N}$ for several values of M . The relationship $P_b = P_{eM}/\log_2 M$, although valid for MASK, is not valid for other schemes due to the specific code structure. One must recompute this relationship for each scheme.

Trade-off between Power and Bandwidth

To maintain a given information rate, the pulse transmission rate in the M -ary case is reduced by the factor $k = \log_2 M$. This means the bandwidth of the M -ary case is reduced by the same factor $k = \log_2 M$. But to maintain the same P_{eM} , Eqs. (13.52) show that the power transmitted (which is proportional to E_b) increases roughly as $M^2/\log_2 M = 2^{2k}/k$. On the other hand, if we maintain a given bandwidth, the information rate in the M -ary case is increased by the factor $k = \log_2 M$. The transmitted power is equal to E_b times the bit rate. Hence, an increased data rate also necessitates increased power by the factor $(M^2/\log_2 M)(\log_2 M) = 2^{2k}$. Thus, the power increases exponentially with the increase in information rate by a factor of k . In high-powered radio systems, such a power increase may not be tolerable. Multi-amplitude systems are attractive where bandwidth is at a premium. Because the voice channels of a telephone network have a fixed bandwidth, multi-amplitude (or multiphase, or a combination of both) signaling appears to be an attractive method of increasing the information rate, particularly because telephone lines can be made to have a high degree of channel stability and low additive noise.

All the results derived here apply to baseband as well as modulated digital systems with coherent detection. For noncoherent detection, similar relationships exist between the binary and M -ary systems.*

Multiphase Signaling

In binary PSK (BPSK), the two basic pulses are $p'(t) \cos \omega_c t$ and $p'(t) \cos(\omega_c t + \pi)$. We generalize the idea for the M -ary case (MPSK) using M pulses with the k th pulse $p'(t) \cos[\omega_c t + (2\pi/M)k]$. Thus, the phases of successive pulses are $2\pi/M$ radians apart (Fig. 13.21). From the symmetry of this scheme it is evident that if, because of channel noise, the phase of any pulse deviates by more than π/M radians, an error is made. Hence, the error probability P_{eM} is the probability that the phase of any pulse deviates by more than π/M radians.

At the receiver, coherent phase reference is available. The basic function of the receiver is to detect the phase of the received pulse. This can be done by two phase detectors with cos

$$\begin{aligned} g_1 &= b_1 \\ g_k &= b_k \oplus b_{k-1} \quad k \geq 2 \end{aligned}$$

Thus for $n = 3$, the binary code **000, 001, 010, 011, 100, 101, 110, 111** is transformed into the Gray code **000, 001, 011, 010, 110, 111, 101, 100**.

* For the noncoherent case, the baseband pulses must be of the same polarity; for example, 0 , $p(t)$, $2p(t)$, $\dots$, $(M-1)p(t)$.

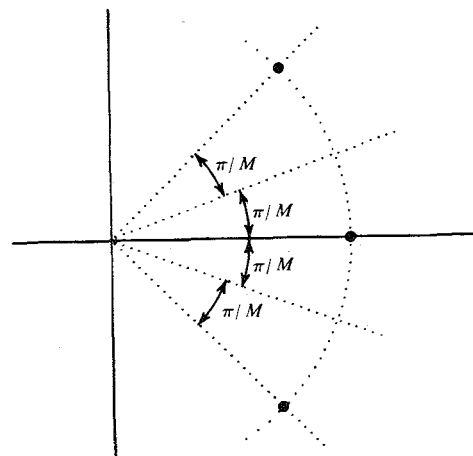


Figure 13.21 MPSK signals.

$\omega_c t$ and $\sin \omega_c t$ as respective references, along with a logic circuit to determine the ratio of the two detected components.*

To compute the error probability P_{eM} , we note that a detection error results if the phase of any pulse deviates by more than π/M . The PDF $p_\Theta(\theta)$ of the phase Θ of a sinusoid plus a bandpass gaussian noise is found in Eq. (11.64d). Hence,

$$P_{eM} = 1 - \int_{-\pi/M}^{\pi/M} p_\Theta(\theta) d\theta$$

The PDF $p_\Theta(\theta)$ in Eq. (11.64d) involves A (the sinusoid amplitude) and σ_n^2 (the noise variance). Assuming matched filtering and white noise [see Eq. (13.9c)],

$$\frac{A^2}{\sigma_n^2} = \frac{2E_p}{\mathcal{N}} = \frac{2E_b \log_2 M}{\mathcal{N}}$$

Hence,

$$P_{eM} = 1 - \frac{1}{2\pi} \int_{-\pi/M}^{\pi/M} e^{-\frac{E_b \log_2 M}{\mathcal{N}}} \left\{ 1 + \sqrt{\frac{4\pi E_b \log_2 M}{\mathcal{N}}} \cos \theta e^{\frac{E_b \cos^2 \theta \log_2 M}{\mathcal{N}}} \right. \\ \left. \times \left[1 - Q \left(\sqrt{\frac{2E_b \log_2 M}{\mathcal{N}}} \cos \theta \right) \right] \right\} d\theta \quad (13.53)$$

Figure 13.22 shows the plot of P_{eM} as a function of $E_b/\mathcal{N}$. For $E_b/\mathcal{N} \gg 1$ (weak noise) and $M \gg 2$, Eq. (13.53) can be approximated as¹⁸

$$P_{eM} \simeq 2Q \left(\sqrt{\frac{2E_b \log_2 M}{\mathcal{N}}} \sin \frac{\pi}{M} \right) \quad (13.54a)$$

* This can be done by a parallel bank of two filters matched to $p'(t) \sin \omega_c t$ and $p'(t) \cos \omega_c t$, respectively. If the input to this bank is $p'(t) \cos (\omega_c t + \theta)$, then the outputs of the two matched filters are proportional to $\sin \theta$ and $\cos \theta$, respectively. This is because the input $p'(t) \cos (\omega_c t + \theta)$ can be expressed as a sum of two orthogonal components $\sin \theta p'(t) \sin \omega_c t$ and $\cos \theta p'(t) \cos \omega_c t$. The ratio of the outputs of these filters is $\tan \theta$.

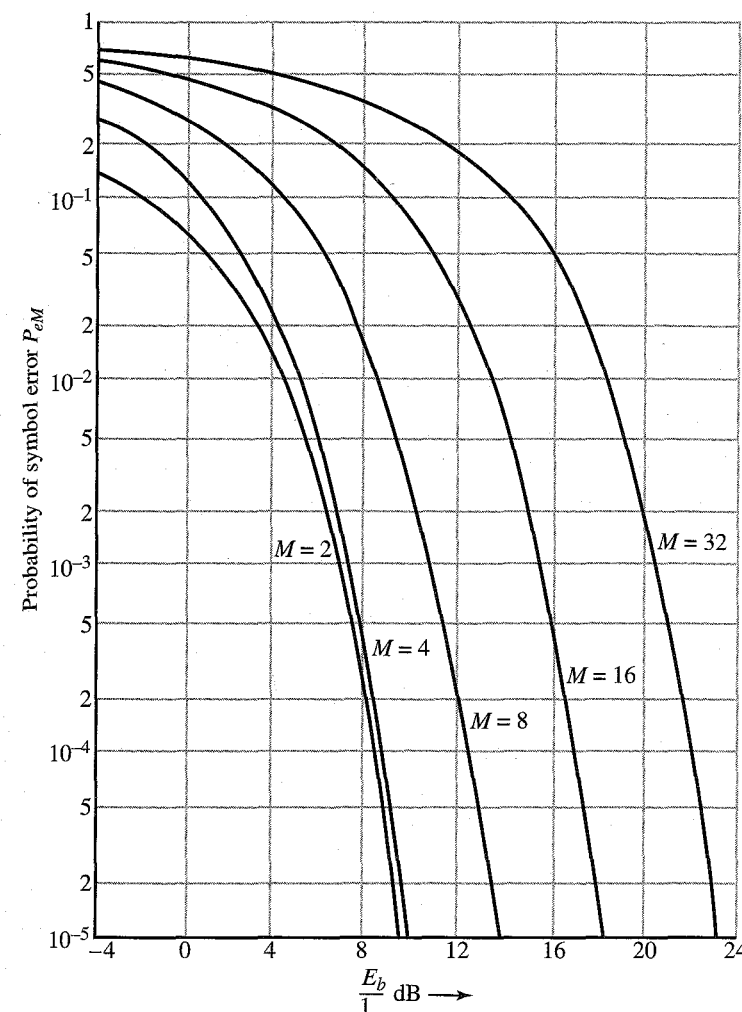


Figure 13.22 Error probability of MPSK.

$$\simeq 2Q \left(\sqrt{\frac{2\pi^2 E_b \log_2 M}{M^2 \mathcal{N}}} \right) \quad (13.54b)$$

An alternate expression for P_{eM} is derived in Chapter 14 [Eq. (14.59b)]. A comparison of Eq. (13.54b) with Eq. (13.52d) shows that the behavior of MPSK is very similar to that of MASK, at least for large $E_b/\mathcal{N}$ and $M \gg 2$. Hence, the comments made in reference to MASK apply to MPSK also.

One of the schemes frequently used in practice is the case of $M = 4$. This is four-phase, or quadriphase, PSK (QPSK). This case can be thought of as two binary PSK systems in parallel in which the carriers are in phase quadrature. Thus, one system uses pulses $\pm p(t) \cos \omega_c t$, and the other uses pulses $\pm p(t) \sin \omega_c t$. Because the carriers are in quadrature, both of these signals can be transmitted over the same channel without interference. Addition of these two streams yields four possible pulses, $\pm \sqrt{2} p(t) \cos (\omega_c t \pm \pi/4)$. This is precisely QPSK. Thus, QPSK doubles the transmission rate without increasing the bandwidth.

Bit Error Rate (BER): As in the case of MASK, the BER for MPSK also is given by $P_b = P_{eM} / \log_2 M$. This is because in MASK and MPSK, the type of errors that predominate are those where a symbol is mistaken for its immediate neighbor. Using Gray code we can assign the adjacent symbols codes that differ in just one digit.

An example of QPSK is found in the SPADE multiple-access communication system used for PCM via the Intelsat satellite communication global network, which attains a rate of 64 kbit/s over a transmission bandwidth of 38 kHz. The scheme uses Nyquist's criterion with a roll-off factor of 19% (see Example 7.1). Regular telephone lines with a bandwidth on the order of 2400 Hz (from 600 to 3000 Hz) are used to transmit binary data at a rate of 1200 digits per second. This rate can be increased by using MPSK. For example, to transmit data at respective rates of 2400 and 4800 digits per second, QPSK and 8-ary PSK are used. Higher rates (9600 digits per second) are obtained by using 16-ary QAM (discussed next).

Quadrature Amplitude Modulation

The quadrature amplitude modulation (QAM), also called amplitude phase-shift keying (APK), is a combination of MASK and MPSK. A QAM signal can be written as

$$p_i(t) = p'(t)(a_i \cos \omega_c t + b_i \sin \omega_c t)$$
$$= p'(t)[r_i \cos(\omega_c t + \theta_i)] \quad i = 1, 2, \dots, M \quad (13.55)$$

where $r_i = \sqrt{a_i^2 + b_i^2}$ and $\theta_i = -\tan^{-1} b_i/a_i$. In MASK, $\theta_i = 0$ for all i ; only r_i is different. In MPSK, r_i is constant for all i ; only θ_i is different. In APK, both r_i and θ_i vary with i . Graphically, signal $p_i(t)$ can be mapped into a point (a_i, b_i) or (r_i, θ_i) . One such signal set is shown in Fig. 13.23. Because each M -ary pulse consists of a sum of pulses modulated by carriers in quadrature (quadrature multiplexing), this is the case of QAM.

Techniques for analyzing such general signals are discussed in Chapter 14. For the 16-ary case in Fig. 13.23, it can be shown that (see Example 14.3)

$$P_{eM} \simeq 3Q \left(\sqrt{\frac{4E_b}{5N}} \right) \quad (13.56)$$

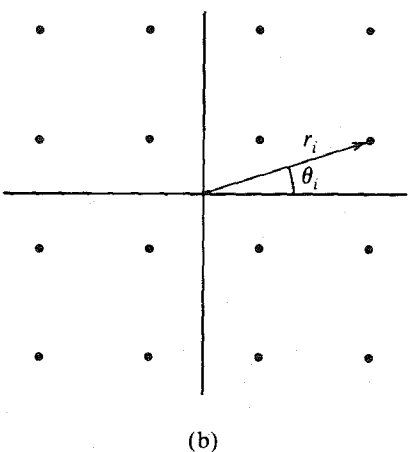


Figure 13.23 16-ary (or 16-point) QAM.

General analysis of MQAM (M -ary QAM) is rather involved and requires advanced techniques discussed in Chapter 14, where a 16-ary QAM is analyzed in Example 14.3. It can be shown that¹⁸ an upper bound on P_{eM} is given by

$$P_{eM} = 4Q \left(\sqrt{\frac{2E_b}{N}} \eta_M \right)$$

where η_M is an efficiency factor, normalized to the efficiency of antipodal signaling, and depends on the constellation size and shape with $\eta_M \leq 1$. Wilson¹⁸ gives the following values of η_M for various values of M for MASK and MQAM:

MASK	MQAM	η_{dB}
4	16	-4
—	32	-6
8	64	-8.5
—	128	-10.2
16	256	-13.3

Multitone Signaling (MFSK)

In this case, M symbols are transmitted by M orthogonal pulses of frequencies $\omega_1, \omega_2, \dots, \omega_M$, each of duration T_M . Thus, the M transmitted pulses are of the form* $\sqrt{2}p'(t) \cos \omega_k t$, where $\omega_k = 2\pi(N+k)/T_M$. The receiver (Fig. 13.24) is a simple extension of the binary receiver. The incoming pulse is multiplied by the corresponding references $\sqrt{2} \cos \omega_i t$ ($i = 1, 2, \dots, M$). The filter $H(\omega)$ is matched to the baseband pulse $p'(t)$. The same result is obtained if in the

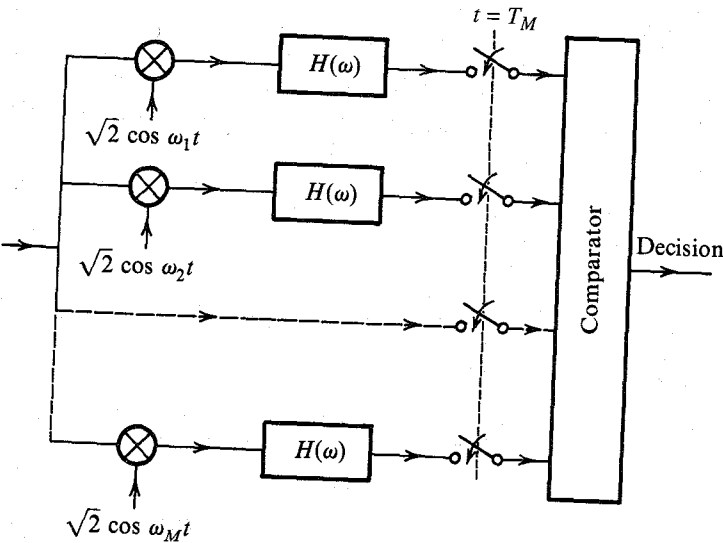


Figure 13.24 Coherent MFSK receiver.

* A better scheme is to use pulses $\sqrt{2}p'(t) \sin \omega_k t$ and $\sqrt{2}p'(t) \cos \omega_k t$, $k = 1, 2, \dots, M/2$. Because these are all orthogonal pulses, we required only $M/2$ distinct frequencies, and the bandwidth is reduced by a factor of 2. This scheme cannot be used for noncoherent detection for obvious reasons.

i th bank, instead of using a multiplier and $H(\omega)$, we use a filter matched to the RF pulse $p'(t) \cos \omega_i t$. The M bank outputs sampled at $t = T_M$ are $r_1, r_2, \dots, r_M$. They are compared, and the decision is $m = j$ if $r_j > r_i$ for all $i \neq j$. If a pulse $p'(t) \cos \omega_i t$ is received, it will cause an output only in the i th bank and will be completely suppressed in the outputs of all the remaining banks if all the M RF pulses are orthogonal.*

The bandwidth of MFSK increases with M . When $m = 1$ is transmitted, the corresponding sampler output will be $A_p + n_1$, and the other sampler outputs are $n_2, n_3, \dots, n_M$, where $n_1, n_2, \dots, n_M$ are all mutually independent gaussian variables with the same variance σ_n^2 . An error is made if $n_j > A_p + n_1$ ($j \neq 1$). Suppose the value of $r_1 = r_1$. Now the correct decision will be made if $n_2, n_3, \dots, n_M$ are all less than r_1 . Because all these variables have identical PDFs (gaussian with variance σ_n), the probability of $n_j < r_1$ is

$$\int_{-\infty}^{r_1} \frac{1}{\sigma_n \sqrt{2\pi}} e^{-n_j^2/2\sigma_n^2} dn_j$$

If $P(C|m = 1)$ is the probability of making a correct decision given that $m = 1$ is transmitted, then

$$\begin{aligned} P(C|m = 1) &= P(r_1 < \infty, n_2 < r_1, n_3 < r_1, \dots, n_M < r_1) \\ &= \int_{-\infty}^{\infty} p_{r_1}(r_1) dr_1 \left(\prod_{j=2}^M \int_{-\infty}^{r_1} \frac{1}{\sigma_n \sqrt{2\pi}} e^{-n_j^2/2\sigma_n^2} dn_j \right) \\ &= \frac{1}{(2\pi\sigma_n^2)^{M/2}} \int_{-\infty}^{\infty} e^{-(r_1-A_p)^2/2\sigma_n^2} dr_1 \left(\int_{-\infty}^{r_1} e^{-x^2/2\sigma_n^2} dx \right)^{M-1} \\ &= \frac{1}{\sigma_n \sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(r_1-A_p)^2/2\sigma_n^2} \left[1 - Q\left(\frac{r_1}{\sigma_n}\right) \right]^{M-1} dr_1 \end{aligned}$$

Because of symmetry, $P(C|m = 1) = P(C|m = 2) = \dots = P(C|m = M)$. Hence, P_{CM} , the probability of a correct decision, is $P(C|m = 1)$ provided all symbols are equiprobable. Because the filters are matched [Eq. (13.9)],

$$\frac{A_p^2}{\sigma_n^2} = \frac{2E_p}{\mathcal{N}} = \frac{2E_b \log_2 M}{\mathcal{N}}$$

and we can express

$$P_{eM} = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(y - \sqrt{2E_b \log_2 M / \mathcal{N}})^2 / 2} [1 - Q(y)]^{M-1} dy \quad (13.57)$$

* This can be shown as follows: The filter matched to $p(t)$ has impulse response $h(t) = p(T_M - t)$. If $q(t)$ is applied at the input of this filter, the response $r(t)$ is

$$r(t) = h(t) * q(t) = \int_{-\infty}^{\infty} p(T_M - x)q(t - x) dx$$

and

$$r(T_M) = \int_{-\infty}^{\infty} p(T_M - x)q(T_M - x) dx = \int_{-\infty}^{\infty} p(t)q(t) dt$$

If $p(t)$ and $q(t)$ are orthogonal, $r(T_M) = 0$.

The integral appearing on the right-hand side of Eq. (13.57) is computed and plotted in Fig. 13.25 (P_{eM} vs. $E_b/\mathcal{N}$). This plot shows an interesting behavior for the case of $M = \infty$. By properly taking the limit of P_{eM} in Eq. (13.57) as $M \rightarrow \infty$, it can be shown that³

$$\lim_{M \rightarrow \infty} P_{eM} = \begin{cases} 1 & E_b/\mathcal{N} < \log_e 2 \\ 0 & E_b/\mathcal{N} \geq \log_e 2 \end{cases}$$

Because the signal power $S_i = E_b R_b$, where R_b is the bit rate, it follows that for error-free communication,

$$\frac{E_b}{\mathcal{N}} \geq \log_e 2 = \frac{1}{1.44} \quad \text{or} \quad \frac{S_i}{\mathcal{N} R_b} \geq \frac{1}{1.44}$$

Hence,

$$R_b \leq 1.44 \frac{S_i}{\mathcal{N}} \text{ bit/s} \quad (13.58)$$

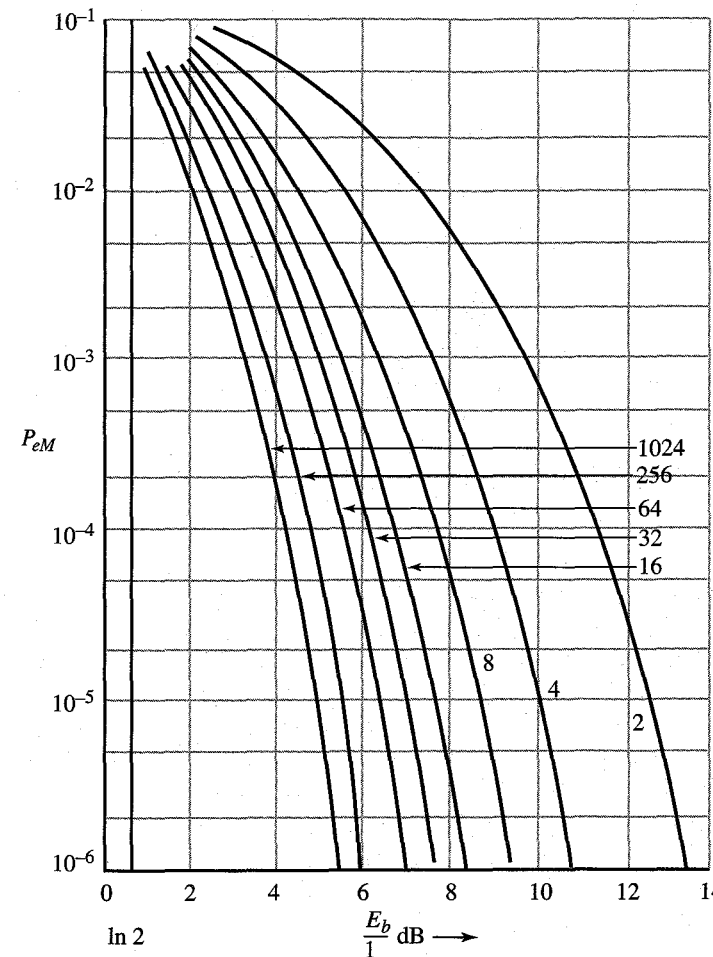


Figure 13.25 Error probability of coherent MFSK.

This shows that M -ary orthogonal signaling can transmit data error-free at a rate of up to $1.44 S_i/\mathcal{N}$ bit/s as $M \rightarrow \infty$ (see Fig. 13.25).

Bit Error Rate (BER): For MASK and MPSK, we showed that $P_b = P_{eM}/\log_2 M$. This result is not valid for MFSK because in MASK and MPSK, the type of errors that predominate are those where a symbol is mistaken for its immediate neighbor. Using Gray code we can assign the adjacent symbols codes that differ in just one digit. In MFSK, on the other hand, a symbol is equally likely to be mistaken for any of the remaining $M - 1$ symbols. Hence, $P(\epsilon)$, the probability of mistaking one M -ary symbol for another, is

$$P(\epsilon) = \frac{P_{eM}}{M - 1} = \frac{P_{eM}}{2^k - 1}$$

If an M -ary symbol differs by 1 bit from N_1 number of symbols, and differs by 2 bits from N_2 number of symbols, and so on, then $\bar{N}_\epsilon$, the average number of bits in error in reception of an M -ary symbol, is

$$\begin{aligned} \bar{N}_\epsilon &= \sum_{n=1}^k n N_n P(\epsilon) \\ &= \sum_{n=1}^k n N_n \frac{P_{eM}}{2^k - 1} \\ &= \frac{P_{eM}}{2^k - 1} \sum_{n=1}^k n \binom{k}{n} \\ &= k 2^{k-1} \frac{P_{eM}}{2^k - 1} \end{aligned}$$

This is an average number of bits in error in a sequence of k bits (one M -ary symbol). Consequently, the BER, P_b is this figure divided by k ,

$$P_b = \frac{2^{k-1}}{2^k - 1} P_{eM} \approx \frac{P_{eM}}{2} \quad k \gg 1$$

Noncoherent MFSK

From the practical point of view, the phase coherence of M frequencies is difficult to maintain. Hence in practice, coherent MFSK is rarely used. Noncoherent MFSK is more common. The receiver for noncoherent MFSK is similar to that for binary noncoherent FSK (Fig. 13.14), but with M banks corresponding to M frequencies. The filter $H_i(\omega)$ is matched to the RF pulse $p(t) \cos \omega_i t$. The analysis is straightforward. If $m = 1$ is transmitted, then r_1 is the envelope of a sinusoid of amplitude A_p plus bandpass gaussian noise, and r_j ($j = 2, 3, \dots, M$) is the envelope of the bandpass gaussian noise. Hence, r_1 has rician density, and $r_2, r_3, \dots, r_M$ have Rayleigh density. Using the same arguments as in the coherent case, we have

$$\begin{aligned} P_{CM} &= P(C|m = 1) = P(0 \leq r_1 < \infty, n_2 < r_1, n_3 < r_1, \dots, n_M < r_1) \\ &= \int_0^\infty \frac{r_1}{\sigma_n^2} I_0\left(\frac{r_1 A_p}{\sigma_n^2}\right) e^{-(r_1^2 + A_p^2)/2\sigma_n^2} dr_1 \left(\int_0^{r_1} \frac{x}{\sigma_n^2} e^{-x^2/2\sigma_n^2} dx\right)^{M-1} dx \end{aligned}$$

$$= \int_0^\infty \frac{r_1}{\sigma_n^2} I_0\left(\frac{r_1 A_p}{\sigma_n^2}\right) e^{-(r_1^2 + A_p^2)/2\sigma_n^2} (1 - e^{-r_1^2/2\sigma_n^2})^{M-1} dr_1$$

Substituting $r_1^2/2\sigma_n^2 = x$ and $(A_p/\sigma_n)^2 = 2E_p/\mathcal{N} = 2E_b \log M/\mathcal{N}$, we obtain

$$P_{CM} = e^{-\frac{E_b \log_2 M}{\mathcal{N}}} \int_0^\infty e^{-x} (1 - e^{-x})^{M-1} I_0\left(2\sqrt{\frac{x E_b \log_2 M}{\mathcal{N}}}\right) dx \quad (13.59a)$$

Expanding $(1 - e^{-x})^{M-1}$ using the binomial theorem, we obtain

$$(1 - e^{-x})^{M-1} = \sum_{m=0}^{M-1} \binom{M-1}{m} (-1)^m e^{-mx}$$

Substitution of this result into Eq. (13.59a) and recognizing that

$$\int_0^\infty y e^{-ay^2} I_0(by) dy = \frac{1}{2a} e^{-b^2/4a}$$

we obtain (after interchanging the order of summation and integration)

$$P_{CM} = \sum_{m=0}^{M-1} \binom{M-1}{m} \frac{(-1)^m}{m+1} e^{-m E_b \log_2 M / \mathcal{N}(m+1)} \quad (13.59b)$$

and

$$P_{eM} = 1 - P_{CM} = \sum_{m=1}^{M-1} \binom{M-1}{m} \frac{(-1)^{m+1}}{m+1} e^{-m E_b \log_2 M / \mathcal{N}(m+1)} \quad (13.59c)$$

The error probability P_{eM} is shown in Fig. 13.26 as a function of $E_b/\mathcal{N}$. It can be seen that the performance of noncoherent MFSK is only slightly inferior to that of coherent MFSK, particularly for large M .

Comments on M -ary Signaling

M -ary signaling provides us with additional means of exchanging, or trading, the transmission rate, transmission bandwidth, and transmitted power. It provides us a flexibility in designing a proper communication system. Thus, for a given rate of transmission, we can trade the transmission bandwidth for transmitted power. We can also increase the information rate by a factor of k ($k = \log_2 M$) by paying a suitable price in terms of the transmission bandwidth or the transmitted power. Figures 13.25 and 13.26 show that in multitone signaling the transmitted power decreases with M . However, the transmission bandwidth increases linearly with M , or exponentially with the rate increase factor k ($M = 2^k$). Thus, multitone signaling is radically different from multi-amplitude or multiphase signaling. In the latter, the bandwidth is independent of M , but the transmitted power increases as $M^2/\log_2 M = 2^{2k}/k$; that is, the power increases exponentially with the information-rate increase factor k . Thus, in the multitone case, the bandwidth increases exponentially with k , and in the multi-amplitude or multiphase case, the power increases exponentially with k . Hence, we should use multi-amplitude or multiphase signaling if the bandwidth is at a premium (as in telephone lines) and use multitone signaling when power is at a premium (as in space communication). A compromise exists between these two extremes. Let us investigate the possibility of increasing

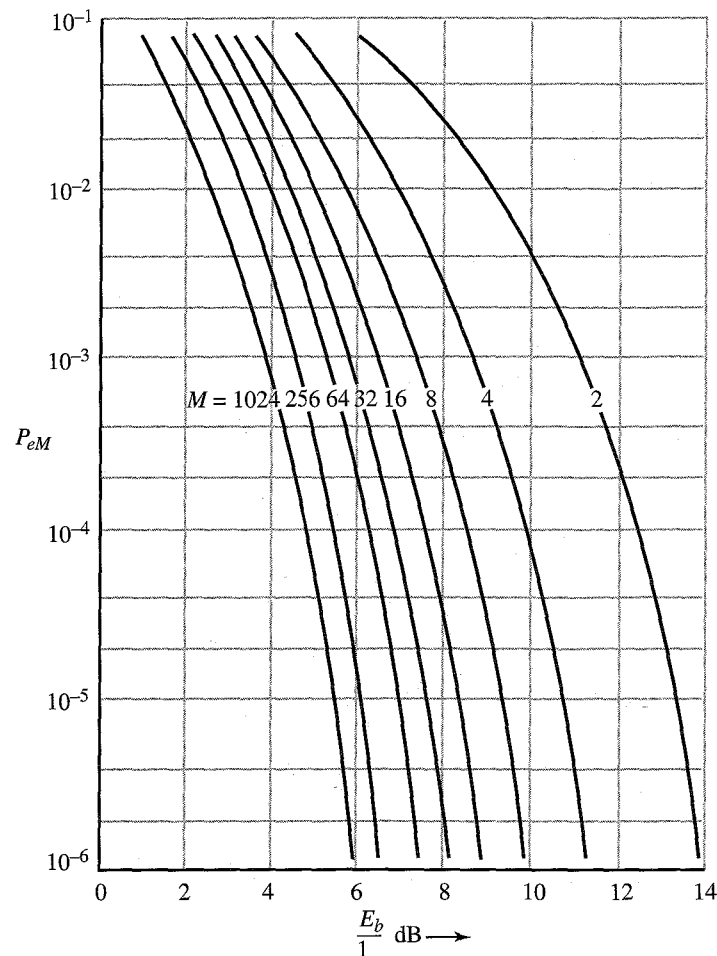


Figure 13.26 Error probability of noncoherent MFSK.

the information rate by a factor k simply by increasing the number of binary pulses transmitted by a factor k . In this case, the transmitted power increases linearly with k . Also because the bandwidth is proportional to the pulse rate, the transmission bandwidth increases linearly with k . Thus, in this case, we can increase the information rate by a factor of k by increasing both the transmission bandwidth and the transmitted power linearly with k , thus avoiding the phantom of the exponential increase that was required in the M -ary system. But here we must increase both the bandwidth and the power, whereas in the M -ary case the increase in information rate can be achieved by increasing either the bandwidth or the power. We have thus a great flexibility in trading various parameters and thus in our ability to match our resources to our requirements.

EXAMPLE 13.1 It is required to transmit 2.08×10^6 binary digits per second with $P_b \leq 10^{-6}$. Three possible schemes are considered:

(a) Binary

(b) 16-ary ASK

(c) 16-ary PSK

The channel noise PSD is $S_n(\omega) = 10^{-8}$. Determine the transmission bandwidth and the signal power required at the receiver input in each case.

(a) *Binary*: We shall consider polar signaling (the most efficient scheme),

$$P_b = P_e = 10^{-6} = Q\left(\sqrt{\frac{2E_b}{N}}\right)$$

This yields $E_b/N = 11.35$. The signal power $S_i = E_b R_b$. Hence,

$$S_i = 11.35 N R_b = 11.35(2 \times 10^{-8})(2.08 \times 10^6) = 0.47 \text{ W}$$

Assuming raised-cosine baseband pulses, the bandwidth B_T is

$$B_T = R_b = 2.08 \text{ MHz}$$

(b) *16-ary ASK*: Because each 16-ary symbol carries the information equivalent of $\log_2 16 = 4$ binary digits, we need to transmit only $R_M = (2.08 \times 10^6)/4 = 0.52 \times 10^6$ 16-ary pulses per second. This requires a bandwidth B_T of 520 kHz for baseband pulses and 1.04 MHz for modulated pulses (assuming raised-cosine pulses). Also,

$$P_b = 10^{-6} = \frac{P_{eM}}{\log_2 16}$$

Therefore,

$$P_{eM} = 4 \times 10^{-6} = 2 \left(\frac{M-1}{M} \right) Q \left[\sqrt{\frac{6E_b \log_2 16}{N(M^2-1)}} \right]$$

For $M = 16$, this yields $E_b = 0.499 \times 10^{-5}$. If the M -ary pulse rate is R_M , then

$$\begin{aligned} S_i &= E_{pM} R_M = E_b \log_2 M R_M \\ &= 0.499 \times 10^{-5} \times 4 \times (0.52 \times 10^6) = 9.34 \text{ W} \end{aligned}$$

(c) *16-ary PSK*: We need to transmit only $R_M = 0.52 \times 10^6$ pulses per second. For baseband pulses, this will require a bandwidth of 520 kHz. But PSK is a modulated signal, and the required bandwidth is $2(0.52 \times 10^6) = 1.04 \text{ MHz}$. Also,

$$P_{eM} = 4P_b = 4 \times 10^{-6} \simeq 2Q \left[\sqrt{\frac{2\pi^2 E_b \log_2 16}{256N}} \right]$$

This yields $E_b = 137.8 \times 10^{-8}$ and

$$\begin{aligned} S_i &= E_b \log_2 16 R_M \\ &= (137.8 \times 10^{-8}) \times 4 \times (0.52 \times 10^6) = 2.86 \text{ W} \end{aligned}$$

13.6 SYNCHRONIZATION

In synchronous, or coherent, detection, we need to achieve synchronization at three different levels: (1) carrier synchronization, (2) bit synchronization, and (3) word synchronization. For noncoherent detection, we need only the second and third levels of synchronization—which were discussed in Chapter 7. Here we shall consider only carrier synchronization.

Carrier synchronization is similar to bit synchronization, only more difficult. In bit synchronization, the problem is to achieve synchronism from bit interval to bit interval—which is of the order T_b . In carrier synchronization, we must achieve synchronism within a fraction of a cycle, and because the duration of one carrier cycle is $1/f_c \ll T_b$, the problem is severe. It should be remembered that the phase error θ that can be tolerated is much less than $\pi/2$. For example, if we are transmitting data at a rate of 2 Mbit/s, the bit interval is 0.5 μ s. If this data is transmitted by PSK with a carrier frequency of 100 MHz, a phase of $\pi/2$ corresponds to 2.5 ns; that is, the synchronization must be achieved within an interval of much less than 2.5 ns!

Carrier synchronization is achieved by three general methods that are similar to those used for bit synchronization (see timing extraction in Sec. 7.5):

1. Using a primary or a secondary standard (i.e., transmitter and receiver slaved to a master timing source)
2. Transmitting a separate synchronizing signal (a pilot)
3. Self-synchronization, where the timing information is extracted from the received signal itself

The first method is expensive and is suitable only for large data systems, not for point-to-point systems.

The second method uses part of the channel capacity to transmit timing information and causes some degradation in performance (see Prob. 13.2-1). But this is a widely used method for point-to-point communication systems. A pilot may be transmitted by frequency-division multiplexing (by choosing a pilot of frequency at which the signal PSD has a null) or by time-division multiplexing (in which the modulated signal is interrupted for a short period of time, during which the synchronizing signal is transmitted).

A baseband signaling scheme with a dc null—such as bipolar, duobinary, or split-phase—is preferred, because such signals after modulation have a spectral null at the carrier frequency. This facilitates the separation of the pilot at the receiver.

The self-synchronization method extracts the carrier by squaring the incoming signal or by using a Costas loop, as discussed in Sec. 4.7. But because these methods yield sign ambiguities, they cannot be used for PSK unless differential coding is used.

REFERENCES

1. S. Pasupathy, "Minimum Shift Keying: A Spectrally Efficient Modulation," *IEEE Commun. Soc. Mag.*, vol. 17, pp. 14–22, July 1979.
2. J. J. Spilker, *Digital Communications by Satellite*, Prentice-Hall, Englewood Cliffs, NJ, 1977.
3. A. J. Viterbi, *Principles of Coherent Communication*, McGraw-Hill, New York, 1966.

4. J. M. Wozenraft and I. M. Jacobs, *Principles of Communication Engineering*, Wiley, New York, 1965, chap. 7.
5. J. I. Marcum, "Statistical Theory of Target Detection by Pulsed Radar," *IRE Trans. Inf. Theory*, vol. IT-6, pp. 259–267, April 1960.
6. M. Schwartz, W. R. Bennett, and S. Stein, *Communication Systems and Techniques*, McGraw-Hill, New York, 1966.
7. D. V. Sarwate and M. B. Pursley, "Crosscorrelation Properties of Pseudorandom and Related Sequences," *Proc. IEEE*, vol. 68, pp. 593–619, May 1980.
8. M. B. Pursley, "Performance Analysis for Phase-Coded Spread-Spectrum Multiple-Access Communication—Part I: System Analysis," *IEEE Trans. Commun.*, vol. COM-25, pp. 795–799, Aug. 1977.
9. E. Geraniotis and M. B. Pursley, "Error Probabilities for Direct Sequence Spread Spectrum Multiple Access Communications—Part II: Approximations," *IEEE Trans. Commun.*, vol. COM-30, pp. 985–995, May 1982.
10. J. S. Lehnert, "An Efficient Technique for Evaluating Direct-Sequence Spread-Spectrum Communications," *IEEE Trans. Commun.*, vol. 37, pp. 851–858, Aug. 1989.
11. R. K. Morrow and J. S. Lehnert, "Bit-to-Bit Error Dependence in Slotted DS/SSMA Packet Systems with Random Signal Sequences," *IEEE Trans. Commun.*, vol. COM-37, pp. 1052–1061, Oct. 1989.
12. J. M. Holtzman, "A Simple Accurate Method to Calculate Spread-Spectrum Multiple Access Error Probabilities," *IEEE Trans. Commun.*, vol. COM-40, pp. 461–464, March 1992.
13. B. D. Woerner and R. Cameron, "An Analysis of CDMA with Imperfect Power Control," *Proc. 42nd IEEE Vehicular Technology Conf.*, pp. 977–980, May 1992.
14. D. L. Schilling et al., "Broadband CDMA for Personal Communications Systems," *IEEE Commun. Mag.*, vol. 29, pp. 86–93, Nov. 1991.
15. G. L. Turin, "Introduction to Spread-Spectrum Antimultipath Techniques and Their Application to Urban Digital Radio," *Proc. IEEE*, vol. 68, pp. 328–353, March 1980.
16. R. Price and P. E. Green, Jr., "A Communication Technique for Multipath Channels," *Proc. IRE*, vol. 46, pp. 555–570, March 1958.
17. E. O. Geraniotis and M. B. Pursley, "Error Probabilities for Slow-Frequency-Hopped Spread-Spectrum Multiple-Access Communications Over Fading Channels," *IEEE Transactions on Communications*, vol. Com-30, no. 5, pp. 996–1009, May 1982.
18. S. G. Wilson, *Digital Modulation and Coding*, Prentice Hall, Upper Saddle River, NJ, 1996.

PROBLEMS

13.1-1 The so-called integrate-and-dump filter is shown in Fig. P13.1-1. The feedback amplifier is an ideal integrator. The switch s_1 closes momentarily and then opens at the instant $t = T_b$, thus dumping all the charge on C and causing the output to go to zero. The switch s_2 samples the output immediately before the dumping action.

(a) Sketch the output $p_o(t)$ when a square pulse $p(t)$ is applied to the input of this filter.

(b) Sketch the output $p_o(t)$ of the filter matched to the square pulse $p(t)$.

(c) Show that the performance of the integrate-and-dump filter is identical to that of the matched filter; that is, show that ρ in both cases is identical.

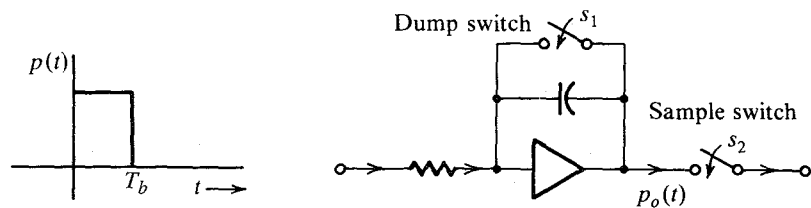


Figure P13.1-1

13.1-2 An alternative to the optimum filter is a suboptimum filter, where we assume a particular filter form and adjust its parameters to maximize ρ . Such filters are inferior to the optimum filter but may be simpler to design.

For a rectangular pulse $p(t)$ of height A and width T_b at the input (Fig. P13.1-2), determine $\rho_{\max}$ if, instead of the matched filter, a one-stage RC filter with $H(\omega) = 1/(1 + j\omega RC)$ is used. Assume a white gaussian noise of PSD $\mathcal{N}/2$. Show that the optimum performance is achieved when $1/RC = 1.26/T_b$. *Hint:* Set $d\rho^2/dx = 0$ ($x = T_b/RC$).

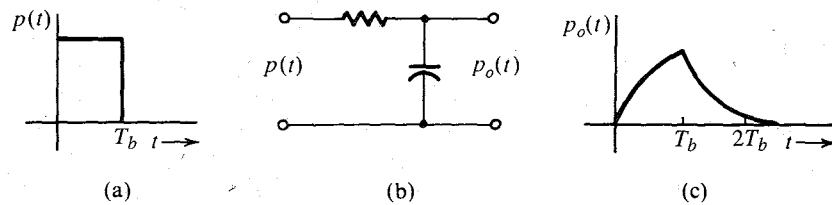


Figure P13.1-2 Suboptimum filter.

13.2-1 As discussed in Sec. 13.3, in coherent schemes, a small pilot is added for synchronization. Because the pilot does not carry information, it causes degradation in P_b . Consider coherent PSK using the following two pulses of duration T_b each:

$$p(t) = A\sqrt{1 - m^2} \cos \omega_c t + Am \sin \omega_c t$$
$$q(t) = -A\sqrt{1 - m^2} \cos \omega_c t + Am \sin \omega_c t$$

where $Am \sin \omega_c t$ is the pilot. Show that when the channel noise is white gaussian,

$$P_b = Q \left[\sqrt{\frac{2E_b(1 - m^2)}{\mathcal{N}}} \right]$$

Hint: Use Eq. (13.18b).

13.2-2 For polar binary communication systems, each error in the decision has some cost. Suppose that when $m = 1$ is transmitted and we read it as $m = 0$ at the receiver, a quantitative penalty, or cost, C_{10} is assigned to such an error, and, similarly, a cost C_{01} is assigned when $m = 0$ is transmitted and we read it as $m = 1$. For the polar case where $P_m(0) = P_m(1) = 0.5$, show that for white gaussian channel noise the optimum threshold that minimizes the overall cost is not 0 but is a_o , given by

$$a_o = \frac{\mathcal{N}}{4} \ln \frac{C_{01}}{C_{10}}$$

Hint: See Hint for Prob. 10.2-11.

13.2-3 For a polar binary system with unequal message probabilities, show that the optimum decision threshold a_o is given by

$$a_o = \frac{\mathcal{N}}{4} \ln \frac{P_m(0)C_{01}}{P_m(1)C_{10}}$$

where C_{01} and C_{10} are the cost of the errors as explained in Prob. 13.2-2, and $P_m(0)$ and $P_m(1)$ are the probabilities of transmitting 0 and 1, respectively. *Hint:* See Hint for Prob. 10.2-11.

13.5-1 For 3-ary communication, messages are chosen from any one of three symbols, m_{-1} , m_0 , and m_1 , which are transmitted by pulses $-p(t)$, 0 , and $p(t)$, respectively. A filter matched to $p(t)$ is used at the receiver. If r is the matched filter output at T_M , plot $p_r(r|m_i)$ ($i = -1, 0$, and 1), and if $P(m_{-1}) = P(m_0) = P(m_1)$, determine the optimum decision thresholds and the error probability P_e . The energy of the pulse $p(t)$ is E_p and the channel noise PSD is $S_n(\omega) = \mathcal{N}/2$.

13.5-2 Binary data is transmitted by using a pulse $p(t)$ for 0 and a pulse $3p(t)$ for 1. Show that the optimum receiver for this case is a filter matched to $p(t)$ with a detection threshold $2E_p$, as shown in Fig. P13.5-2. Determine the error probability P_b of this receiver as a function of $E_b/\mathcal{N}$ if 0 and 1 are equiprobable.

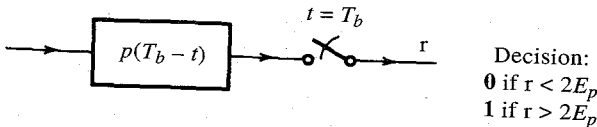


Figure P13.5-2

13.5-3 A binary source emits data at a rate of 256,000 bit/s. Multiamplitude-shift keying (MASK) with $M = 2, 16$, and 32 is considered. In each case, determine the signal power required at the receiver input and the minimum transmission bandwidth required if $S_n(\omega) = 10^{-8}$ and the bit error rate P_b is required to be less than 10^{-7} .

13.5-4 Repeat Prob. 13.5-3 for M -ary PSK.

14 OPTIMUM SIGNAL DETECTION

Our scope of the discussion of signal detection in Chapter 13 was rather limited. In the first place, we restricted ourselves primarily to binary systems with a receiver of assumed (linear) form. Although we optimized the linear type of receiver, we have no assurance that there does not exist another type of receiver that might be superior to the optimum linear receiver derived in Chapter 13. Second, we did not derive the optimum linear receiver for the M -ary case in general. In Chapter 13 we only analyzed a few M -ary schemes with an assumed receiver structure.

In this chapter, we shall analyze the problem of digital signal detection from a more fundamental point of view. We shall determine the optimum receiver (in the sense of minimizing the error probability) for general M -ary signaling in the presence of additive white gaussian noise (AWGN). We shall place no constraints on the receiver. Rather, we shall try to answer the question: What receiver will yield the minimum error probability?

Such an analysis is greatly facilitated by a geometrical representation of signals. We shall now show that a signal is in reality an n -dimensional vector and can be represented by a point in an n -dimensional hyperspace. The foundations for such a viewpoint were laid in Chapter 2. Here we shall refine and concretize our argument. The reader may want to review the concepts of vector representation of a signal in Chapter 2.

14.1 GEOMETRICAL REPRESENTATION OF SIGNALS: SIGNAL SPACE

We are used to 3-dimensional physical space and 3-dimensional vectors in this space. There is no reason, however, to restrict ourselves to three dimensions only. We can extend the concept to an n -dimensional space, although it may be hard to visualize the space for $n > 3$.

To begin with, we note that the familiar 3-dimensional vectors are nothing but entities specified by three numbers (x_1, x_2, x_3) in a certain order (ordered 3-tuple)—nothing more, nothing less. Extending this concept, we say that any entity specified by n numbers in a certain order (ordered n -tuple) is an n -dimensional vector. Thus, if an entity is specified by an

ordered n -tuple ($x_1, x_2, \dots, x_n$), it is an n -dimensional vector $\mathbf{x}$. We define n unit vectors $\phi_1, \phi_2, \dots, \phi_n$ as

$$\begin{aligned}\phi_1 &= (1, 0, 0, \dots, 0) \\ \phi_2 &= (0, 1, 0, \dots, 0) \\ &\dots\dots\dots \\ \phi_n &= (0, 0, 0, \dots, 1)\end{aligned}\quad (14.1)$$

Any vector $\mathbf{x} = (x_1, x_2, \dots, x_n)$ can be expressed as a linear combination of n unit vectors,

$$\mathbf{x} = x_1\phi_1 + x_2\phi_2 + \dots + x_n\phi_n \quad (14.2a)$$

$$= \sum_{k=1}^n x_k\phi_k \quad (14.2b)$$

We define the scalar product $\mathbf{x} \cdot \mathbf{y}$ as

$$\mathbf{x} \cdot \mathbf{y} = \sum_{k=1}^n x_k y_k \quad (14.3)$$

where $\mathbf{y} = (y_1, y_2, \dots, y_n)$ is another n -dimensional vector in the same space. Vectors $\mathbf{x}$ and $\mathbf{y}$ are said to be **orthogonal** if

$$\mathbf{x} \cdot \mathbf{y} = 0 \quad (14.4)$$

The **length** of a vector $\mathbf{x}$ is $|\mathbf{x}|$, defined by

$$|\mathbf{x}|^2 = \mathbf{x} \cdot \mathbf{x} = \sum_{k=1}^n x_k^2 \quad (14.5)$$

A set of n -dimensional vectors is said to be independent if none of the vectors in this set can be represented as a linear combination of the remaining vectors in the set. Thus, if $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ is an independent set, then it is impossible to find constants $a_1, a_2, \dots, a_m$ (not all zero) such that

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \dots + a_m\mathbf{x}_m = 0 \quad (14.6)$$

An n -dimensional space can have at most n independent vectors. If a space has a maximum of n independent vectors, then every vector $\mathbf{x}$ in this space can be expressed as a linear combination of these n independent vectors. If not, then $\mathbf{x}$ becomes a member of the independent vector space. But this is not possible with the assumption that there exists a maximum of n independent vectors. Thus, any vector in this space can be specified by n numbers and, hence, is an n -dimensional space. It also follows from this discussion that the dimensionality of a space can be, at most, as large as the total number of vectors in the space.

A subset of vectors in a given n -dimensional space can have dimensionality less than n . Thus, in a 3-dimensional space, all vectors lying in one plane can be specified by two dimensions, and all vectors lying along a line can be specified by one dimension.

A set of n independent vectors in an n -dimensional space are called **basis vectors** because every vector in this space can be expressed as a linear combination of these n independent

vectors. A set of basis vectors form coordinate axes, and they are not unique. The n unit vectors in Eq. (14.1) are independent and can serve as basis vectors. These vectors have an additional property in that they are all mutually **orthogonal**, that is,

$$\boldsymbol{\varphi}_j \cdot \boldsymbol{\varphi}_k = \begin{cases} 0 & j \neq k \\ 1 & j = k \end{cases} \quad (14.7)$$

Such a set is an **orthonormal** set of vectors because in addition to being orthogonal, their lengths are unity.

A vector $\mathbf{x}$ ($x_1, x_2, \dots, x_n$) can be represented as

$$\mathbf{x} = x_1 \boldsymbol{\varphi}_1 + x_2 \boldsymbol{\varphi}_2 + \dots + x_n \boldsymbol{\varphi}_n$$

If we know $\mathbf{x}$, its components x_k can be found by taking the scalar product of both sides with $\boldsymbol{\varphi}_k$. Using Eq. (14.7) we get

$$\mathbf{x} \cdot \boldsymbol{\varphi}_k = x_k \quad k = 1, 2, \dots, n \quad (14.8)$$

One final observation: in order to represent an n -dimensional vector completely, we need n basis vectors (independent vectors). In general, an n -dimensional vector cannot be completely represented by less than n basis vectors of the space. A set of n basis vectors forms a **complete set**.

Signal as a Vector

Any entity that can be represented by an n -tuple is an n -dimensional vector. If a signal $x(t)$ can be specified by an n -tuple, it, too, is a vector. As we know, a signal $x(t)$ band-limited to B Hz can be specified by the values of its Nyquist samples. Consequently, such a signal, too, is a vector. We can express $x(t)$ as [see Eq. (6.10)]

$$x(t) = \sum_k x \left(\frac{k}{2B} \right) \text{sinc}(2\pi Bt - k\pi) \quad (14.9)$$

Thus, $x(t)$ can be specified by an n -tuple (its Nyquist sample values). With this introduction, let us now turn to a systematic development of the signal-space concept.

We define n signals $\varphi_1(t), \varphi_2(t), \dots, \varphi_n(t)$ as independent if none of these n signals can be represented by a linear combination of the remaining $n - 1$ signals. This means it is impossible to find constants $a_1, a_2, \dots, a_n$ (not all zero) such that

$$a_1 \varphi_1(t) + a_2 \varphi_2(t) + \dots + a_n \varphi_n(t) = 0 \quad (14.10)$$

Suppose that a signal $x(t)$ can be represented by a linear combination of n independent signals $\{\varphi_k(t)\}$ as

$$x(t) = x_1 \varphi_1(t) + x_2 \varphi_2(t) + \dots + x_n \varphi_n(t) \quad (14.11a)$$

$$= \sum_{k=1}^n x_k \varphi_k(t) \quad (14.11b)$$

If every signal in a certain signal space can be represented by a linear combination of n independent signals $\{\varphi_k(t)\}$, then we have an n -dimensional signal space.

The signal set $\{\varphi_k(t)\}$ will be assumed to be *orthogonal*, that is,*

$$\int_{-\infty}^{\infty} \varphi_j(t) \varphi_k(t) dt = \begin{cases} 0 & j \neq k \\ k_j & j = k \end{cases} \quad (14.12)$$

If $k_j = 1$ for all j , then the set is *orthonormal*. For an orthonormal set, the coefficients x_k in Eq. (14.11) can be obtained by multiplying both sides of Eq. (14.11) by $\varphi_k(t)$, then integrating and using Eq. (14.12) (with $k_j = 1$):

$$x_k = \int_{-\infty}^{\infty} x(t) \varphi_k(t) dt \quad (14.13)$$

Once the basis signals $\{\varphi_k(t)\}$ are specified, we can represent a signal $x(t)$ by an n -tuple $(x_1, x_2, \dots, x_n)$. Alternately we may represent this signal geometrically by a point $(x_1, x_2, \dots, x_n)$ in an n -dimensional space. We can now associate a vector $\mathbf{x}$ ($x_1, x_2, \dots, x_n$) with the signal $x(t)$. Note that the basis signal $\varphi_1(t)$ is represented by the corresponding basis vector $\boldsymbol{\varphi}_1(1, 0, 0, \dots, 0)$, and $\varphi_2(t)$ is represented by $\boldsymbol{\varphi}_2(0, 1, 0, \dots, 0)$, and so on.

Turning our attention to Eq. (14.9), we see that in this case the basis signals are $\{\text{sinc}(2\pi Bt - k\pi)\}$. This set is orthogonal because (see Prob. 3.7-4)

$$\int_{-\infty}^{\infty} \text{sinc}(2\pi Bt - j\pi) \text{sinc}(2\pi Bt - k\pi) dt = \begin{cases} 0 & j \neq k \\ 1/2B & j = k \end{cases} \quad (14.14)$$

We can normalize this set by multiplying each member by $\sqrt{2B}$. Thus, if we define

$$\varphi_k(t) = \sqrt{2B} \text{sinc}(2\pi Bt - k\pi) \quad (14.15)$$

then

$$x(t) = \sum_k x_k \varphi_k(t) \quad (14.16a)$$

where

$$x_k = \frac{1}{\sqrt{2B}} x \left(\frac{k}{2B} \right) \quad (14.16b)$$

and $\{\varphi_k(t)\}$ is an orthonormal set of signals. Thus, any band-limited signal $x(t)$ can be represented by a point

$$(\dots, x_{-k}, \dots, x_{-2}, x_{-1}, x_0, x_1, x_2, \dots, x_k, \dots)$$

where x_k is the k th Nyquist sample of $x(t)$ divided by $1/\sqrt{2B}$. To determine the dimensionality of this signal space, recall that a band-limited signal cannot be time-limited (i.e., it exists over an infinite time interval). Hence, the total number of samples at a Nyquist rate of $2B$ samples

* If $\{\varphi_k(t)\}$ is complex, orthogonality implies

$$\int_{-\infty}^{\infty} \varphi_j(t) \varphi_k^*(t) dt = 0$$

and Eq. (14.13) becomes

$$x_k = \int_{-\infty}^{\infty} x(t) \varphi_k^*(t) dt$$

per second will be infinite, and the dimensionality is infinite. Higher dimensions, however, can be ignored, because their contribution is negligible. For example, a band-limited signal with a finite energy may exist over an infinite time interval. But its amplitude for large values of t must approach 0, or the energy

$$E_x = \int_{-\infty}^{\infty} x^2(t) dt$$

will not be finite. Hence, the amplitude of every band-limited signal is negligible beyond some $|t| = T/2$. Such a signal is essentially time-limited to T seconds. We can argue in the same way that a signal time-limited to T seconds is essentially band-limited to B Hz. Because the Nyquist rate is $2B$ samples per second, the total number of samples over T seconds is $2BT + 1$, where T and B are interpreted as before. If $x(t)$ is band-limited to B , then T is its essential time duration, and if $x(t)$ is time-limited to T , then it is essentially band-limited to B Hz. A rigorous development of this result, as well as an estimation of the error in ignoring higher dimensions (those beyond $2BT + 1$), can be found in Landau and Pollak.¹

Just as there are an infinite number of possible sets of basis vectors for a vector space, there are an infinite number of possible sets of basis signals for a given signal space. For a band-limited signal space, $\{\text{sinc}(2\pi Bt - k\pi)\}$ is one possible set of basis signals.

Scalar Product

In a certain signal space, let $x(t)$ and $y(t)$ be two signals represented by vectors $\mathbf{x}$ ($x_1, x_2, \dots, x_n$) and $\mathbf{y}$ ($y_1, y_2, \dots, y_n$). If $\{\phi_k(t)\}$ are the orthonormal basis signals, then

$$x(t) = \sum_i x_i \phi_i(t)$$

$$y(t) = \sum_j y_j \phi_j(t)$$

Hence,

$$\int_{-\infty}^{\infty} x(t)y(t) dt = \int_{-\infty}^{\infty} \left[\sum_i x_i \phi_i(t) \right] \left[\sum_j y_j \phi_j(t) \right] dt$$

Because the basis signals are orthonormal, from Eq. (14.12) with $k_j = 1$, we obtain

$$\int_{-\infty}^{\infty} x(t)y(t) dt = \sum_k x_k y_k \quad (14.17a)$$

The right-hand side of Eq. (14.17a), however, is by definition the scalar product of vectors $\mathbf{x}$ and $\mathbf{y}$;

$$\mathbf{x} \cdot \mathbf{y} = \sum_k x_k y_k$$

Hence, we have

$$\mathbf{x} \cdot \mathbf{y} = \int_{-\infty}^{\infty} x(t)y(t) dt \quad (14.17b)$$

* Including end samples.

If $x(t)$ and $y(t)$ are mutually orthogonal, then it follows from Eq. (14.17b) that the corresponding vectors $\mathbf{x}$ and $\mathbf{y}$ are also orthogonal. We conclude that the integral of the product of two signals is equal to the scalar product of the corresponding vectors.

Energy of a Signal

For a signal $x(t)$, the energy E_x is given by

$$E_x = \int_{-\infty}^{\infty} x^2(t) dt$$

It follows from Eq. (14.17b) that

$$E_x = \mathbf{x} \cdot \mathbf{x} = |\mathbf{x}|^2 \quad (14.18)$$

where $|\mathbf{x}|^2$ is the square of the length of the vector $\mathbf{x}$. Hence, the signal energy is given by the square of the length of the corresponding vector.

EXAMPLE 14.1 A signal space consists of four signals $s_1(t)$, $s_2(t)$, $s_3(t)$, and $s_4(t)$, as shown in Fig. 14.1. Determine a suitable set of basis vectors and the dimensionality of the signals. Represent these signals geometrically in the vector space.

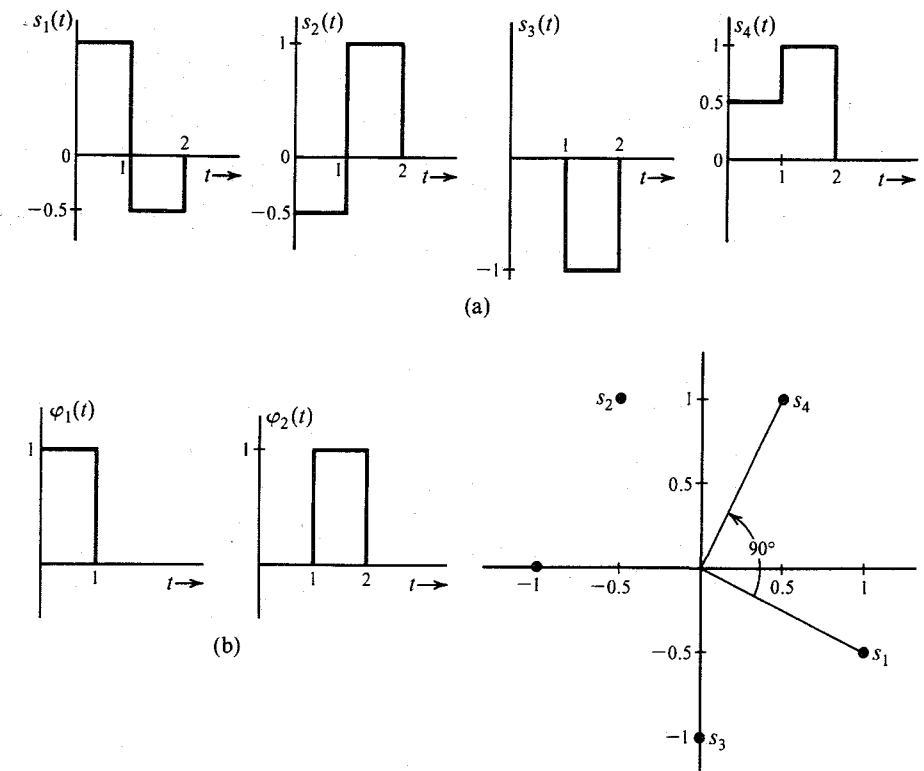


Figure 14.1 Signals and their representation in signal space

The two rectangular pulses $\varphi_1(t)$ and $\varphi_2(t)$ in Fig. 14.1b are suitable as a basis signal set. In terms of this set, the vectors s_1, s_2, s_3 , and s_4 corresponding to signals $s_1(t), s_2(t), s_3(t)$, and $s_4(t)$ are $s_1 = (1, -0.5)$, $s_2 = (-0.5, 1)$, $s_3 = (0, -1)$, and $s_4 = (0.5, 1)$. These points are plotted in Fig. 14.1c. Observe that

$$s_1 \cdot s_4 = 0.5 - 0.5 = 0$$

Hence, s_1 and s_4 are orthogonal. This result may be verified from the fact that

$$\int_{-\infty}^{\infty} s_1(t)s_4(t) dt = 0$$

Note that each point in the signal space in Fig. 14.1c corresponds to some waveform.

Systematic Determination of an Orthogonal Basis Set

We have shown that the dimensionality of a vector space is equal to the maximum number of independent vectors in the space. Thus, in an n -dimensional space, there can be no more than n vectors that are independent. Alternatively, it is always possible to find a set of n vectors that are independent. Once such a set (basis set) is chosen, any vector in this space can be expressed in terms of (as a linear combination of) the vectors in this set. This set of n independent vectors is by no means unique. The reader is familiar with this fact in the physical space of three dimensions, where one can find an infinite number of independent sets of three vectors, each forming a valid coordinate system. If we are given a set of n independent vectors, it is possible to obtain from this set another set of n independent vectors that is orthogonal. This is done by the **Gram-Schmidt orthogonalization process** discussed in Appendix C.

A deterministic signal can be represented by one point in a signal space. A random process, on the other hand, consists of an ensemble of waveforms, each of which maps into a point in a signal space. Hence, a random process appears as an ensemble of points in a signal space. Let us consider in detail the gaussian random process and its representation in a signal space.

14.2 GAUSSIAN RANDOM PROCESS

In order to specify a gaussian random process, we need to familiarize ourselves first with jointly gaussian random variables.

Jointly Gaussian Random Variables

Random variables (RVs) $x_1, x_2, \dots, x_n$ are said to be jointly gaussian if their joint PDF is given by

$$p_{x_1 x_2 \dots x_n}(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{n/2} \sqrt{|K|}} \exp \left[\frac{-1}{2|K|} \sum_i \sum_j \Delta_{ij} (x_i - \bar{x}_i)(x_j - \bar{x}_j) \right] \quad (14.19)$$

where K is the covariance matrix

$$K = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix} \quad (14.20a)$$

and the covariance of x_i and x_j is

$$\sigma_{ij} = \overline{(x_i - \bar{x}_i)(x_j - \bar{x}_j)} \quad (14.20b)$$

$|K|$ is the determinant of matrix K , and Δ_{ij} is the cofactor for the element σ_{ij} . Note that $\sigma_{ii} = \sigma_i^2$.

Gaussian variables are important not only because they are frequently observed, but also because they have certain properties that simplify many mathematical operations that are practically impossible or very difficult for other types of RVs. These properties are as follows:

Property 1: The gaussian density is completely specified by only the first and second moments (means and covariances). This follows from Eq. (14.19). It can be shown that, in general, the probability density of an RV depends on all the moments $\bar{x}^n$ ($n = 1, 2, \dots$) of the variable. A gaussian variable is a special case where only the first two moments (means) and the second moments (covariances) are necessary to specify the distribution.

Property 2: If n jointly gaussian variables $x_1, x_2, \dots, x_n$ are uncorrelated, they are independent.

If the n variables are uncorrelated, $\sigma_{ij} = 0$ ($i \neq j$), and Eq. (14.19) reduces to

$$p_{x_1 x_2 \dots x_n}(x_1, x_2, \dots, x_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \left[\frac{-(x_i - \bar{x}_i)^2}{2\sigma_i^2} \right] \quad (14.21a)$$

$$= p_{x_1}(x_1) p_{x_2}(x_2) \dots p_{x_n}(x_n) \quad (14.21b)$$

As we observed earlier, independent variables are always uncorrelated, but uncorrelated variables are not necessarily independent. For the case of jointly gaussian RVs, however, uncorrelatedness implies independence.

Property 3: When $x_1, x_2, \dots, x_n$ are jointly gaussian, all the marginal densities, such as $p_{x_i}(x_i)$, and all the conditional densities, such as $p_{x_i x_j | x_k x_l \dots x_p}(x_i, x_j | x_k, x_l, \dots, x_p)$, are gaussian. This property can be readily verified (see Prob. 10.2-9).

Property 4: Linear combinations of jointly gaussian variables are also jointly gaussian. Thus, if we form m variables $y_1, y_2, \dots, y_m$ ($m \leq n$) such that

$$y_i = \sum_{k=1}^n a_{ik} x_k$$

then $y_1, y_2, \dots, y_m$ are also jointly gaussian variables.

Definition of a Gaussian Random Process

A random process $x(t)$ is said to be Gaussian if the RVs $x(t_1), x(t_2), \dots, x(t_n)$ are jointly gaussian [Eq. (14.19)] for every n and for every set $(t_1, t_2, \dots, t_n)$.

For convenience, the RV $x(t_i)$ will be denoted by x_i . Hence, the joint PDF of RVs $x_1, x_2, \dots, x_n$ of a gaussian random process is given by Eq. (14.19). Observe that

$$\begin{aligned}\sigma_{ij} &= \overline{(x_i - \bar{x}_i)(x_j - \bar{x}_j)} \\ &= \overline{x_i x_j} - \bar{x}_i \bar{x}_j - \bar{x}_i \bar{x}_j + \bar{x}_i \bar{x}_j \\ &= \overline{x_i x_j} - \bar{x}_i \bar{x}_j \\ &= \overline{x(t_i)x(t_j)} - \bar{x}_i \bar{x}_j \\ &= R_x(t_i, t_j) - \overline{x(t_i)} \overline{x(t_j)}\end{aligned}\quad (14.22)$$

This shows that a gaussian random process is completely specified by its autocorrelation function $R_x(t_i, t_j)$ and its mean value $\overline{x(t)}$.

Stationary Gaussian Random Process

Our discussion of the gaussian process thus far applies to stationary and nonstationary processes. We have shown that the gaussian process is completely specified by its autocorrelation function $R_x(t_i, t_j)$ and its mean value function.* An important corollary of this statement is that if the autocorrelation function $R_x(t_i, t_j)$ and the means are unaffected by a shift of the time origin, then all of the statistics of the process are unaffected by a shift of the time origin. In other words, the process is stationary. Thus, if

$$R_x(t_i, t_j) = R_x(t_i - t_j) \quad (14.23)$$

and

$$\overline{x(t)} = \text{constant for all } t$$

the process is stationary if it is gaussian. But the condition in Eq. (14.23) defines wide-sense stationarity. Hence, *for a gaussian process, wide-sense stationarity implies stationarity in the strict sense.*

For a stationary gaussian process, the covariance σ_{ij} becomes

$$\sigma_{ij} = R_x(t_j - t_i) - \bar{x}^2 \quad (14.24)$$

We shall once again stress the point that distinguishes the gaussian from the nongaussian process. The complete statistics of a gaussian process are determined from its autocorrelation function (which is a second-order parameter) and its mean value. This is the property that simplifies the study of gaussian processes. In general, for nongaussian processes higher order statistics cannot be determined from lower order statistics. For the gaussian process, however, the mean and the second-order parameter σ_{ij} determine all the higher order statistics. This very property also enables us to state that a gaussian process is strictly stationary if it is wide-sense stationary.

Transmission of a Gaussian Process through a Linear System

Another significant property of the gaussian process is that the response of a linear system to a gaussian process is also a gaussian process. This can be shown as follows. Let $x(t)$ be a

* For nonstationary processes (or processes that are not wide-sense stationary), the mean value is a function of t .

gaussian process applied to the input of a linear system whose unit impulse response is $h(t)$. If $y(t)$ is the output (response) process, then

$$\begin{aligned}y(t) &= \int_{-\infty}^{\infty} x(t - \tau)h(\tau) d\tau \\ &= \lim_{\Delta\tau \rightarrow 0} \sum_{k=-\infty}^{\infty} x(t - \tau_k)h(\tau_k) \Delta\tau \quad \tau_k = k\Delta\tau\end{aligned}$$

Because $x(t)$ is a gaussian process, all the variables $x(t - \tau_k)$ are jointly gaussian (by definition). Hence, the variables $y(t_1), y(t_2), \dots, y(t_n)$ for all n and every set $(t_1, t_2, \dots, t_n)$ are linear combinations of variables that are jointly gaussian. Therefore, the variables $y(t_1), y(t_2), \dots, y(t_n)$ must be jointly gaussian, according to the earlier discussion. It follows that the process $y(t)$ is a gaussian process.

To summarize, the gaussian random process has the following properties:

1. A gaussian random process is completely specified by its autocorrelation function and mean value.
2. If a gaussian random process is wide-sense stationary, then it is stationary in the strict sense.
3. The response of a linear system to a gaussian random process is also a gaussian random process.

Geometrical Representation of a Random Process

Let $x(t)$ be a random process and $\varphi_1(t), \varphi_2(t), \dots, \varphi_n(t)$ a complete set of orthonormal basis signals for this space. We can express $x(t)$ as

$$\begin{aligned}x(t) &= x_1\varphi_1(t) + x_2\varphi_2(t) + \dots + x_n\varphi_n(t) \\ &= \sum_{k=1}^n x_k\varphi_k(t)\end{aligned}\quad (14.25)$$

where

$$x_k = \int_{-\infty}^{\infty} x(t)\varphi_k(t) dt \quad (14.26)$$

Note that $x(t)$ is a random process, consisting of an ensemble of sample functions. The coefficients x_k in Eq. (14.25) will be different for each sample function. Consequently, $x_1, x_2, \dots, x_n$ appearing in Eq. (14.25) are RVs. Each sample function will have a specific set $(x_1, x_2, \dots, x_n)$ and will map into one point in the signal space. Hence, the ensemble of sample functions will map into an ensemble of points in the signal space, as shown in Fig. 14.2. This figure shows only a 3-dimensional graph because it is not possible to show a higher dimensional one. But it is sufficient to indicate the idea.

Each time the random experiment is repeated, the outcome (the sample function) is a certain point $\mathbf{x}$. The ensemble of points in the signal space appears as a cloud, with the density of points directly proportional to the probability of observing $\mathbf{x}$ in that region. If we denote the joint PDF of $x_1, x_2, \dots, x_n$ by $p_{\mathbf{x}}(\mathbf{x})$, then

$$p_{\mathbf{x}}(\mathbf{x}) = p_{x_1 x_2 \dots x_n}(x_1, x_2, \dots, x_n)$$

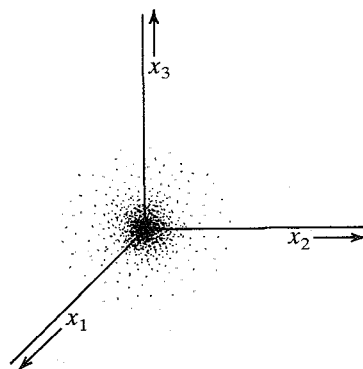


Figure 14.2 Geometrical representation of a gaussian random process.

Thus, $p_{\mathbf{x}}(\mathbf{x})$ has a certain value at each point in the signal space, and $p_{\mathbf{x}}(\mathbf{x})$ represents the relative probability (cloud density) of observing $\mathbf{x} = \mathbf{x}$.

A Word about Notation: Let us clarify the notation used here to avoid confusion later. As before, we use roman type to denote an RV or a random process. Thus, x or $x(t)$ represents an RV or a random process. A particular value assumed by the RV in a certain trial is denoted by italic type. Thus, x represents the value assumed by x . Similarly, $x(t)$ represents a particular sample function of the random process $x(t)$. In the case of random vectors, we follow the same convention; a random vector is denoted by roman boldface type, and a particular value assumed by the vector in a certain trial is represented by boldface italic type. Thus, $\mathbf{r}$ denotes a random vector representing a random process $\mathbf{r}(t)$, but $\mathbf{r}$ is a particular value of $\mathbf{r}$ and represents a particular received waveform (sample function) $\mathbf{r}(t)$ in some trial. Note that roman type represents random entities and italic type represents particular values (which are, of course, nonrandom).

White Gaussian Noise

Consider a white noise process $n_w(t)$ with PSD $\mathcal{N}/2$. Let $\varphi_1(t)$, $\varphi_2(t)$, ... be a complete set of orthonormal basis signals for this space. We can express $n_w(t)$ as

$$\begin{aligned} n_w(t) &= n_1\varphi_1(t) + n_2\varphi_2(t) + \cdots \\ &= \sum_k n_k\varphi_k(t) \end{aligned}$$

White noise has infinite bandwidth. Consequently, the dimensionality of the signal space is infinity.

We shall now show that RVs n_1 , n_2 , ... are independent, with variance $\mathcal{N}/2$ each. Because [see Eq. (14.13)]

$$\begin{aligned} n_k &= \int_{-\infty}^{\infty} n_w(t)\varphi_k(t) dt \\ \overline{n_j n_k} &= \int_{-\infty}^{\infty} \overline{n_w(\alpha)\varphi_j(\alpha)} d\alpha \int_{-\infty}^{\infty} n_w(\beta)\varphi_k(\beta) d\beta \end{aligned}$$

$$\begin{aligned} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \overline{n_w(\alpha)n_w(\beta)} \varphi_j(\alpha)\varphi_k(\beta) d\alpha d\beta \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_{n_w}(\beta - \alpha) \varphi_j(\alpha)\varphi_k(\beta) d\alpha d\beta \end{aligned}$$

Because $R_{n_w}(\tau) = \mathcal{F}^{-1}(\mathcal{N}/2) = (\mathcal{N}/2)\delta(\tau)$,

$$\begin{aligned} \overline{n_j n_k} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\mathcal{N}}{2} \delta(\beta - \alpha) \varphi_j(\alpha)\varphi_k(\beta) d\alpha d\beta \\ &= \frac{\mathcal{N}}{2} \int_{-\infty}^{\infty} \varphi_j(\alpha)\varphi_k(\alpha) d\alpha \\ &= \begin{cases} 0 & j \neq k \\ \frac{\mathcal{N}}{2} & j = k \end{cases} \end{aligned} \quad (14.27)$$

Hence, n_j and n_k are uncorrelated gaussian RVs, each with variance $\mathcal{N}/2$. Since they are gaussian, uncorrelatedness implies independence. This proves the result.

For the time being, assume that we are considering an N -dimensional case. The joint PDF of independent joint gaussian RVs $n_1, n_2, \dots, n_N$, each with zero mean and variance $\mathcal{N}/2$, is [see Eq. (14.21)]

$$\begin{aligned} p_{\mathbf{n}}(\mathbf{n}) &= \prod_{j=1}^N \frac{1}{\sqrt{2\pi\mathcal{N}/2}} e^{-n_j^2/2(\mathcal{N}/2)} \\ &= \frac{1}{(\pi\mathcal{N})^{N/2}} e^{-(n_1^2 + n_2^2 + \cdots + n_N^2)/\mathcal{N}} \end{aligned} \quad (14.28a)$$

$$= \frac{1}{(\pi\mathcal{N})^{N/2}} e^{-|\mathbf{n}|^2/\mathcal{N}} \quad (14.28b)$$

This shows that the PDF $p_{\mathbf{n}}(\mathbf{n})$ depends only on $|\mathbf{n}|$, the magnitude of the noise vector $\mathbf{n}$ in the hyperspace, and is therefore spherically symmetrical if plotted in the N -dimensional hyperspace.

14.3 OPTIMUM RECEIVER²⁻⁵

We shall now consider, from a more fundamental point of view, the problem of M -ary communication in the presence of **additive white gaussian channel noise (AWGN)**. No constraint shall be placed on the optimum structure, and we shall try to answer the fundamental question: What receiver will yield the minimum error probability?

The comprehension of the signal-detection problem is greatly facilitated by geometrical representation of signals. In a signal space, we can represent a signal by a fixed point (or a vector). A random process can be represented by a random point (or a random vector). The region in which the random point may lie will be shown shaded, with the shading intensity proportional to the probability of observing the signal in that region. In the M -ary scheme, we use M symbols, or messages, $m_1, m_2, \dots, m_M$. Each of these symbols is represented by a

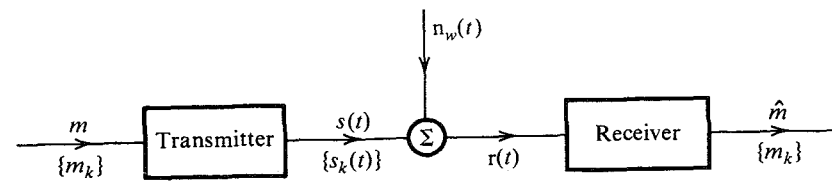


Figure 14.3 *M*-ary communication system.

specified waveform. Let the corresponding waveforms be $s_1(t)$, $s_2(t)$, ..., $s_M(t)$. Thus, the symbol (or message) m_k is sent by transmitting the waveform $s_k(t)$. These waveforms are corrupted by AWGN $n_w(t)$ (Fig. 14.3) with PSD

$$S_{n_w}(\omega) = \frac{\mathcal{N}}{2}$$

At the receiver, the received signal $r(t)$ consists of one of the M message waveforms $s_k(t)$ plus the channel noise,

$$r(t) = s_k(t) + n_w(t) \tag{14.29a}$$

We can represent $r(t)$ in a signal space by letting $\mathbf{r}$, $\mathbf{s}_k$, and $\mathbf{n}_w$ be the points (or vectors) representing signals $r(t)$, $s_k(t)$, and $n_w(t)$, respectively. Then it is evident that

$$\mathbf{r} = \mathbf{s}_k + \mathbf{n}_w \tag{14.29b}$$

The vector $\mathbf{s}_k$ is a fixed vector, because the waveform $s_k(t)$ is nonrandom. The vector $\mathbf{n}_w$ (or point $\mathbf{n}_w$) is random. Hence, the vector $\mathbf{r}$ is also random. Because $n_w(t)$ is a gaussian white noise, the probability distribution of $\mathbf{n}_w$ has spherical symmetry in the signal space (see Sec. 14.2). Hence, the distribution of $\mathbf{r}$ is a spherical distribution centered at a fixed point $\mathbf{s}_k$, as shown in Fig. 14.4. Whenever the message m_k is transmitted, the probability of observing the received signal $r(t)$ in a given region is indicated by the intensity of the shading in Fig. 14.4. Actually, because the noise is white, the space has an infinite number of dimensions. For simplicity, however, we have shown the space to be 3-dimensional. This will suffice to indicate our line of reasoning. We can draw similar regions for various points $\mathbf{s}_1$, $\mathbf{s}_2$, ..., $\mathbf{s}_M$. Figure 14.5a shows the regions for two messages m_j and m_k when $\mathbf{s}_j$ and $\mathbf{s}_k$ are widely separated in signal

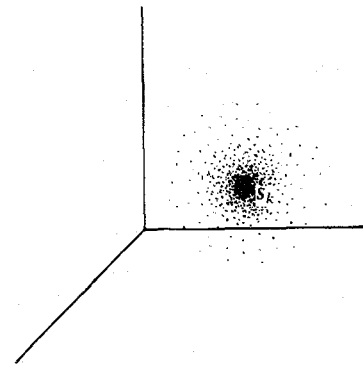


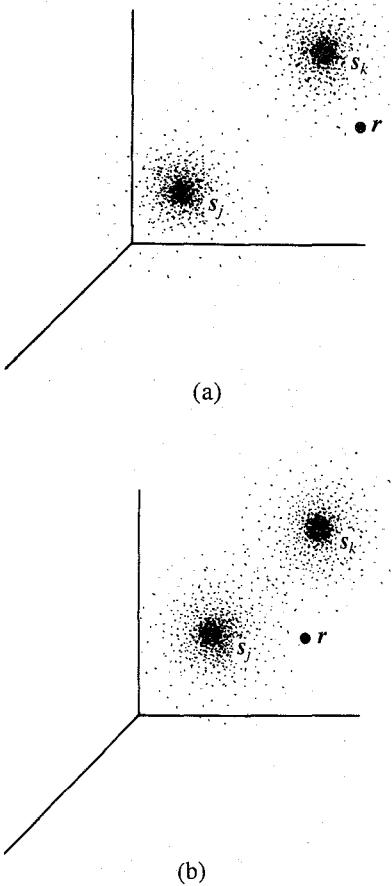
Figure 14.4 Effect of gaussian channel noise on the received signal.

space. In this case, there is virtually no overlap between the two regions. If either m_j or m_k is transmitted, the received signal will lie in one of the two regions. From the position of the received signal, one can decide with a very small probability of error whether m_j or m_k was transmitted. Note that theoretically each region extends to infinity, although the probability of observing the received signal diminishes rapidly as one moves away from the center. Hence, there will always be an overlap between the two regions, resulting in a nonzero error probability. In Fig. 14.5a, the received signal $\mathbf{r}$ is much closer to $\mathbf{s}_k$ than to $\mathbf{s}_j$. It is therefore more likely that m_k was transmitted.

Figure 14.5b illustrates the case when the points $\mathbf{s}_j$ and $\mathbf{s}_k$ are spaced closely together. In this case, there is a considerable overlap between the two regions. Because the received signal $\mathbf{r}$ is closer to $\mathbf{s}_j$ than to $\mathbf{s}_k$, it is more likely that m_j was transmitted. But in this case there is also a considerable probability that m_k may have been transmitted. Hence in this situation, there will be a much higher probability of error in any decision scheme.

The optimum receiver must decide which message has been transmitted from a knowledge of $\mathbf{r}$. The signal space must be divided into M nonoverlapping, or disjoint, regions

Figure 14.5 Binary communication in the presence of noise.



$R_1, R_2, \dots, R_M$, corresponding to the M messages $m_1, m_2, \dots, m_M$. If r falls in the region R_k , the decision is m_k . The problem of designing the receiver then reduces to choosing the boundaries of these regions $R_1, R_2, \dots, R_M$ such that the probability of error in decision making is minimum.

To recapitulate: A transmitter transmits a sequence of messages from a set of M messages $m_1, m_2, \dots, m_M$. These messages are represented by finite energy waveforms $s_1(t), s_2(t), \dots, s_M(t)$. One waveform is transmitted every T_M seconds. We assume that the receiver is time synchronized with the transmitter. The waveforms are corrupted during transmissions by an AWGN of PSD $\mathcal{N}/2$. Knowing the received waveform, the receiver must make a decision as to which waveform was transmitted. The merit criterion of the receiver is the minimum probability of error in making this decision.

Let us now discuss the dimensionality of the signal space in our problem. If there was no noise, we would be dealing with only M waveforms $s_1(t), s_2(t), \dots, s_M(t)$. In this case a signal space of, at most, M dimensions would suffice. This is because the dimensionality of a signal space is always equal to or less than the number of independent signals in the space (see Sec. 14.1). For the sake of generality we shall assume the space to have N dimensions ($N \leq M$). Let $\varphi_1(t), \varphi_2(t), \dots, \varphi_N(t)$ be the orthonormal basis set for this space. Such a set can be constructed by using the Gram-Schmidt procedure discussed in Appendix C. We can then represent the signal waveform $s_k(t)$ as

$$\begin{aligned} s_k(t) &= s_{k1}\varphi_1(t) + s_{k2}\varphi_2(t) + \dots + s_{kN}\varphi_N(t) \\ &= \sum_{j=1}^N s_{kj}\varphi_j(t) \end{aligned} \quad (14.30a)$$

where

$$s_{kj} = \int_{-\infty}^{\infty} s_k(t)\varphi_j(t) dt \quad (14.30b)$$

Now consider the white gaussian channel noise $n_w(t)$. This signal has an infinite bandwidth ($B = \infty$). It has an infinite number of dimensions and obviously cannot be represented in the N -dimensional signal space discussed earlier. We can, however, split $n_w(t)$ into two components: (1) the projection of $n_w(t)$ on the N -dimensional signal space, and (2) the remaining component, which will be orthogonal to the N -dimensional signal space. Let us denote the two components by $n(t)$ and $n_0(t)$. Thus,

$$n_w(t) = n(t) + n_0(t) \quad (14.31)$$

where

$$n(t) = \sum_{k=1}^N n_k\varphi_k(t) \quad (14.32a)$$

and

$$n_j = \int_{-\infty}^{\infty} n(t)\varphi_j(t) dt \quad (14.32b)$$

Because $n_0(t)$ is orthogonal to the N -dimensional space, it is orthogonal to every signal in that space. Hence,

$$\int_{-\infty}^{\infty} n_0(t)\varphi_j(t) dt = 0 \quad j = 1, 2, \dots, N$$

Therefore,

$$\begin{aligned} n_j &= \int_{-\infty}^{\infty} [n(t) + n_0(t)]\varphi_j(t) dt \\ &= \int_{-\infty}^{\infty} n_w(t)\varphi_j(t) dt \end{aligned} \quad (14.33)$$

From Eqs. (14.33) and (14.32a) it is evident that we can filter out the component $n_0(t)$ from $n_w(t)$. This can be seen from the fact that the received signal, $r(t)$, can be expressed as

$$\begin{aligned} r(t) &= s_k(t) + n_w(t) \\ &= s_k(t) + n(t) + n_0(t) \\ &= q(t) + n_0(t) \end{aligned} \quad (14.34)$$

where $q(t)$ is the projection of $r(t)$ on the N -dimensional space. Thus,

$$q(t) = s_k(t) + n(t) \quad (14.35)$$

We can obtain the projection $q(t)$ from $r(t)$ by observing that [see Eqs. (14.30) and (14.32a)]

$$q(t) = \sum_{j=1}^N (s_{kj} + n_j)\varphi_j(t) \quad (14.36)$$

From Eqs. (14.30b), (14.33), and (14.36) it follows that if we feed the received signal $r(t)$ into the system shown in Fig. 14.6, the resultant outcome will be $q(t)$. Thus, the orthogonal noise component can be filtered out without disturbing the message signal. The question here is: Would such filtering help in our decision making? We can easily show that it cannot hurt us. The noise $n_w(t)$ is independent of the signal waveform $s_k(t)$. Therefore, its component $n_0(t)$ is also independent of $s_k(t)$. Thus, $n_0(t)$ contains no information about the transmitted signal, and discarding such a component from the received signal $r(t)$ will not cause any loss of information regarding the signal waveform $s_k(t)$. This, however, is not enough. We must also make sure that the noise being discarded [$n_0(t)$] is not in any way related to the remaining noise component $n(t)$. If $n_0(t)$ and $n(t)$ are related in any way, it will be possible to obtain some information about $n(t)$ from $n_0(t)$, thus enabling us to detect that signal with less error probability. If the components $n_0(t)$ and $n(t)$ are independent random processes, the component $n_0(t)$ does not carry any information about $n(t)$ and can be discarded. Under these conditions, $n_0(t)$ is **irrelevant** to the decision making at the receiver.

The process $n(t)$ is represented by components $n_1, n_2, \dots, n_N$ along $\varphi_1(t), \varphi_2(t), \dots, \varphi_N(t)$, and $n_0(t)$ is represented by the remaining components (infinite number) along the remaining basis signals in the complete set, $\{\varphi_k(t)\}$. From Eq. (14.27) we observe that all the components are independent. Hence, the components representing $n_0(t)$ are independent of the components representing $n(t)$. Consequently, $n_0(t)$ is independent of $n(t)$ and is irrelevant data.

The received signal $r(t)$ is now reduced to the signal $q(t)$, which contains the desired signal waveform and the projection of the channel noise on the N -dimensional signal space.

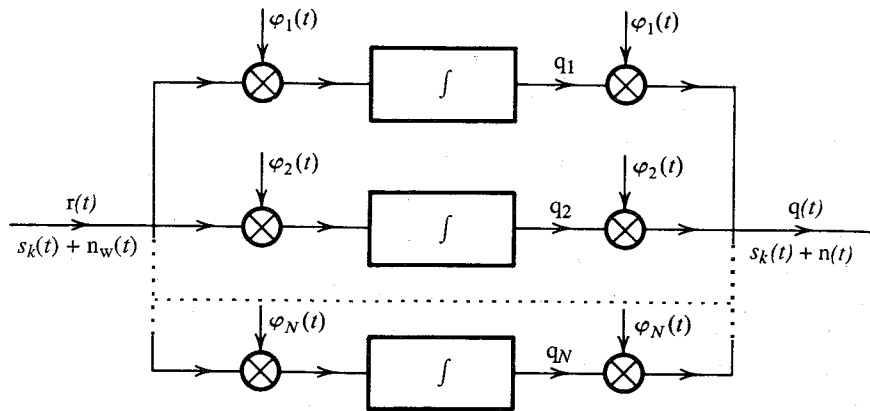


Figure 14.6 Eliminating the noise orthogonal to signal space.

Thus, the signal $q(t)$ can be completely represented in the signal space. Let the vectors representing $n(t)$ and $q(t)$ be denoted by $\mathbf{n}$ and $\mathbf{q}$. Thus,

$$\mathbf{q} = \mathbf{s} + \mathbf{n}$$

where $\mathbf{s}$ may be any one of vectors $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_M$.

The random vector $\mathbf{n}$ ($n_1, n_2, \dots, n_N$) is represented by N independent Gaussian variables, each with zero mean and variance $\sigma_n^2 = \mathcal{N}/2$. The joint PDF of vector $\mathbf{n}$ in such a case has a spherical symmetry, as shown in Eq. (14.28b),

$$p_{\mathbf{n}}(\mathbf{n}) = \frac{1}{(\pi \mathcal{N})^{N/2}} e^{-|\mathbf{n}|^2/\mathcal{N}} \quad (14.37a)$$

Note that this is actually a compact notation for

$$p_{n_1, n_2, \dots, n_N}(n_1, n_2, \dots, n_N) = \frac{1}{(\pi \mathcal{N})^{N/2}} e^{-(n_1^2 + n_2^2 + \dots + n_N^2)/\mathcal{N}} \quad (14.37b)$$

Decision Procedure

Our problem is now considerably simplified. The irrelevant noise component has been filtered out. The residual signal $q(t)$ can be represented in an N -dimensional signal space. We proceed to determine the M decision regions $R_1, R_2, \dots, R_M$ in this space. The regions must be so chosen that the probability of error in making the decision is minimized.

Suppose the received vector $\mathbf{q} = \mathbf{q}$. Then if the receiver decides $\hat{m} = m_k$, the conditional probability of making the correct decision, given that $\mathbf{q} = \mathbf{q}$, is

$$P(C|\mathbf{q} = \mathbf{q}) = P(m_k|\mathbf{q} = \mathbf{q}) \quad (14.38)$$

where $P(C|\mathbf{q} = \mathbf{q})$ is the conditional probability of making the correct decision given $\mathbf{q} = \mathbf{q}$, and $P(m_k|\mathbf{q} = \mathbf{q})$ is the conditional probability that m_k was transmitted given $\mathbf{q} = \mathbf{q}$. The unconditional probability $P(C)$ is given by

$$P(C) = \int_{\mathbf{q}} P(C|\mathbf{q} = \mathbf{q}) p_{\mathbf{q}}(\mathbf{q}) d\mathbf{q} \quad (14.39)$$

where the integration is performed over the entire region occupied by $\mathbf{q}$. Note that this is an N -fold integration with respect to the variables $q_1, q_2, \dots, q_N$ over the range $(-\infty, \infty)$. Also, because $p_{\mathbf{q}}(\mathbf{q}) \geq 0$, this integral is maximum when $P(C|\mathbf{q} = \mathbf{q})$ is maximum. From Eq. (14.38) it now follows that if a decision $\hat{m} = m_k$ is made, the error probability is minimized if

$$P(m_k|\mathbf{q} = \mathbf{q})$$

is maximized. The probability $P(m_k|\mathbf{q} = \mathbf{q})$ is called the **a posteriori probability** of m_k . This is because it represents the probability that m_k was transmitted when $\mathbf{q}$ is received.

The decision procedure is now clear. Once we receive $\mathbf{q} = \mathbf{q}$, we evaluate all M a posteriori probabilities. Then we make the decision in favor of that message for which the a posteriori probability is highest—that is, the receiver decides that $\hat{m} = m_k$ if

$$P(m_k|\mathbf{q} = \mathbf{q}) > P(m_j|\mathbf{q} = \mathbf{q}) \quad \text{for all } j \neq k \quad (14.40)$$

Thus, the detector that minimizes the error probability is the **maximum a posteriori probability (MAP) detector**.

We can use Bayes' mixed rule [Eq. (10.49b)] to determine the a posteriori probabilities. We have

$$P(m_k|\mathbf{q} = \mathbf{q}) = \frac{P(m_k) p_{\mathbf{q}}(\mathbf{q}|m_k)}{p_{\mathbf{q}}(\mathbf{q})} \quad (14.41)$$

Hence, the receiver decides $\hat{m} = m_k$ if the decision function

$$\frac{P(m_i) p_{\mathbf{q}}(\mathbf{q}|m_i)}{p_{\mathbf{q}}(\mathbf{q})} \quad i = 1, 2, \dots, M$$

is maximum for $i = k$.

Note that the denominator $p_{\mathbf{q}}(\mathbf{q})$ is common to all decision functions and, hence, may be ignored. Thus, the receiver sets $\hat{m} = m_k$ if the decision function

$$P(m_i) p_{\mathbf{q}}(\mathbf{q}|m_i) \quad i = 1, 2, \dots, M \quad (14.42)$$

is maximum for $i = k$.

Thus, once $\mathbf{q}$ is obtained, we compute the decision function [Eq. (14.42)] for all messages $m_1, m_2, \dots, m_M$ and decide that the message for which the function is maximum is the one most likely to have been sent.

We now turn our attention to computing the decision functions. The a priori probability $P(m_i)$ represents the probability that the message m_i will be transmitted. These probabilities must be known if the criterion discussed is to be used.* The term $p_{\mathbf{q}}(\mathbf{q}|m_i)$ represents the PDF of $\mathbf{q}$ when $s(t) = s_i(t)$. Under this condition,

$$\mathbf{q} = \mathbf{s}_i + \mathbf{n}$$

and

$$\mathbf{n} = \mathbf{q} - \mathbf{s}_i$$

The point $\mathbf{s}_i$ is constant, and $\mathbf{n}$ is a random point. Obviously, $\mathbf{q}$ is a random point with the same distribution as $\mathbf{n}$ but centered at the points $\mathbf{s}_i$.

* In case these probabilities are unknown, one must use other merit criteria, such as maximum likelihood or minimax, as will be discussed in later sections.

Alternately, the probability $\mathbf{q} = \mathbf{q}$ (given $m = m_i$) is the same as the probability $\mathbf{n} = \mathbf{q} - \mathbf{s}_i$. Hence [Eq. (14.37a)],

$$p_{\mathbf{q}}(\mathbf{q}|m_i) = p_{\mathbf{n}}(\mathbf{q} - \mathbf{s}_i) = \frac{1}{(\pi \mathcal{N})^{N/2}} e^{-|\mathbf{q} - \mathbf{s}_i|^2 / \mathcal{N}} \quad (14.43)$$

The decision function in Eq. (14.42) now becomes

$$\frac{P(m_i)}{(\pi \mathcal{N})^{N/2}} e^{-|\mathbf{q} - \mathbf{s}_i|^2 / \mathcal{N}} \quad (14.44)$$

Note that the decision function is always nonnegative for all values of i . Hence, comparing these functions is equivalent to comparing their logarithms, because the logarithm is a monotone function for the positive argument. Hence, for convenience, the decision function will be chosen as the logarithm of Eq. (14.44). In addition, the factor $(\pi \mathcal{N})^{N/2}$ is common for all i and can be left out. Hence, the decision function is

$$\ln P(m_i) - \frac{1}{\mathcal{N}} |\mathbf{q} - \mathbf{s}_i|^2 \quad (14.45)$$

Note that $|\mathbf{q} - \mathbf{s}_i|^2$ is the square of the length of the vector $\mathbf{q} - \mathbf{s}_i$. Hence,

$$\begin{aligned} |\mathbf{q} - \mathbf{s}_i|^2 &= (\mathbf{q} - \mathbf{s}_i) \cdot (\mathbf{q} - \mathbf{s}_i) \\ &= |\mathbf{q}|^2 + |\mathbf{s}_i|^2 - 2\mathbf{q} \cdot \mathbf{s}_i \end{aligned} \quad (14.46)$$

Hence, the decision function in Eq. (14.45) becomes (after multiplying throughout by $\mathcal{N}/2$)

$$\frac{\mathcal{N}}{2} \ln P(m_i) - \frac{1}{2} [|\mathbf{q}|^2 + |\mathbf{s}_i|^2 - 2\mathbf{q} \cdot \mathbf{s}_i] \quad (14.47)$$

Note that the term $|\mathbf{s}_i|^2$ is the square of the length of $\mathbf{s}_i$ and represents E_i , the energy of signal $\mathbf{s}_i(t)$. The terms $\mathcal{N} \ln P(m_i)$ and E_i are constants in the decision function. Let

$$a_i = \frac{1}{2} [\mathcal{N} \ln P(m_i) - E_i] \quad (14.48)$$

Now the decision function in Eq. (14.47) becomes

$$a_i + \mathbf{q} \cdot \mathbf{s}_i - \frac{1}{2} |\mathbf{q}|^2$$

The term $\frac{1}{2} |\mathbf{q}|^2$ is common to all M decision functions and can be omitted for the purpose of comparison. Thus, the new decision function b_i is

$$b_i = a_i + \mathbf{q} \cdot \mathbf{s}_i \quad (14.49)$$

We compute this function b_i for $i = 1, 2, \dots, N$, and the receiver decides that $\hat{m} = m_k$ if this function is the largest for $i = k$. If the signal $q(t)$ is applied at the input terminals of a system whose impulse response is $h(t)$, the output at $t = T_M$ is given by

$$\int_{-\infty}^{\infty} q(\lambda) h(T_M - \lambda) d\lambda$$

If we choose a filter matched to $\mathbf{s}_i(t)$, that is, $h(t) = \mathbf{s}_i(T_M - t)$,

$$h(T_M - \lambda) = \mathbf{s}_i(\lambda)$$

and the output is

$$\int_{-\infty}^{\infty} q(\lambda) \mathbf{s}_i(\lambda) d\lambda = \mathbf{q} \cdot \mathbf{s}_i$$

Hence, $\mathbf{q} \cdot \mathbf{s}_i$ is the output at $t = T_M$ of a filter matched to $\mathbf{s}_i(t)$ when $q(t)$ is applied to its input.

Actually, we do not have $q(t)$. The incoming signal $r(t)$ is given by

$$\begin{aligned} r(t) &= \mathbf{s}_i(t) + n_w(t) \\ &= \underbrace{\mathbf{s}_i(t) + n(t)}_{q(t)} + \underbrace{n_0(t)}_{\text{irrelevant noise}} \end{aligned}$$

where $n_0(t)$ is the (irrelevant) component of $n_w(t)$ orthogonal to the N -dimensional signal space. Because $n_0(t)$ is orthogonal to this space, it is orthogonal to every signal in this space. Hence, it is orthogonal to the signal $\mathbf{s}_i(t)$, and

$$\int_{-\infty}^{\infty} n_0(t) \mathbf{s}_i(t) dt = 0$$

and

$$\begin{aligned} \mathbf{q} \cdot \mathbf{s}_i &= \int_{-\infty}^{\infty} q(t) \mathbf{s}_i(t) dt + \int_{-\infty}^{\infty} n_0(t) \mathbf{s}_i(t) dt \\ &= \int_{-\infty}^{\infty} [q(t) + n_0(t)] \mathbf{s}_i(t) dt \\ &= \int_{-\infty}^{\infty} r(t) \mathbf{s}_i(t) dt \end{aligned} \quad (14.50)$$

Hence, it is immaterial whether we use $q(t)$ or $r(t)$ at the input. We thus apply the incoming signal $r(t)$ to a parallel bank of matched filters, and the output of the filters is sampled at $t = T_M$. Then a constant a_i is added to the i th filter output sample, and the resulting outputs are compared. The decision is made in favor of the signal for which this output is the largest. The receiver implementation for this decision procedure is shown in Fig. 14.7a. As shown in Chapter 13, a matched filter is equivalent to a correlator. One may therefore use correlators instead of matched filters. Such an arrangement is shown in Fig. 14.7b.

We have shown that in the presence of AWGN, the matched-filter receiver is the optimum receiver when the merit criterion is the minimum error probability. Note that the system is linear, although it was not constrained to be so. Therefore, for white gaussian noise the optimum receiver happens to be linear. The matched filter obtained in Chapter 13 and the decision procedure are identical to those derived here.

The optimum receiver can be implemented in another way. From Eq. (14.50), we have

$$\mathbf{q} \cdot \mathbf{s}_i = \mathbf{r} \cdot \mathbf{s}_i$$

From Eq. (14.3), we can rewrite this as

$$\mathbf{q} \cdot \mathbf{s}_i = \sum_{j=1}^N r_j s_{ij}$$

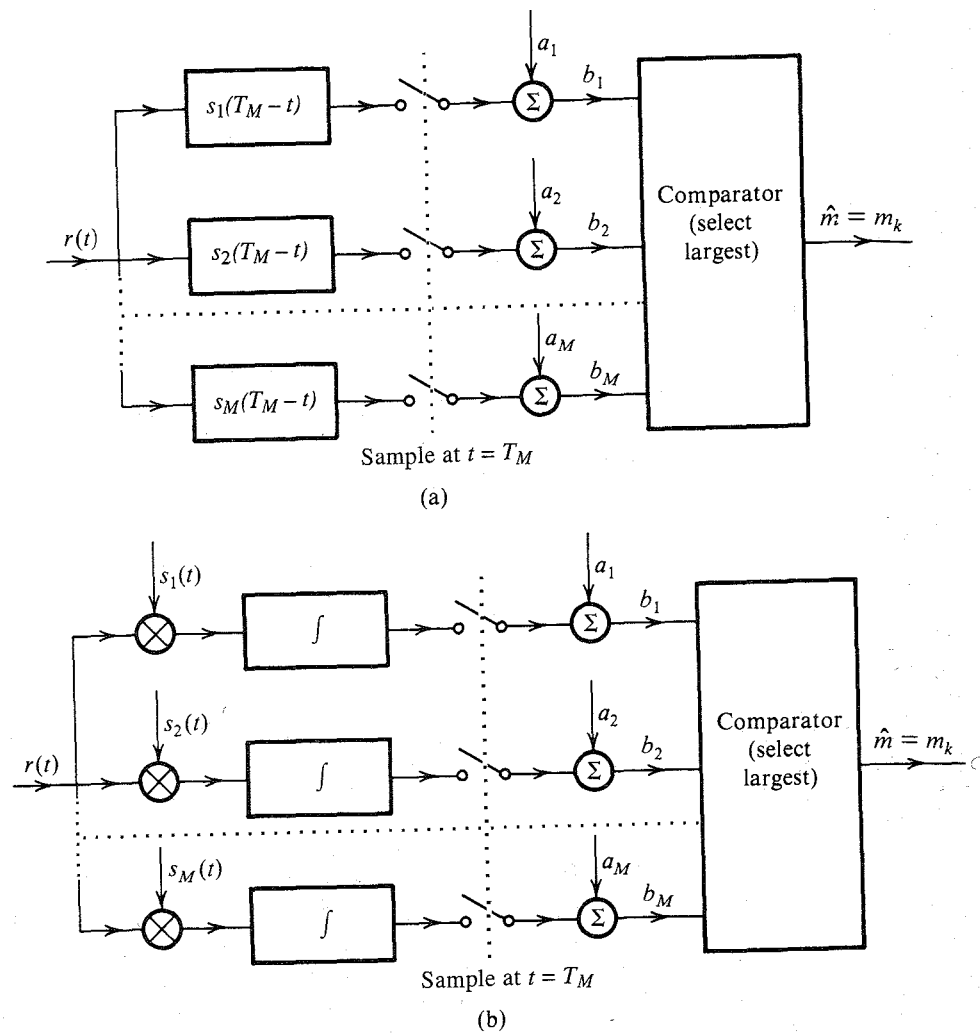


Figure 14.7 Optimum M -ary receiver. (a) Matched-filter detector. (b) Correlation detector.

The term $q \cdot s_i$ is computed according to this equation by first generating r_j 's and then computing the sum of $r_j s_{ij}$ (remember that the s_{ij} 's are known), as shown in Fig. 14.8a. The M correlator detectors in Fig. 14.7b can be replaced by N filters matched to $\varphi_1(t)$, $\varphi_2(t)$, ..., $\varphi_N(t)$, as shown in Fig. 14.8b. Both types of optimum receivers (Figs. 14.7 and 14.8) perform identically. The choice will depend on the circumstances. For example, if $N < M$ and signals $\{\varphi_j(t)\}$ are easier to generate than $\{s_j(t)\}$, then the choice of Fig. 14.8 is obvious.

Decision Regions and Error Probability

In order to compute the error probability of the optimum receiver, we need to determine decision regions in the signal space first. As mentioned earlier, the signal space is divided into M nonoverlapping, or disjoint, decision regions $R_1, R_2, \dots, R_M$, corresponding to M

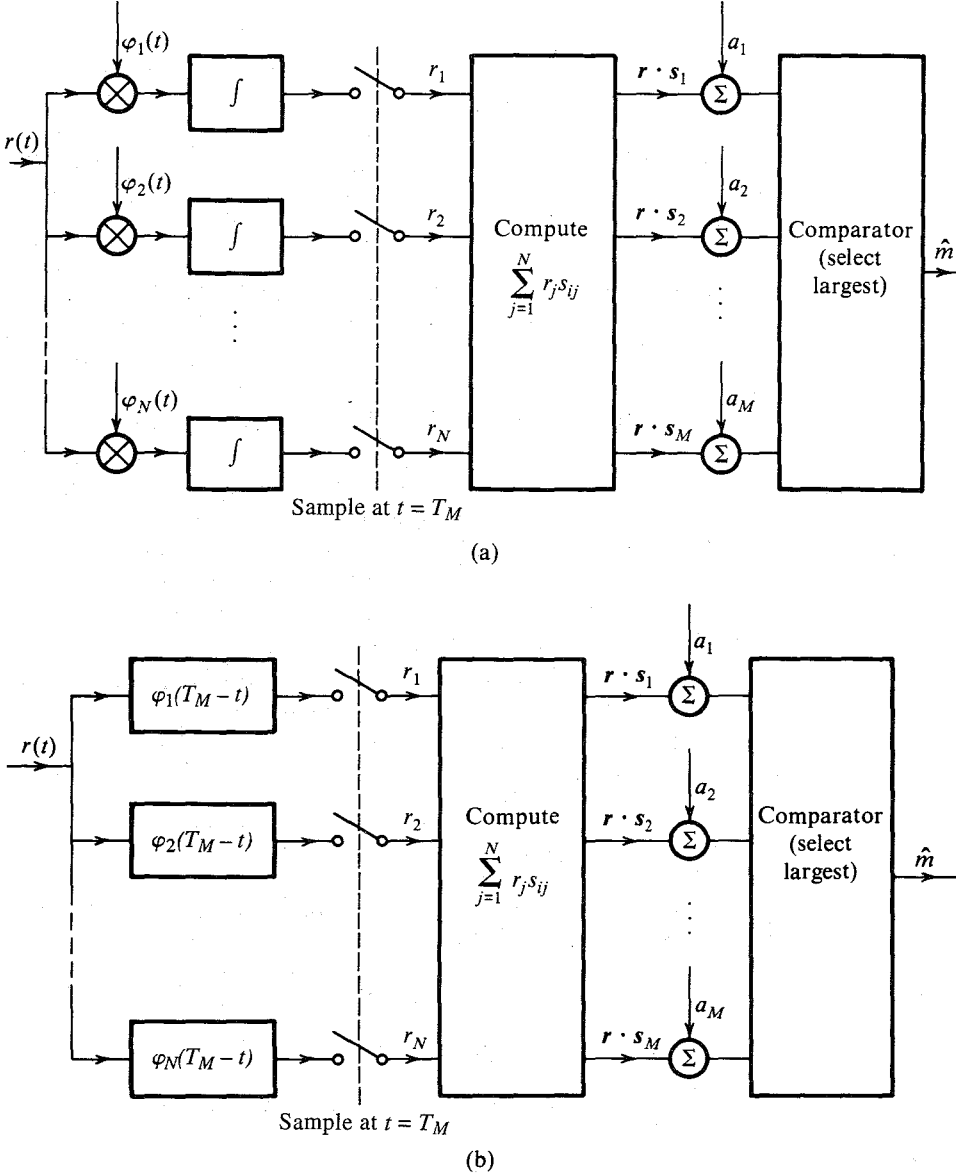


Figure 14.8 Another form of optimum M -ary receiver. (a) Correlation detector. (b) Matched-filter detector.

messages. If q falls in the region R_k , the decision is that m_k was transmitted. The decision regions are so chosen that the probability of error of the receiver is minimum. In light of this geometrical representation, we shall now try to interpret how the optimum receiver sets these decision regions.

The decision function is given by Eq. (14.45). The optimum receiver sets $\hat{m} = m_k$ if the decision function

$$\mathcal{N} \ln P(m_i) - |\mathbf{q} - \mathbf{s}_i|^2$$

is maximum for $i = k$. This equation defines the decision regions.

For simplicity, let us first consider the case of equiprobable messages; that is, $P(m_i) = 1/M$ for all i . In this case, the first term in the decision function is the same for all i and, hence, can be dropped. Thus, the receiver decides that $\hat{m} = m_k$ if the term $-|\mathbf{q} - \mathbf{s}_i|^2$ is largest (numerically the smallest) for $i = k$. Alternatively, this may also be stated as follows: the receiver decides that $\hat{m} = m_k$ if the decision function $|\mathbf{q} - \mathbf{s}_i|^2$ is minimum for $i = k$. Note that $|\mathbf{q} - \mathbf{s}_i|$ is the distance of point $\mathbf{q}$ from point $\mathbf{s}_i$. Thus, the decision procedure in this case has a simple interpretation in geometrical space. The decision is made in favor of that signal which is closest to $\mathbf{q}$, the projection of $\mathbf{r}$ [the component of $\mathbf{r}(t)$] in the signal space. This result is expected on qualitative grounds for gaussian noise, because the gaussian noise has a spherical symmetry. If, however, the messages are not equiprobable, we cannot go too far on purely qualitative grounds. Nevertheless, we can draw certain broad conclusions. If a particular message m_i is more likely than the others, one will be safer in deciding more often in favor of m_i than other messages. Hence, in such a case the decision regions will be biased, or weighted, in favor of m_i . This is shown by the appearance of the term $\ln P(m_i)$ in the decision function. To better understand this point, let us consider a 2-dimensional signal space and two signals $\mathbf{s}_1$ and $\mathbf{s}_2$, as shown in Fig. 14.9a. In this figure, the decision regions R_1 and R_2 are shown for equiprobable messages; $P(m_1) = P(m_2) = 0.5$. The boundary of the decision region is the perpendicular bisector of the line joining points $\mathbf{s}_1$ and $\mathbf{s}_2$. Note that any point on the boundary is equidistant from $\mathbf{s}_1$ and $\mathbf{s}_2$. If $\mathbf{q}$ happens to fall on the boundary, we just "flip a coin" and decide whether to select m_1 or m_2 . Figure 14.9b shows the case when the two messages are not equiprobable. To delineate the boundary of the decision regions, we use Eq. (14.45). The decision is m_1 if

$$|\mathbf{q} - \mathbf{s}_1|^2 - \mathcal{N} \ln P(m_1) < |\mathbf{q} - \mathbf{s}_2|^2 - \mathcal{N} \ln P(m_2)$$

Otherwise, the decision is m_2 .

Note that $|\mathbf{q} - \mathbf{s}_1|$ and $|\mathbf{q} - \mathbf{s}_2|$ represent d_1 and d_2 , the distance of $\mathbf{q}$ from $\mathbf{s}_1$ and $\mathbf{s}_2$, respectively. Thus, the decision is m_1 if

$$d_1^2 - d_2^2 < \mathcal{N} \ln \frac{P(m_1)}{P(m_2)}$$

The right-hand side of this inequality is a constant c :

$$c = \mathcal{N} \ln \frac{P(m_1)}{P(m_2)}$$

Thus, the decision is m_1 if

$$d_1^2 - d_2^2 < c$$

The decision is m_2 if

$$d_1^2 - d_2^2 > c$$

On the boundary of the decision regions,

$$d_1^2 - d_2^2 = c$$

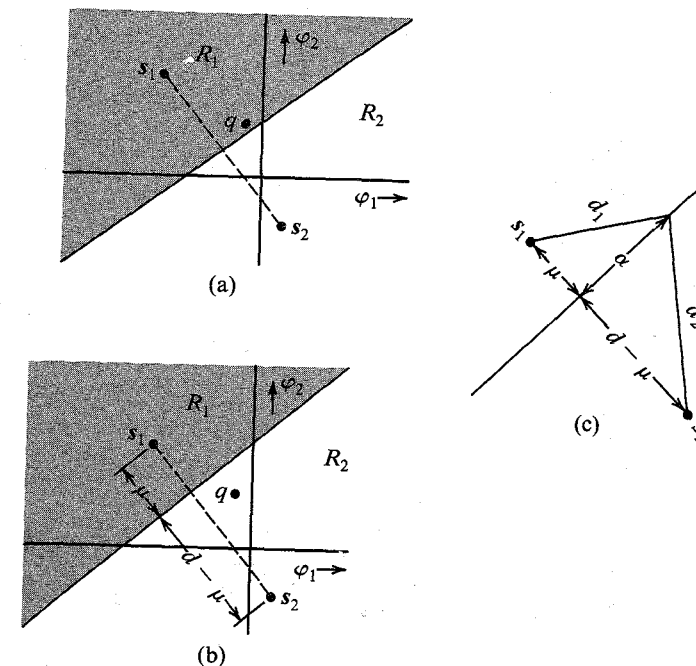


Figure 14.9 Determining optimum decision regions in a binary case.

We now show that such a boundary is given by a straight line perpendicular to line $\mathbf{s}_1\mathbf{s}_2$ and passing through $\mathbf{s}_1\mathbf{s}_2$ at a distance μ from $\mathbf{s}_1$, where

$$\mu = \frac{c + d^2}{2d} = \frac{\mathcal{N}}{2d} \ln \left[\frac{P(m_1)}{P(m_2)} \right] + \frac{d}{2} \quad (14.51)$$

where d is the distance between $\mathbf{s}_1$ and $\mathbf{s}_2$. To prove this, we redraw the pertinent part of Fig. 14.9b as Fig. 14.9c. It is evident from this figure that

$$d_1^2 = \alpha^2 + \mu^2$$

$$d_2^2 = \alpha^2 + (d - \mu)^2$$

Hence,

$$d_1^2 - d_2^2 = 2d\mu - d^2 = c$$

Therefore,

$$\mu = \frac{c + d^2}{2d}$$

This is the desired result. Thus, along the decision boundary $d_1^2 - d_2^2$ is constant and equal to c . The boundaries of the decision regions for $M > 2$ may be determined along similar lines. The decision regions for the case of three equiprobable 2-dimensional signals are shown in Fig. 14.10. The boundaries of the decision regions are perpendicular bisectors of the lines

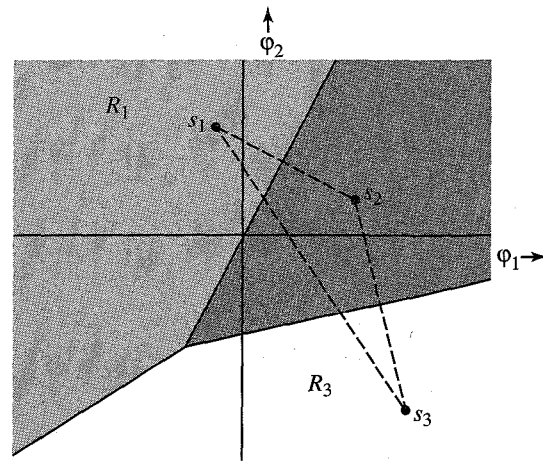


Figure 14.10 Determining optimum decision regions.

joining the original transmitted signals. If the signals are not equiprobable, then the boundaries will be shifted away from the signals with larger probabilities of occurrence.

For signals in N -dimensional space, the decision regions will be N -dimensional hypercones.

If there are M messages $m_1, m_2, \dots, m_M$ with decision regions $R_1, R_2, \dots, R_M$, respectively, then $P(C|m_i)$, the probability of a correct decision when m_i is transmitted, is given by

$$P(C|m_i) = P(q \text{ lies in } R_i) \quad (14.52)$$

and $P(C)$, the probability of a correct decision, is given by

$$P(C) = \sum_{i=1}^M P(m_i) P(C|m_i) \quad (14.53a)$$

and P_{eM} , the probability of error, is given by

$$P_{eM} = 1 - P(C) \quad (14.53b)$$

EXAMPLE 14.2 Binary data is transmitted using polar signaling over an AWGN channel with noise PSD $\mathcal{N}/2$. The two signals used are

$$s_1(t) = -s_2(t) = \sqrt{E}\varphi(t)$$

The symbol probabilities $P(m_1)$ and $P(m_2)$ are unequal. Design the optimum receiver and determine the corresponding error probability.

The two signals are represented graphically in Fig. 14.11a. If the energy of each signal is E , the distance of each signal from the origin is $\sqrt{E}$. The distance d between the two signals is

$$d = 2\sqrt{E}$$

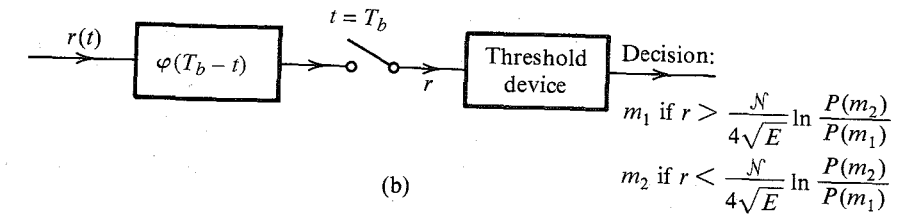
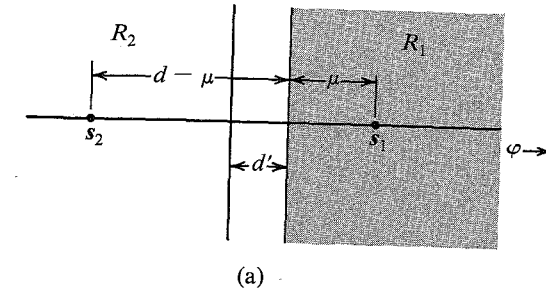


Figure 14.11 Decision regions for a binary case in Example 14.2.

The decision regions R_1 and R_2 are shown in Fig. 14.11a. The distance μ is given by Eq. (14.51). Also,

$$\begin{aligned} P(C|m = m_1) &= P(\text{noise vector originating at } s_1 \text{ remains in } R_1) \\ &= P(n > -\mu) \\ &= 1 - Q\left(\frac{\mu}{\sigma_n}\right) \\ &= 1 - Q\left(\frac{\mu}{\sqrt{\mathcal{N}/2}}\right) \end{aligned}$$

Similarly,

$$P(C|m = m_2) = 1 - Q\left(\frac{d - \mu}{\sqrt{\mathcal{N}/2}}\right)$$

and

$$\begin{aligned} P(C) &= P(m_1) \left[1 - Q\left(\frac{\mu}{\sqrt{\mathcal{N}/2}}\right) \right] + P(m_2) \left[1 - Q\left(\frac{d - \mu}{\sqrt{\mathcal{N}/2}}\right) \right] \\ &= 1 - P(m_1) Q\left(\frac{\mu}{\sqrt{\mathcal{N}/2}}\right) - P(m_2) Q\left(\frac{d - \mu}{\sqrt{\mathcal{N}/2}}\right) \end{aligned}$$

and

$$P_e = 1 - P(C) = P(m_1) Q\left(\frac{\mu}{\sqrt{\mathcal{N}/2}}\right) + P(m_2) Q\left(\frac{d - \mu}{\sqrt{\mathcal{N}/2}}\right) \quad (14.54a)$$

where

$$d = 2\sqrt{E} \quad (14.54b)$$

and

$$\mu = \frac{\mathcal{N}}{4\sqrt{E}} \ln \frac{P(m_1)}{P(m_2)} + \sqrt{E} \quad (14.54c)$$

When $P(m_1) = P(m_2) = 0.5$, $\mu = \sqrt{E} = d/2$, and Eq. (14.54a) reduces to

$$P_e = Q\left(\sqrt{\frac{2E}{\mathcal{N}}}\right) \quad (14.54d)$$

In this problem, because $N = 1$ and $M = 2$, the receiver in Fig. 14.8 is preferable to that in Fig. 14.7. For this case the receiver of the form in Fig. 14.8b reduces to that shown in Fig. 14.11b. The decision threshold d' as seen from Fig. 14.11a is

$$d' = \sqrt{E} - \mu = \frac{\mathcal{N}}{4\sqrt{E}} \ln \frac{P(m_2)}{P(m_1)}$$

Note that d' is the decision threshold. Thus, in Fig. 14.11b, if the receiver output $r > d'$, the decision is m_1 . Otherwise the decision is m_2 .

When $P(m_1) = P(m_2) = 0.5$, the decision threshold is zero. This is precisely the result derived in Chapter 13 for polar signaling.

EXAMPLE 14.3 QAM

Design the optimum receiver and compute the corresponding error probability for the 16-point QAM configuration shown in Fig. 14.12a, assuming all signals to be equiprobable and assuming an AWGN channel.

Let us first calculate the error probability. The first quadrant of the signal space is reproduced in Fig. 14.12b. Because all the signals are equiprobable, the decision region boundaries will be perpendicular bisectors joining various signals, as shown in Fig. 14.12b.

From Fig. 14.12b it follows that

$$\begin{aligned} P(C|m_1) &= P(\text{noise vector originating at } s_1 \text{ lies within } R_1) \\ &= P\left(n_1 > -\frac{d}{2}, n_2 > -\frac{d}{2}\right) \\ &= P\left(n_1 > -\frac{d}{2}\right) P\left(n_2 > -\frac{d}{2}\right) \\ &= \left[1 - Q\left(\frac{d/2}{\sigma_n}\right)\right]^2 \\ &= \left[1 - Q\left(\frac{d}{\sqrt{2\mathcal{N}}}\right)\right]^2 \end{aligned}$$

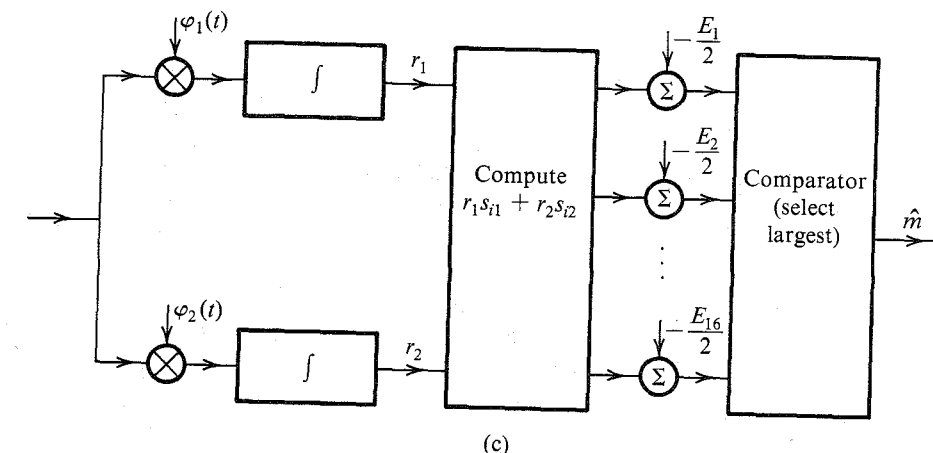
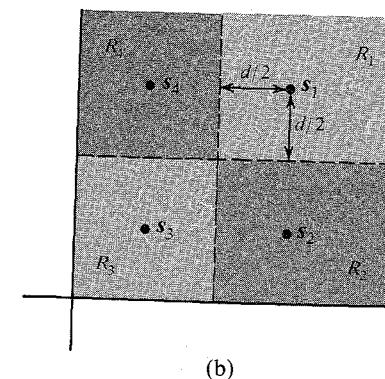
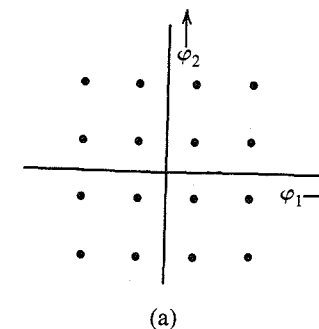


Figure 14.12 16-ary QAM.

For convenience, let us define

$$p = 1 - Q\left(\frac{d}{\sqrt{2\mathcal{N}}}\right) \quad (14.55)$$

Hence,

$$P(C|m_1) = p^2$$

Using similar arguments, we have

$$\begin{aligned} P(C|m_2) &= P(C|m_4) = \left[1 - Q\left(\frac{d}{\sqrt{2\mathcal{N}}}\right)\right] \left[1 - 2Q\left(\frac{d}{\sqrt{2\mathcal{N}}}\right)\right] \\ &= p(2p - 1) \end{aligned}$$

and

$$P(C|m_3) = (2p - 1)^2$$

Because of the symmetry of the signals in all four quadrants, we get similar probabilities for the four signals in each quadrant. Hence,

$$\begin{aligned}
 P(C) &= \sum_{i=1}^{16} P(C|m_i)P(m_i) \\
 &= \frac{1}{16} \sum_{i=1}^{16} P(C|m_i) \\
 &= \frac{1}{16} [4p^2 + 4p(2p-1) + 4p(2p-1) + 4(2p-1)^2] \\
 &= \frac{1}{4} [9p^2 - 6p + 1] \\
 &= \left(\frac{3p-1}{2} \right)^2 \quad (14.56a)
 \end{aligned}$$

and

$$P_{eM} = 1 - P(C) = \frac{9}{4} \left(p + \frac{1}{3} \right) (1-p)$$

In practice, $P_{eM} \rightarrow 0$, and, hence, $P(C) \rightarrow 1$. This means $p \simeq 1$ [see Eq. (14.56a)], and

$$P_{eM} \simeq 3(1-p) = 3Q \left(\frac{d}{\sqrt{2\mathcal{N}}} \right) \quad (14.56b)$$

To express this in terms of the received power S_i , we determine $\bar{E}$, the average energy of the signal set in Fig. 14.12. Because E_k , the energy of s_k , is the square of the distance of s_k from the origin,

$$\begin{aligned}
 E_1 &= \left(\frac{3d}{2} \right)^2 + \left(\frac{3d}{2} \right)^2 = \frac{9}{2}d^2 \\
 E_2 &= \left(\frac{3d}{2} \right)^2 + \left(\frac{d}{2} \right)^2 = \frac{5}{2}d^2
 \end{aligned}$$

Similarly,

$$E_3 = \frac{d^2}{2} \quad \text{and} \quad E_4 = \frac{5}{2}d^2$$

Hence,

$$\bar{E} = \frac{1}{4} \left[\frac{9}{2}d^2 + \frac{5}{2}d^2 + \frac{d^2}{2} + \frac{5}{2}d^2 \right] = \frac{5}{2}d^2$$

and $d^2 = 0.4\bar{E}$. Moreover, for $M = 16$, each symbol carries the information of $\log_2 16 = 4$ bits. Hence, the energy per bit E_b is

$$E_b = \frac{\bar{E}}{4}$$

and

$$\frac{E_b}{\mathcal{N}} = \frac{\bar{E}}{4\mathcal{N}} = \frac{5d^2}{8\mathcal{N}}$$

Hence,

$$\begin{aligned}
 P_{eM} &= 3Q \left(\frac{d}{\sqrt{2\mathcal{N}}} \right) \\
 &= 3Q \left(\sqrt{\frac{4E_b}{5\mathcal{N}}} \right) \quad (14.57)
 \end{aligned}$$

A comparison of this with binary PSK [Eq. (13.21b)] shows that 16-point QAM requires almost 2.5 times as much power as does binary PSK; but the rate of transmission is increased by a factor of $\log_2 M = 4$. This comparison does not take into account the fact that P_b , the BER, is somewhat smaller than P_{eM} . In this case, $N = 2$ and $M = 16$. Hence, the receiver in Fig. 14.8 is preferable. Such a receiver is shown in Fig. 14.12c. Note that because all signals are equiprobable,

$$a_i = \frac{-E_i}{2}$$

For QAM, $\varphi_1(t) = \sqrt{2/T_M} \cos \omega_o(t)$ and $\varphi_2(t) = \sqrt{2/T_M} \sin \omega_o(t)$, where $\omega_o = 2\pi/T_M$.

EXAMPLE 14.4 MPSK

Determine the error probability of the optimum receiver for equiprobable MPSK signals, each with energy E .

Figure 14.13a shows the MPSK signal configuration for $M = 8$. Because all the signals are equiprobable, the decision regions are conical, as shown. The message m_1 is transmitted by a signal $s_1(t)$ represented by the vector $\mathbf{s}_1(s_1, 0)$. If the projection in the signal space of the received signal $\mathbf{r}$ is $\mathbf{q}(q_1, q_2)$, and the noise is $\mathbf{n}(n_1, n_2)$, then

$$\mathbf{q} = (s_1 + n_1, n_2) = (\underbrace{\sqrt{E} + n_1}_{q_1}, \underbrace{n_2}_{q_2})$$

Also,

$$P(C|m_1) = P(\mathbf{q} \text{ lies in } R_1)$$

This is simply the volume under the conical region of the joint PDF of q_1 and q_2 . Because n_1 and n_2 are independent gaussian RVs with variance $\mathcal{N}/2$, q_1 and q_2 are independent gaussian variables with means $\sqrt{E}$ and 0, respectively, and each with variance $\mathcal{N}/2$. Hence,

$$P_{q_1 q_2}(q_1, q_2) = \left[\frac{1}{\sqrt{\pi\mathcal{N}}} e^{-(q_1 - \sqrt{E})^2/\mathcal{N}} \right] \left[\frac{1}{\sqrt{\pi\mathcal{N}}} e^{-q_2^2/\mathcal{N}} \right]$$

and

$$P(C|m_1) = \frac{1}{\pi\mathcal{N}} \iint_{R_1} e^{-[(q_1 - \sqrt{E})^2 + q_2^2]/\mathcal{N}} dq_1 dq_2 \quad (14.58a)$$

$$= \frac{1}{\pi\mathcal{N}} \int_{q_1} \left(\int_{q_2} e^{-q_2^2/\mathcal{N}} dq_2 \right) e^{-(q_1 - \sqrt{E})^2/\mathcal{N}} dq_1 \quad (14.58b)$$

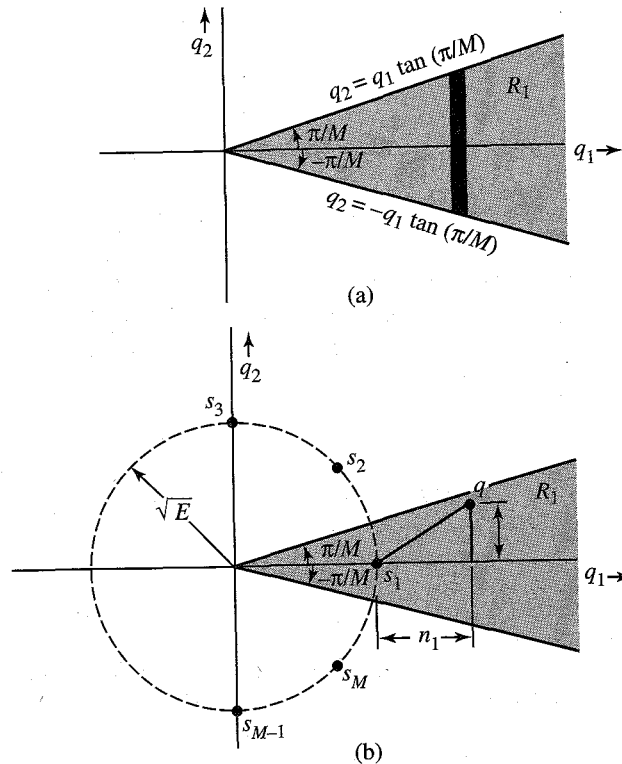


Figure 14.13 MPSK signals.

To integrate over R_1 , we first integrate over the (dark) shaded strip in Fig. 14.13b. Along the border of R_1 ,

$$q_2 = \pm \left(\tan \frac{\pi}{M} \right) q_1$$

Hence,

$$\begin{aligned} P(C|m_1) &= \frac{1}{\pi \mathcal{N}} \int_0^\infty \left(\int_{-q_1 \tan(\pi/M)}^{q_1 \tan(\pi/M)} e^{-q_2^2/\mathcal{N}} dq_2 \right) e^{-(q_1 - \sqrt{E})^2/\mathcal{N}} dq_1 \\ &= \frac{1}{\sqrt{\pi \mathcal{N}}} \int_0^\infty \left[1 - 2Q \left(\frac{q_1 \tan(\pi/M)}{\sqrt{\mathcal{N}/2}} \right) \right] e^{-(q_1 - \sqrt{E})^2/\mathcal{N}} dq_1 \end{aligned}$$

Changing the variable to $x = \sqrt{2/\mathcal{N}} q_1$, we get

$$P(C|m_1) = \frac{1}{\sqrt{2\pi}} \int_0^\infty \left[1 - 2Q \left(x \tan \frac{\pi}{M} \right) \right] e^{-(x - \sqrt{2E/\mathcal{N}})^2/2} dx \quad (14.59a)$$

Using the fact that E_b , the energy per bit, is $E/\log_2 M$, we have

$$P(C|m_1) = \frac{1}{\sqrt{2\pi}} \int_0^\infty \left[1 - 2Q \left(x \tan \frac{\pi}{M} \right) \right] e^{-[x - \sqrt{(2 \log_2 M) E_b/\mathcal{N}}]^2/2} dx \quad (14.59b)$$

The integration can also be performed in cylindrical coordinates using the transformation $q_1 = \rho \sqrt{\mathcal{N}/2} \cos \theta$ and $q_2 = \rho \sqrt{\mathcal{N}/2} \sin \theta$. The limits on ρ are $(0, \infty)$ and those on θ are $-\pi/M$ to π/M . Hence,

$$P(C|m_1) = \frac{1}{2\pi} \int_{-\pi/M}^{\pi/M} d\theta \int_0^\infty \rho e^{-(\rho^2 - 2\rho \sqrt{2E/\mathcal{N}} \cos \theta + 2E/\mathcal{N})/2} d\rho \quad (14.60a)$$

$$= \frac{1}{2\pi} \int_{-\pi/M}^{\pi/M} d\theta \int_0^\infty \rho e^{-[\rho^2 - 2\rho \sqrt{(2 \log_2 M) E_b/\mathcal{N}} \cos \theta + (2 \log_2 M) E_b/\mathcal{N}]/2} d\rho \quad (14.60b)$$

Because of the symmetry of the signal configuration, $P(C|m_i)$ is the same for all i . Hence,

$$P(C) = P(C|m_1)$$

and

$$P_{eM} = 1 - P(C|m_1)$$

P_{eM} is computed numerically. The plot is shown in Fig. 13.22. Still another form of the integral for P_{eM} was found in Chapter 13 [Eq. (13.53)]. For MPSK, $s_i(t) = \sqrt{2E/T_M} \cos(\omega_o t + \theta_i)$, where $\omega_o = 2\pi/T_M$, $\theta_i = 2\pi i/M$, and the optimum receiver turns out to be just a phase detector similar to that shown in Fig. 14.12 (see Prob. 14.3-3).

General Expression for Error Probability

Thus far we have considered rather simple schemes where the decision regions can be found easily. The method of computing error probabilities from the knowledge of decision regions has also been discussed. When the dimensions of the signal space increase, it becomes difficult to visualize the decision regions graphically, and as a result the method loses its power. We now develop an analytical expression for computing error probability for a general M -ary scheme.

From the structure of the optimum receiver in Fig. 14.7, we observe that if m_1 is transmitted, then the correct decision will be made only if

$$b_1 > b_2, b_3, \dots, b_M$$

In other words,

$$P(C|m_1) = \text{probability}(b_1 > b_2, b_3, \dots, b_M|m_1) \quad (14.61)$$

If m_1 is transmitted, then (Fig. 14.7)

$$b_k = \int_0^{T_M} [s_1(t) + n(t)] s_k(t) dt + a_k \quad (14.62)$$

Let

$$\rho_{ij} = \int_0^{T_M} s_i(t) s_j(t) dt \quad i, j = 1, 2, \dots, M \quad (14.63)$$

ρ_{ij} are known as **crosscorrelation coefficients**. Thus (if m_1 is transmitted),

$$b_k = \rho_{1k} + \int_0^{T_M} n(t) s_k(t) dt + a_k \quad (14.64a)$$

$$= \rho_{1k} + a_k + \sum_{j=1}^N s_{kj} n_j \quad (14.64b)$$

where n_j is the component of $n(t)$ along $\varphi_j(t)$. Note that $\rho_{1k} + a_k$ is a constant, and variables n_j ($j = 1, 2, \dots, N$) are independent jointly gaussian variables, each with zero mean and a variance of $\mathcal{N}/2$. Thus, variables b_k are a linear combination of jointly gaussian variables. It follows that the variables $b_1, b_2, \dots, b_M$ are also jointly gaussian. The probability of making a correct decision when m_1 is transmitted can be computed from Eq. (14.61). Note that b_1 can lie anywhere in the range $(-\infty, \infty)$. More precisely, if $p(b_1, b_2, \dots, b_M | m_1)$ is the joint PDF of $b_1, b_2, \dots, b_M$, then Eq. (14.61) can be expressed as

$$P(C|m_1) = \int_{-\infty}^{\infty} \int_{-\infty}^{b_1} \dots \int_{-\infty}^{b_1} p(b_1, b_2, \dots, b_M | m_1) db_1, db_2, \dots, db_M \quad (14.65a)$$

where the limits of integration of b_1 are $(-\infty, \infty)$, and for the remaining variables the limits are $(-\infty, b_1)$. Thus,

$$P(C|m_1) = \int_{-\infty}^{\infty} db_1 \int_{-\infty}^{b_1} db_2 \dots \int_{-\infty}^{b_1} p(b_1, b_2, \dots, b_M | m_1) db_M \quad (14.65b)$$

Similarly, $P(C|m_2), \dots, P(C|m_M)$ can be computed, and

$$P(C) = \sum_{j=1}^M P(C|m_j) P(m_j)$$

and

$$P_{eM} = 1 - P(C)$$

EXAMPLE 14.5 Orthogonal Signal Set

In this set all M equal-energy signals $s_1(t), s_2(t), \dots, s_M(t)$ are mutually orthogonal. As an example, a signal set for $M = 3$ is shown in Fig. 14.14.

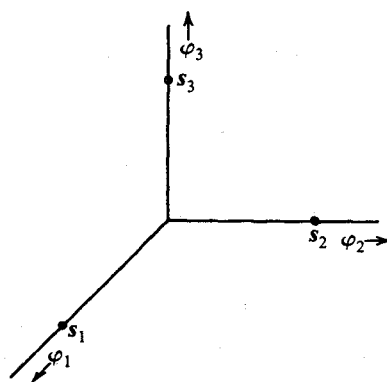


Figure 14.14 Orthogonal signals.

The orthogonal set $\{s_k(t)\}$ is characterized by the fact that

$$s_j \cdot s_k = \begin{cases} 0 & j \neq k \\ E & j = k \end{cases} \quad (14.66)$$

Hence,

$$\rho_{ij} = s_i \cdot s_j = \begin{cases} 0 & i \neq j \\ E & i = j \end{cases} \quad (14.67)$$

Further, we shall assume all signals to be equiprobable. This yields

$$\begin{aligned} a_k &= \frac{1}{2} \left[\mathcal{N} \ln \left(\frac{1}{M} \right) - E_k \right] \\ &= -\frac{1}{2} (\mathcal{N} \ln M + E) \end{aligned}$$

where E is the energy of each signal. Note that a_k is the same for all values of k . Because the constants a_k enter the expression only for the sake of comparison (see Fig. 14.8b), they can be ignored, and hence, we can let $a_k = 0$. Also for an orthogonal set,

$$s_k(t) = \sqrt{E} \varphi_k(t) \quad (14.68)$$

Therefore,

$$s_{kj} = \begin{cases} \sqrt{E} & k = j \\ 0 & k \neq j \end{cases} \quad (14.69)$$

Hence, from Eqs. (14.64b), (14.67), and (14.69), we have (when m_1 is transmitted)

$$b_k = \begin{cases} E + \sqrt{E} n_1 & k = 1 \\ \sqrt{E} n_k & k = 2, 3, \dots, M \end{cases} \quad (14.70)$$

Note that $n_1, n_2, \dots, n_M$ are independent gaussian variables, each with zero mean and variance $\mathcal{N}/2$. Variables b_k that are of the form $(\alpha n_k + \beta)$ are also independent gaussian variables. Equation (14.70) shows that the variable b_1 has the mean E and variance $(\sqrt{E})^2 (\mathcal{N}/2) = \mathcal{N}E/2$. Hence,

$$\begin{aligned} p_{b_1}(b_1) &= \frac{1}{\sqrt{\pi \mathcal{N}E}} e^{-(b_1 - E)^2 / \mathcal{N}E} \\ p_{b_k}(b_k) &= \frac{1}{\sqrt{\pi \mathcal{N}E}} e^{-b_k^2 / \mathcal{N}E} \quad k = 2, 3, \dots, M \end{aligned}$$

Because $b_1, b_2, \dots, b_M$ are independent, the joint probability density is the product of the individual densities:

$$p(b_1, b_2, \dots, b_M | m_1) = \frac{1}{\sqrt{\pi \mathcal{N}E}} e^{-(b_1 - E)^2 / \mathcal{N}E} \prod_{k=2}^M \left(\frac{1}{\sqrt{\pi \mathcal{N}E}} e^{-b_k^2 / \mathcal{N}E} \right)$$

and

$$\begin{aligned}
P(C|m_1) &= \frac{1}{\sqrt{\pi \mathcal{N}E}} \int_{-\infty}^{\infty} db_1 [e^{-(b_1-E)^2/\mathcal{N}E}] \times \prod_{k=2}^M \left(\int_{-\infty}^{b_1} \frac{1}{\sqrt{\pi \mathcal{N}E}} e^{-b_k^2/\mathcal{N}E} db_k \right) \\
&= \frac{1}{\sqrt{\pi \mathcal{N}E}} \int_{-\infty}^{\infty} db_1 [e^{-(b_1-E)^2/\mathcal{N}E}] \times \left(\int_{-\infty}^{b_1} \frac{1}{\sqrt{\pi \mathcal{N}E}} e^{-x^2/\mathcal{N}E} dx \right)^{M-1} \\
&= \frac{1}{\sqrt{\pi \mathcal{N}E}} \int_{-\infty}^{\infty} \left[1 - Q \left(\frac{b_1}{\sqrt{\mathcal{N}E/2}} \right) \right]^{M-1} \times e^{-(b_1-E)^2/\mathcal{N}E} db_1 \quad (14.71a)
\end{aligned}$$

Changing the variable so that $b_1/\sqrt{\mathcal{N}E/2} = y$, and recognizing that $E/\mathcal{N} = (\log_2 M) E_b/\mathcal{N}$, we obtain

$$P(C|m_1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(y-\sqrt{2E/\mathcal{N}})^2/2} [1 - Q(y)]^{M-1} dy \quad (14.71b)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-[y-\sqrt{(2\log_2 M)E_b/\mathcal{N}}]^2/2} [1 - Q(y)]^{M-1} dy \quad (14.71c)$$

Note that because this signal set is geometrically symmetrical,

$$P(C|m_1) = P(C|m_2) = \dots = P(C|m_M)$$

Hence,

$$P(C) = P(C|m_1)$$

and

$$\begin{aligned}
P_{eM} &= 1 - P(C) \\
&= 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-[y-\sqrt{(2\log_2 M)E_b/\mathcal{N}}]^2/2} [1 - Q(y)]^{M-1} dy \quad (14.71d)
\end{aligned}$$

This confirms our earlier result in Eq. (13.57). Figure 13.25 shows the plot of P_{eM} vs. $E_b/\mathcal{N}$.

Significance of the Geometrical Configuration of Signals

From this discussion, one very interesting fact emerges. Whenever the optimum receiver is used, the error probability does not depend on specific signal waveforms but it depends only on their geometrical configuration in the signal space. A given configuration in a signal space can represent infinite possible signal sets, depending on the choice of the basis signals used. A given configuration *does* specify the average energy of the set. This means the error probability depends on signal waveforms only through the average energy of the set. Thus, the average signal energy (or power) emerges as a fundamental parameter that determines the error probability.

This discussion also vividly demonstrates the insight provided by the use of signal space in the study of optimum receivers.

Bandwidth of M -ary Signals

As discussed in Sec. 14.1, the dimensionality of a signal is $2BT_M + 1$, where T_M is the signal duration and B is its essential bandwidth. It follows that for an N -dimensional signal space ($N \leq M$), the bandwidth is $B = (N - 1)/2T_M$. Thus, reducing N reduces the bandwidth.

We can verify that N -dimensional signals can be transmitted over $(N - 1)/2T_M$ Hz by constructing a specific signal set. Let us choose the signals

$$\begin{aligned}
\varphi_0(t) &= \frac{1}{\sqrt{T_M}} \\
\varphi_1(t) &= \sqrt{\frac{2}{T_M}} \sin \omega_o t \\
\varphi_2(t) &= \sqrt{\frac{2}{T_M}} \cos \omega_o t \quad \omega_o = \frac{2\pi}{T_M} \\
\varphi_3(t) &= \sqrt{\frac{2}{T_M}} \sin 2\omega_o t \quad 0 \leq t \leq T_M \\
\varphi_4(t) &= \sqrt{\frac{2}{T_M}} \cos 2\omega_o t \\
&\dots \dots \dots \\
\varphi_{k-1}(t) &= \sqrt{\frac{2}{T_M}} \sin \left(\frac{k}{2} \omega_o t \right) \\
\varphi_k(t) &= \sqrt{\frac{2}{T_M}} \cos \left(\frac{k}{2} \omega_o t \right)
\end{aligned} \quad (14.72)$$

These are $k + 1$ orthogonal pulses with a total bandwidth of $(k/2)(\omega_o/2\pi) = k/2T_M$ Hz. Hence, when $k + 1 = N$, the bandwidth* is $(N - 1)/2T_M$. Thus, $N = 2T_M B + 1$.

To attain a given error probability, there is a trade-off between the average energy of the signal set and its bandwidth. If we reduce the signal space dimensionality, the transmission bandwidth is reduced. But the signals are now closer together, because of the reduced dimensionality. This will increase P_{eM} . Hence, to maintain a given P_{eM} , we must move the signals farther apart; that is, increase energy. Thus, the cost of reduced bandwidth is paid in terms of increased energy. In Chapter 15 we shall study the ideal trade-off between bandwidth and signal energy and compare the performance of various signal schemes in light of this relationship.

* Here we are ignoring the band spreading at the edge. This spread is about $1/T_M$ Hz. The actual bandwidth is larger than $(N - 1)/2T_M$ by this amount.

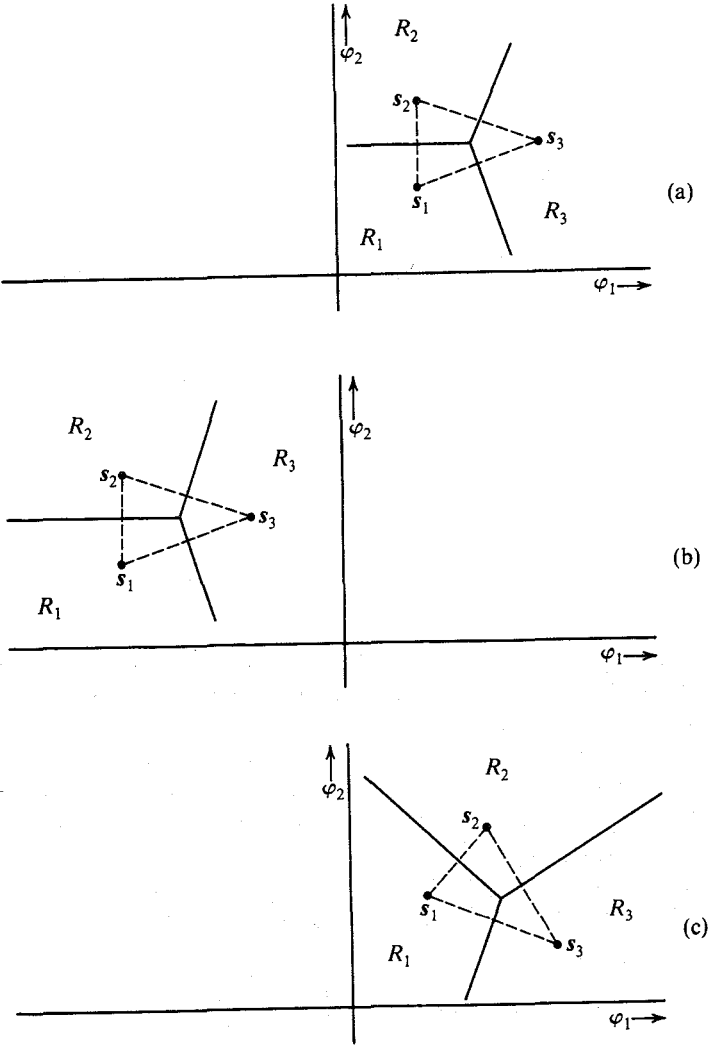
14.4 EQUIVALENT SIGNAL SETS

The computation of error probabilities is greatly facilitated by the translation and rotation of coordinate axes. We now show that such operations are permissible.

Consider a signal set with its corresponding decision regions, as shown in Fig. 14.15a. The conditional probability $P(C|m_1)$ is the probability that the noise vector drawn from s_1 lies within R_1 . Note that this probability does not depend on the origin of the coordinate system. We may translate the coordinate system any way we wish. This is equivalent to translating the signal set and the corresponding decision regions. Thus, the $P(C|m_i)$ for the translated system shown in Fig. 14.15b is identical to that of the system in Fig. 14.15a.

In the case of gaussian noise, we make another important observation. The rotation of the coordinate system does not affect the error probability because the noise-vector probability

Figure 14.15 Translation and rotation of coordinate axes.



density has spherical symmetry. To show this we shall consider Fig. 14.15c, where the signal set in Fig. 14.15a is shown translated and rotated. Note that a rotation of the coordinate system is equivalent to a rotation of the signal set in the opposite sense. Here for convenience we rotate the signal set instead of the coordinate system. It can be seen that the probability that the noise vector $\mathbf{n}$ drawn from s_1 lies in R_1 is the same in Fig. 14.15a and c, because this probability is given by the integral of the noise probability density $p_n(\mathbf{n})$ over the region R_1 . Because $p_n(\mathbf{n})$ has a spherical symmetry for gaussian noise, the probability will remain unaffected by a rotation of the region R_1 . Clearly, for additive gaussian channel noise, translation and rotation of the coordinate system (or translation and rotation of the signal set) do not affect the error probability. Note that when we rotate or translate a set of signals, the resulting set represents an entirely different set of signals. Yet the error probabilities of the two sets are identical. Such sets are called **equivalent sets**.

The following example demonstrates the utility of translation and rotation of a signal set in the computation of error probability.

EXAMPLE 14.6 A quaternary PSK (QPSK) signal set is shown in Fig. 14.16a,

$$\begin{aligned}s_1 &= -s_2 = \sqrt{E} \varphi_1 \\ s_3 &= -s_4 = \sqrt{E} \varphi_2\end{aligned}$$

Assuming all symbols to be equiprobable, determine P_{eM} for an AWGN channel with noise PSD $\mathcal{N}/2$.

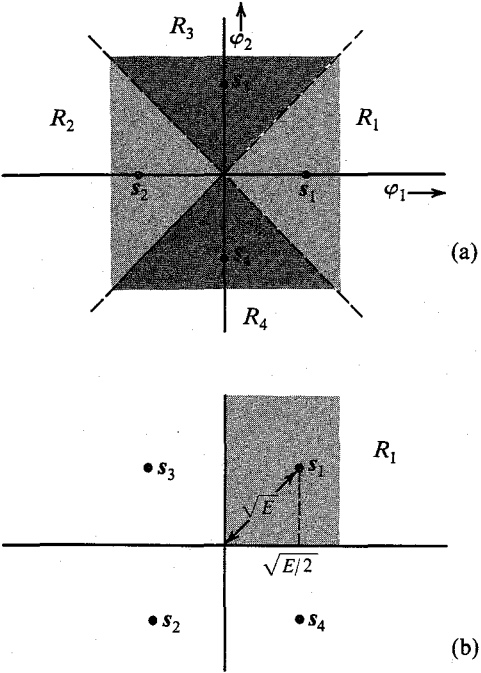


Figure 14.16 Analysis of QPSK.

This problem has already been solved in Example 14.4 for a general value of M . Here we shall solve it for $M = 4$ to demonstrate the power of the rotation of axes.

Because all the symbols are equiprobable, the decision region boundaries will be perpendicular bisectors of lines joining various signal points (Fig. 14.16a). Now

$$P(C|m_1) = P(\text{noise vector originating at } s_1 \text{ remains in } R_1) \quad (14.73)$$

This can be found by integrating the joint PDF of components n_1 and n_2 (originating at s_1) over the region R_1 . This double integral can be found by using suitable limits, as in Eq. (14.59). The problem is greatly simplified, however, if we rotate the signal set by 45° , as shown in Fig. 14.16b. The decision regions are rectangular, and if n_1 and n_2 are noise components along ϕ_1 and ϕ_2 , then Eq. (14.73) can be expressed as

$$\begin{aligned} P(C|m_1) &= P\left(n_1 > -\sqrt{\frac{E}{2}}, n_2 > -\sqrt{\frac{E}{2}}\right) \\ &= P\left(n_1 > -\sqrt{\frac{E}{2}}\right) P\left(n_2 > -\sqrt{\frac{E}{2}}\right) \\ &= \left[1 - Q\left(\sqrt{\frac{E}{2\sigma_n^2}}\right)\right]^2 \\ &= \left[1 - Q\left(\sqrt{\frac{E}{\mathcal{N}}}\right)\right]^2 \end{aligned} \quad (14.74a)$$

$$= \left[1 - Q\left(\sqrt{\frac{2E_b}{\mathcal{N}}}\right)\right]^2 \quad (14.74b)$$

Minimum-Energy Signal Set

As noted earlier, an infinite number of possible equivalent signal sets exist. Because signal energy depends on its distance from the origin, however, equivalent sets do not necessarily have the same average energy. Thus, among the infinite possible equivalent signal sets, the one in which the signals are closest to the origin has the minimum average signal energy (or transmitted power).

Let $m_1, m_2, \dots, m_M$ be M messages with waveforms $s_1(t), s_2(t), \dots, s_M(t)$ represented by points $s_1, s_2, \dots, s_M$ in the signal space. The mean energy of these signals is $\bar{E}$, given by

$$\bar{E} = \sum_{i=1}^M P(m_i) |s_i|^2$$

Translation of this signal set is equivalent to subtracting some vector $\mathbf{a}$ from each signal. Let this operation yield the minimum-mean-energy set. We now wish to find the vector $\mathbf{a}$ such that the new mean energy

$$\bar{E} = \sum_{i=1}^M P(m_i) |s_i - \mathbf{a}|^2 \quad (14.75)$$

is minimum. We can show that $\mathbf{a}$ must be the center of gravity of M points located at $s_1, s_2, \dots, s_M$ with masses $P(m_1), P(m_2), \dots, P(m_M)$, respectively:

$$\mathbf{a} = \sum_{i=1}^M P(m_i) s_i = \bar{s} \quad (14.76)$$

To prove this, suppose the mean energy is minimum for some translation $\mathbf{b}$. Then

$$\begin{aligned} \bar{E} &= \sum_{i=1}^M P(m_i) |s_i - \mathbf{b}|^2 \\ &= \sum_{i=1}^M P(m_i) |(s_i - \mathbf{a}) + (\mathbf{a} - \mathbf{b})|^2 \\ &= \sum_{i=1}^M P(m_i) |s_i - \mathbf{a}|^2 + 2(\mathbf{a} - \mathbf{b}) \cdot \sum_{i=1}^M P(m_i) (s_i - \mathbf{a}) + \sum_{i=1}^M P(m_i) |\mathbf{a} - \mathbf{b}|^2 \end{aligned}$$

Observe that the second term in the above expression vanishes because of the relationship in Eq. (14.76). Hence,

$$\bar{E} = \sum_{i=1}^M P(m_i) |s_i - \mathbf{a}|^2 + \sum_{i=1}^M P(m_i) |\mathbf{a} - \mathbf{b}|^2$$

This is minimum when $\mathbf{b} = \mathbf{a}$. Note that the rotation of the coordinates does not change the energy, and, hence, there is no need to rotate the signal set to minimize the energy.

EXAMPLE 14.7 For the binary orthogonal signal set shown in Fig. 14.17a, determine the minimum-energy equivalent signal set.

The minimum-energy set for this case is shown in Fig. 14.17b. The origin lies at the center of gravity of the signals. We have also rotated the signals for convenience. The distances k_1 and k_2 must be such that

$$k_1 + k_2 = d$$

and

$$k_1 P(m_1) = k_2 P(m_2)$$

Solution of these two equations yields

$$k_1 = P(m_2)d$$

and

$$k_2 = P(m_1)d$$

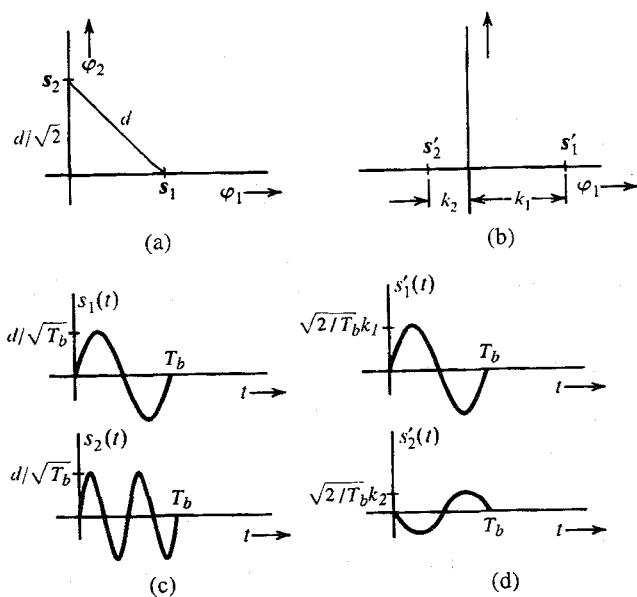


Figure 14.17 Equivalent signal sets.

Both signal sets (Fig. 14.17a and b) have the same error probability, but the latter has a smaller mean energy. If $\bar{E}$ and $\bar{E}'$ are the respective mean energies of the two sets, then

$$\bar{E} = P(m_1) \frac{d^2}{2} + P(m_2) \frac{d^2}{2} = \frac{d^2}{2}$$

and

$$\begin{aligned} \bar{E}' &= P(m_1)k_1^2 + P(m_2)k_2^2 \\ &= P(m_1)P^2(m_2)d^2 + P(m_2)P^2(m_1)d^2 \\ &= P(m_1)P(m_2)d^2 \end{aligned}$$

Note that the product $P(m_1)P(m_2)$ is maximum when $P(m_1) = P(m_2) = 1/2$, in which case

$$P(m_1)P(m_2) = \frac{1}{4}$$

and

$$\bar{E}' \leq \frac{d^2}{4}$$

Therefore,

$$\bar{E}' \leq \frac{\bar{E}}{2}$$

and for the case of equiprobable signals,

$$\bar{E}' = \frac{\bar{E}}{2}$$

In this case,

$$k_1 = k_2 = \frac{d}{2}$$

$$\bar{E} = \frac{d^2}{2} \quad \text{and} \quad \bar{E}' = \frac{d^2}{4}$$

The signals in Fig. 14.17b are called **antipodal signals** when $k_1 = k_2$. The error probability of the signal set in Fig. 14.17a (and Fig. 14.17b) is equal to that in Fig. 14.11a and can be found from Eq. (14.54a).

As a concrete example, let us choose the basis signals as sinusoids of frequency $\omega_o = 2\pi/T_M$:

$$\begin{aligned} \phi_1(t) &= \sqrt{\frac{2}{T_M}} \sin \omega_o t \\ &0 \leq t < T_M \end{aligned}$$

$$\phi_2(t) = \sqrt{\frac{2}{T_M}} \sin 2\omega_o t$$

Hence,

$$\begin{aligned} s_1(t) &= \frac{d}{\sqrt{2}} \phi_1(t) = \frac{d}{\sqrt{T_M}} \sin \omega_o t \\ &0 \leq t < T_M \end{aligned}$$

$$s_2(t) = \frac{d}{\sqrt{2}} \phi_2(t) = \frac{d}{\sqrt{T_M}} \sin 2\omega_o t$$

The signals $s_1(t)$ and $s_2(t)$ are shown in Fig. 14.17c, and the geometrical representation is shown in Fig. 14.17a. Both signals are located at a distance $d/\sqrt{2}$ from the origin, and the distance between the signals is d .

The minimum energy signals $s'_1(t)$ and $s'_2(t)$ for this set are given by

$$\begin{aligned} s'_1(t) &= \sqrt{\frac{2}{T_M}} P(m_2) d \sin \omega_o t \\ &0 \leq t < T_M \end{aligned}$$

$$s'_2(t) = \sqrt{\frac{2}{T_M}} P(m_1) d \sin \omega_o t$$

These signals are sketched in Fig. 14.17d.

EXAMPLE 14.8 Simplex Signal Set

A minimum-energy equivalent set of an equiprobable orthogonal set is called a **simplex**, or **transorthogonal, signal set**. Derive the simplex set corresponding to the orthogonal set in Eq. (14.66).

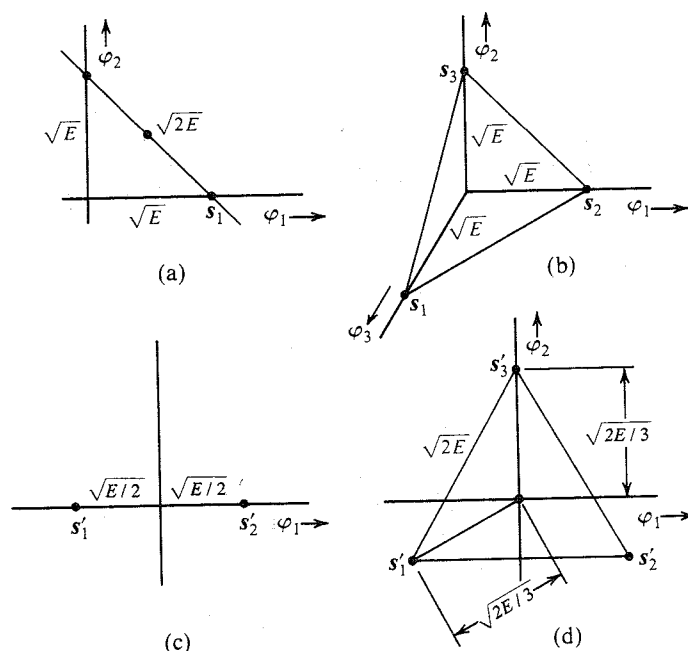


Figure 14.18 Simplex signals.

To obtain the minimum-energy set, the origin should be shifted to the center of gravity of the signal set. For the 2-dimensional case (Fig. 14.18a), the simplex set is shown in Fig. 14.18c, and for the 3-dimensional case (Fig. 14.18b), the simplex set is shown in Fig. 14.18d. Note that the dimensionality of the simplex signal set is less than that of the orthogonal set by 1. This is true in general for any value of M . It can be shown that the simplex signal set is the optimum (minimum error probability) for the case of equiprobable signals embedded in white gaussian noise when energy is constrained.^{6,7}

We can calculate the mean energy of the simplex set by noting that it is obtained by translating the orthogonal set by a vector $\mathbf{a}$, given by Eq. (14.76),

$$\mathbf{a} = \frac{1}{M} \sum_{i=1}^M \mathbf{s}_i$$

For orthogonal signals,

$$\mathbf{s}_i = \sqrt{E} \boldsymbol{\varphi}_i$$

Therefore,

$$\mathbf{a} = \frac{\sqrt{E}}{M} \sum_{i=1}^M \boldsymbol{\varphi}_i$$

where E is the energy of each signal in the orthogonal set and $\boldsymbol{\varphi}_i$ is the unit vector along the i th coordinate axis. The signals in the simplex set are given by

$$\begin{aligned} \mathbf{s}'_k &= \mathbf{s}_k - \mathbf{a} \\ &= \sqrt{E} \boldsymbol{\varphi}_k - \frac{\sqrt{E}}{M} \sum_{i=1}^M \boldsymbol{\varphi}_i \end{aligned} \quad (14.77)$$

The energy E' of signal $\mathbf{s}'_k$ is given by $|\mathbf{s}'_k|^2$,

$$E' = \mathbf{s}'_k \cdot \mathbf{s}'_k \quad (14.78)$$

Substituting Eq. (14.77) into Eq. (14.78) and observing that the set $\boldsymbol{\varphi}_i$ is orthonormal, we have

$$\begin{aligned} E' &= E - \frac{E}{M} \\ &= E \left(1 - \frac{1}{M} \right) \end{aligned} \quad (14.79)$$

Hence, for the same performance (error probability), the mean energy of the simplex signal set is $1 - 1/M$ times that of the orthogonal signal set. For $M \gg 1$, the difference is not significant. For this reason and because they are easier to generate, orthogonal rather than simplex signals are used in practice whenever M exceeds 4 or 5.

In the next chapter, we shall show that in the limit as $M \rightarrow \infty$, the orthogonal (as well as the simplex) signals attain the upper bound of performance predicted by Shannon's theorem.

14.5 NONWHITE (COLORED) CHANNEL NOISE

Thus far we have restricted our discussion exclusively to white gaussian channel noise. Our discussion can be extended with relative ease to nonwhite, or colored, gaussian channel noise.

If the noise PSD $S_n(\omega)$ is not white, we use a noise-whitening filter $H(\omega)$ at the input of the receiver, where

$$|H(\omega)| = \frac{1}{\sqrt{S_n(\omega)}}$$

Consider a signal set $\{s_i(t)\}$ and a channel noise $n(t)$ that is not white [$S_n(\omega)$ is not constant]. At the input of the receiver, we use a noise-whitening filter $H(\omega)$ that transforms the colored noise into white noise (Fig. 14.19). But it also alters the signal set $\{s_i(t)\}$ to $\{s'_i(t)\}$, where

$$s'_i(t) = s_i(t) * h(t)$$

or

$$S'_i(\omega) = S_i(\omega)H(\omega)$$

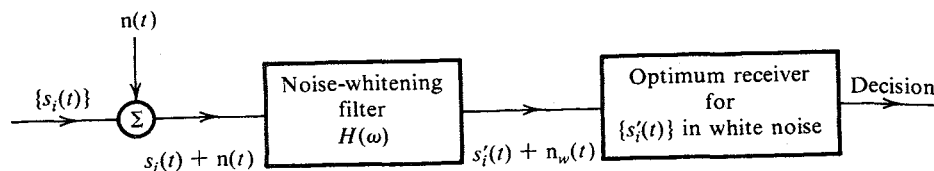


Figure 14.19 Optimum M -ary receiver for nonwhite channel noise.

We now have a new signal set $\{s_i'(t)\}$ mixed with white gaussian noise, for which the optimum receiver and the corresponding error probability can be determined by the method discussed earlier.

In general, a whitening filter may be unrealizable. In such a case we need to allow a sufficient delay in $H(\omega)$ to be able to closely realize the filter.

14.6 OTHER USEFUL PERFORMANCE CRITERIA

The optimum receiver uses the decision strategy that makes the best possible use of the observed data and any a priori information available. The strategy will also depend on the weights assigned to various types of errors. In this chapter we have thus far assumed that all errors have equal weight (or equal cost). This assumption may not be justified in some cases, and we may therefore have to alter the decision rule.

Generalized Bayes Receiver

If we are given a priori probabilities and the cost functions of various types of errors, the receiver that minimizes the average cost of decision is called the **Bayes receiver**, and the decision rule is **Bayes' decision rule**. Note that the receiver that has been discussed so far is the Bayes receiver under the condition that all errors have equal cost (equal weight). To generalize this rule, let

$$C_{kj} = \text{cost of deciding that } \hat{m} = m_k \text{ when } m_j \text{ was transmitted} \quad (14.80)$$

and, as usual,

$$P(m_i|q) = \text{conditional probability that } m_i \text{ was transmitted when } q \text{ is received}$$

If q is received, then the probability that m_j was transmitted is $P(m_j|q)$ for all $j = 1, 2, \dots, M$. Hence, the average cost of the decision $\hat{m} = m_k$ is β_k , given by

$$\begin{aligned} \beta_k &= C_{k1}P(m_1|q) + C_{k2}P(m_2|q) + \dots + C_{kM}P(m_M|q) \\ &= \sum_{j=1}^M C_{kj}P(m_j|q) \end{aligned} \quad (14.81)$$

Thus, if q is received, the optimum receiver decides that $\hat{m} = m_k$ if

$$\beta_k < \beta_i \quad \text{for all } i \neq k$$

or

$$\sum_{j=1}^M C_{kj}P(m_j|q) < \sum_{j=1}^M C_{ij}P(m_j|q) \quad \text{for all } i \neq k \quad (14.82)$$

Use of Bayes' mixed rule in Eq. (14.82) yields

$$\sum_{j=1}^M C_{kj}P(m_j)p_q(q|m_j) < \sum_{j=1}^M C_{ij}P(m_j)p_q(q|m_j) \quad \text{for all } i \neq k \quad (14.83)$$

Note that C_{kk} is the cost of setting $\hat{m} = m_k$ when m_k is transmitted. This cost is generally zero. If we assign equal weight to all other errors, then

$$C_{kj} = \begin{cases} 0 & k = j \\ 1 & k \neq j \end{cases}$$

and the decision rule in Eq. (14.83) reduces to the rule in Eq. (14.42), as expected. The generalized Bayes receiver for $M = 2$, assuming $C_{11} = C_{22} = 0$, sets $\hat{m} = m_1$ if

$$C_{12}P(m_2)p_q(q|m_2) < C_{21}P(m_1)p_q(q|m_1)$$

Otherwise, the receiver decides that $\hat{m} = m_2$.

Maximum-Likelihood Receiver

The strategy used in the Bayes receiver discussed in the preceding subsection is general, except that it can be implemented only when the a priori probabilities $P(m_1), P(m_2), \dots, P(m_M)$ are known. Frequently this information is not available. Under these conditions various possibilities exist, depending on the assumptions made. When, for example, there is no reason to expect any one signal to be more likely than any other, we may assign equal probabilities to all the messages:

$$P(m_1) = P(m_2) = \dots = P(m_M) = \frac{1}{M}$$

Bayes' rule [Eq. (14.42)] in this case becomes: set $\hat{m} = m_k$ if

$$p_q(q|m_k) > p_q(q|m_i) \quad \text{for all } i \neq k \quad (14.84)$$

Observe that $p_q(q|m_k)$ represents the probability of observing q when m_k is transmitted. Thus, the receiver chooses that signal which, when transmitted, will maximize the likelihood (probability) of observing the received q . Hence, this receiver is called the **maximum-likelihood receiver**. Note that the maximum-likelihood receiver is a Bayes receiver under the condition that the a priori message probabilities are equal. In terms of geometrical concepts, the maximum-likelihood receiver decides in favor of that signal which is closest to the received data q . The practical implementation of the maximum-likelihood receiver is the same as that of the Bayes receiver (Figs. 14.7 and 14.8) under the condition that all a priori probabilities are equal to $1/M$.

If the signal set is geometrically symmetrical, and if all a priori probabilities are equal (maximum-likelihood receiver), then the decision regions for various signals are congruent.

In this case, because of symmetry, the conditional probability of a correct decision is the same no matter which signal is transmitted, that is,

$$P(C|m_i) = \text{constant} \quad \text{for all } i$$

Because

$$P(C) = \sum_{i=1}^M P(m_i) P(C|m_i)$$

in this case

$$P(C) = P(C|m_i) \quad (14.85)$$

Thus, the error probability of the maximum-likelihood receiver is independent of the actual source statistics $P(m_i)$ for the case of symmetrical signal sets. It should, however, be realized that if the actual source statistics were known beforehand, one could design a better receiver using Bayes' decision rule.

It is apparent that if the source statistics are not known, the maximum-likelihood receiver proves very attractive for a symmetrical signal set. In such a receiver one can specify the error probability independently of the actual source statistics.

Minimax Receiver

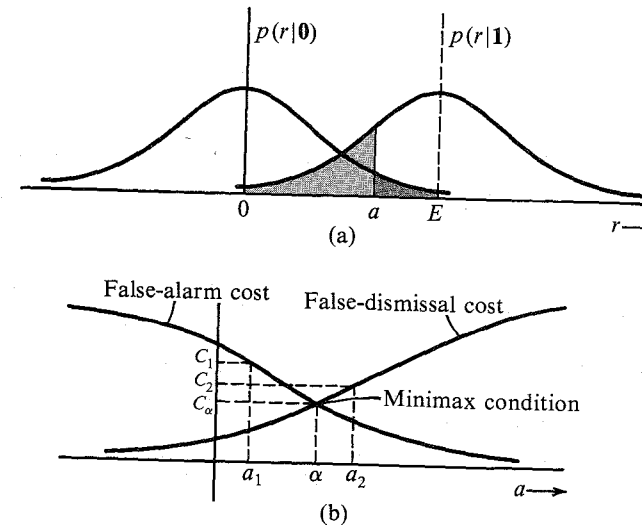
Designing a receiver with a certain decision rule completely specifies the conditional probabilities $P(C|m_i)$. The probability of error is given by

$$\begin{aligned} P_{eM} &= 1 - P(C) \\ &= 1 - \sum_{i=1}^M P(m_i) P(C|m_i) \end{aligned}$$

Thus, in general, for a given receiver (with some specified decision rule) the error probability depends on the source statistics $P(m_i)$. The error probability is the largest for some source statistics. The error probability in the worst possible case is $[P_{eM}]_{\max}$ and represents the upper bound on the error probability of the given receiver. This upper bound $[P_{eM}]_{\max}$ serves as an indication of the quality of the receiver. Each receiver (with a certain decision rule) will have a certain $[P_{eM}]_{\max}$. The receiver that has the smallest upper bound on the error probability, that is, the minimum $[P_{eM}]_{\max}$, is called the **minimax receiver**.

We shall illustrate the minimax concept for a binary receiver with on-off signaling. The conditional PDFs of the receiving-filter output sample r at $t = T_b$ are $p(r|1)$ and $p(r|0)$. These are the PDFs of r for the "on" and the "off" pulse (i.e., no pulse), respectively. Figure 14.20a shows these PDFs with a certain threshold a . If we receive $r \geq a$, we choose the hypothesis "signal present" (1), and the shaded area to the right of a is the probability of **false alarm** (deciding "signal present" when in fact the signal is not present). If $r < a$, we choose the hypothesis "signal absent" (0), and the shaded area to the left of a is the probability of **false dismissal** (deciding "signal absent" when in fact the signal is present). It is obvious that the larger is the threshold a , the larger is the false dismissal error and the smaller is the false alarm error (see Fig. 14.20b).

Figure 14.20 Explanation of minimax concept.



We shall now find the minimax condition for this receiver. For the minimax receiver, we consider all possible receivers (all possible values of a in this case) and find the maximum error probability (or cost) that occurs under the worst possible a priori probability distribution. Let us choose $a = a_1$, as shown in Fig. 14.20b. In this case the worst possible case occurs when $P(0) = 1$ and $P(1) = 0$, that is, when the signal $s(t)$ is always absent. The type of error in this case is false alarm. These errors have a cost C_1 . On the other hand, if we choose $a = a_2$, the worst possible case occurs when $P(0) = 0$ and $P(1) = 1$, that is, when the signal is always present, causing only the false-dismissal type of errors. These errors have a cost C_2 . It is evident that for the setting $a = \alpha$, the cost of false alarm and false dismissal are equal, namely, C_α . Hence, for all possible source statistics the cost is C_α . Because $C_\alpha < C_1$ and C_2 , this cost is the *minimum* of the maximum possible cost (because the worst cases are considered) that accrues for all values of a . Hence, $a = \alpha$ represents the minimax setting.

It follows from this discussion that the minimax receiver is rather conservative. It is designed under the pessimistic assumption that the worst possible source statistics exist. The maximum-likelihood receiver, on the other hand, is designed on the assumption that all messages are equally likely. It can, however, be shown that for a symmetrical signal set, the maximum-likelihood receiver is in fact the minimax receiver. This can be proved by observing that for a symmetrical set, the probability of error of a maximum-likelihood receiver (equal a priori probabilities) is independent of the source statistics [Eq. (14.85)]. Hence, for a symmetrical set, the error probability $P_{eM} = \alpha$ of a maximum-likelihood receiver is also equal to its $[P_{eM}]_{\max}$. We now show that no other receiver exists whose $[P_{eM}]_{\max}$ is less than the α of a maximum-likelihood receiver for a symmetrical signal set. This is seen from the fact that for equiprobable messages, the maximum-likelihood receiver is optimum by definition. All other receivers must have $P_{eM} > \alpha$ for equiprobable messages. Hence, $[P_{eM}]_{\max}$ for these receivers can never be less than α . This proves that the maximum-likelihood receiver is indeed the minimax receiver for a symmetrical signal set.

REFERENCES

1. H. J. Landau and H. O. Pollak, "Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty, III: The Dimensions of Space of Essentially Time- and Band-Limited Signals," *Bell Syst. Tech. J.*, vol. 41, pp. 1295–1336, July 1962.
2. J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*, Wiley, New York, 1965.
3. E. Arthurs and H. Dym, "On Optimum Detection of Digital Signals in the Presence of White Gaussian Noise—A Geometric Interpretation and a Study of Three Basic Data Transmission Systems," *IRE Trans. Commun. Syst.*, vol. CS-10, pp. 336–372, Dec. 1962.
4. B. P. Lathi, *An Introduction to Random Signals and Communication Theory*, International Textbook Co., Scranton, PA, 1968.
5. H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, vols. I, II, and III, Wiley, New York, 1968–1971.
6. H. J. Landau and D. Slepian, "On the Optimality of the Regular Simplex Code," *Bell Syst. Tech. J.*, vol. 45, pp. 1247–1272, Oct. 1966.
7. A. V. Balakrishnan, "Contribution to the Sphere-Packing Problem of Communication Theory," *J. Math. Anal. Appl.*, vol. 3, pp. 485–506, Dec. 1961.

PROBLEMS

14.1-1 Find at least three different sets of orthonormal basis signals for a 5-dimensional signal space.

14.1-2 The basis signals of a 3-dimensional signal space are given by $\varphi_1(t) = p(t)$, $\varphi_2(t) = p(t - T_o)$, and $\varphi_3(t) = p(t - 2T_o)$, where

$$p(t) = \frac{1}{\sqrt{T_o}}[u(t) - u(t - T_o)]$$

Sketch the waveforms of the signals represented by $(1, 1, 0)$, $(2, -1, 1)$, $(3, 2, -\frac{1}{2})$, and $(-\frac{1}{2}, -1, 1)$ in this space.

14.1-3 Repeat Prob. 14.1-2 if

$$\varphi_1(t) = \frac{1}{\sqrt{T_o}}$$

$$\varphi_2(t) = \sqrt{\frac{2}{T_o}} \sin \frac{2\pi}{T_o} t \quad 0 \leq t \leq T_o$$

$$\varphi_3(t) = \sqrt{\frac{2}{T_o}} \cos \frac{2\pi}{T_o} t$$

14.1-4 If $p(t)$ is as in Prob. 14.1-2 and

$$\varphi_k(t) = p[t - (k - 1)T_o] \quad k = 1, 2, 3, 4, 5$$

- (a) Sketch the signals represented by $(-1, 2, 3, 1, 4)$, $(2, 1, -4, -4, 2)$, $(3, -2, 3, 4, 1)$, and $(-2, 4, 2, 2, 0)$ in this space.
- (b) Find the energy of each signal.
- (c) Find the pairs of signals that are orthogonal.

14.2-1 For a certain stationary gaussian random process $x(t)$, it is given that $R_x(\tau) = e^{-|\tau|}$. Determine the joint PDF of RVs $x(t)$, $x(t + 1)$, and $x(t + 2)$.

14.3-1 A source emits M equiprobable messages, which are assigned signals $s_1, s_2, \dots, s_M$, as shown in Fig. P14.3-1. Determine the optimum receiver and the corresponding error probability P_{eM} for an AWGN channel as a function of E_b .

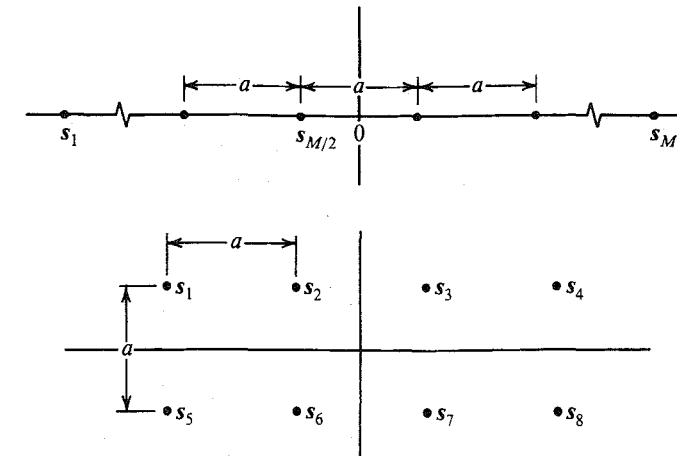


Figure P14.3-1 / Figure P14.3-2

14.3-2 A source emits eight equiprobable messages, which are assigned signals $s_1, s_2, \dots, s_8$, as shown in Fig. P14.3-2.

- (a) Find the optimum receiver for an AWGN channel.
- (b) Determine the decision regions and the error probability P_{eM} of the optimum receiver as a function of E_b .

14.3-3 Show that for MPSK, the optimum receiver of the form in Fig. 14.8a is equivalent to a phase comparator. Assume all messages equiprobable and an AWGN channel.

14.3-4 The vertices of an N -dimensional hypercube are a set of 2^N signals

$$s_k(t) = \frac{d}{2} \sum_{j=1}^N a_{kj} \varphi_j(t)$$

where $\{\varphi_1(t), \varphi_2(t), \dots, \varphi_N(t)\}$ is a set of N orthonormal signals, and a_{kj} is either 1 or -1 . Note that all the N signals are at a distance of $\sqrt{Nd}/2$ from the origin and form the vertices of the N -dimensional cube.

- (a) Sketch the signal configuration in the signal space for $N = 1, 2$, and 3 .
- (b) For each configuration in part (a), sketch one possible set of waveforms.
- (c) If all the 2^N symbols are equiprobable, find the optimum receiver and determine the error probability P_{eM} of the optimum receiver as a function of E_b assuming an AWGN channel.

14.3-5 A ternary signal configuration is shown in Fig. P14.3-5.

- (a) If $P(m_0) = 0.5$ and $P(m_1) = P(m_{-1}) = 0.25$, determine the optimum decision regions and P_{eM} of the optimum receiver as a function of $\bar{E}$. Assume an AWGN channel.
- (b) Find P_{eM} as a function of $\bar{E}/\mathcal{N}$.

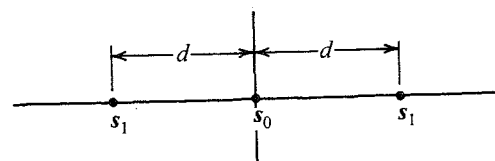


Figure P14.3-5

- 14.3-6** A 16-ary signal configuration is shown in Fig. P14.3-6. Write the expression (do not evaluate various integrals) for the P_{eM} of the optimum receiver, assuming all symbols to be equiprobable. Assume an AWGN channel.

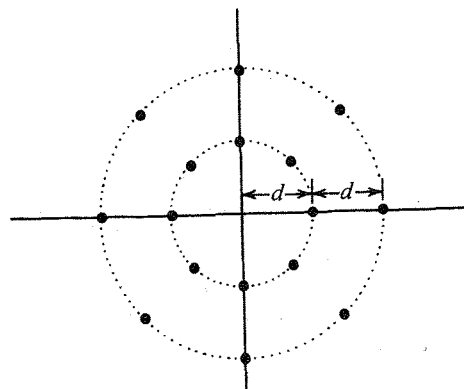


Figure P14.3-6

- 14.3-7** A five-signal configuration in a 2-dimensional space is shown in Fig. P14.3-7.

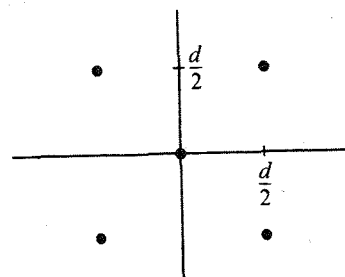


Figure P14.3-7

- Choose the appropriate $\phi_1(t)$ and $\phi_2(t)$ and sketch the waveforms of the five signals.
- In the signal space, sketch the optimum decision regions, assuming an AWGN channel.
- Determine the error probability P_{eM} as a function of $\bar{E}$ of the optimum receiver.

- 14.3-8** A 16-point QAM signal configuration is shown in Fig. P14.3-8. Assuming that all symbols are equiprobable, determine the error probability P_{eM} as a function of E_b of the optimum receiver for an AWGN channel.

Compare the performance of this scheme with that in Example 14.3.

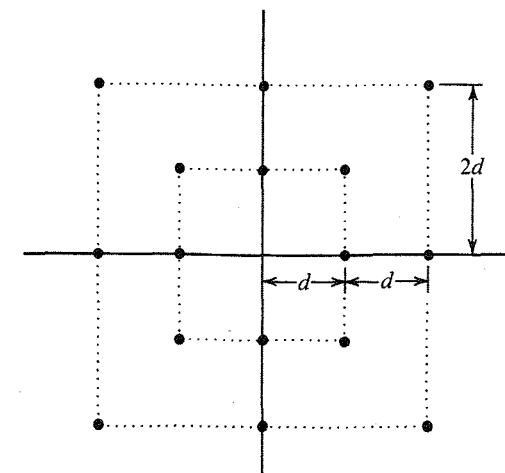


Figure P14.3-8

- 14.3-9** An orthogonal signal set is given by

$$s_k(t) = \sqrt{E} \phi_k(t) \quad k = 1, 2, \dots, N$$

A biorthogonal signal set is formed from the orthogonal set by augmenting it with the negative of each signal. Thus, we add to the orthogonal set another set

$$s_{-k}(t) = -\sqrt{E} \phi_k(t)$$

This gives $2N$ signals in an N -dimensional space. Assuming all signals to be equiprobable and an AWGN channel, obtain the error probability of the optimum receiver. How does the bandwidth of the biorthogonal set compare with that of the orthogonal set?

- 14.4-1** (a) What is the minimum-energy equivalent signal set of a binary on-off signal set?

(b) Using geometrical signal space concepts, explain why the binary on-off and the binary orthogonal sets have identical error probabilities and why the binary polar energy requirements are 3 dB lower than those of the on-off or the orthogonal set.

- 14.4-2** A source emits three equiprobable messages m_1 , m_2 , and m_3 , encoded by signals $s_1(t)$, $s_2(t)$, and $s_3(t)$, respectively, where

$$\left. \begin{aligned} s_1(t) &= 20\sqrt{2} \sin \frac{2\pi}{T_M} t \\ s_2(t) &= 10\sqrt{2} \cos \frac{2\pi}{T_M} t \\ s_3(t) &= -10\sqrt{2} \cos \frac{2\pi}{T_M} t \end{aligned} \right\} \quad T_M = \frac{1}{20}$$

Each of these signal durations is $0 \leq t \leq T_M$ and is zero outside this interval. The signals are transmitted over AWGN channels.

- Represent these signals in a signal space.
- Determine the decision regions.
- Obtain an equivalent minimum-energy signal set.
- Determine the optimum receiver.

- 14.4-3** A quaternary signaling scheme uses four waveforms,

$$\begin{aligned}
 s_1(t) &= (\sqrt{3} - 1) \varphi_1(t) \\
 s_2(t) &= -2\varphi_1(t) + (\sqrt{3} - 1) \varphi_2(t) \\
 s_3(t) &= -(\sqrt{3} + 1) \varphi_1(t) - 2\varphi_2(t) \\
 s_4(t) &= -(\sqrt{3} + 1) \varphi(t)
 \end{aligned}$$

where $\varphi_1(t)$ and $\varphi_2(t)$ are orthonormal basis signals. All the signals are equiprobable, and the channel noise is white gaussian with PSD $S_n(\omega) = 0.2$.

- Represent these signals in the signal space, and determine the optimum decision regions.
- Compute the error probability of the optimum receiver.
- Find the minimum-energy equivalent signal set.

14.4-4 A ternary signaling scheme ($M = 3$) uses the three waveforms $s_1(t)$, $s_2(t)$, and $s_3(t)$ shown in Fig. P14.4-4. The transmission rate is 500 symbols per second. All three messages are equiprobable, and the channel noise is white gaussian with PSD $S_n(\omega) = 10^{-5}$.

- Determine the error probability of the optimum receiver.
- Determine the minimum-energy signal set and sketch the waveforms.
- Compute the mean energies of the signal set in Fig. P14.4-4 and its minimum-energy equivalent set, found in part (b).

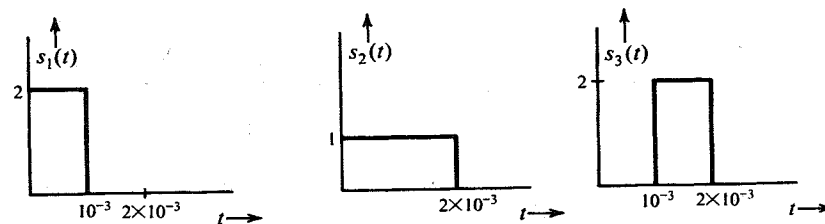


Figure P14.4-4

14.4-5 Repeat Prob. 14.4-4 if $P(m_1) = 0.25$, $P(m_2) = 0.5$, and $P(m_3) = 0.25$.

14.4-6 A binary signaling scheme uses the two waveforms $s_1(t)$ and $s_2(t)$ shown in Fig. P14.4-6. The signaling rate is 1000 pulses per second. Both signals are equally likely, and the channel noise is white gaussian with PSD $S_n(\omega) = 2.5 \times 10^{-6}$.

- Determine the minimum-energy equivalent signal set.
- Determine the error probability of the optimum receiver.
- Represent these signals as vectors using a suitable orthogonal signal space. *Hint:* Use Gram-Schmidt orthogonalization to determine the appropriate basis signals $\varphi_1(t)$ and $\varphi_2(t)$.

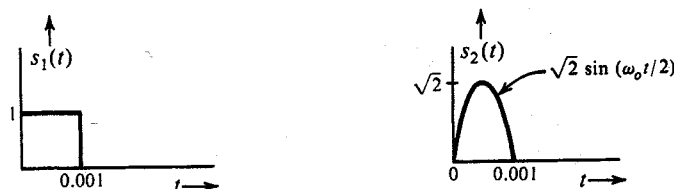


Figure P14.4-6

15 INTRODUCTION TO INFORMATION THEORY

In all the modes of communication discussed thus far, the communication is not error-free. We may be able to improve the accuracy in digital signals by reducing the error probability P_e . But it appears that as long as a channel noise exists, the communication cannot be error-free. For example, in all the digital systems discussed thus far, P_e varies as e^{-kE_b} asymptotically. By increasing E_b , the energy per bit, we can reduce P_e to any desired level. Now, the signal power is $S_i = E_b R_b$, where R_b is the bit rate. Hence, increasing E_b means either increasing the signal power S_i (for a given bit rate) or decreasing the bit rate R_b (for a given power), or both. Because of physical limitations, however, S_i cannot be increased beyond a certain limit. Hence, to reduce P_e further, we must reduce R_b , the rate of transmission of information digits. Thus, the price to be paid for reducing P_e is a reduction in the transmission rate. To make $P_e \rightarrow 0$, $R_b \rightarrow 0$. Hence, it appears that in the presence of channel noise it is impossible to achieve error-free communication. Thus thought communication engineers until the publication of Shannon's classical paper¹ in 1948. Shannon showed that for a given channel, as long as the rate of information digits per second to be transmitted is maintained within a certain limit (known as the channel capacity), it is possible to achieve error-free communication. That is, to attain $P_e \rightarrow 0$, it is not necessary to make $R_b \rightarrow 0$. Such a goal ($P_e \rightarrow 0$) can be attained by maintaining $R_b < C$, the channel capacity (per second). The gist of Shannon's paper is that the presence of random disturbance in a channel does not, by itself, set any limit on transmission accuracy. Instead, it sets a limit on the information rate for which arbitrarily small error probability ($P_e \rightarrow 0$) can be achieved.

We have been using the phrase "rate of information transmission" as if information could be measured. This is indeed so. We shall now discuss the information content of a message as understood by our "common sense" and also as it is understood in the "engineering sense." Surprisingly, both approaches yield the same measure of information in a message.

15.1 MEASURE OF INFORMATION

Common-Sense Measure of Information

Consider the following three hypothetical headlines in a morning paper:

1. Tomorrow the sun will rise in the east.
2. United States invades Cuba.
3. Cuba invades the United States.

The reader will hardly notice the first headline. He or she will be very, very interested in the second. But what really catches the reader's fancy is the third one. This item will attract much more attention than the previous two headlines. From the viewpoint of "common sense," the first headline conveys hardly any information, the second conveys a large amount of information, and the third conveys yet a larger amount of information. If we look at the probabilities of occurrence of these three events, we find that the probability of occurrence of the first event is unity (a certain event), that of the second is very low (an event of small but finite probability), and that of the third is practically zero (an almost impossible event). If an event of low probability occurs, it causes greater surprise and, hence, conveys more information than the occurrence of an event of larger probability. Thus, the information is connected with the element of surprise, which is a result of uncertainty, or unexpectedness. The more unexpected the event, the greater the surprise, and hence the more information. The probability of occurrence of an event is a measure of its unexpectedness and, hence, is related to the information content. Thus, from the point of view of common sense, the amount of information received from a message is directly related to the uncertainty or inversely related to the probability of its occurrence. If P is the probability of occurrence of a message and I is the information gained from the message, it is evident from the preceding discussion that when $P \rightarrow 1$, $I \rightarrow 0$ and when $P \rightarrow 0$, $I \rightarrow \infty$, and, in general a smaller P gives a larger I . This suggests the following model:

$$I \sim \log \frac{1}{P} \quad (15.1)$$

Engineering Measure of Information

We now show that from an engineering point of view, the information content of a message is identical to that obtained on an intuitive basis [Eq. (15.1)]. What do we mean by an engineering point of view? An engineer is responsible for the efficient transmission of messages. For this service the engineer will charge a customer an amount proportional to the information to be transmitted. But in reality the engineer will charge the customer in proportion to the time required to transmit the message. In short, from an engineering point of view, the information in a message is proportional to the (minimum) time required to transmit the message. We shall now show that this concept of information also leads to Eq. (15.1). This implies that a message with higher probability can be transmitted in a shorter time than that required for a message with lower probability. This fact may be verified by the example of the transmission of alphabetic symbols in the English language using Morse code. This code is made up of various combinations of two symbols (such as a mark and a space, or pulses of height A and $-A$ volts). Each letter is represented by a certain combination of these symbols, called the **code word**, which has a certain length. Obviously, for efficient transmission, shorter code words are assigned to the letters e , t , a , and o , which occur more frequently. The longer code words are assigned to letters x , k , q , and z , which occur less frequently. Each letter may be considered as a message. It is obvious that the letters that occur more frequently (with higher probability of occurrence) need a shorter time to transmit (shorter code words) as compared to those with

smaller probability of occurrence. We shall now show that on the average, the time required to transmit a symbol (or a message) with probability of occurrence P is indeed proportional to $\log (1/P)$.

For the sake of simplicity, let us begin with the case of binary messages m_1 and m_2 , which are equally likely to occur. We may use binary digits to encode these messages. Messages m_1 and m_2 may be represented by digits **0** and **1**, respectively. Clearly, we must have a minimum of one binary digit (which can assume two values) to represent each of the two equally likely messages. Next, consider the case of the four equiprobable messages m_1 , m_2 , m_3 , and m_4 . If these messages are encoded in binary form, we need a minimum of two binary digits per message. Each binary digit can assume two values. Hence, a combination of two binary digits can form the four code words **00**, **01**, **10**, **11**, which can be assigned to the four equiprobable messages m_1 , m_2 , m_3 , and m_4 , respectively. It is clear that each of these four messages takes twice as much transmission time as that required by each of the two equiprobable messages and, hence, contains twice as much information. Similarly, we can encode any one of eight equiprobable messages with a minimum of three binary digits. This is because three binary digits form eight distinct code words, which can be assigned to each of the eight messages. It can be seen that, in general, we need $\log_2 n$ binary digits to encode each of n equiprobable messages.* Because all the messages are equiprobable, P , the probability of any one message occurring, is $1/n$. Hence, each message (with probability P) needs $\log_2(1/P)$ binary digits for encoding. Thus, from the engineering viewpoint, the information I contained in a message with probability of occurrence P is proportional to $\log_2(1/P)$,

$$I = k \log_2 \frac{1}{P} \quad (15.2)$$

where k is a constant to be determined. Once again, we come to the conclusion (from the engineering viewpoint) that the information content of a message is proportional to the logarithm of the reciprocal of the probability of the message.

We shall now define the information conveyed by a message according to Eq. (15.2). The constant of proportionality is taken as unity for convenience, and the information is then in terms of binary units, abbreviated **bit** (binary unit),

$$I = \log_2 \frac{1}{P} \text{ bits} \quad (15.3)$$

According to this definition, the information I in a message can be interpreted as the minimum number of binary digits required to encode the message. This is given by $\log_2(1/P)$, where P is the probability of occurrence of the message. Although here we have shown this result for the special case of equiprobable messages, we shall show in the next section that this is true for nonequiprobable messages also.

Next, we shall consider the case of r -ary digits instead of binary digits for encoding. Each of the r -ary digits can assume r values (**0**, **1**, **2**, ..., $r-1$). Each of n messages (encoded by r -ary digits) can then be transmitted by a particular sequence of r -ary signals. Because each r -ary digit can assume r values, k r -ary digits can form a maximum of r^k distinct code words. Hence, to encode each of the n equiprobable messages, we need a minimum of $k = \log_r n$ r -ary

* Here we are assuming that the number n is such that $\log_2 n$ is an integer. Later on we shall observe that this restriction is not necessary.

digits.* But $n = 1/P$, where P is the probability of occurrence of each message. Obviously, we need a minimum of $\log_r(1/P)$ r -ary digits. The information I per message is therefore

$$I = \log_r \frac{1}{P} \quad r\text{-ary units} \quad (15.4)$$

From Eqs. (15.3) and (15.4) it is evident that

$$I = \log_2 \frac{1}{P} \text{ bits} = \log_r \frac{1}{P} \quad r\text{-ary units}$$

Hence,†

$$1 \text{ } r\text{-ary unit} = \log_2 r \text{ bits} \quad (15.5)$$

A Note on the Unit of Information: Although it is tempting to use the r -ary unit as a general unit of information, the binary unit bit ($r = 2$) is commonly used in the literature. There is, of course, no loss of generality in using $r = 2$. These units can always be converted into any other units by using Eq. (15.5). Henceforth, we shall always use the binary unit (bit) for information, unless otherwise stated. The bases of the logarithmic functions will be omitted, but will be understood to be 2.

Average Information per Message: Entropy of a Source

Consider a memoryless source m emitting messages $m_1, m_2, \dots, m_n$ with probabilities $P_1, P_2, \dots, P_n$, respectively ($P_1 + P_2 + \dots + P_n = 1$). A **memoryless source** implies that each message emitted is independent of the previous message(s). By the definition in Eq. (15.3) [or Eq. (15.4)], the information content of message m_i is I_i , given by

$$I_i = \log \frac{1}{P_i} \text{ bits} \quad (15.6)$$

The probability of occurrence of m_i is P_i . Hence, the mean, or average, information per message emitted by the source is given by $\sum_{i=1}^n P_i I_i$ bits. The average information per message of a source m is called its **entropy**, denoted by $H(m)$. Hence,

$$H(m) = \sum_{i=1}^n P_i I_i \text{ bits}$$

* Here again we are assuming that n is such that $\log_r n$ is an integer. As we shall see later, this restriction is not necessary.

† In general,

$$1 \text{ } r\text{-ary unit} = \log_s r \text{ } s\text{-ary units}$$

The 10-ary unit of information is called the **hartley** in honor of R. V. L. Hartley,² who was one of the pioneers (along with Nyquist³ and Carson) in the area of information transmission in the 1920s. The rigorous mathematical foundations of information theory, however, were established by E. Shannon¹ in 1948:

$$1 \text{ hartley} = \log_2 10 = 3.32 \text{ bits}$$

Sometimes the unit **nat** is used:

$$1 \text{ nat} = \log_2 e = 1.44 \text{ bits}$$

$$= \sum_{i=1}^n P_i \log \frac{1}{P_i} \text{ bits} \quad (15.7a)$$

$$= - \sum_{i=1}^n P_i \log P_i \text{ bits} \quad (15.7b)$$

The entropy of a source is a function of the message probabilities. It is interesting to find the message probability distribution that yields the maximum entropy. Because the entropy is a measure of uncertainty, the probability distribution that generates the maximum uncertainty will have the maximum entropy. On qualitative grounds, one expects entropy to be maximum when all the messages are equiprobable. We shall now show that this is indeed true.

Because $H(m)$ is a function of $P_1, P_2, \dots, P_n$, the maximum value of $H(m)$ is found from the equation $dH(m)/dP_i = 0$ for $i = 1, 2, \dots, n$, with the constraint that

$$P_n = 1 - (P_1 + P_2 + \dots + P_{n-1}) \quad (15.8)$$

Because

$$H(m) = - \sum_{i=1}^n P_i \log P_i \quad (15.9)$$

we need consider only the terms $-P_i \log P_i$ and $-P_n \log P_n$ [because P_n is a function of P_i , as seen from Eq. (15.8)]. Hence,

$$\begin{aligned} \frac{dH(m)}{dP_i} &= \frac{d}{dP_i} (-P_i \log P_i - P_n \log P_n) \\ &= -P_i \left(\frac{1}{P_i} \right) \log e - \log P_i + P_n \left(\frac{1}{P_n} \right) \log e + \log P_n \\ &= \log \frac{P_n}{P_i} \end{aligned}$$

which is zero if $P_i = P_n$. Because this is true for all i , we have

$$P_1 = P_2 = \dots = P_n = \frac{1}{n} \quad (15.10)$$

To show that Eq. (15.10) yields $[H(m)]_{\max}$ and not $[H(m)]_{\min}$, we note that when $P_1 = 1$ and $P_2 = P_3 = \dots = P_n = 0$, $H(m) = 0$, whereas the probabilities in Eq. (15.10) yield

$$\begin{aligned} H(m) &= - \sum_{i=1}^n \frac{1}{n} \log \frac{1}{n} \\ &= \log n \end{aligned} \quad (15.11)$$

The Intuitive (Common-Sense) and the Engineering Interpretation of Entropy:

Earlier we observed that both the intuitive and the engineering viewpoints lead to the same definition of the information associated with a message. The conceptual bases, however, are entirely different for the two points of view. Consequently, we have two physical interpretations

of information. According to the engineering point of view, the information content of any message is equal to the minimum number of digits required to encode the message, and, therefore, the entropy $H(m)$ is equal to the minimum number of digits per message required, on the average, for encoding. From the intuitive standpoint, on the other hand, information is thought of as being synonymous with the amount of surprise, or uncertainty, associated with the event (or message). A smaller probability of occurrence implies more uncertainty about the event. Uncertainty is, of course, associated with surprise. Hence intuitively, the information associated with a message is a measure of the uncertainty (unexpectedness) of the message. Therefore, $\log(1/P_i)$ is a measure of the uncertainty of the message m_i , and $\sum_{i=1}^n P_i \log(1/P_i)$ is the average uncertainty (per message) of the source that generates messages $m_1, m_2, \dots, m_n$ with probabilities $P_1, P_2, \dots, P_n$. Both these interpretations prove useful in the qualitative understanding of the mathematical definitions and results in information theory. Entropy may also be viewed as a function associated with a random variable m that assumes values $m_1, m_2, \dots, m_n$ with probabilities $P(m_1), P(m_2), \dots, P(m_n)$:

$$\begin{aligned} H(m) &= \sum_{i=1}^n P(m_i) \log \frac{1}{P(m_i)} \\ &= \sum_{i=1}^n P_i \log \frac{1}{P_i} \end{aligned}$$

Thus, we can associate an entropy with every discrete random variable.

If the source is not memoryless (i.e., a message emitted at any time is not independent of the previous messages emitted), then the source entropy will be less than $H(m)$ in Eq. (15.9). This is because the dependence of a message on previous messages reduces its uncertainty.

15.2 SOURCE ENCODING

The minimum number of binary digits required to encode a message was shown to be equal to the source entropy $\log(1/P)$ if all the messages of the source are equiprobable (each message probability is P). We shall now generalize this result to the case of nonequiprobable messages. We shall show that the average number of binary digits per message required for encoding is given by $H(m)$ (in bits) for an arbitrary probability distribution of messages.

Let a source m emit messages $m_1, m_2, \dots, m_n$ with probabilities $P_1, P_2, \dots, P_n$, respectively. Consider a sequence of N messages with $N \rightarrow \infty$. Let k_i be the number of times message m_i occurs in this sequence. Then according to the relative frequency interpretation (or law of large numbers),

$$\lim_{N \rightarrow \infty} \frac{k_i}{N} = P_i$$

Thus, the message m_i occurs NP_i times in a sequence of N messages (provided $N \rightarrow \infty$). Therefore, in a typical sequence of N messages, m_1 will occur NP_1 times, m_2 will occur NP_2 times, $\dots$, m_n will occur NP_n times. All other compositions are extremely unlikely to occur ($P \rightarrow 0$). Thus, any typical sequence (where $N \rightarrow \infty$) has the same proportion of the n messages, although in general the order will be different. We shall assume a zero-memory source, that is, the message is emitted from the source independently of the previous messages.

Consider now a typical sequence S_N of N messages from the source. Because the n messages (of probability $P_1, P_2, \dots, P_n$) occur $NP_1, NP_2, \dots, NP_n$ times, and because each message is independent, the probability of occurrence of a typical sequence S_N is given by

$$P(S_N) = (P_1)^{NP_1} (P_2)^{NP_2} \dots (P_n)^{NP_n} \quad (15.12)$$

Because all possible sequences of N messages from this source have the same composition, all the sequences (of N messages) are equiprobable, with probability $P(S_N)$. We can consider these long sequences as new messages (which are now equiprobable). To encode one such sequence we need L_N binary digits, where

$$L_N = \log \left[\frac{1}{P(S_N)} \right] \quad \text{binary digits} \quad (15.13)$$

Substituting Eq. (15.12) into Eq. (15.13), we obtain

$$\begin{aligned} L_N &= N \sum_{i=1}^n P_i \log \frac{1}{P_i} \\ &= NH(m) \quad \text{binary digits} \end{aligned}$$

Note that L_N is the length (number of binary digits) of the code word required to encode N messages in sequence. Hence, L , the average number of digits required per message, is L_N/N and is given by

$$L = \frac{L_N}{N} = H(m) \quad \text{binary digits} \quad (15.14)$$

This is the desired result, which states that it is possible to encode the messages emitted by a source using, on the average, $H(m)$ number of binary digits per message, where $H(m)$ is the entropy of the source (in bits). Although it does not prove that, on the average, this is the minimum number of digits required, one can show that $H(m)$ is indeed the minimum. It is not possible to find any uniquely decodable code whose average length is less than $H(m)$.^{4,5}

Compact Codes

The source encoding theorem says that to encode a source with entropy $H(m)$, we need, on the average, a minimum of $H(m)$ binary digits per message, or $H_r(m)$ r -ary digits per message, where $H_r(m)$ is the entropy in Eq. (15.9) computed with r as the base of the logarithm. The number of digits in the code word is the **length** of the code word. Thus, the average word length of an optimum code is $H(m)$. Unfortunately, to attain this length, in general, we have to encode a sequence of N messages ($N \rightarrow \infty$) at a time. If we wish to encode each message directly without using longer sequences, then, in general, the average length of the code word per message will be greater than $H(m)$. In practice, it is not desirable to use long sequences, as they cause transmission delay and add to equipment complexity. Hence, it is preferable to encode messages directly, even if the price has to be paid in terms of increased word length. In most cases, the price turns out to be small. The following is a procedure, given without proof, for finding the optimum source code, called the Huffman code. The proof that this code is optimum can be found elsewhere.⁴⁻⁶

We shall illustrate the procedure with an example using a binary code. We first arrange the messages in the order of descending probability, as shown in Table 15.1. Here we have

Table 15.1

Original Source		Reduced Sources			
Messages	Probabilities	S_1	S_2	S_3	S_4
m_1	0.30	0.30	0.30	0.43	0.57
m_2	0.25	0.25	0.27	0.30	0.43
m_3	0.15	0.18	0.25	0.27	
m_4	0.12	0.15	0.18		
m_5	0.10	0.12			
m_6	0.08				

six messages with probabilities 0.30, 0.25, 0.15, 0.12, 0.10, and 0.08, respectively. We now combine the last two messages into one message with probability $P_5 + P_6 = 0.18$. This leaves five messages with probabilities, 0.30, 0.25, 0.15, 0.12, and 0.18. These messages are now rearranged in the second column in the order of descending probability. We repeat this procedure by combining the last two messages in the second column and rearranging them in the order of descending probability. This is done until the number of messages is reduced to 2. These two (reduced) messages are now assigned 0 and 1 as their first digits in the code sequence. We now go back and assign the numbers 0 and 1 to the second digit for the two messages that were combined in the previous step. We keep regressing this way until the first column is reached. The code finally obtained (for the first column) can be shown to be optimum. The complete procedure is shown in Tables 15.1 and 15.2.

The optimum (Huffman) code obtained this way is also called a **compact code**. The average length of the compact code in the present case is given by

$$L = \sum_{i=1}^n P_i L_i = 0.3(2) + 0.25(2) + 0.15(3) + 0.12(3) + 0.1(3) + 0.08(3) = 2.45 \text{ binary digits}$$

The entropy $H(m)$ of the source is given by

$$H(m) = \sum_{i=1}^n P_i \log_2 \frac{1}{P_i} = 2.418 \text{ bits}$$

Table 15.2

Original Source			Reduced Sources			
Messages	Probabilities	Code	S_1	S_2	S_3	S_4
m_1	0.30	00	0.30	00	0.43	1
m_2	0.25	10	0.25	01	0.30	00
m_3	0.15	010	0.18	11	0.27	01
m_4	0.12	011	0.15	10		
m_5	0.10	110	0.12	011		
m_6	0.08	111				

Hence, the minimum possible length (attained by an infinitely long sequence of messages) is equal to 2.418 binary digits. Using direct coding (the Huffman code), it is possible to attain an average length of 2.45 bits in the example given. This is a close approximation of the optimum performance attainable. Thus, little is gained by complex coding using long sequences of messages in this case.

The merit of any code is measured by its average length in comparison to $H(m)$ (the average minimum length). We define the **code efficiency** η as

$$\eta = \frac{H(m)}{L}$$

where L is the average length of the code. In our present example,

$$\eta = \frac{2.418}{2.45} = 0.976$$

The **redundancy** γ is defined as

$$\gamma = 1 - \eta = 0.024$$

Huffman code is uniquely decodable. If we receive a sequence of Huffman-coded messages, it can be decoded only one way, that is, without ambiguity. For instance, if the source in this example were to emit the following message sequence: $m_1 m_5 m_2 m_1 m_4 m_3 m_6 \dots$, it would be encoded as **001101000011010111**. . . . The reader may verify that this message sequence can be decoded only one way, viz, $m_1 m_5 m_2 m_1 m_4 m_3 m_6 \dots$, even if there is no demarcation between individual messages.

A similar procedure is used to find a compact r -ary code. In this case we arrange the messages in descending order of probability, combine the last r messages into one message, and rearrange the new set (reduced set) in the order of descending probability. We repeat the procedure until the final set reduces to r messages. Each of these messages is now assigned one of the r numbers 0, 1, 2, . . . , $r - 1$. We now regress in exactly the same way as in the binary case until each of the original messages is assigned a code.

For an r -ary code, we will have exactly r messages left in the last reduced set if, and only if, the total number of original messages is equal to $r + k(r - 1)$, where k is an integer. This is because each reduction decreases the number of messages by $r - 1$. Hence, if there is a total of k reductions, the total number of original messages must be $r + k(r - 1)$. In case the original messages do not satisfy this condition, we must add some dummy messages with zero probability of occurrence until this condition is fulfilled. As an example, if $r = 4$ and the number of messages n is 6, then we must add one dummy message with zero probability of occurrence to make the total number of messages 7, that is, $[4 + 1(4 - 1)]$, and proceed as usual. The procedure is illustrated in Example 15.1.

EXAMPLE 15.1 A zero-memory source emits six messages with probabilities 0.3, 0.25, 0.15, 0.12, 0.1, and 0.08. Find the 4-ary (quaternary) Huffman code. Determine its average word length, the efficiency, and the redundancy.

In this case, we need to add one dummy message to satisfy the required condition of $r + k(r - 1)$ messages and proceed as usual. The Huffman code is found in Table 15.3. The length L of this code is

$$L = 0.3(1) + 0.25(1) + 0.15(1) + 0.12(2) + 0.1(2) + 0.08(2) + 0(2)$$
$$= 1.3 \text{ 4-ary digits}$$

Also,

$$H_4(m) = - \sum_{i=1}^6 P_i \log_4 P_i$$
$$= 1.209 \text{ 4-ary units}$$

The code efficiency η is given by

$$\eta = \frac{1.209}{1.3} = 0.93$$

The redundancy $\gamma = 1 - \eta = 0.07$.

Table 15.3

Original Source			
Messages	Probabilities	Code	Reduced Sources
m_1	0.30	0	0.30 0
m_2	0.25	2	0.30 1
m_3	0.15	3	0.25 2
m_4	0.12	10	0.15 3
m_5	0.10	11	
m_6	0.08	12	
m_7	0.00	13	

To achieve code efficiency $\eta \rightarrow 1$, we need $N \rightarrow \infty$. The Huffman code uses $N = 1$, but its efficiency is, in general, less than 1. A compromise exists between these two extremes of $N = 1$ and $N = \infty$. We can use $N = 2$ or 3. In most cases, the use of $N = 2$ or 3 can yield an efficiency close to 1, as the following example shows.

EXAMPLE 15.2 A zero-memory source emits messages m_1 and m_2 with probabilities 0.8 and 0.2, respectively. Find the optimum (Huffman) binary code for this source as well as for its **second-** and **third-order extensions** (that is, for $N = 2$ and 3). Determine the code efficiencies in each case.

The Huffman code for the source is simply 0 and 1, giving $L = 1$, and

$$H(m) = -(0.8 \log 0.8 + 0.2 \log 0.2)$$
$$= 0.72 \text{ bit}$$

Hence,

$$\eta = 0.72$$

For the second-order extension of the source ($N = 2$), there are four possible composite messages, m_1m_1, m_1m_2, m_2m_1 , and m_2m_2 , with probabilities 0.64, 0.16, 0.16, and 0.04, respectively. The Huffman code is obtained in Table 15.4.

Table 15.4

Original Source			
Messages	Probabilities	Code	Reduced Source
m_1m_1	0.64	0	0.64 0
m_1m_2	0.16	11	0.20 10
m_2m_1	0.16	100	0.16 11
m_2m_2	0.04	101	

In this case the average word length L' is

$$L' = 0.64(1) + 0.16(2) + 0.16(3) + 0.04(3)$$
$$= 1.56$$

This is the word length for two messages of the original source. Hence L , the word length per message, is

$$L = \frac{L'}{2} = 0.78$$

and

$$\eta = \frac{0.72}{0.78} = 0.923$$

If we proceed with $N = 3$ (the third-order extension of the source), we have eight possible messages, and following the Huffman procedure, we find the code as shown in Table 15.5.

Table 15.5

Messages	Probabilities	Code
$m_1m_1m_1$	0.512	0
$m_1m_1m_2$	0.128	100
$m_1m_2m_1$	0.128	101
$m_2m_1m_1$	0.128	110
$m_1m_2m_2$	0.032	11100
$m_2m_1m_2$	0.032	11101
$m_2m_2m_1$	0.032	11110
$m_2m_2m_2$	0.008	11111

The word length L'' is

$$L'' = (0.512)1 + (0.128 + 0.128 + 0.128)3$$
$$+ (0.032 + 0.032 + 0.032)5 + (0.008)5$$
$$= 2.184$$

Then,

$$L = \frac{L''}{3} = 0.728$$

and

$$\eta = \frac{0.72}{0.728} = 0.989$$

15.3 ERROR-FREE COMMUNICATION OVER A NOISY CHANNEL

As seen in the previous section, messages of a source with entropy $H(m)$ can be encoded by using an average of $H(m)$ digits per message. This encoding has zero redundancy. Hence, if we transmit these coded messages over a noisy channel, some of the information will be received erroneously. There is absolutely no possibility of error-free communication over a noisy channel when messages are encoded with zero redundancy. The use of redundancy, in general, helps combat noise. This can be seen from a simple example of a **single parity-check code**, in which an extra binary digit is added to each code word to ensure that the total number of 1's in the resulting code word is always even (or odd). If a single error occurs in the received code word, the parity is violated, and the receiver requests retransmission. This is a rather simple example to demonstrate the utility of redundancy. More complex coding procedures, which can correct up to n digits, will be discussed in the next chapter.

The addition of an extra digit increases the average word length to $H(m) + 1$, giving $\eta = H(m)/[H(m) + 1]$, and the redundancy is $1 - \eta = 1/[H(m) + 1]$. Thus, the addition of an extra check digit increases redundancy, but it also helps combat noise. Immunity against channel noise can be increased by increasing the redundancy. Shannon has shown that it is possible to achieve error-free communication by adding sufficient redundancy. For example, if we have a **binary symmetric channel (BSC)** with an error probability P_e , then for error-free communication over this channel, messages from a source with entropy $H(m)$ must be encoded by binary codes with a word length of at least $H(m)/C_s$, where (see Sec. 15.4)

$$C_s = 1 - \left[P_e \log \frac{1}{P_e} + (1 - P_e) \log \frac{1}{1 - P_e} \right] \tag{15.15}$$

The parameter C_s ($C_s < 1$) is called the **channel capacity**.

The efficiency of these codes is never greater than C_s . If a certain binary channel has $C_s = 0.4$, a code that can achieve error-free communication must have at least $2.5 H(m)$ binary digits per message, which is two-and-one-half times as many digits as required for coding without redundancy. This means there are $1.5 H(m)$ redundant digits per message. Thus, on the average, for every 2.5 digits transmitted, one digit is the information digit and 1.5 digits are redundant, or check, digits, giving a redundancy of $1 - C_s = 0.6$.

As discussed in the beginning of this chapter, P_e , the error probability of binary signaling, varies as e^{-kE_b} and, hence, to make $P_e \rightarrow 0$, either $S_i \rightarrow \infty$ or $R_b \rightarrow 0$. Because S_i must be finite, $P_e \rightarrow 0$ only if $R_b \rightarrow 0$. But Shannon's results state that it is really not necessary to let $R_b \rightarrow 0$ for error-free communication. All that is required is to hold R_b below C , the channel capacity per second ($C = 2BC_s$). Where is the discrepancy? To answer this question let us investigate carefully the role of redundancy in error-free communication. Although the discussion here is with reference to a binary scheme, it is quite general and can be extended to the M -ary case.

Consider a simple method of reducing P_e by repeating a given digit an odd number of times. For example, we can transmit **0** and **1** as **000** and **111**. The receiver uses the majority rule to make the decision; that is, if at least two out of three digits are **1**, the decision is **1**, and if at least two out of three digits are **0**, the decision is **0**. Thus, even if one out of three digits is in error, the information is received error-free. This scheme will fail if two out of three digits are in error. In order to correct two errors, we need five repetitions. In any case, repetitions cause redundancy but improve P_e (see Example 10.8). It will be instructive to understand this situation from a graphic point of view. Consider the case of three repetitions. We can show all eight possible sequences of three binary digits graphically as the vertices of a cube (Fig. 15.1). It is convenient to map binary sequences as shown in Fig. 15.1 and to talk in terms of what is called the **Hamming distance** between binary sequences. If two binary sequences of the same length differ in j places (j digits), then the Hamming distance between the sequences is considered to be j . Thus, the Hamming distance between **000** and **010** (or **001** and **101**) is 1, and between **000** and **111** it is 3. In the case of three repetitions, we transmit binary **1** by **111** and binary **0** by **000**. The Hamming distance between these two sequences is 3. Observe that of the eight possible vertices, we are occupying only two (**000** and **111**) for transmitted messages. At the receiver, however, because of channel noise, we are liable to receive any one of the eight sequences. The majority decision rule can be interpreted as a rule that decides in favor of that message (**000** or **111**), which is at the closest Hamming distance from the received sequence. Sequences **000**, **001**, **010**, and **100** are within 1 unit of Hamming distance from **000** but are at least 2 units away from **111**. Hence, when we receive any one of these four sequences, our decision is binary **0**. Similarly, when any one of the sequences **110**, **111**, **011**, or **101** is received, the decision is binary **1**.

We can now see why the error probability is reduced in this scheme. Of the eight possible vertices, we have used only two, which are separated by 3 Hamming units. If we draw a Hamming sphere of unit radius around each of these two vertices (**000** and **111**), the two Hamming spheres* are nonoverlapping. The channel noise can cause a distance between the received sequence and the transmitted sequence, and as long as this distance is equal to or less than 1 unit, we can still detect the message without error. In a similar way, the case of five repetitions can be represented by a hypercube of five dimensions. The transmitted sequences

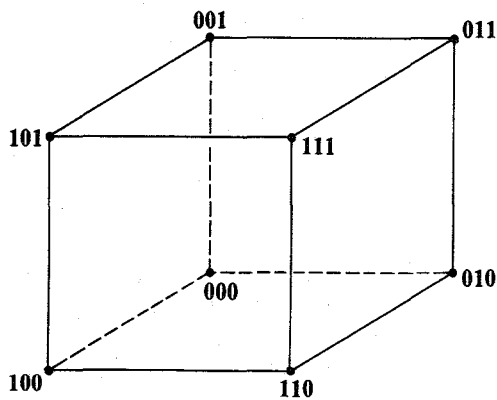


Figure 15.1 3-dimensional cube in Hamming space.

* Note that the Hamming sphere is not a true geometrical hypersphere because the Hamming distance is not a true geometrical distance (e.g., sequences **001**, **010**, and **100** lie on a Hamming sphere centered at **111** and having a radius 2).

00000 and 11111 occupy two vertices separated by five units, and the Hamming spheres of 2-unit radius drawn around each of these two vertices would be nonoverlapping. In this case, even if channel noise causes two errors, we can still detect the message correctly. Hence, the reason for the reduction in error probability is that we have not used all the available vertices for messages. Had we occupied all the available vertices for messages (as is the case without redundancy, or repetition), then if channel noise caused an error (even one), the received sequence would occupy a vertex assigned to another transmitted sequence, and we are certain to make a wrong decision. Precisely because we have left the neighboring vertices of the transmitted sequence unoccupied, are we able to detect the sequence correctly, despite channel errors (within a certain limit). The smaller the fraction of vertices used, the smaller the error probability. It should also be remembered that redundancy (or repetition) is what makes it possible to have unoccupied vertices.

If we continue to increase n , the number of repetitions, we will reduce P_e , but we will also reduce R_b by the factor n . But no matter how large we make n , the error probability never becomes zero. The trouble with this scheme is that it is inefficient because we are adding redundant (or check) digits to each information digit. To give an analogy, redundant (or check) digits are like guards protecting the information digit. To hire guards for each information digit is somewhat similar to a case of families living on a certain street hit by several burglaries. Each family panics and hires a guard. This is obviously expensive and inefficient. A better solution would be for all the families on the street to hire one guard and share the expenses. One guard can check on all the houses on the street to hire one guard and share the expenses. If the street is too long, it might be necessary to hire more than one guard. But it is certainly not necessary to hire one guard per house. In using repetitions, we had a similar situation. Redundant (or repeated) digits were used to check on only one transmitted digit. Using the clue from the preceding analogy, it might be more efficient if we used redundant digits not to check (guard) any one individual transmitted digit but, rather, a block of digits. Herein lies the key to our problem. Let us consider a group of information digits over a certain time interval of T seconds, and let us add some redundant digits to check on all these digits.

Suppose we need to transmit α binary information digits per second. Then over a period of T seconds, we have a block of αT binary information digits. If to this block of information digits we add $(\beta - \alpha)T$ check digits ($\beta - \alpha$ check digits, or redundant digits, per second), then we need to transmit βT ($\beta > \alpha$) digits for every αT information digits. Therefore over a T -second interval, we have

$$\begin{aligned} \alpha T & \text{ information digits} \\ \beta T & \text{ total transmitted digits } (\beta > \alpha) \\ (\beta - \alpha)T & \text{ check digits} \end{aligned}$$

Thus, instead of transmitting one binary digit every $1/\alpha$ seconds, we let αT digits accumulate over T seconds. Now consider this as a message to be transmitted. There are a total of $2^{\alpha T}$ such supermessages. Thus, every T seconds we need to transmit one of the $2^{\alpha T}$ possible supermessages. These supermessages are transmitted by a sequence of βT binary digits. There are in all $2^{\beta T}$ possible sequences of βT binary digits, and they can be represented as vertices of a βT -dimensional hypercube. Because we have only $2^{\alpha T}$ messages to be transmitted whereas $2^{\beta T}$ vertices are available, we occupy only a $2^{-(\beta-\alpha)T}$ fraction of the vertices of the βT -

dimensional hypercube. Observe that we have reduced the transmission rate by a factor of α/β . This rate-reduction factor α/β is independent of T . The fraction of the vertices occupied (occupancy factor) by transmitted messages is $2^{-(\beta-\alpha)T}$ and can be made as small as possible simply by increasing T . In the limit as $T \rightarrow \infty$, the occupancy factor approaches 0. This will make the error probability go to 0, and we have the possibility of error-free communication.

One important question, however, remains to be answered. What must be the rate reduction ratio α/β for this dream to come true? To answer this question, we observe that increasing T increases the length of the transmitted sequence (βT digits). If P_e is the digit error probability, then it can be seen from the relative frequency definition (or the law of large numbers) that as $T \rightarrow \infty$, the total number of digits in error in a sequence of βT digits ($\beta T \rightarrow \infty$) is exactly $\beta T P_e$. Hence, the received sequences will be at a Hamming distance of $\beta T P_e$ from the transmitted sequences. Therefore, for error-free communication, we must leave all the vertices unoccupied within spheres of radius $\beta T P_e$ drawn around each of the $2^{\alpha T}$ occupied vertices. In short, we must be able to pack $2^{\alpha T}$ nonoverlapping spheres, each of radius $\beta T P_e$, into the Hamming space of βT dimensions. This means that for a given β , α cannot be increased beyond some limit without causing overlap in the spheres and the consequent failure of the scheme. Shannon's theorem states that for this scheme to work, α/β must be less than the constant (channel capacity) C_s , which is a function of the channel noise and the signal power:

$$\frac{\alpha}{\beta} < C_s \quad (15.16)$$

It must be remembered that such perfect, error-free communication is not practical. In this system we accumulate the information digits for T seconds before encoding them, and because $T \rightarrow \infty$, for error-free communication we must wait until eternity before we start encoding. Hence, there will be an infinite delay at the transmitter and an additional delay of the same amount at the receiver. Second, the equipment needed for the storage, encoding, and decoding sequence of infinite digits would be monstrous. Needless to say that in practice the dream of error-free communication cannot be achieved. Then what is the use of Shannon's result? For one thing, it indicates the upper limit on the rate of error-free communication that can be achieved on a channel. This in itself is monumental. Second, it indicates the way to reduce the error probability with only a small reduction in the rate of transmission of information digits. We can therefore seek a compromise between error-free communication with infinite delay and virtually error-free communication with a finite delay.

15.4 CHANNEL CAPACITY OF A DISCRETE MEMORYLESS CHANNEL

In this section, discrete memoryless channels will be considered. Let a source emit symbols $x_1, x_2, \dots, x_r$. The receiver receives symbols $y_1, y_2, \dots, y_s$. The set of symbols $\{y_k\}$ may or may not be identical to the set $\{x_k\}$, depending on the nature of the receiver. If we use the types of receivers discussed in Chapter 14, the set of received symbols will be the same as the set transmitted. This is because the optimum receiver, upon receiving a signal, decides which of the r symbols $x_1, x_2, \dots, x_r$ has been transmitted. Here we shall be more general and shall not constrain the set $\{y_k\}$ to be identical to the set $\{x_k\}$.

If the channel is noiseless, then the reception of some symbol y_j uniquely determines the message transmitted. Because of noise, however, there is a certain amount of uncertainty regarding the transmitted symbol when y_j is received. If $P(x_i|y_j)$ represents the conditional probabilities that x_i was transmitted when y_j is received, then there is an uncertainty of $\log [1/P(x_i|y_j)]$ about x_i when y_j is received. When this uncertainty is averaged over all x_i and y_j , we obtain $H(x|y)$, which is the average uncertainty about a transmitted symbol when a symbol is received. Thus,

$$H(x|y) = \sum_i \sum_j P(x_i, y_j) \log \frac{1}{P(x_i|y_j)} \quad \text{bits per symbol} \quad (15.17)$$

If the channel were noiseless, the uncertainty would be zero.* Obviously, this uncertainty, $H(x|y)$, is caused by channel noise. Hence, it is the average loss of information about a transmitted symbol when a symbol is received. We call $H(x|y)$ the **equivocation** of x with respect to y .

Note that $P(y_j|x_i)$ represents the probability that y_j is received when x_i is transmitted. This is a characteristic of the channel and the receiver. Thus, a given channel (with its receiver) is specified by the **channel matrix**:

		Outputs				
		y_1	y_2	$\dots$	y_s	
Inputs	x_1	(	$P(y_1 x_1)$	$P(y_2 x_1)$	$\dots$	$P(y_s x_1)$
	x_2		$P(y_1 x_2)$	$P(y_2 x_2)$	$\dots$	$P(y_s x_2)$
	$\dots$		$\dots$	$\dots$	$\dots$	$\dots$
	x_r		$P(y_1 x_r)$	$P(y_2 x_r)$	$\dots$	$P(y_s x_r)$
		)				

We can obtain the reverse conditional probabilities $P(x_i|y_j)$ using Bayes' rule:

$$P(x_i|y_j) = \frac{P(y_j|x_i)P(x_i)}{P(y_j)} \quad (15.18a)$$

$$= \frac{P(y_j|x_i)P(x_i)}{\sum_i P(x_i, y_j)} \quad (15.18b)$$

$$= \frac{P(y_j|x_i)P(x_i)}{\sum_i P(x_i)P(y_j|x_i)} \quad (15.18c)$$

Thus, if the input symbol probabilities $P(x_i)$ and the channel matrix are known, the reverse conditional probabilities can be computed from Eqs. (15.18). The reverse conditional probability $P(x_i|y_j)$ is the probability that x_i was transmitted when y_j is received.

If the channel were noise-free, the average amount of information received would be $H(x)$ bits (entropy of the source) per received symbol. Note that $H(x)$ is the average information transmitted over the channel per symbol. Because of channel noise, we lose an average of

* This can be verified from the fact that for a noiseless channel all the probabilities in Eq. (15.17) are either 0 or 1. If $P(x_i|y_j) = 1$, then $\log [1/P(x_i|y_j)] = 0$ and if $P(x_i|y_j) = 0$, then $P(x_i, y_j) = P(y_j)P(x_i|y_j) = 0$. This shows that $H(x|y) = 0$.

$H(x|y)$ bits of information per symbol. Therefore, in this transaction the amount of information the receiver receives is, on the average, $I(x; y)$ bits per received symbol, where

$$I(x; y) = H(x) - H(x|y) \quad \text{bits per symbol} \quad (15.19)$$

$I(x; y)$ is called the **mutual information** of x and y . Because

$$H(x) = \sum_i P(x_i) \log \frac{1}{P(x_i)} \quad \text{bits}$$

we have

$$I(x; y) = \sum_i P(x_i) \log \frac{1}{P(x_i)} - \sum_i \sum_j P(x_i, y_j) \log \frac{1}{P(x_i|y_j)}$$

Also because

$$\sum_j P(x_i, y_j) = P(x_i)$$

We have

$$\begin{aligned} I(x; y) &= \sum_i \sum_j P(x_i, y_j) \log \frac{1}{P(x_i)} - \sum_i \sum_j P(x_i, y_j) \log \frac{1}{P(x_i|y_j)} \\ &= \sum_i \sum_j P(x_i, y_j) \log \frac{P(x_i|y_j)}{P(x_i)} \end{aligned} \quad (15.20a)$$

$$= \sum_i \sum_j P(x_i, y_j) \log \frac{P(x_i, y_j)}{P(x_i)P(y_j)} \quad (15.20b)$$

Alternatively, using Bayes' rule in Eq. (15.20a), $I(x; y)$ may be expressed as

$$I(x; y) = \sum_i \sum_j P(x_i, y_j) \log \frac{P(y_j|x_i)}{P(y_j)} \quad (15.20c)$$

or we may substitute Eq. (15.18c) into Eq. (15.20a):

$$I(x; y) = \sum_i \sum_j P(x_i)P(y_j|x_i) \log \frac{P(y_j|x_i)}{\sum_i P(x_i)P(y_j|x_i)} \quad (15.20d)$$

Equation (15.20d) expresses $I(x; y)$ in terms of the input symbol probabilities and the channel matrix.

The units of $I(x; y)$ should be carefully noted. $I(x; y)$ is the average amount of information received per symbol transmitted. Hence, its units are bits per symbol. If we use binary digits at the input, then the symbol is a binary digit, and the units of $I(x; y)$ are bits per binary digit.

Because $I(x; y)$ in Eq. (15.20b) is symmetrical with respect to x and y , it follows that

$$I(x; y) = I(y; x) \quad (15.21a)$$

$$= H(y) - H(y|x) \quad (15.21b)$$

The quantity $H(y|x)$ is the equivocation of y with respect to x and is the average uncertainty about the received symbol when the transmitted symbol is known. Equation (15.21b) can be rewritten as

$$H(x) - H(x|y) = H(y) - H(y|x) \quad (15.21c)$$

From Eq. (15.20d) it is clear that $I(x; y)$ is a function of the transmitted symbol probabilities $P(x_i)$ and the channel matrix. For a given channel, $I(x; y)$ will be maximum for some set of probabilities $P(x_i)$. This maximum value is the **channel capacity** C_s ,

$$C_s = \max_{P(x_i)} I(x; y) \quad \text{bits per symbol} \quad (15.22)$$

Thus, C_s represents the maximum information that can be transmitted by one symbol over the channel. These ideas will become clear from the following example of a binary symmetric channel (BSC).

EXAMPLE 15.3 Find the channel capacity of the BSC shown in Fig. 15.2.

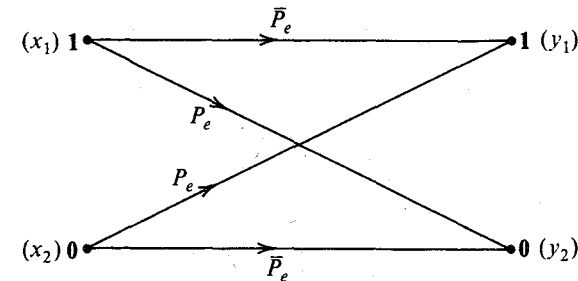


Figure 15.2 Binary symmetric channel.

Let $P(x_1) = \alpha$ and $P(x_2) = \bar{\alpha} = (1 - \alpha)$. Also,

$$P(y_1|x_2) = P(y_2|x_1) = P_e$$

$$P(y_1|x_1) = P(y_2|x_2) = \bar{P}_e = 1 - P_e$$

Substitution of these probabilities into Eq. (15.20d) gives

$$\begin{aligned} I(x; y) &= \alpha \bar{P}_e \log \left(\frac{\bar{P}_e}{\alpha \bar{P}_e + \bar{\alpha} P_e} \right) + \alpha P_e \log \left(\frac{P_e}{\alpha P_e + \bar{\alpha} \bar{P}_e} \right) \\ &\quad + \bar{\alpha} P_e \log \left(\frac{P_e}{\alpha \bar{P}_e + \bar{\alpha} P_e} \right) + \bar{\alpha} \bar{P}_e \log \left(\frac{\bar{P}_e}{\alpha P_e + \bar{\alpha} \bar{P}_e} \right) \\ &= (\alpha P_e + \bar{\alpha} \bar{P}_e) \log \left(\frac{1}{\alpha P_e + \bar{\alpha} \bar{P}_e} \right) \\ &\quad + (\alpha \bar{P}_e + \bar{\alpha} P_e) \log \left(\frac{1}{\alpha \bar{P}_e + \bar{\alpha} P_e} \right) \\ &\quad - \left(P_e \log \frac{1}{P_e} + \bar{P}_e \log \frac{1}{\bar{P}_e} \right) \end{aligned}$$

If we define

$$\Omega(z) = z \log \frac{1}{z} + \bar{z} \log \frac{1}{\bar{z}}$$

with $\bar{z} = 1 - z$, then

$$I(x; y) = \Omega(\alpha P_e + \bar{\alpha} \bar{P}_e) - \Omega(P_e) \quad (15.23)$$

The function $\Omega(z)$ vs. z is shown in Fig. 15.3. It can be seen that $\Omega(z)$ is maximum at $z = \frac{1}{2}$. (Note that we are interested in the region $0 < z < 1$ only.) For a given P_e , $\Omega(P_e)$ is fixed. Hence from Eq. (15.23) it follows that $I(x; y)$ is maximum when $\Omega(\alpha P_e + \bar{\alpha} \bar{P}_e)$ is maximum. This occurs when

$$\alpha P_e + \bar{\alpha} \bar{P}_e = 0.5$$

or

$$\alpha P_e + (1 - \alpha)(1 - P_e) = 0.5$$

This equation is satisfied when

$$\alpha = 0.5 \quad (15.24)$$

For this value of α , $\Omega(\alpha P_e + \bar{\alpha} \bar{P}_e) = 1$ and

$$\begin{aligned} C_s &= \max_{P(x_i)} I(x; y) = 1 - \Omega(P_e) \\ &= 1 - \left[P_e \log \frac{1}{P_e} + (1 - P_e) \log \left(\frac{1}{1 - P_e} \right) \right] \end{aligned} \quad (15.25)$$

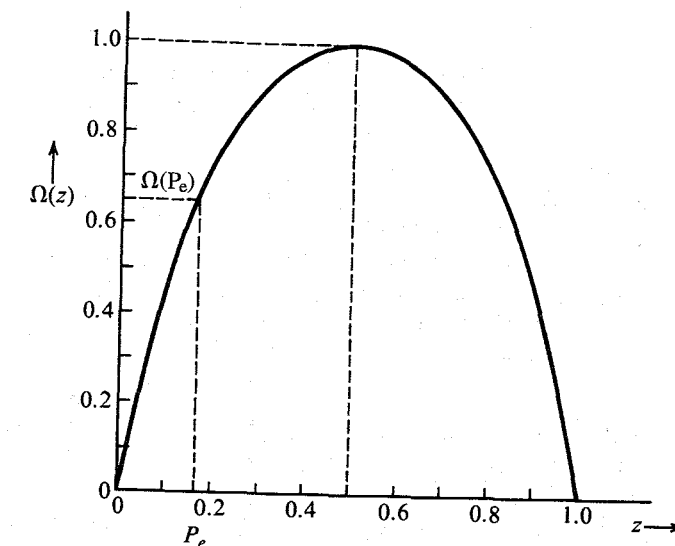


Figure 15.3 Plot of $\Omega(z)$.

Figure 15.4 shows C_s vs. P_e . From this figure it follows that the maximum value of C_s is unity. This means we can transmit at most 1 bit of information per binary digit. This is the expected result, because one binary digit can convey one of the two equiprobable messages. The information content of one of the two equiprobable messages is $\log_2 2 = 1$ bit. Second, we observe that C_s is maximum when the error probability $P_e = 0$ or $P_e = 1$. When the error probability $P_e = 0$, the channel is noiseless, and we expect C_s to be maximum. But surprisingly, C_s is also maximum when $P_e = 1$. This is easy to explain, because a channel that consistently and with certainty makes errors is as good as a noiseless channel. All we have to do is reverse the decision that is made, and we have error-free reception; that is, if 0 is received, we decide that 1 was actually sent, and vice versa. The channel capacity C_s is zero (minimum) when $P_e = \frac{1}{2}$. If the error probability is $\frac{1}{2}$, then the transmitted symbols and the received symbols are statistically independent. If we received 0, for example, either 1 or 0 is equally likely to have been transmitted, and the information received is zero.

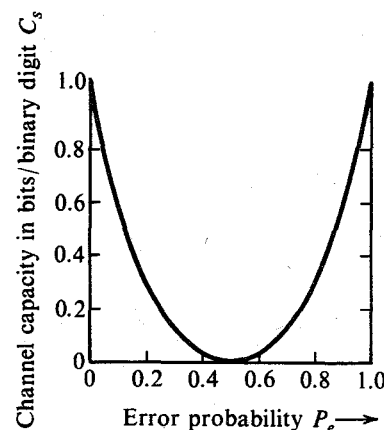


Figure 15.4 Binary symmetric channel capacity as a function of error probability P_e .

Channel Capacity per Second

The channel capacity C_s in Eq. (15.22) gives the maximum possible information transmitted when one symbol (digit) is transmitted. If K symbols are being transmitted per second, then the maximum rate of transmission of information per second is $K C_s$. This is the channel capacity in information units per seconds and will be denoted by C (in bits per second):

$$C = K C_s \quad \text{bit/s}$$

A Comment on Channel Capacity: Channel capacity is the property of a particular physical channel over which the information is transmitted. This is true provided the term **channel** is correctly interpreted. A channel means not only the transmission medium but it also includes the specifications of the kind of signals (binary, r -ary, etc., or orthogonal, simplex, etc.) and the kind of receiver used (the receiver determines the error probability). All these specifications are included in the channel matrix. A channel matrix completely specifies a

channel. If we decide to use, for example, 4-ary digits instead of binary digits over the same physical channel, the channel matrix changes (it becomes a 4×4 matrix), as does the channel capacity. Similarly, a change in the receiver or the signal power or noise power will change the channel matrix and, hence, the channel capacity.

Magnitude of the Channel Capacity

The channel capacity C_s is the maximum value of $H(x) - H(x|y)$; clearly, $C_s \leq H(x)$ [because $H(x|y) \geq 0$]. But $H(x)$ is the average information per input symbol. Hence, C_s is always less than (or equal to) the average information per input symbol. If we use binary symbols at the input, the maximum value of $H(x)$ is 1 bit, occurring when $P(x_1) = P(x_2) = \frac{1}{2}$. Hence, for a binary channel, $C_s \leq 1$ bit per binary digit. If we use r -ary symbols, the maximum value of $H_r(x)$ is 1 r -ary unit. Hence, $C_s \leq 1$ r -ary unit per symbol.

Verification of Error-Free Communication over a BSC

We have shown that over a noisy channel, C_s bits of information can be transmitted per symbol. If we consider a binary channel, this means that for each binary digit (symbol) transmitted, the received information is C_s bits ($C_s \leq 1$). Thus, to transmit 1 bit of information, we need to transmit at least $1/C_s$ binary digits. This gives a code efficiency C_s and redundancy $1 - C_s$. When the transmission of information is implied, it means error-free transmission, because $I(x; y)$ was defined as the transmitted information minus the loss of information caused by channel noise.

The problem with this derivation is that it is based on a certain speculative definition of information [Eq. (15.1)]. And based on this definition, we defined the information lost during the transmission over the channel. We really have no direct proof that the information lost over the channel will oblige us in this way. Hence, the only way to ensure that this whole speculative structure is sound is to verify it. If we can show that C_s bits of error-free information can be transmitted per symbol over a channel, the verification will be complete. A general case will be discussed later. Here we shall verify the results for a BSC.

Let us consider a binary source emitting messages at a rate of α digits per second. We accumulate these information digits over T seconds to give a total of αT digits. Because αT digits form $2^{\alpha T}$ possible combinations, our problem is now to transmit one of these $2^{\alpha T}$ supermessages every T seconds. These supermessages are transmitted by a code of word length βT digits, with $\beta > \alpha$ to ensure redundancy. Because βT digits can form $2^{\beta T}$ distinct patterns (vertices of a βT -dimensional hypercube), and we have only $2^{\alpha T}$ messages, we are utilizing only a $2^{-(\beta-\alpha)T}$ fraction of the vertices. The remaining vertices are deliberately unused in order to combat noise. If we let $T \rightarrow \infty$, the fraction of vertices used approaches 0. Because there are βT digits in each transmitted sequence, the number of digits received in error will be exactly $\beta T P_e$ when $T \rightarrow \infty$. We now construct Hamming spheres of radius $\beta T P_e$ each around the $2^{\alpha T}$ vertices used for the messages. When any message is transmitted, the received message will be on the Hamming sphere surrounding the vertex corresponding to that message. We use the following decision rule: If a received sequence falls inside or on a sphere surrounding message m_i , then the decision is " m_i is transmitted." If $T \rightarrow \infty$, the decision will be without error if all the $2^{\alpha T}$ spheres are nonoverlapping.

Of all the possible sequences of βT digits, the number of sequences that differ from a given sequence by exactly j digits is $\binom{\beta T}{j}$ (see Example 10.6). Hence, K , the total number of sequences that differ from a given sequence by less than or equal to $\beta T P_e$ digits, is

$$K = \sum_{j=0}^{\beta T P_e} \binom{\beta T}{j} \quad (15.26)$$

Here we use an inequality often used in information theory:^{4,7}

$$\sum_{j=0}^{\beta T P_e} \binom{\beta T}{j} \leq 2^{\beta T \Omega(P_e)} \quad P_e < 0.5$$

Hence,

$$K \leq 2^{\beta T \Omega(P_e)} \quad (15.27)$$

with

$$\Omega(P_e) = P_e \log \frac{1}{P_e} + (1 - P_e) \log \frac{1}{1 - P_e}$$

From the $2^{\beta T}$ possible vertices we choose $2^{\alpha T}$ vertices to be assigned to the supermessages. How shall we select these vertices? From the decision procedure it is clear that if we assign a particular vertex to a supermessage, then none of the other vertices lying within a sphere of radius $\beta T P_e$ can be assigned to another supermessage. Thus, when we choose a vertex for m_1 , the corresponding K vertices [Eq. (15.26)] become ineligible for consideration. From the remaining $2^{\beta T} - K$ vertices we choose another vertex for m_2 . We proceed this way until all the $2^{\beta T}$ vertices are exhausted. This is a rather tedious procedure. Let us see what happens if we choose the required $2^{\alpha T}$ vertices randomly from the $2^{\beta T}$ vertices. In this procedure there is a danger that we may select more than one vertex lying within a distance $\beta T P_e$. If, however, α/β is sufficiently small, the probability of making such a choice is extremely small as $T \rightarrow \infty$. The probability of choosing any particular vertex s_1 as one of the $2^{\alpha T}$ vertices from $2^{\beta T}$ vertices is $2^{\alpha T}/2^{\beta T} = 2^{-(\beta-\alpha)T}$.

Remembering that K vertices lie within a distance of $\beta T P_e$ digits from s_1 , the probability that we may also choose another vertex s_2 that is within the distance $\beta T P_e$ from s_1 is

$$P = K 2^{-(\beta-\alpha)T}$$

From Eq. (15.27) it follows that

$$P \leq 2^{-[\beta[1-\Omega(P_e)]-\alpha]T}$$

Hence, as $T \rightarrow \infty$, $P \rightarrow 0$ if

$$\beta[1 - \Omega(P_e)] > \alpha$$

that is, if

$$\frac{\alpha}{\beta} < 1 - \Omega(P_e) \quad (15.28a)$$

But $1 - \Omega(P_e)$ is C_s , the channel capacity of a BSC [Eq. (15.25)]. Therefore,

$$\frac{\alpha}{\beta} < C_s \quad (15.28b)$$

Hence, the probability of choosing two sequences randomly within a distance $\beta T P_e$ approaches 0 as $T \rightarrow \infty$ provided $\alpha/\beta < C_s$, and we have error-free communication. We can choose $\alpha/\beta = C_s - \epsilon$, where ϵ is arbitrarily small.

15.5 CHANNEL CAPACITY OF A CONTINUOUS CHANNEL*

For a discrete random variable x taking on values $x_1, x_2, \dots, x_n$ with probabilities $P(x_1), P(x_2), \dots, P(x_n)$, the entropy $H(x)$ was defined as

$$H(x) = \sum_{i=1}^n P(x_i) \log P(x_i) \quad (15.29)$$

For analog data, we have to deal with continuous random variables. Therefore, we must extend the definition of entropy to continuous random variables. One is tempted to state that $H(x)$ for continuous random variables is obtained by using the integral instead of discrete summation in Eq. (15.29)[†]:

$$H(x) = \int_{-\infty}^{\infty} p(x) \log \frac{1}{p(x)} dx \quad (15.30)$$

We shall see that Eq. (15.30) is indeed the meaningful definition of entropy for a continuous random variable. We cannot accept this definition, however, unless we show that it has the meaningful interpretation as uncertainty. A random variable x takes a value in the range $(n\Delta x, (n+1)\Delta x)$ with probability $p(n\Delta x)\Delta x$ in the limit as $\Delta x \rightarrow 0$. The error in the approximation will vanish in the limit as $\Delta x \rightarrow 0$. Hence $H(x)$, the entropy of a continuous random variable x , is given by

$$\begin{aligned} H(x) &= \lim_{\Delta x \rightarrow 0} \sum_n p(n\Delta x)\Delta x \log \frac{1}{p(n\Delta x)\Delta x} \\ &= \lim_{\Delta x \rightarrow 0} \left[\sum_n p(n\Delta x)\Delta x \log \frac{1}{p(n\Delta x)} - \sum_n p(n\Delta x)\Delta x \log \Delta x \right] \\ &= \int_{-\infty}^{\infty} p(x) \log \frac{1}{p(x)} dx - \lim_{\Delta x \rightarrow 0} \log \Delta x \int_{-\infty}^{\infty} p(x) dx \\ &= \int_{-\infty}^{\infty} p(x) \log \frac{1}{p(x)} dx - \lim_{\Delta x \rightarrow 0} \log \Delta x \end{aligned} \quad (15.31)$$

In the limit as $\Delta x \rightarrow 0$, $\log \Delta x \rightarrow -\infty$. It therefore appears that the entropy of a continuous random variable is infinite. This is quite true. The magnitude of uncertainty associated with a continuous random variable is infinite. This fact is also apparent intuitively. A continuous random variable assumes a nonnumerably infinite number of values, and, hence, the uncertainty is on the order of infinity. Does this mean that there is no meaningful definition of entropy for a continuous random variable? On the contrary, we shall see that the first term in Eq. (15.31) serves as a meaningful measure of the entropy (average information) of a continuous random variable x . This may be argued as follows. We can consider $\int p(x) \log [1/p(x)] dx$ as a relative entropy with $-\log \Delta x$ serving as a datum, or reference. The information transmitted over a channel is actually the difference between the two terms $H(x)$ and $H(x|y)$. Obviously, if we have a common datum for both $H(x)$ and $H(x|y)$, the difference $H(x) - H(x|y)$ will

* The channel is assumed to be memoryless.

† Throughout this discussion, the PDF $p_x(x)$ will be abbreviated as $p(x)$, because it causes no ambiguity and improves the clarity of equations.

be the same as the difference between their relative entropies. We are therefore justified in considering the first term in Eq. (15.31) as the **differential** entropy of x . We must, however, always remember that this is a relative entropy and not the absolute entropy. Failure to realize this subtle point generates many apparent fallacies, one of which is given in Example 15.4.

Based on this argument, we define $H(x)$, the differential entropy of a continuous random variable x , as

$$H(x) = \int_{-\infty}^{\infty} p(x) \log \frac{1}{p(x)} dx \quad \text{bits} \quad (15.32a)$$

$$= - \int_{-\infty}^{\infty} p(x) \log p(x) dx \quad \text{bits} \quad (15.32b)$$

Although $H(x)$ is the differential (relative) entropy of x , we shall call it the entropy of random variable x .

EXAMPLE 15.4 A signal amplitude x is a random variable uniformly distributed in the range $(-1, 1)$. This signal is passed through an amplifier of gain 2. The output y is also a random variable, uniformly distributed in the range $(-2, 2)$. Determine the (differential) entropies $H(x)$ and $H(y)$.

We have

$$P(x) = \begin{cases} \frac{1}{2} & |x| < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$P(y) = \begin{cases} \frac{1}{4} & |y| < 2 \\ 0 & \text{otherwise} \end{cases}$$

Hence,

$$H(x) = \int_{-1}^1 \frac{1}{2} \log 2 dx = 1 \text{ bit}$$

$$H(y) = \int_{-2}^2 \frac{1}{4} \log 4 dx = 2 \text{ bits}$$

The entropy of the random variable y is twice that of x . This result may come as a surprise, since a knowledge of x uniquely determines y , and vice versa, because $y = 2x$. Hence, the average uncertainty of x and y should be identical. Amplification itself can neither add nor subtract information. Why, then, is $H(y)$ twice as large as $H(x)$? This becomes clear when we remember that $H(x)$ and $H(y)$ are differential (relative) entropies, and they will be equal if and only if their datum (or reference) entropies are equal. The reference entropy R_1 for x is $-\log \Delta x$, and the reference entropy R_2 for y is $-\log \Delta y$ (in the limit as $\Delta x, \Delta y \rightarrow 0$),

$$R_1 = \lim_{\Delta x \rightarrow 0} -\log \Delta x$$

$$R_2 = \lim_{\Delta y \rightarrow 0} -\log \Delta y$$

and

$$R_1 - R_2 = \lim_{\Delta x, \Delta y \rightarrow 0} \log \left(\frac{\Delta y}{\Delta x} \right)$$

$$= \log \left(\frac{dy}{dx} \right)$$

$$= \log 2 = 1 \text{ bit}$$

Thus, R_1 , the reference entropy of x , is higher than the reference entropy R_2 for y . Hence, if x and y have equal absolute entropies, their differential (relative) entropies must differ by 1 bit.

Maximum Entropy for a Given Mean Square Value of x

For discrete random variables, we observed that the entropy is maximum when all the outcomes (messages) were equally likely (uniform probability distribution). For continuous random variables, there also exists a PDF $p(x)$ that maximizes $H(x)$ in Eqs. (15.32). In the case of a continuous distribution, however, we may have additional constraints on x . Either the maximum value of x or the mean square value of x may be given. We shall find here the PDF $p(x)$ that will yield maximum entropy when $\overline{x^2}$ is given to be a constant σ^2 . The problem, then, is to maximize $H(x)$:

$$H(x) = \int_{-\infty}^{\infty} p(x) \log \frac{1}{p(x)} dx \quad (15.33)$$

with the constraints

$$\int_{-\infty}^{\infty} p(x) dx = 1 \quad (15.34a)$$

$$\int_{-\infty}^{\infty} x^2 p(x) dx = \sigma^2 \quad (15.34b)$$

To solve this problem, we use a theorem from the calculus of variation. Given the integral I ,

$$I = \int_a^b F(x, p) dx \quad (15.35)$$

subject to the following constraints:

$$\int_a^b \varphi_1(x, p) dx = \lambda_1$$

$$\int_a^b \varphi_2(x, p) dx = \lambda_2$$

$$\vdots$$

$$\int_a^b \varphi_k(x, p) dx = \lambda_k \quad (15.36)$$

where $\lambda_1, \lambda_2, \dots, \lambda_k$ are given constants. The result from the calculus of variation states that the form of $p(x)$ that maximizes I in Eq. (15.35) with the constraints in Eq. (15.36) is found from the solution of the equation

$$\frac{\partial F}{\partial p} + \alpha_1 \frac{\partial \varphi_1}{\partial p} + \alpha_2 \frac{\partial \varphi_2}{\partial p} + \dots + \alpha_k \frac{\partial \varphi_k}{\partial p} = 0 \quad (15.37)$$

The quantities $\alpha_1, \alpha_2, \dots, \alpha_k$ are adjustable constants, called **undetermined multipliers**, which can be found by substituting the solution of $p(x)$ [obtained from Eq. (15.37)] in Eq. (15.36). In the present case,

$$F(p, x) = p \log \frac{1}{p}$$

$$\varphi_1(x, p) = p$$

$$\varphi_2(x, p) = x^2 p$$

Hence, the solution for p is given by

$$\frac{\partial}{\partial p} \left(p \log \frac{1}{p} \right) + \alpha_1 + \alpha_2 \frac{\partial}{\partial p} x^2 p = 0$$

or

$$-(1 + \log p) + \alpha_1 + \alpha_2 x^2 = 0$$

Solving for p , we have

$$p = e^{(\alpha_1 - 1)} e^{\alpha_2 x^2} \quad (15.38)$$

Substituting Eq. (15.38) into Eq. (15.34a), we have

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} e^{\alpha_1 - 1} e^{\alpha_2 x^2} dx \\ &= 2e^{\alpha_1 - 1} \int_0^{\infty} e^{\alpha_2 x^2} dx \\ &= 2e^{\alpha_1 - 1} \left(\frac{1}{2} \sqrt{\frac{\pi}{-\alpha_2}} \right) \end{aligned}$$

provided α_2 is negative, or

$$e^{\alpha_1 - 1} = \sqrt{\frac{-\alpha_2}{\pi}} \quad (15.39)$$

Next we substitute Eqs. (15.38) and (15.39) into Eq. (15.34b):

$$\begin{aligned} \sigma^2 &= \int_{-\infty}^{\infty} x^2 \sqrt{\frac{-\alpha_2}{\pi}} e^{\alpha_2 x^2} dx \\ &= 2 \sqrt{\frac{-\alpha_2}{\pi}} \int_0^{\infty} x^2 e^{\alpha_2 x^2} dx \\ &= -\frac{1}{2\alpha_2} \end{aligned}$$

or

$$\alpha_2 = -\frac{1}{2\sigma^2} \quad (15.40a)$$

and

$$e^{\alpha_1 - 1} = \sqrt{\frac{1}{2\pi\sigma^2}} \quad (15.40b)$$

Substituting Eqs. (15.40) into Eq. (15.38), we have

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} \quad (15.41)$$

We therefore conclude that for a given mean square value, the maximum entropy (or maximum uncertainty) is obtained when the distribution of x is gaussian. This maximum entropy, or uncertainty, is given by

$$H(x) = \int_{-\infty}^{\infty} p(x) \log_2 \frac{1}{p(x)} dx$$

Note that

$$\begin{aligned} \log \frac{1}{p(x)} &= \log \left(\sqrt{2\pi\sigma^2} e^{x^2/2\sigma^2} \right) \\ &= \frac{1}{2} \log (2\pi\sigma^2) + \frac{x^2}{2\sigma^2} \log e \end{aligned}$$

Hence,

$$\begin{aligned} H(x) &= \int_{-\infty}^{\infty} p(x) \left[\frac{1}{2} \log (2\pi\sigma^2) + \frac{x^2}{2\sigma^2} \log e \right] dx \\ &= \frac{1}{2} \log (2\pi\sigma^2) \int_{-\infty}^{\infty} p(x) dx + \frac{\log e}{2\sigma^2} \int_{-\infty}^{\infty} x^2 p(x) dx \\ &= \frac{1}{2} \log (2\pi\sigma^2) + \frac{\log e}{2\sigma^2} \sigma^2 \\ &= \frac{1}{2} \log (2\pi e \sigma^2) \end{aligned} \quad (15.42a)$$

$$= \frac{1}{2} \log (17.1 \sigma^2) \quad (15.42b)$$

To reiterate, for a given mean square value $\overline{x^2}$, the entropy is maximum for a gaussian distribution, and the corresponding entropy is $\frac{1}{2} \log (2\pi e \sigma^2)$.

The reader can similarly show that if x is constrained to some peak value M ($-M < x < M$), then the entropy is maximum when x is uniformly distributed:

$$p(x) = \begin{cases} \frac{1}{2M} & -M < x < M \\ 0 & \text{otherwise} \end{cases}$$

Entropy of a Band-Limited White Gaussian Noise

Consider a band-limited white gaussian noise $n(t)$ with PSD $\mathcal{N}/2$. Because

$$R_n(\tau) = \mathcal{N}B \operatorname{sinc}(2\pi B\tau)$$

we know that $\operatorname{sinc}(2\pi B\tau)$ is zero at $\tau = \pm k/2B$ (k integer). Therefore,

$$R_n\left(\frac{k}{2B}\right) = 0 \quad k = \pm 1, \pm 2, \pm 3, \dots$$

Hence,

$$R_n\left(\frac{k}{2B}\right) = \overline{n(t)n\left(t + \frac{k}{2B}\right)} = 0 \quad k = \pm 1, \pm 2, \dots$$

Because $n(t)$ and $n(t + k/2B)$ ($k = \pm 1, \pm 2, \dots$) are Nyquist samples of $n(t)$, it follows that Nyquist samples of $n(t)$ are all uncorrelated. Because $n(t)$ is gaussian, uncorrelatedness implies independence. Hence, all Nyquist samples of $n(t)$ are independent. Note that

$$\overline{n^2} = R_n(0) = \mathcal{N}B$$

Hence, the variance of each Nyquist sample is $\mathcal{N}B$. From Eq. (15.42a) it follows that the entropy $H(n)$ of each Nyquist sample of $n(t)$ is

$$H(n) = \frac{1}{2} \log(2\pi e \mathcal{N}B) \quad \text{bits per sample} \quad (15.43a)$$

Because $n(t)$ is completely specified by $2B$ Nyquist samples per second, the entropy per second of $n(t)$ is the entropy of $2B$ Nyquist samples. Because all the samples are independent, knowledge of one sample gives no information about any other sample. Hence, the entropy of $2B$ Nyquist samples is the sum of the entropies of the $2B$ samples, and

$$H'(n) = B \log(2\pi e \mathcal{N}B) \quad \text{bit/s} \quad (15.43b)$$

where $H'(n)$ is the entropy per second of $n(t)$.

From the results derived thus far, we can draw one significant conclusion. Among all signals band-limited to B Hz and constrained to have a certain mean square value σ^2 , the white gaussian band-limited signal has the largest entropy per second. The reason for this lies in the fact that for a given mean square value, gaussian samples have the largest entropy; moreover, all the $2B$ samples of a gaussian band-limited process are independent. Hence, the entropy per second is the sum of the entropies of all the $2B$ samples. In processes that are not white, the Nyquist samples are correlated, and, hence, the entropy per second is less than the sum of the entropies of the $2B$ samples. If the signal is not gaussian, then its samples are not gaussian, and, hence, the entropy per sample is also less than the maximum possible entropy for a given mean square value. To reiterate, for a class of band-limited signals constrained to a certain mean square value, the white gaussian signal has the largest entropy per second, or the largest amount of uncertainty. This is also the reason why white gaussian noise is the worst possible noise in terms of interference with signal transmission.

Mutual Information $I(x; y)$

The ultimate test of any concept is its usefulness. We shall now show that the relative entropy defined in Eqs. (15.32) does lead to meaningful results when we consider $I(x; y)$, the mutual information of continuous random variables x and y . We wish to transmit a random variable

x over a channel. Each value of x in a given continuous range is now a message that may be transmitted, for example, as a pulse of height x . The message recovered by the receiver will be a continuous random variable y . If the channel were noise-free, the received value y would uniquely determine the transmitted value x . But channel noise introduces a certain uncertainty about the true value of x . Consider the event that at the transmitter, a value of x in the interval $(x, x + \Delta x)$ has been transmitted ($\Delta x \rightarrow 0$). The probability of this event is $p(x)\Delta x$ in the limit $\Delta x \rightarrow 0$. Hence, the amount of information transmitted is $\log[1/p(x)\Delta x]$. Let the value of y at the receiver be y and let $p(x|y)$ be the conditional probability density of x when $y = y$. Then $p(x|y)\Delta x$ is the probability that x will lie in the interval $(x, x + \Delta x)$ when $y = y$ (provided $\Delta x \rightarrow 0$). Obviously, there is an uncertainty about the event that x lies in the interval $(x, x + \Delta x)$. This uncertainty, $\log[1/p(x|y)\Delta x]$, arises because of channel noise and therefore represents a loss of information. Because $\log[1/p(x)\Delta x]$ is the information transmitted and $\log[1/p(x|y)\Delta x]$ is the information lost over the channel, the net information received is $I(x; y)$ given by

$$I(x; y) = \log \frac{p(x|y)}{p(x)} \quad (15.44)$$

Note that this relation is true in the limit $\Delta x \rightarrow 0$. Therefore, $I(x; y)$, represents the information transmitted over a channel when we receive y ($y = y$) when x is transmitted ($x = x$). We are interested in finding the average information transmitted over a channel when some x is transmitted and a certain y is received. We must therefore average $I(x; y)$ over all values of x and y . The average information transmitted will be denoted by $I(x; y)$, where

$$I(x; y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) I(x; y) dx dy \quad (15.45a)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \log \frac{p(x|y)}{p(x)} dx dy \quad (15.45b)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \log \frac{1}{p(x)} dx dy$$

$$+ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \log p(x|y) dx dy$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x) p(y|x) \log \frac{1}{p(x)} dx dy$$

$$+ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \log p(x|y) dx dy$$

$$= \int_{-\infty}^{\infty} p(x) \log \frac{1}{p(x)} dx \int_{-\infty}^{\infty} p(y|x) dy$$

$$+ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \log p(x|y) dx dy$$

Note that

$$\int_{-\infty}^{\infty} p(y|x) dy = 1 \quad \text{and} \quad \int_{-\infty}^{\infty} p(x) \log \frac{1}{p(x)} dx = H(x)$$

Hence,

$$I(x; y) = H(x) + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \log p(x|y) dx dy \quad (15.46a)$$

$$= H(x) - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \log \frac{1}{p(x|y)} dx dy \quad (15.46b)$$

The integral on the right-hand side is the average over x and y of $\log [1/p(x|y)]$. But $\log [1/p(x|y)]$ represents the uncertainty about x when y is received. This, as we have seen, is the information lost over the channel. The average of $\log [1/p(x|y)]$ is the average loss of information when some x is transmitted and some y is received. This, by definition, is $H(x|y)$, the equivocation of x with respect to y ,

$$H(x|y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \log \frac{1}{p(x|y)} dx dy \quad (15.47)$$

Hence,

$$I(x; y) = H(x) - H(x|y) \quad (15.48)$$

Thus, when some value of x is transmitted and some value of y is received, the average information transmitted over the channel is $I(x; y)$, given by Eq. (15.48). We can define the channel capacity C_s as the maximum amount of information that can be transmitted, on the average, per sample or per value transmitted:

$$C_s = \max I(x; y) \quad (15.49)$$

For a given channel, $I(x; y)$ is a function of the input probability density $p(x)$ alone. This can be shown as follows:

$$p(x, y) = p(x)p(y|x) \quad (15.50)$$

$$\begin{aligned} \frac{p(x|y)}{p(x)} &= \frac{p(y|x)}{p(y)} \\ &= \frac{p(y|x)}{\int_{-\infty}^{\infty} p(x, y) dx} \\ &= \frac{p(y|x)}{\int_{-\infty}^{\infty} p(x)p(y|x) dx} \end{aligned} \quad (15.51)$$

Substituting Eqs. (15.50) and (15.51) into Eq. (15.45b), we obtain

$$I(x; y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x)p(y|x) \log \left(\frac{p(y|x)}{\int_{-\infty}^{\infty} p(x)p(y|x) dx} \right) dx dy \quad (15.52)$$

The conditional probability density $p(y|x)$ is characteristic of a given channel. Hence, for a given channel, $I(x; y)$ is a function of the input probability density $p(x)$ alone. Thus,

$$C_s = \max_{p(x)} I(x; y)$$

If the channel allows the transmission of K values per second, then C , the channel capacity per second, is given by

$$C = KC_s \text{ bit/s} \quad (15.53)$$

Just as in the case of discrete variables, $I(x; y)$ is symmetrical with respect to x and y for continuous random variables. This can be seen by rewriting Eq. (15.45b) as

$$I(x; y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy \quad (15.54)$$

This equation shows that $I(x; y)$ is symmetrical with respect to x and y . Hence,

$$I(x; y) = I(y; x)$$

From Eq. (15.48) it now follows that

$$I(x; y) = H(x) - H(x|y) = H(y) - H(y|x) \quad (15.55)$$

Capacity of a Band-Limited AWGN Channel

The channel capacity C is, by definition, the maximum rate of information transmission over a channel. The mutual information $I(x; y)$ is given by Eq. (15.55):

$$I(x; y) = H(y) - H(y|x) \quad (15.56)$$

The channel capacity C is the maximum value of the mutual information $I(x; y)$ per second. Let us first find the maximum value of $I(x; y)$ per sample. We shall find here the capacity of a channel band-limited to B Hz and disturbed by a white gaussian noise of PSD $\mathcal{N}/2$. In addition, we shall constrain the signal power (or its mean square value) to S . The disturbance is assumed to be additive; that is, the received signal $y(t)$ is given by

$$y(t) = x(t) + n(t) \quad (15.57)$$

Because the channel is band-limited, both the signal $x(t)$ and the noise $n(t)$ are band-limited to B Hz. Obviously, $y(t)$ is also band-limited to B Hz. All these signals can therefore be completely specified by samples taken at the uniform rate of $2B$ samples per second. Let us find the maximum information that can be transmitted per sample. Let x , n , and y represent samples of $x(t)$, $n(t)$, and $y(t)$, respectively. The information $I(x; y)$ transmitted per sample is given by Eq. (15.56):

$$I(x; y) = H(y) - H(y|x)$$

We shall now find $H(y|x)$. By definition [Eq. (15.47)],

$$\begin{aligned} H(y|x) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \log \frac{1}{p(y|x)} dx dy \\ &= \int_{-\infty}^{\infty} p(x) dx \int_{-\infty}^{\infty} p(y|x) \log \frac{1}{p(y|x)} dy \end{aligned}$$

Because

$$y = x + n$$

for a given x , y is equal to n plus a constant (x). Hence, the distribution of y when x has a given value is identical to that of n except for a translation by x . If $p_n(\cdot)$ represents the PDF of noise sample n , then

$$p(y|x) = p_n(y-x) \quad (15.58)$$

$$\int_{-\infty}^{\infty} p(y|x) \log \frac{1}{p(y|x)} dy = \int_{-\infty}^{\infty} p_n(y-x) \log \frac{1}{p_n(y-x)} dy$$

Letting $y-x = z$, we have

$$\int_{-\infty}^{\infty} p(y|x) \log \frac{1}{p(y|x)} dy = \int_{-\infty}^{\infty} p_n(z) \log \frac{1}{p_n(z)} dz$$

The right-hand side is the entropy $H(n)$ of the noise sample n . Hence,

$$H(y|x) = H(n) \int_{-\infty}^{\infty} p(x) dx$$

$$= H(n) \quad (15.59)$$

In deriving Eq. (15.59), we made no assumptions about the noise. Hence, Eq. (15.59) is very general and applies to all types of noise. The only condition is that the noise disturb the channel in an additive fashion. Thus,

$$I(x; y) = H(y) - H(n) \quad \text{bits per sample} \quad (15.60)$$

We have assumed that the mean square value of the signal $x(t)$ is constrained to have a value S , and the mean square value of the noise is N . We shall also assume that the signal $x(t)$ and the noise $n(t)$ are independent. In such a case, the mean square value of y will be the sum of the mean square values of x and n . Hence,

$$\overline{y^2} = S + N$$

For a given noise [given $H(n)$], $I(x; y)$ is maximum when $H(y)$ is maximum. We have seen that for a given mean square value of y ($\overline{y^2} = S + N$), $H(y)$ will be maximum if y is gaussian, and the maximum entropy $H_{\max}(y)$ is then given by

$$H_{\max}(y) = \frac{1}{2} \log [2\pi e(S + N)] \quad (15.61)$$

Because

$$y = x + n$$

and n is gaussian, y will be gaussian only if x is gaussian. As the mean square value of x is S , this implies that

$$p(x) = \frac{1}{\sqrt{2\pi S}} e^{-x^2/2S}$$

and

$$I_{\max}(x; y) = H_{\max}(y) - H(n)$$

$$= \frac{1}{2} \log [2\pi e(S + N)] - H(n)$$

For a white gaussian noise with mean square value N ,

$$H(n) = \frac{1}{2} \log 2\pi eN \quad N = \mathcal{N}B$$

and

$$C_s = I_{\max}(x; y) = \frac{1}{2} \log \left(\frac{S + N}{N} \right) \quad (15.62a)$$

$$= \frac{1}{2} \log \left(1 + \frac{S}{N} \right) \quad (15.62b)$$

The channel capacity per second will be the maximum information that can be transmitted per second. Equations (15.62) represent the maximum information transmitted per sample. If all the samples are statistically independent, the total information transmitted per second will be $2B$ times C_s . If the samples are not independent, then the total information will be less than $2BC_s$. Because the channel capacity C represents the maximum possible information transmitted per second,

$$C = 2B \left[\frac{1}{2} \log \left(1 + \frac{S}{N} \right) \right]$$

$$= B \log \left(1 + \frac{S}{N} \right) \quad \text{bit/s} \quad (15.63)$$

The samples of a band-limited gaussian signal are independent if and only if the signal PSD is uniform over the band (see Example 11.2 and Prob. 11.2-3). Obviously, to transmit information at the maximum rate [Eq. (15.63)], the PSD of signal $y(t)$ must be uniform. The PSD of y is given by

$$S_y(\omega) = S_x(\omega) + S_n(\omega)$$

Because $S_n(\omega) = \mathcal{N}/2$, the PSD of $x(t)$ must also be uniform. Thus, the maximum rate of transmission (C bit/s) is attained when $x(t)$ is also a white gaussian signal.

To recapitulate, when the channel noise is additive, white, gaussian with mean square value N ($N = \mathcal{N}B$), the channel capacity C of a band-limited channel under the constraint of a given signal power S is given by

$$C = B \log \left(1 + \frac{S}{N} \right) \quad \text{bit/s}$$

where B is the channel bandwidth in hertz. The maximum rate of transmission (C bit/s) can be realized only if the input signal is a white gaussian signal.

Capacity of a Channel of Infinite Bandwidth: Superficially, Eq. (15.63) seems to indicate that the channel capacity goes to ∞ as the channel's bandwidth B goes to ∞ . This, however, is not true. For white noise, the noise power $N = \mathcal{N}B$. Hence, as B increases, N also increases. It can be shown that in the limit as $B \rightarrow \infty$, C approaches a limit:

$$C = B \log \left(1 + \frac{S}{N} \right)$$

$$= B \log \left(1 + \frac{S}{\mathcal{N}B} \right)$$

$$\begin{aligned}\lim_{B \rightarrow \infty} C &= \lim_{B \rightarrow \infty} B \log \left(1 + \frac{S}{\mathcal{N}B} \right) \\ &= \lim_{B \rightarrow \infty} \frac{S}{\mathcal{N}} \left[\frac{\mathcal{N}B}{S} \log \left(1 + \frac{S}{\mathcal{N}B} \right) \right]\end{aligned}$$

This limit can be found by noting that

$$\lim_{x \rightarrow \infty} x \log_2 \left(1 + \frac{1}{x} \right) = \log_2 e = 1.44$$

Hence,

$$\lim_{B \rightarrow \infty} C = 1.44 \frac{S}{\mathcal{N}} \quad \text{bit/s} \quad (15.64)$$

Thus, for a white gaussian channel noise, the channel capacity C approaches a limit of $1.44S/\mathcal{N}$ as $B \rightarrow \infty$. The variation of C with B is shown in Fig. 15.5. It is evident that the capacity can be made infinite only by increasing the signal power S to infinity. For finite signal and noise powers, the channel capacity always remains finite.

Verification of Error-Free Communication over a Continuous Channel

Using the concepts of information theory, we have shown that it is possible to transmit error-free information at a rate of $B \log_2 (1 + S/\mathcal{N})$ bit/s over a channel band-limited to B Hz. The signal power is S , and the channel noise is white gaussian with power $\mathcal{N}$. This theorem can be verified in a way similar to that used for the verification of the channel capacity of a discrete case. This verification using signal space is so general that it is in reality an alternate proof of the capacity theorem.

Let us consider M -ary communication with M equiprobable messages $m_1, m_2, \dots, m_M$ transmitted by signals $s_1(t), s_2(t), \dots, s_M(t)$. All signals are time-limited with duration T and have an essential bandwidth B Hz. Their powers are less than or equal to S . The channel is band-limited to B , and the channel noise is white gaussian with power $\mathcal{N}$.

All the signals and noise waveforms have $2BT + 1$ dimensions. In the limit we shall let $T \rightarrow \infty$. Hence $2BT \gg 1$, and the number of dimensions will be taken as $2BT$ in our future discussion. Because the noise power is $\mathcal{N}$, the energy of the noise waveform of T -second duration is NT . The maximum signal energy is ST . Because signals and noise are independent, the maximum received energy is $(S + \mathcal{N})T$. Hence, all the received signals will lie in a $2BT$ -dimensional hypersphere of radius $\sqrt{(S + \mathcal{N})T}$ (Fig. 15.6a). A typical received

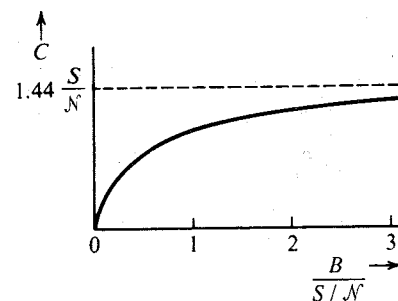


Figure 15.5 Channel capacity vs. bandwidth for a channel with white gaussian noise and fixed power.

signal $s_i(t) + n(t)$ has an energy $(S_i + \mathcal{N})T$, and the point $\mathbf{r}$ representing this signal lies at a distance of $\sqrt{(S_i + \mathcal{N})T}$ from the origin (Fig. 15.6a). The signal vector $\mathbf{s}_i$, the noise vector $\mathbf{n}$, and the received vector $\mathbf{r}$ are shown in Fig. 15.6a. Because

$$|\mathbf{s}_i| = \sqrt{S_i T}, \quad |\mathbf{n}| = \sqrt{\mathcal{N}T}, \quad |\mathbf{r}| = \sqrt{(S_i + \mathcal{N})T}$$

it follows that vectors $\mathbf{s}_i$, $\mathbf{n}$, and $\mathbf{r}$ form a right triangle. Also, $\mathbf{n}$ lies on the sphere of radius $\sqrt{\mathcal{N}T}$, centered at $\mathbf{s}_i$. Note that because $\mathbf{n}$ is random, it can lie anywhere on the sphere centered at $\mathbf{s}_i$.

We have M possible transmitted vectors located inside the big sphere. For each possible s , we can draw a sphere of radius $\sqrt{\mathcal{N}T}$ around s . If a received vector $\mathbf{r}$ lies on one of the small spheres, the center of that sphere is the transmitted waveform. If we pack the big sphere with M nonoverlapping and nontouching spheres, each of radius $\sqrt{\mathcal{N}T}$ (Fig. 15.6b), and use the centers of these M spheres for the transmitted waveforms, we will be able to detect all these M waveforms correctly at the receiver simply by using the maximum-likelihood receiver. The maximum-likelihood receiver looks at the received signal point $\mathbf{r}$ and decides that the transmitted signal is that one of the M possible transmitted points that is closest to $\mathbf{r}$ (smallest error vector). Every received point $\mathbf{r}$ will lie on the surface of one of the M nonoverlapping spheres, and using the maximum-likelihood criterion, the transmitted signal will be chosen correctly as the point lying at the center of the sphere on which $\mathbf{r}$ lies.

To compute the number of small spheres that can be packed into the big sphere, we must determine the volume of a sphere of D dimensions.

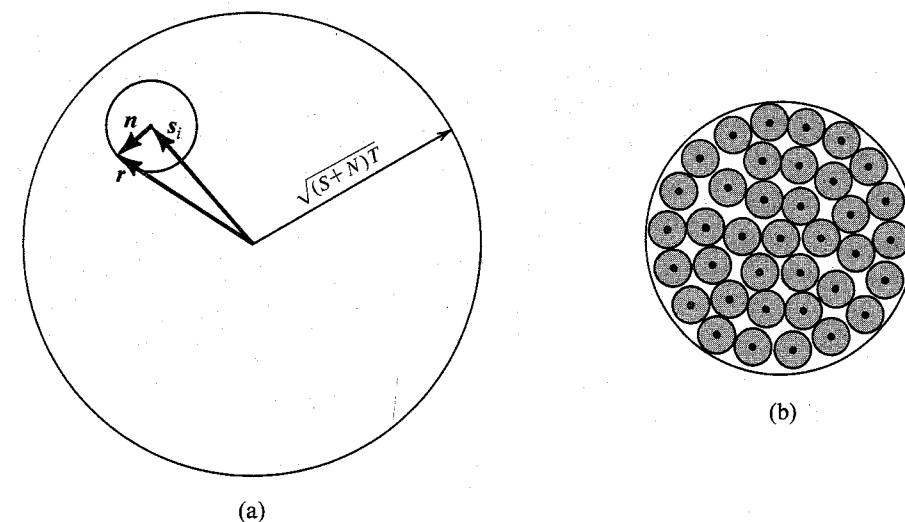


Figure 15.6 (a) Signal space representation of transmitted and received signals and noise signal. (b) Choice of signals for error-free communication.

* Because $\mathcal{N}$ is the average noise power, the energy over an interval T is $NT + \epsilon$, where $\epsilon \rightarrow 0$ as $T \rightarrow \infty$. Hence, we can assume that $\mathbf{n}$ lies on the sphere.

Volume of a D -Dimensional Sphere: A D -dimensional sphere is described by the equation

$$x_1^2 + x_2^2 + \cdots + x_D^2 = R^2$$

where R is the radius of the sphere. We can show that the volume $V(R)$ of a sphere of radius R is given by

$$V(R) = R^D V(1) \quad (15.65)$$

where $V(1)$ is the volume of a D -dimensional sphere of unit radius and, thus, is constant. To prove this, we have by definition

$$V(R) = \iiint_{x_1^2 + x_2^2 + \cdots + x_D^2 \leq R^2} dx_1 dx_2 \cdots dx_D$$

Letting $y_j = x_j/R$, we have

$$\begin{aligned} V(R) &= R^D \iiint_{y_1^2 + y_2^2 + \cdots + y_D^2 \leq 1} dy_1 dy_2 \cdots dy_D \\ &= R^D V(1) \end{aligned}$$

Hence, the ratio of the volumes of two spheres of radii $\hat{R}$ and R is

$$\frac{V(\hat{R})}{V(R)} = \left(\frac{\hat{R}}{R}\right)^D$$

A direct consequence of this result is that when D is large, almost all of the volume of the sphere is concentrated at the surface. This is because if $\hat{R}/R < 1$, then $(\hat{R}/R)^D \rightarrow 0$ as $D \rightarrow \infty$. This ratio approaches zero even if $\hat{R}$ differs from R by a very small amount Δ (Fig. 15.7). This means that no matter how small Δ is, the volume within radius $\hat{R}$ is a negligible fraction of the total volume within radius R if D is large enough. Hence, for a large D , almost all of the volume of a D -dimensional sphere is concentrated at the surface. Such a result sounds strange, but a little reflection will show that it is reasonable. This is because the volume is proportional to the D th power of the radius. Thus, for large D , a small increase in R can increase the volume tremendously, and all the increase comes from a tiny increase in R near the surface of

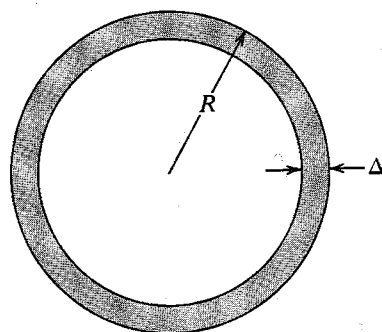


Figure 15.7 Volume of a shell of a D -dimensional hypersphere.

the sphere. This means that most of the volume must be concentrated at the surface.

The number of nonoverlapping spheres of radius $\sqrt{NT}$ that can be packed into a sphere of radius $\sqrt{(S+N)T}$ is bounded by the ratio of the volume of the signal sphere to the volume of the noise sphere. Hence,

$$M \leq \frac{[\sqrt{(S+N)T}]^{2BT} V(1)}{(\sqrt{NT})^{2BT} V(1)} = \left(1 + \frac{S}{N}\right)^{BT} \quad (15.66)$$

Each of the M -ary signals carries the information of $\log_2 M$ binary digits. Hence, the transmission of one of the M signals every T seconds is equivalent to the information rate C given by

$$C = \frac{\log M}{T} \leq B \log \left(1 + \frac{S}{N}\right) \text{ bit/s} \quad (15.67)$$

This equation gives the upper limit of C . To show that we can actually receive error-free information at a rate of $B \log(1 + S/N)$, we use the argument proposed by Shannon.⁸ Instead of choosing the M transmitted messages at the centers of nonoverlapping spheres (Fig. 15.6b), Shannon proposed to select the M points randomly located in the signal sphere I_s of radius $\sqrt{ST}$ (Fig. 15.8). Consider one particular transmitted signal s_k . Because the signal energy is assumed to be $\leq S$, point s_k will lie somewhere inside the signal sphere I_s of radius $\sqrt{ST}$. Because all the M signals are picked randomly from this sphere, the probability of finding a signal within a volume ΔV is $M \Delta V / V_s$, where V_s is the volume of I_s . But because for large D all of the volume of the sphere is concentrated at the surface, all M signal points selected randomly would lie near the surface of I_s . Figure 15.8 shows the transmitted signal s_k , the received signal r , and the noise n . We draw a sphere of radius $\sqrt{NT}$ with r as the center. This sphere intersects the sphere I_s and forms a common lens-shaped region. The signal s_k lies on the surface of both spheres. We shall use a maximum-likelihood receiver. This means that when r is received, we shall make the decision that " s_k was transmitted" provided none of

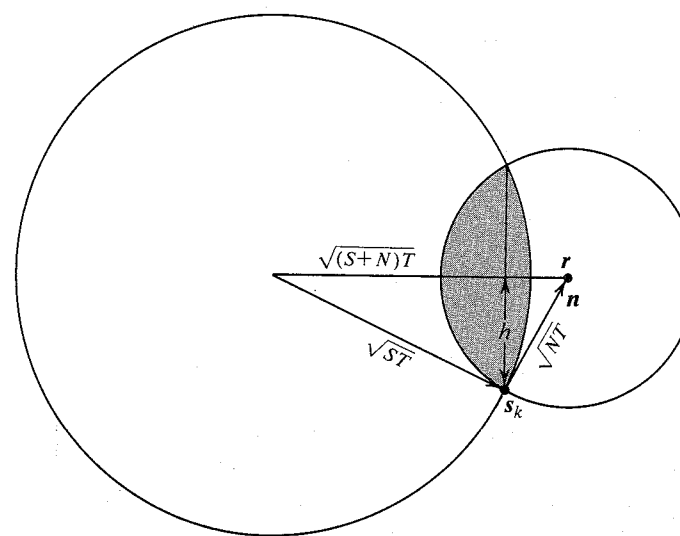


Figure 15.8 Derivation of channel capacity.

the remaining $M - 1$ signal points are closer to $\mathbf{r}$ than $\mathbf{s}_k$. The probability of finding any one signal in the lens is V_{lens}/V_s . Hence P_e , the error probability in the detection of $\mathbf{s}_k$ when $\mathbf{r}$ is received, is

$$P_e = (M - 1) \frac{V_{\text{lens}}}{V_s} < M \frac{V_{\text{lens}}}{V_s}$$

From Fig. 15.8, we observe that $V_{\text{lens}} < V(h)$, where $V(h)$ is the volume of the D -dimensional sphere of radius h . Because $\mathbf{r}$, $\mathbf{s}_k$, and $\mathbf{n}$ form a right triangle,

$$h\sqrt{(S + N)T} = \sqrt{(ST)(NT)} \quad \text{and} \quad h = \sqrt{\frac{SNT}{S + N}}$$

Hence,

$$V(h) = \left(\frac{SNT}{S + N} \right)^{BT} V(1)$$

Also,

$$V_s = (ST)^{BT} V(1)$$

and

$$P_e < M \left(\frac{N}{S + N} \right)^{BT}$$

If we choose

$$M = \left[k \left(1 + \frac{S}{N} \right) \right]^{BT}$$

then

$$P_e < [k]^{BT}$$

If we let $k = 1 - \Delta$, where Δ is a positive number chosen as small as we wish, then

$$P_e \rightarrow 0 \quad \text{as} \quad BT \rightarrow \infty$$

This means that P_e can be made arbitrarily small by increasing T , provided M is chosen arbitrarily close to $(1 + S/N)^{BT}$. Thus,

$$C = \frac{1}{T} \log_2 M = \left[B \log \left(1 + \frac{S}{N} \right) - \epsilon \right] \quad \text{bit/s} \quad (15.68)$$

where ϵ is a positive number chosen as small as we please. This proves the desired result. A more rigorous derivation of this result can be found in Wozencraft and Jacobs.⁹

Because the M signals are selected randomly from the signal space, they tend to acquire the statistics of white noise⁸ (i.e., a white gaussian random process).

Comments on Channel Capacity: According to the result derived in this chapter, theoretically we can communicate error-free up to C bit/s. There are, however, practical difficulties in achieving this rate. In proving the capacity formula, we assumed that communication is effected by signals of duration T . This means we must wait T seconds to accumulate the input data and then encode it by one of the waveforms of duration T . Because the capacity rate is achieved only in the limit as $T \rightarrow \infty$, we have to wait a long time at the receiver to get the information. Moreover, because the number of possible messages that can be transmitted over interval T increases exponentially with T , the transmitter and receiver structures increase in complexity beyond imagination as $T \rightarrow \infty$.

The channel capacity indicated by Shannon's equation [Eq. (15.68)] is the maximum error-free communication rate achievable on an optimum system without any restrictions (except for bandwidth B , signal power S , and gaussian white channel noise power N). If we have any other restrictions, this maximum rate will not be achieved. For example, if we consider a binary channel (a channel restricted to transmit only binary signals), we will not be able to attain Shannon's rate, even if the channel is optimum. The channel capacity formula [Eq. (15.68)] indicates that the transmission rate is a monotonically increasing function of the signal power S . If we use a binary channel, however, we know that increasing the transmitted power beyond a certain point buys very little advantage (see Fig. 12.14). Hence, on a binary channel, increasing S will not increase the error-free communication rate beyond some value. This does not mean that the channel capacity formula has failed. It simply means that when we have a large amount of power (with a finite bandwidth) available, the binary scheme is not the optimum communication scheme.

One last comment: Shannon's results tell us the upper theoretical limit of error-free communication. But they do not tell us precisely how this can be achieved. To quote the words of Abramson, "[This is one of the problems] which has persisted to mock information theorists since Shannon's original paper in 1948. Despite an enormous amount of effort spent since that time in quest of this Holy Grail of information theory, a *deterministic* method of generating the codes promised by Shannon is still to be found."⁴

15.6 PRACTICAL COMMUNICATION SYSTEMS IN LIGHT OF SHANNON'S EQUATION

It would be instructive to determine the ideal law for the exchange between the SNR and the transmission bandwidth using the channel capacity equation. Consider a message of bandwidth B that is used for modulation (or coding), with the resulting modulated signal of bandwidth B_T . This signal is received at the input of an ideal demodulator with signal and noise powers of S_i and N_i , respectively* (Fig. 15.9). The demodulator output bandwidth is B , and the SNR is S_o/N_o . Because an SNR S/N and a bandwidth B can transmit ideally $B \log(1 + S/N)$ bits of information, the ideal information rates of the signals at the input and the output of the demodulator are $B_T \log(1 + S_i/N_i)$ bits and $B \log(1 + S_o/N_o)$ bits, respectively. Because the demodulator neither creates nor destroys information, the two rates should be equal, that is,

$$B_T \log \left(1 + \frac{S_i}{N_i} \right) = B \log \left(1 + \frac{S_o}{N_o} \right)$$

* An additive white gaussian channel noise is assumed.

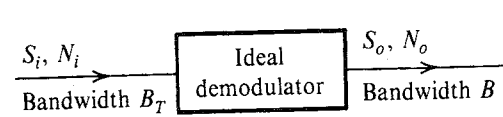


Figure 15.9 Ideal exchange between SNR and bandwidth.

and

$$\left(1 + \frac{S_o}{N_o}\right) = \left(1 + \frac{S_i}{N_i}\right)^{B_T/B} \quad (15.69a)$$

In practice, for the majority of systems, S_o/N_o as well as $S_i/N_i \gg 1$, and

$$\frac{S_o}{N_o} \approx \left(\frac{S_i}{N_i}\right)^{B_T/B} \quad (15.69b)$$

Also,

$$\begin{aligned} \frac{S_i}{N_i} &= \frac{S_i}{\mathcal{N}B_T} \\ &= \left(\frac{S_i}{\mathcal{N}B}\right) \left(\frac{B}{B_T}\right) = \frac{B}{B_T} \gamma \quad \gamma = \frac{S_i}{\mathcal{N}B} \end{aligned}$$

Hence, Eqs. (15.69) become

$$\frac{S_o}{N_o} = \left(1 + \frac{\gamma}{B_T/B}\right)^{B_T/B} - 1 \quad (15.70a)$$

$$\approx \left(\frac{\gamma}{B_T/B}\right)^{B_T/B} \quad (15.70b)$$

Equations (15.69) and (15.70) give the ideal law of exchange between the SNR and the bandwidth. The output SNR S_o/N_o is plotted in Fig. 15.10 as a function of γ for various values of B_T/B .

The output SNR increases exponentially with the bandwidth expansion factor B_T/B . This means that to maintain a given output SNR, the transmitted signal power can be reduced exponentially with the bandwidth expansion factor. Thus, for a small increase in bandwidth, we can reduce the transmitted power considerably. On the other hand, for a small reduction in bandwidth, we need to increase the transmitted power considerably. Hence, in practice, the trade is in the sense of reducing the transmitted power at the cost of increased transmission bandwidth and rarely the other way.

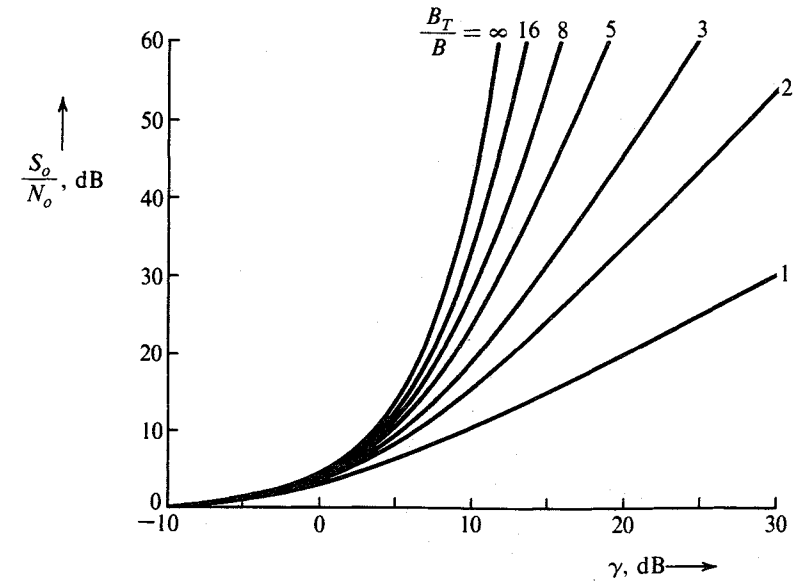
Let us now investigate various systems studied thus far and see how they fare in comparison with the ideal system.

AM

For baseband and SSB-SC systems, $B_T/B = 1$, and Eqs. (15.70) yield

$$\frac{S_o}{N_o} = \gamma \quad (15.71)$$

which is exactly the performance of these systems [see Eqs. (12.3) and (12.12)]. For DSB-SC, $B_T/B = 2$, and Eqs. (15.70) predict

Figure 15.10 Ideal behavior of SNR vs. γ for various ratios of B_T/B .

$$\frac{S_o}{N_o} \approx \frac{\gamma^2}{4} \quad (15.72)$$

Thus DSB-SC, for which $S_o/N_o = \gamma$, falls short of ideal performance. If, however, we consider the fact that quadrature multiplexing can be used to transmit two DSB signals simultaneously, with the effective bandwidth per signal as B rather than $2B$, DSB-SC has ideal performance. Because for AM quadrature multiplexing is not used,* AM performance [Eq. (12.14)] falls considerably short of the ideal performance [Eq. (15.72)].

The ideal performance of the baseband and SSB-SC or DSB-SC systems is really an empty boast, because these systems do not exchange SNR for bandwidth.

FM

For FM, we have [Eq. (12.37)]

$$\frac{S_o}{N_o} = 3\beta^2 \gamma \left(\frac{\overline{m^2}}{m_p^2}\right)$$

Because $B_T/B = 2(\beta + 1)$,

$$\frac{S_o}{N_o} = 3 \left[\frac{1}{2} \left(\frac{B_T}{B} \right) - 1 \right]^2 \left(\frac{\overline{m^2}}{m_p^2} \right) \gamma \quad (15.73)$$

Figure 15.11 shows the plots of S_o/N_o for $B_T/B = 8$ and 20, assuming $m_p^2/\overline{m^2} = 2$. The ideal S_o/N_o plots for $B_T/B = 8$ and 20 are also shown in the figure for comparison.

* Because this requires a phase reference at the receiver, it defeats the purpose of AM.

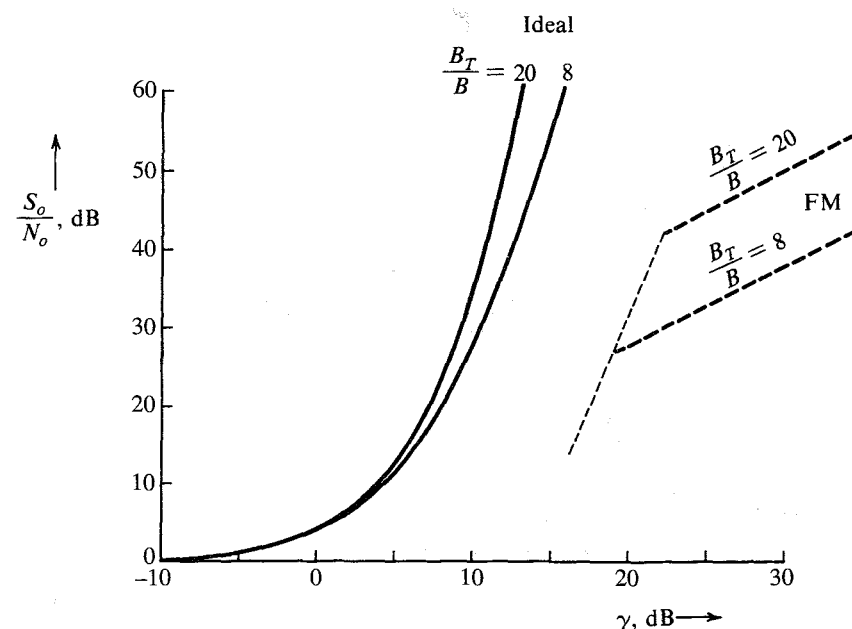


Figure 15.11 Comparison of ideal behavior with FM system behavior.

FM performance falls far below ideal performance. Even if a gain of 13 dB resulting from preemphasis and deemphasis is added to these plots, FM is several decibels inferior to ideal curves. The comparison between FM and the ideal system gets progressively worse as γ increases. If we observe the behavior of FM at the threshold, however, it does not fare as badly. It can be shown that for FM, when optimum demodulation (phase-locked loop) is used, the behavior of the SNR at the threshold (the dashed line* in Fig. 15.11) is close to ideal.¹⁰

PCM

As seen earlier, M -ary PCM shows a saturation effect unless we go to higher values of M as γ increases. If the message signal is quantized in L levels, then each sample can be encoded by $\log_M L$ number of M -ary pulses. If B is the bandwidth of the message signal, we need to transmit $2B$ samples per second. Consequently, R_M , the number of M -ary pulses per second, is

$$R_M = 2B \log_M L$$

Also, the transmission bandwidth B_T is half the number of (M -ary) pulses per second. Hence,

$$B_T = \frac{R_M}{2} = B \log_M L \quad (15.74a)$$

From Eq. (13.51a), the power S_i is found as

$$S_i = \frac{M^2 - 1}{3} E_p R_M \quad (15.74b)$$

* The dashed threshold line shown in Fig. 15.11 is for a frequency discriminator. For optimum demodulation using a phase-locked loop, the threshold line is shifted left by 3 to 5 dB.

Also,

$$N_i = \mathcal{N} B_T = \frac{\mathcal{N} R_M}{2} \quad (15.75)$$

Each of the M -ary pulses carries the information of $\log_2 M$ bits, and we are transmitting $2B \log_M L$ number of M -ary pulses per second. Hence, we are transmitting information at a rate of R_b bits, where

$$\begin{aligned} R_b &= (2B \log_M L)(\log_2 M) \\ &= 2B_T \log_2 M \\ &= B_T \log_2 M^2 \quad \text{bit/s} \end{aligned}$$

Substitution of Eqs. (15.74b) and (15.75) into this equation yields

$$R_b = B_T \log_2 \left(1 + \frac{3\mathcal{N} S_i}{2E_p N_i} \right) \quad \text{bit/s} \quad (15.76)$$

We are transmitting the information equivalent of R_b binary digits per second over the M -ary PCM channel. The reception is not error-free, however. The pulses are detected with an error probability P_{eM} given in Eq. (13.52b). If P_{eM} is on the order of 10^{-6} , we could consider the reception to be essentially error-free. From Eq. (13.52b),

$$P_{eM} \simeq 2Q \left(\sqrt{\frac{2E_p}{\mathcal{N}}} \right) = 10^{-6} \quad M \gg 1$$

This gives

$$\frac{2E_p}{\mathcal{N}} = 24$$

Substitution of this value in Eq. (15.76) gives

$$R_b = B_T \log_2 \left(1 + \frac{1}{8} \frac{S_i}{N_i} \right) \quad \text{bit/s} \quad (15.77)$$

Thus, over a channel of bandwidth B_T with an SNR of S_i/N_i , a PCM system can transmit information at a rate of R_b in Eq. (15.77). The ideal channel with bandwidth B_T and SNR S_i/N_i transmits information at a rate of C bit/s, where

$$C = B_T \log_2 \left(1 + \frac{S_i}{N_i} \right) \quad \text{bit/s} \quad (15.78)$$

It follows that PCM uses roughly eight times (9 dB) as much power as the ideal system. This performance is still much superior to that of FM. Figure 15.12 shows R_b/B_T as a function of S_i/N_i . For the ideal system

$$\frac{R_b}{B_T} = \frac{C}{B_T} = \log_2 \left(1 + \frac{S_i}{N_i} \right)$$

PCM at the threshold is 9 dB inferior to the ideal curve.

When PCM is in saturation, the detection error probability approaches 0. Each M -ary pulse transmits $\log_2 M$ bits, and there are $2B_T$ pulses per second. Hence,

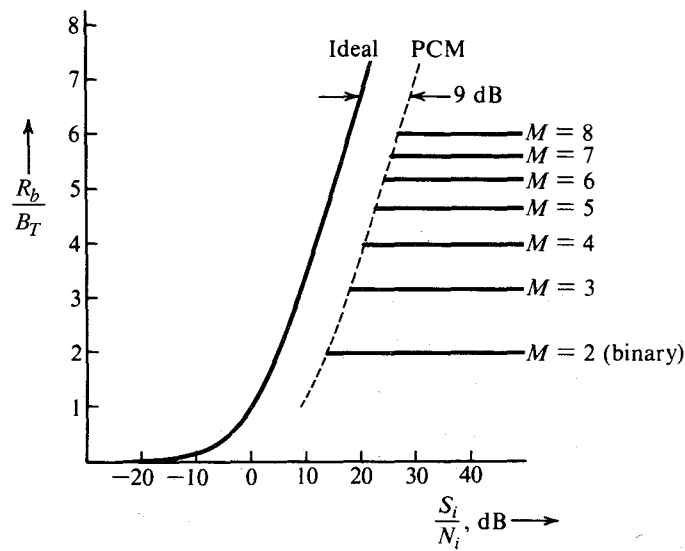


Figure 15.12 Comparison of ideal system behavior with that of PCM.

$$R_b = 2B_T \log_2 M$$

or

$$\frac{R_b}{B_T} = 2 \log_2 M$$

This is clearly seen in Fig. 15.12.

Orthogonal Signaling

We have already shown that [see Eq. (13.58)] for M -ary orthogonal signaling, the error-free communication rate is

$$R_b \leq 1.44 \frac{S_i}{N} \quad \text{bit/s} \tag{15.79}$$

We have shown in Eq. (15.64) that this is precisely the rate of error-free communication over an ideal channel with infinite bandwidth. Therefore, as $M \rightarrow \infty$, the bandwidth of an M -ary scheme approaches ∞ , and its rate of communication approaches that of an ideal channel.

REFERENCES

1. C. E. Shannon, "Mathematical Theory of Communication," *Bell Sys. Tech. J.*, vol. 27, pp. 379–423, July 1948; pp. 623–656, Oct. 1948.
2. R. V. L. Hartley, "Transmission of Information," *Bell Syst. Tech. J.*, vol. 7, pp. 535–563, July 1928.
3. H. Nyquist, "Certain Factors Affecting Telegraph Speed," *Bell Syst. Tech. J.*, vol. 3, pp. 324–346, April 1924.
4. N. Abramson, *Information Theory and Coding*, McGraw-Hill, New York, 1963.
5. R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, New York, 1968.

PROBLEMS

6. D. A. Huffman, "A Method for Construction of Minimum Redundancy Codes," *Proc. IRE*, vol. 40, pp. 1098–1101, Sept. 1952.
7. R. W. Hamming, *Coding and Information Theory*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 1986.
8. C. E. Shannon, "Communication in the Presence of Noise," *Proc. IRE*, vol. 37, pp. 10–21, Jan. 1949.
9. J. M. Wozencraft and I. A. Jacobs, *Principles of Communication Engineering*, Wiley, New York, 1965, chap. 5.
10. A. J. Viterbi, *Principles of Coherent Communication*, McGraw-Hill, New York, 1966.

- 15.1-1 A message source generates one of four messages randomly every microsecond. The probabilities of these messages are 0.4, 0.3, 0.2, and 0.1. Each emitted message is independent of the other messages in the sequence.
- (a) What is the source entropy?
- (b) What is the rate of information generated by this source (in bits per second)?
- 15.1-2 A television picture is composed of approximately 300,000 basic picture elements (about 600 picture elements in a horizontal line and 500 horizontal lines per frame). Each of these elements can assume 10 distinguishable brightness levels (such as black and shades of gray) with equal probability. Find the information content of a television picture frame.
- 15.1-3 A radio announcer describes a television picture orally in 1000 words out of his vocabulary of 10,000 words. Assume that each of the 10,000 words in his vocabulary is equally likely to occur in the description of this picture (a crude approximation, but good enough to give an idea). Determine the amount of information broadcast by the announcer in describing the picture. Would you say the announcer can do justice to the picture in 1000 words? Is the old adage "a picture is worth a thousand words" an exaggeration or an underrating of the reality? Use data in Prob. 15.1-2 to estimate the information of a picture.
- 15.1-4 From the Old North Church Tower in Boston, Paul Revere's friend was to show him one lantern if the British army began advancing overland and two lanterns if they chose to cross the bay in boats.
- (a) Assume Revere had no way of guessing ahead of time what route the British might choose. How much information did he receive when he saw two lanterns?
- (b) What if Revere were 90% sure the British would march overland? Then, how much information would the two lanterns have conveyed?
- 15.1-5 Estimate the information per letter in the English language by various methods, assuming each character is independent of the others. (This is not true, but is good enough to get a rough idea.)
- (a) In the first method, assume that all 27 characters (26 letters and a space) are equiprobable. This is a gross approximation, but is good for a quick answer.
- (b) In the second method, use the table of probabilities of various characters (Table P15.1-5).
- (c) Lastly, use Zipf's law relating the word rank to its probability. In English prose, if we order words according to the frequency of usage so that the most frequently used word (*the*) is word number 1 (rank 1), the next most probable word (*of*) is number 2 (rank 2), and so on, then empirically it is found that $P(r)$, the probability of the r th word (rank r) is very nearly

$$P(r) = \frac{0.1}{r}$$

Using Zipf's law, compute the entropy per word. Assume that there are 8727 words. The reason for this number is that the probabilities $P(r)$ sum to 1 for r from 1 to 8727. Zipf's law, surprisingly,

gives reasonably good results. Assuming there are 5.5 letters (including space) on average per word, determine the entropy or information per letter.

Table P15.1.5
Probability of occurrence of letters in the English language

Letter	Probability	$-\log P_i$	Letter	Probability	$-\log P_i$
Space	0.187	2.46	M	0.02075	5.60
E	0.1073	3.22	U	0.02010	5.64
T	0.0856	3.84	G	0.01633	5.94
A	0.0668	3.90	Y	0.01623	5.95
O	0.0654	3.94	P	0.01623	5.95
N	0.0581	4.11	W	0.01620	6.32
R	0.0559	4.16	B	0.01179	6.42
I	0.0519	4.27	V	0.00752	7.06
S	0.0499	4.33	K	0.00344	8.20
H	0.04305	4.54	X	0.00136	9.54
D	0.03100	5.02	J	0.00108	9.85
L	0.02775	5.17	Q	0.00099	9.98
F	0.02395	5.38	Z	0.00063	10.63
C	0.02260	5.45			

- 15.2-1 A source emits seven messages with probabilities 1/2, 1/4, 1/8, 1/16, 1/32, 1/64, and 1/64, respectively. Find the entropy of the source. Obtain the compact binary code and find the average length of the code word. Determine the efficiency and the redundancy of the code.
- 15.2-2 A source emits seven messages with probabilities 1/3, 1/3, 1/9, 1/9, 1/27, 1/27, and 1/27, respectively. Find the entropy of the source. Obtain the compact 3-ary code and find the average length of the code word. Determine the efficiency and the redundancy of the code.
- 15.2-3 A source emits one of four messages randomly every 1 microsecond. The probabilities of these messages are 0.5, 0.3, 0.1, and 0.1. Messages are generated independently.

(a) What is the source entropy?

(b) Obtain a compact binary code and determine the average length of the code word, the efficiency, and the redundancy of the code.

(c) Repeat part (b) for a compact ternary code.
- 15.2-4 For the messages in Prob. 15.2-1, obtain the compact 3-ary code and find the average length of the code word. Determine the efficiency and the redundancy of this code.
- 15.2-5 For the messages in Prob. 15.2-2, obtain the compact binary code and find the average length of the code word. Determine the efficiency and the redundancy of this code.
- 15.2-6 A source emits three equiprobable messages randomly and independently.

(a) Find the source entropy.

(b) Find a compact ternary code, the average length of the code word, the code efficiency, and the redundancy.

(c) Repeat part (b) for a binary code.

(d) To improve the efficiency of a binary code, we now code the second extension of the source. Find a compact binary code, the average length of the code word, the code efficiency, and the redundancy.

15.4-1 A binary channel matrix is given by

Outputs

$$\begin{matrix} & y_1 & y_2 \\ \text{Inputs } x_1 & \begin{bmatrix} 2 \\ 3 \end{bmatrix} & \begin{bmatrix} 1 \\ 3 \end{bmatrix} \\ x_2 & \begin{bmatrix} 1 \\ 10 \end{bmatrix} & \begin{bmatrix} 9 \\ 10 \end{bmatrix} \end{matrix}$$

This means $P_{y|x}(y_1|x_1) = 2/3$, $P_{y|x}(y_2|x_1) = 1/3$, etc. You are also given that $P_x(x_1) = 1/3$ and $P_x(x_2) = 2/3$. Determine $H(x)$, $H(x|y)$, $H(y)$, $H(y|x)$, and $I(x; y)$.

15.4-2 For the ternary channel in Fig. P15.4-2, $P_x(x_1) = P$, $P_x(x_2) = P_x(x_3) = Q$. (Note: $P+2Q = 1$.)

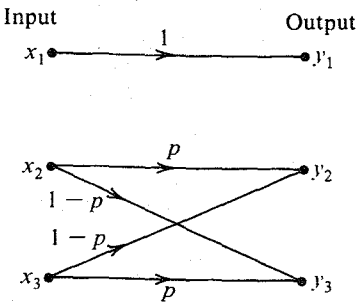


Figure P15.4-2

- (a) Determine $H(x)$, $H(x|y)$, $H(y)$, and $I(x; y)$.
- (b) Show that the channel capacity C_s is given by

$$C_s = \log \left(\frac{\beta + 2}{\beta} \right)$$

$$\text{where } \beta = 2^{-[p \log p + (1-p) \log (1-p)]}$$

15.4-3 Consider the binary symmetric channel shown in Fig. P15.4-3a. The channel matrix is given by

$$M = \begin{bmatrix} 1 - P_e & P_e \\ P_e & 1 - P_e \end{bmatrix}$$

Figure P15.4-3b shows a cascade of two such BSCs.

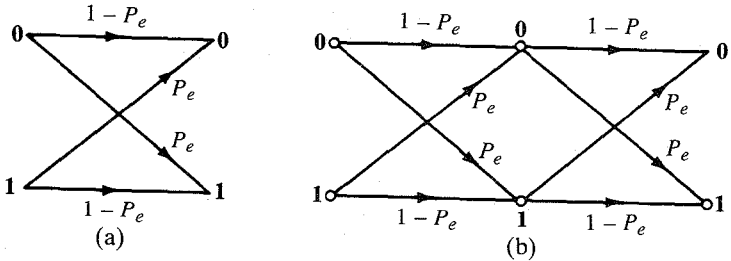


Figure P15.4-3

- (a) Determine the channel matrix for the cascaded channel in Fig. P15.4-3b. Show that this matrix is M^2 .
- (b) If the two BSC channels in Fig. P15.4-3b have error probabilities P_{e1} and P_{e2} , with channel matrices M_1 and M_2 , respectively, show that the channel matrix of the cascade of these two channels is $M_1 M_2$.
- (c) Using the results in part (b), show that the channel matrix for the cascade of k identical BSCs each with channel matrix M is M^k . Verify your answer for $n = 3$ by confirming the results in Example 10.7.
- (d) Use the result in part (c) to determine the channel capacity for a cascade of k identical BSC channels each with error probability P_e .

- 15.4-4 In data communication using error detection code, as soon as an error is detected, an automatic request for retransmission (ARQ) enables retransmission of the data in error. In such a channel the data in error is erased. Hence, there is an erase probability p , but the probability of error is zero. Such a channel, known as a **binary erasure channel (BEC)**, can be modeled as shown in Fig. P15.4-4. Determine $H(x)$, $H(x|y)$, and $I(x; y)$ assuming the two transmitted messages equiprobable.

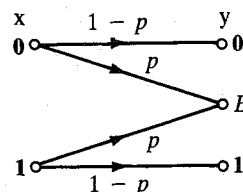


Figure P15.4-4

- 15.4-5 A cascade of two channels is shown in Fig. P15.4-5. The symbols at the source, at the output of the first channel, and at the output of the second channel are denoted by x , y , and z . Show that

$$H(x|z) \geq H(x|y)$$

and

$$I(x; y) \geq I(x; z)$$

This shows that the information that can be transmitted over a cascaded channel can be no greater than that transmitted over one link. In effect, information channels tend to leak information. *Hint:* For a cascaded channel, observe that

$$P(z_k|y_j, x_i) = P(z_k|y_j)$$

Hence, by Bayes' rule,

$$P(x_k|y_j, z_k) = P(x_i|y_j)$$

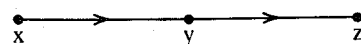


Figure P15.4-5

- 15.5-1 For a continuous random variable x constrained to a peak magnitude M ($-M < x < M$), show that the entropy is maximum when x is uniformly distributed in the range $(-M, M)$ and has zero probability density outside this range. Show that the maximum entropy is given by $\log 2M$.
- 15.5-2 For a continuous random variable x constrained to only positive values $0 < x < \infty$ and a mean value A , show that the entropy is maximum when

$$P_x(x) = \frac{1}{A} e^{-x/A} u(x)$$

Show that the corresponding entropy is

$$H(x) = \log eA$$

- 15.5-3 A television transmission requires 30 frames of 300,000 picture elements each to be transmitted per second. Using the data in Prob. 15.1-2, estimate the theoretical bandwidth of the AWGN channel if the SNR at the receiver is required to be at least 50 dB.
- 15.5-4 Show that the channel capacity of a band-limited channel disturbed by a colored gaussian noise under the constraint of a given signal power is

$$C = B \log [S_s(\omega) + S_n(\omega)] - \int_{f_1}^{f_2} \log S_n(\omega) df \quad df = \frac{d\omega}{2\pi}$$

where B is the channel bandwidth (in hertz) over the frequency range (f_1, f_2) ($f_2 - f_1 = B$). $S_s(\omega)$ and $S_n(\omega)$ are the signal and the noise power densities, respectively. Show that this maximum rate of information transmission is attained if the desired signal is gaussian and its PSD satisfies the condition

$$S_s(\omega) + S_n(\omega) = \alpha \quad (\text{a constant})$$

Hint: Consider a narrow-band Δf in the range (f_1, f_2) . The maximum rate of transmission over this band is given by

$$\Delta f \log \left[\frac{S_s(\omega) \Delta f + S_n(\omega) \Delta f}{S_n(\omega) \Delta f} \right] = \Delta f \log \left[\frac{S_s(\omega) + S_n(\omega)}{S_n(\omega)} \right]$$

provided the signal over this band is gaussian. The rate of transmission over the entire band is given by

$$\int_{f_1}^{f_2} \log \left[\frac{S_s(\omega) + S_n(\omega)}{S_n(\omega)} \right] df$$

Now maximize this under the constraint

$$2 \int_{f_1}^{f_2} S_s(\omega) df = S$$

- 15.5-5 Using the results in Prob. 15.5-4, show that the worst kind of gaussian noise is a white gaussian noise that is constrained to a given mean square value. *Hint:* Use the expression for channel capacity in Prob. 15.5-4. The first term in this expression is a constant. Now show that the second term attains a maximum value when $S_n(\omega)$ is a constant under the constraint

$$2 \int_{f_1}^{f_2} S_n(\omega) df = N$$

16 ERROR CORRECTING CODES

As seen from the discussion in Chapter 15, the key to realizing error-free communication is the use of appropriate redundancy. The addition of a single parity-check digit to detect an odd number of errors is a good example. Since Shannon's pioneering paper, a great deal of work has been carried out in the area of error correcting codes.

16.1 INTRODUCTION

In this chapter we shall discuss two important types of codes: block codes and convolutional codes. The information coming from the data or message source will be assumed to be in binary form (a sequence of binary digits).

In **block codes**, a block of k data digits is encoded by a code word of n digits ($n > k$). For each sequence of k data digits, there is a distinct code word of n digits. In **convolutional**, or **recurrent codes**, the coded sequence of n digits depends not only on the k data digits but also on the previous $N - 1$ data digits ($N > 1$). Hence, the coded sequence for a certain k data digits is not unique but depends on $N - 1$ earlier data digits. In block codes, k data digits are accumulated and then encoded into an n -digit code word. In convolutional codes, the coding is done on a continuous, or running, basis rather than by accumulating k data digits.

If k data digits are transmitted by a code word of n digits, the number of check digits is $m = n - k$. The **code efficiency** (also known as the **code rate**) is k/n . Such a code is known as an (n, k) code. Data digits $(d_1, d_2, \dots, d_k)$ are a k -tuple, and, hence, this is a k -dimensional vector $\mathbf{d}$. Similarly, a code word $(c_1, c_2, \dots, c_n)$ is an n -dimensional vector $\mathbf{c}$. As a preliminary, we shall determine the minimum number of check digits required to detect or correct t number of errors in an (n, k) code.

A total of 2^n code words (or vertices of an n -dimensional hypercube) is available to assign to 2^k data words. Suppose we wish to find a code that will correct up to t wrong digits. In this case, if we transmit a data word $\mathbf{d}_j$ by one of the code words (or vertices) $\mathbf{c}_j$, then because of channel errors the received word will not be $\mathbf{c}_j$ but will be $\mathbf{c}'_j$. If the channel noise

causes errors in t or less digits, then $\mathbf{c}'_j$ will lie somewhere in the Hamming sphere of radius t centered at $\mathbf{c}_j$. If the code is to correct up to t errors, then the code must have the property that all of the Hamming spheres of radius t centered at the code words are nonoverlapping. This means we may not use vertices (or words) that are within a Hamming distance of t from any code word. If a received word lies within a Hamming sphere of radius t centered at $\mathbf{c}_j$, then we decide that the transmitted code word was $\mathbf{c}_j$. This scheme is capable of correcting up to t errors, and $d_{\min}$, the minimum distance between t error correcting code words, is

$$d_{\min} = 2t + 1 \quad (16.1)$$

Next, in order to find a relationship between n and k , we observe that 2^n vertices, or words, are available for 2^k data words, and $2^n - 2^k$ are redundant vertices. How many vertices, or words, can lie within a Hamming sphere of radius t ? The number of sequences (of n digits) that differ from a given sequence by j digits is the number of the combination of n things taken j at a time and is given by $\binom{n}{j}$ [see Eq. (10.16)]. Hence, the number of ways in which up to t errors can occur is given by $\sum_{j=1}^t \binom{n}{j}$. Thus for each code word, we must leave $\sum_{j=1}^t \binom{n}{j}$ number of words unused. Because we have 2^k code words, we must leave $2^k \sum_{j=1}^t \binom{n}{j}$ words unused. Hence, the total number of words must be at least

$$2^k + 2^k \sum_{j=1}^t \binom{n}{j} = 2^k \sum_{j=0}^t \binom{n}{j}$$

But the total number of words, or vertices, available is 2^n . Hence,

$$2^n \geq 2^k \sum_{j=0}^t \binom{n}{j}$$

or

$$2^{n-k} \geq \sum_{j=0}^t \binom{n}{j} \quad (16.2a)$$

Observe that $n - k = m$ is the number of check digits. Hence, Eq. (16.2a) can be expressed as

$$2^m \geq \sum_{j=0}^t \binom{n}{j} \quad (16.2b)$$

This is known as the **Hamming bound**. It should also be remembered that the Hamming bound is a necessary but not a sufficient condition in general. However, for single-error correcting codes, it is a necessary and sufficient condition. If some m satisfies the Hamming bound, it does not necessarily mean that a t -error correcting code of n digits can be constructed. Table 16.1 shows some examples of error correction codes and their efficiencies.

A code for which the inequalities in Eqs. (16.2) become equalities is known as the **perfect code**. In such a code the Hamming spheres (about all the code words) are not only nonoverlapping but they exhaust all the 2^n vertices, leaving no vertex outside some sphere. An e -error correcting perfect code satisfies the condition that every possible sequence is at

Table 16.1
Some examples of error correcting codes

	n	k	Code	Code Efficiency (or Code Rate)
Single-error correcting, $t = 1$	3	1	(3, 1)	0.33
	4	1	(4, 1)	0.25
Minimum code separation 3	5	2	(5, 2)	0.4
	6	3	(6, 3)	0.5
	7	4	(7, 4)	0.57
	15	11	(15, 11)	0.73
	31	26	(31, 26)	0.838
Double-error correcting, $t = 2$	10	4	(10, 4)	0.4
	15	8	(15, 8)	0.533
Minimum code separation 5				
Triple-error correcting, $t = 3$	10	2	(10, 2)	0.2
	15	5	(15, 5)	0.33
Minimum code separation 7	23	12	(23, 12)	0.52

a distance at most e from some code word. Perfect codes exist in only a comparatively few cases. Binary, single-error correcting, perfect codes are called **Hamming codes**.

For a Hamming code, $t = 1$ and $d_{\min} = 3$, and from Eq. (16.2b) we have

$$2^m = \sum_{j=0}^1 \binom{n}{j} = 1 + n \tag{16.3}$$

and

$$n = 2^m - 1$$

Another way of correcting errors is to design a code to detect (not to correct) up to t errors. When the receiver detects an error, it requests retransmission. Because error detection requires fewer check digits, these codes operate at a higher efficiency.

To detect t errors, code words need to be separated by a Hamming distance of not more than $t + 1$. Suppose a transmitted code word c_j has α number of errors ($\alpha \leq t$). Then the received codeword c'_j is at a distance of α from c_j . Because $\alpha \leq t$, c'_j can never be any other valid code word, because all code words are separated by at least $t + 1$. Thus, the reception of c'_j immediately indicates that an error has been made.

Thus, the minimum distance $d_{\min}$ between t error detecting code words is

$$d_{\min} = t + 1$$

In presenting the theory, we shall use modulo-2 addition, defined in Chapter 7:

$$\begin{aligned} 1 \oplus 1 &= 0 \oplus 0 = 0 \\ 0 \oplus 1 &= 1 \oplus 0 = 1 \end{aligned}$$

Note that the modulo-2 sum of any binary digit with itself is always zero. All the additions in the mathematical development of binary codes presented henceforth are modulo-2.

16.2 LINEAR BLOCK CODES

A code word consists of n digits $c_1, c_2, \dots, c_n$, and a data word consists of k digits $d_1, d_2, \dots, d_k$. Because the code word and the data word are an n -tuple and a k -tuple, respectively, they are n - and k -dimensional vectors. We shall use row matrices to represent these words,

$$c = (c_1, c_2, \dots, c_n), \quad d = (d_1, d_2, \dots, d_k)$$

For the general case of linear block codes, all the n digits of c are formed by linear combinations (modulo-2 additions) of k data digits. A special case where $c_1 = d_1, c_2 = d_2, \dots, c_k = d_k$ and the remaining digits from c_{k+1} to c_n are linear combinations of $d_1, d_2, \dots, d_k$ is known as a **systematic code**.* Thus in a systematic code, the first k digits of a code word are the data digits and the last $m = n - k$ digits are the **parity-check digits**, formed by linear combinations of data digits $d_1, d_2, \dots, d_k$:

$$\begin{aligned} c_1 &= d_1 \\ c_2 &= d_2 \\ &\vdots \\ c_k &= d_k \\ c_{k+1} &= h_{11}d_1 \oplus h_{12}d_2 \oplus \dots \oplus h_{1k}d_k \\ c_{k+2} &= h_{21}d_1 \oplus h_{22}d_2 \oplus \dots \oplus h_{2k}d_k \\ &\vdots \\ c_n &= h_{m1}d_1 \oplus h_{m2}d_2 \oplus \dots \oplus h_{mk}d_k \end{aligned} \tag{16.4a}$$

or

$$c = dG \tag{16.4b}$$

where

$$G = \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & \dots & 0 & h_{11} & h_{21} & \dots & h_{m1} \\ 0 & 1 & 0 & \dots & 0 & h_{12} & h_{22} & \dots & h_{m2} \\ & & & \vdots & & & & \ddots & \\ 0 & 0 & 0 & \dots & 1 & h_{1k} & h_{2k} & \dots & h_{mk} \\ & & & \underbrace{I_k}_{(k \times k)} & & \underbrace{P}_{(k \times m)} & & & \end{array} \right] \tag{16.5}$$

The $k \times n$ matrix G is called the **generator matrix**, which can be partitioned into a $k \times k$ identity matrix I_k and a $k \times m$ matrix P . All the elements of P are either 0 or 1. The code word can be expressed as

$$\begin{aligned} c &= dG \\ &= d[I_k, P] \\ &= [d, dP] \\ &= [d, c_p] \end{aligned} \tag{16.6}$$

* It can be shown that the performance of systematic block codes is identical to that of nonsystematic block codes.

where

$$c_p = dP \tag{16.7}$$

Thus, knowing the data digits, we can calculate the check digits from Eq. (16.7).

EXAMPLE 16.1 For a (6, 3) code, the generator matrix *G* is

$$G = \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{array} \right]$$

For all eight possible data words, find the corresponding code words, and verify that this code is a single-error correcting code.

Table 16.2 shows the eight data words and the corresponding code words found from $c = dG$.

Table 16.2

Data Word <i>d</i>	Code Word <i>c</i>
111	111000
110	110110
101	101011
100	100101
011	011101
010	010011
001	001110
000	000000

Note that the distance between any two code words is at least 3. Hence, the code can correct at least one error. Figure 16.1 shows a possible encoder for this code, using a three-digit shift register and three modulo-2 adders.

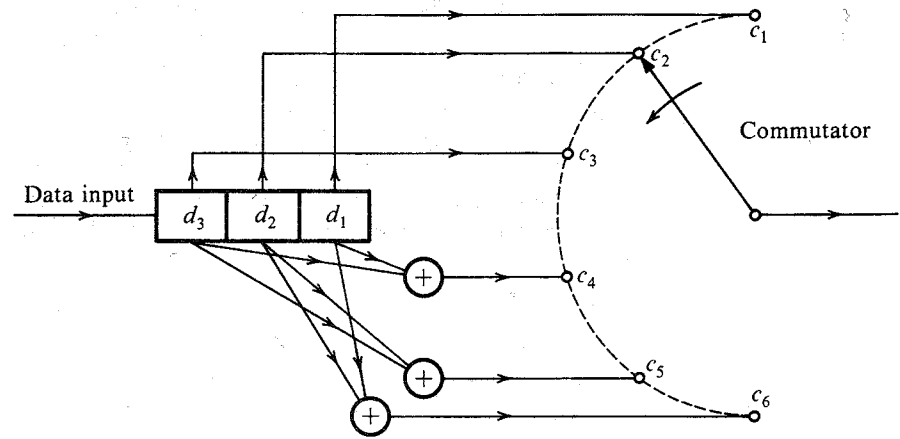


Figure 16.1 Encoder for linear block codes.

Decoding

Let us consider some code word properties that could be utilized for the purpose of decoding. From Eq. (16.7) and the fact that the modulo-2 sum of any sequence with itself is zero, we get

$$dP \oplus c_p = \underbrace{[d \ c_p]}_c \begin{bmatrix} P \\ I_m \end{bmatrix} = 0 \tag{16.8}$$

where I_m is the identity matrix of order $m \times m$ ($m = n - k$). Thus,

$$cH^T = 0 \tag{16.9a}$$

where

$$H^T = \begin{bmatrix} P \\ I_m \end{bmatrix} \tag{16.9b}$$

and its transpose

$$H = [P^T \ I_m] \tag{16.9c}$$

is called the **parity-check matrix**. Every code word must satisfy Eq. (16.9a). This is our clue to decoding. Consider the received word *r*. Because of possible errors caused by channel noise, *r* in general differs from the transmitted code word *c*,

$$r = c \oplus e$$

where the error word (or error vector) *e*, is also a row vector of *n* elements. For example, if the data word **100** in Example 16.1 is transmitted as a code word **100101** (see Table 16.2), and the channel noise causes a detection error in the third digit, then

$$r = 101101$$

$$c = 100101$$

and

$$e = 001000$$

Thus, an element **1** in *e* indicates an error in the corresponding position, and **0** indicates no error. The Hamming distance between *r* and *c* is simply the number of 1's in *e*.

Suppose the transmitted code word is *c_i* and the channel noise causes an error *e_i*, making the received word $r = c_i + e_i$. If there were no errors, that is, if *e_i* = **000000**, then $rH^T = 0$. But because of possible channel errors, rH^T is in general a nonzero row vector *s*, called the **syndrome**,

$$s = rH^T \tag{16.10a}$$

$$\begin{aligned} &= (c_i \oplus e_i)H^T \\ &= c_iH^T \oplus e_iH^T \\ &= e_iH^T \end{aligned} \tag{16.10b}$$

Knowing *r*, we can compute *s* [Eq. (16.10a)] and presumably we can compute *e_i* from

Eq. (16.10b). Unfortunately, knowledge of s does not allow us to solve uniquely for e_i . This is because r can also be expressed in terms of code words other than c_i . Thus,

$$r = c_j \oplus e_j \quad j \neq i$$

Hence,

$$s = (c_j \oplus e_j)H^T = e_j H^T$$

Because there are 2^k possible code words,

$$s = eH^T$$

is satisfied by 2^k error vectors. To give an example, if a data word $d = 100$ is transmitted by a code word **100101** in Example 16.1, and if a detection error is caused in the third digit, then the received word is **101101**. In this case we have $c = 100101$ and $e = 001000$. But the same word could have been received if $c = 101011$ and $e = 000110$, or if $c = 010011$ and $e = 111110$, and so on. Thus, there are eight possible error vectors (2^k error vectors) that satisfy Eq. (16.10b). Which vector shall we choose? For this, we must define our decision criterion. One reasonable criterion is the maximum-likelihood rule where, if we receive r , then we decide in favor of that c for which r is most likely to be received. In other words, we decide " c_i transmitted" if

$$P(r|c_i) > P(r|c_k) \quad \text{all } k \neq i$$

For a BSC, this rule gives a very simple answer. Suppose the Hamming distance between r and c_i is d , that is, the channel noise causes errors in d digits. Then if P_e is the digit error probability of a BSC,

$$P(r|c_i) = P_e^d (1 - P_e)^{n-d} = (1 - P_e)^n \left(\frac{P_e}{1 - P_e} \right)^d$$

If $P_e < 0.5$, then $P(r|c_i)$ is a monotonically decreasing function of d because $P_e/(1 - P_e) < 1$. Hence, to maximize $P(r|c_i)$, we must choose that c_i which is closest to r ; that is, we must choose the error vector with the smallest number of 1's. A vector with the smallest number of 1's is called the **minimum-weight vector**.

EXAMPLE 16.2 A (6, 3) code is generated according to the generating matrix in Example 16.1. The receiver receives $r = 100011$. Determine the corresponding data word if the channel is a BSC and the maximum-likelihood decision is used.

We have

$$\begin{aligned} s &= rH^T \\ &= [1 \ 0 \ 0 \ 0 \ 1 \ 1] \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= [1 \ 1 \ 0] \end{aligned}$$

Because for modulo-2 operation, subtraction is the same as addition, the correct transmitted code word c is given by

$$c = r \oplus e$$

where e satisfies

$$\begin{aligned} s &= [1 \ 1 \ 0] = eH^T \\ &= [e_1 \ e_2 \ e_3 \ e_4 \ e_5 \ e_6] \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned}$$

We see that $e = 001000$ satisfies this equation. But so does $e = 000110$, or **010101**, or **011011**, or **111110**, or **110000**, or **101101**, or **100011**. The suitable choice, the minimum-weight e , is **001000**. Hence,

$$c = 100011 \oplus 001000 = 101011$$

The decoding procedure just described is quite disorganized. A systematic procedure would be to consider all possible syndromes and for each syndrome associate a minimum-weight error vector. For instance, the single-error correcting code in Example 16.1 has a syndrome with three digits. Hence, there are eight possible syndromes. We prepare a table of minimum-weight error vectors corresponding to each syndrome (see Table 16.3). This table can be prepared by considering all possible minimum-weight error vectors and computing s for each of them using Eq. (16.10b). The first minimum-weight error vector **000000** is a trivial case that has the syndrome **000**. Next, we consider all possible unit weight error vectors. There are six such vectors: **100000**, **010000**, **001000**, **000100**, **000010**, **000001**. Syndromes for these can readily be calculated from Eq. (16.10b) and tabulated (Table 16.3). This still leaves one syndrome, **111**, that is not matched with some error vector. Since all unit-weight error vectors are exhausted, we must look for error vectors of weight 2.

We find that for the first seven syndromes (Table 16.3), there is a unique minimum-weight vector e . But for $s = 111$, the error vector e has a minimum weight of 2, and it is not unique. For example, $e = 100010$ or **010100** or **001001** all have $s = 111$, and all three e 's are minimum weight (weight 2). In such a case, we can pick any one of these e 's as a **correctable** error pattern. In Table 16.3, we have picked $e = 100010$ as the double-error correctable pattern. This means the present code can correct all six single-error patterns and one double-error pattern (**100010**). For instance, if $c = 101011$ is transmitted and the channel noise causes the double error **100010**, the received vector $r = 001001$, and

$$s = rH^T = [1 \ 1 \ 1]$$

From Table 16.3 we see that corresponding to $s = 111$ is $e = 100010$, and we immediately decide $c = r \oplus e = 101011$. Note, however, that this code will not correct double-error patterns other than **100010**. Thus, this code not only corrects all single errors but one double-error pattern as well. This extra bonus of one double-error correction occurs because n and

Table 16.3
Decoding table for code in Table 16.2

<i>e</i>	<i>s</i>
000000	000
100000	101
010000	011
001000	110
000100	100
000010	010
000001	001
100010	111

k oversatisfy the Hamming bound [Eq. (16.2b)]. In case *n* and *k* were to satisfy the bound exactly, we would have only single-error correction ability. This is the case for the (7, 4) code, which can correct all single-error patterns only.

Thus for systematic decoding, we prepare a table of all correctable error patterns and the corresponding syndromes. For decoding, we need only calculate $s = rH^T$ and, from the decoding table, find the corresponding *e*. The decision is $c = r \oplus e$.

Because *s* has $m = n - k$ digits, there is a total of 2^{n-k} syndromes, each of $n - k$ digits. There is the same number of correctable error vectors, each of *n* digits. Hence, for the purpose of decoding, we need a storage of $(2n - k)2^{n-k} = (2n - k)2^m$. This storage requirement grows exponentially with *m*, the number of parity-check digits, and can be enormous, even for moderately complex codes.

Because the maximum-likelihood decision is the same as choosing the code word closest to the received word, we could just as well compare the received word with each of the 2^k possible code words of *n* digits each. This involves a storage of $n2^k$, which can be much larger than the $(2n - k)2^m$ storage required earlier.

It is still not clear how to choose coefficients of the generator or parity-check matrix. Unfortunately, there is no systematic way to do this, except for the case of single-error correcting codes, also known as *Hamming codes*. Let us consider a single-error correcting (7, 4) code. This code satisfies the Hamming bound exactly, and we shall see that a proper code can be constructed. In this case $m = 3$, and there are seven nonzero syndromes, and because $n = 7$, there are exactly seven single-error patterns. Hence, we can correct all single-error patterns and no more. Consider the single-error pattern $e = 1000000$. Because

$$s = eH^T$$

eH^T will be simply the first row of H^T . Similarly, for $e = 0100000$, $s = eH^T$ will be the second row of H^T , and so on. Now for unique decodability, we require that all seven syndromes corresponding to the seven single-error patterns be distinct. Conversely, if all the seven syndromes are distinct, we can decode all the single-error patterns. This means that the only requirement on H^T is that all seven of its rows be distinct and nonzero. Note that H^T is an $(n \times n - k)$ matrix (i.e., 7×3 in this case). Because there exist seven nonzero patterns of three digits, it is possible to find seven nonzero rows of three digits each. There are many ways in which these rows can be ordered. But we must remember that the three bottom rows must form identity matrix I_m [see Eq. (16.9b)].

One possible form of H^T is

$$H^T = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} P \\ I_m \end{bmatrix}$$

The corresponding generator matrix *G* is

$$G = [I_k \quad P] = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

Thus when $d = 1011$, the corresponding code word $c = 1011001$, and so forth.

A general (n, k) code has *m*-dimensional syndrome vectors ($m = n - k$). Hence, there are $2^m - 1$ distinct nonzero syndrome vectors that can correct $2^m - 1$ single-error patterns. Because in an (n, k) code there are exactly *n* single-error patterns, all these patterns can be corrected if

$$2^m - 1 \geq n$$

or

$$2^{n-k} \geq n + 1$$

This is precisely the condition in Eq. (16.3) for $t = 1$. Thus, for any (n, k) satisfying this condition, it is possible to construct a single-error correcting code by the procedure discussed.

More discussion on block coding can be found in Peterson and Weldon¹ as well as in Lin and Costello.²

16.3 CYCLIC CODES^{1,2}

Cyclic codes are a subclass of linear block codes. As seen before, a procedure for selecting a generator matrix is relatively easy for single-error correcting codes. This procedure, however, cannot carry us very far in constructing higher order error correcting codes. Cyclic codes have a fair amount of mathematical structure that permits the design of higher order correcting codes. Second, for cyclic codes, encoding and syndrome calculations can be easily implemented using simple shift registers.

Cyclic codes are such that code words are simple lateral shifts of one another. For example, if $c = (c_1, c_2, \dots, c_{n-1}, c_n)$ is a code word, then so are $(c_2, c_3, \dots, c_n, c_1)$ and $(c_3, c_4, \dots, c_n, c_1, c_2)$, and so on. We shall use the following notation. If

$$c = (c_1, c_2, \dots, c_n) \tag{16.11a}$$

is a code vector of a code *C*, then $c^{(i)}$ denotes *c* shifted cyclically *i* places to the left, that is,

$$c^{(i)} = (c_{i+1}, c_{i+2}, \dots, c_n, c_1, c_2, \dots, c_i) \tag{16.11b}$$

Cyclic codes can be described in a polynomial form. This property is extremely useful in the analysis and implementation of these codes. The code vector c in Eq. (16.11a) can be expressed as the $(n - 1)$ -degree polynomial

$$c(x) = c_1x^{n-1} + c_2x^{n-2} + \cdots + c_n \tag{16.12a}$$

The coefficients of the polynomial are either 0 or 1, and they obey the following properties:

$$\begin{aligned} 0 + 0 &= 0 & 0 \times 0 &= 0 \\ 0 + 1 &= 1 + 0 = 1 & 0 \times 1 &= 1 \times 0 = 0 \\ 1 + 1 &= 0 & 1 \times 1 &= 1 \end{aligned}$$

The code polynomial $c^{(i)}(x)$ for the code vector $c^{(i)}$ in Eq. (16.11b) is

$$c^{(i)}(x) = c_{i+1}x^{n-1} + c_{i+2}x^{n-2} + \cdots + c_nx^i + c_1x^{i-1} + \cdots + c_i \tag{16.12b}$$

One of the interesting properties of the code polynomials is that when $x^i c(x)$ is divided by $x^n + 1$, the remainder is $c^{(i)}(x)$. We can verify this property as follows:

$$\begin{array}{r} xc(x) = c_1x^n + c_2x^{n-1} + \cdots + c_nx \\ \underline{c_1} \\ x^n + 1 \overline{) c_1x^n + c_2x^{n-1} + \cdots + c_nx} \\ \phantom{x^n + 1 \overline{) }} \underline{c_1x^n + c_1} \\ \phantom{x^n + 1 \overline{) }} c_2x^{n-1} + c_3x^{n-2} + \cdots + c_nx + c_1 \leftarrow \text{remainder} \end{array}$$

The remainder is clearly $c^{(1)}(x)$. In deriving this result, we have used the fact that subtraction amounts to summation when modulo-2 operations are involved. Continuing in this fashion, we can show that the remainder of $x^i c(x)$ divided by $x^n + 1$ is $c^{(i)}(x)$.

We shall now prove an important theorem in cyclic codes. It says that a cyclic code polynomial $c(x)$ can be generated by the data polynomial $d(x)$ of degree $k - 1$ and a generator polynomial $g(x)$ of degree $n - k$ as

$$c(x) = d(x) g(x) \tag{16.13}$$

where the generator polynomial $g(x)$ is an $(n - k)$ th-order factor of $(x^n + 1)$.

For a data vector $(d_1, d_2, \dots, d_k)$, the data polynomial is

$$d(x) = d_1x^{k-1} + d_2x^{k-2} + \cdots + d_k \tag{16.14}$$

Proof: Consider a polynomial

$$\begin{aligned} c(x) &= d(x) g(x) \\ &= d_1x^{k-1}g(x) + d_2x^{k-2}g(x) + \cdots + d_kg(x) \end{aligned} \tag{16.15}$$

This is a polynomial of degree $n - 1$ or less. There are a total of 2^k such polynomials corresponding to 2^k data vectors. Thus, we have a linear (n, k) code generated by Eq. (16.13). To prove that this code is cyclic, let

$$c(x) = c_1x^{n-1} + c_2x^{n-2} + \cdots + c_n$$

be a code polynomial in this code [Eq. (16.15)]. Then,

$$\begin{aligned} xc(x) &= c_1x^n + c_2x^{n-1} + \cdots + c_nx \\ &= c_1(x^n + 1) + (c_2x^{n-1} + c_3x^{n-2} + \cdots + c_nx + c_1) \\ &= c_1(x^n + 1) + c^{(1)}(x) \end{aligned}$$

Because $xc(x)$ is $xd(x)g(x)$, and $g(x)$ is a factor of $x^n + 1$, $c^{(1)}(x)$ must also be a multiple of $g(x)$ and can also be expressed as $d(x)g(x)$ for some data vector d . Therefore, $c^{(1)}(x)$ is also a code polynomial. Continuing this way, we see that $c^{(2)}(x)$, $c^{(3)}(x)$, ... are all code polynomials generated by Eq. (16.15). Thus, the linear (n, k) code generated by $d(x)g(x)$ is indeed cyclic.

EXAMPLE 16.3 Find a generator polynomial $g(x)$ for a $(7, 4)$ cyclic code, and find code vectors for the following data vectors: **1010**, **1111**, **0001**, and **1000**.

In this case $n = 7$ and $n - k = 3$, and

$$x^7 + 1 = (x + 1)(x^3 + x + 1)(x^3 + x^2 + 1)$$

For a $(7, 4)$ code, the generator polynomial must be of the order $n - k = 3$. In this case, there are two possible choices for $g(x)$: $x^3 + x + 1$ or $x^3 + x^2 + 1$. Let us choose the latter, that is,

$$g(x) = x^3 + x^2 + 1$$

as a possible generator polynomial. For $d = [1 \ 0 \ 1 \ 0]$,

$$d(x) = x^3 + x$$

and the code polynomial is

$$\begin{aligned} c(x) &= d(x)g(x) \\ &= (x^3 + x)(x^3 + x^2 + 1) \\ &= x^6 + x^5 + x^4 + x \end{aligned}$$

Hence,

$$c = \mathbf{1110010}$$

Similarly, code words for other data words can be found (see Table 16.4).

Table 16.4

d	e
1010	1110010
1111	1001011
0001	0001101
1000	1101000

Note the structure of the code words. The first k digits are not necessarily the data digits. Hence, this is not a systematic code.

In a systematic code, the first k digits are data digits, and the last $m = n - k$ digits are the parity-check digits. Systematic codes are a special case of general codes. Our discussion thus far applies to general cyclic codes, of which systematic cyclic codes are a special case. We shall now develop a method of generating systematic cyclic codes.

Systematic Cyclic Codes

We shall show that for a systematic code, the code word polynomial $c(x)$ corresponding to the data polynomial $d(x)$ is given by

$$c(x) = x^{n-k}d(x) + \rho(x) \quad (16.16a)$$

where $\rho(x)$ is the remainder from dividing $x^{n-k}d(x)$ by $g(x)$,

$$\rho(x) = \text{Rem} \frac{x^{n-k}d(x)}{g(x)} \quad (16.16b)$$

To prove this we observe that

$$\frac{x^{n-k}d(x)}{g(x)} = q(x) + \frac{\rho(x)}{g(x)} \quad (16.17a)$$

where $q(x)$ is of degree $k - 1$ or less. We add $\rho(x)/g(x)$ to both sides of Eq. (16.17a), and because $f(x) + f(x) = 0$ under modulo-2 operation, we have

$$\frac{x^{n-k}d(x) + \rho(x)}{g(x)} = q(x) \quad (16.17b)$$

or

$$q(x)g(x) = x^{n-k}d(x) + \rho(x) \quad (16.17c)$$

Because $q(x)$ is on the order of $k - 1$ or less, $q(x)g(x)$ is a code polynomial. As $x^{n-k}d(x)$ represents $d(x)$ shifted to the left by $n - k$ digits, the first k digits of this code word are precisely d , and the last $n - k$ digits corresponding to $\rho(x)$ must be parity-check digits. This will become clear by considering a specific example.

EXAMPLE 16.4 Construct a systematic (7, 4) cyclic code using a generator polynomial (see Example 16.3).

We use

$$g(x) = x^3 + x^2 + 1$$

Consider a data vector $d = 1010$,

$$d(x) = x^3 + x$$

and

$$x^{n-k}d(x) = x^6 + x^4$$

Hence,

$$\begin{array}{r} x^3 + x^2 + 1 \overline{) x^6 + x^4} \quad \leftarrow q(x) \\ \underline{x^6 + x^5 + x^3} \\ x^5 + x^4 + x^3 \\ \underline{x^5 + x^4 + x^2} \\ x^3 + x^2 \\ \underline{x^3 + x^2 + 1} \\ 1 \quad \leftarrow \rho(x) \end{array}$$

Hence, from Eq. (16.16a),

$$\begin{aligned} c(x) &= x^3d(x) + \rho(x) \\ &= x^3(x^3 + x) + 1 \\ &= x^6 + x^4 + 1 \end{aligned}$$

and

$$c = 1010001$$

We could also have found the code word c directly by using Eq. (16.17c). Thus, $c(x) = q(x)g(x) = (x^3 + x^2 + 1)(x^3 + x^2 + 1) = x^6 + x^4 + 1$. We construct the entire code table in this manner (Table 16.5). This is quite a laborious procedure. There is, however, a shortcut, using the code generating matrix G . Using the earlier procedure, we compute the code words corresponding to the data words **1000**, **0100**, **0010**, **0001**. These are **1000110**, **0100011**, **0010111**, **0001101**. Now recognize that these four code words are the four rows of G . This is because $c = dG$, and when $d = 1000$, dG is the first row of G , and so on. Hence,

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}$$

Now, we can construct the rest of the code table using $c = dG$. This is an efficient method because it allows us to construct the entire code table from the knowledge of only n code words.

Table 16.5 shows the complete code. Note that $d_{\min}$, the minimum distance between two code words, is 3. Hence, this is a single-error correcting code, and 14 of these code words can be obtained by successive cyclic shifts of the two code words **1110010** and **1101000**. The remaining two code words **1111111** and **0000000** remain unchanged under cyclic shift.

Cyclic codes can also be described by a generator matrix G (see Probs. 16.3-6 and 16.3-7). It can be shown that Hamming codes are cyclic codes.

Table 16.5

<i>d</i>	<i>c</i>
1111	1111111
1110	1110010
1101	1101000
1100	1100101
1011	1011100
1010	1010001
1001	1001011
1000	1000110
0111	0111001
0110	0110100
0101	0101110
0100	0100011
0011	0011010
0010	0010111
0001	0001101
0000	0000000

Cyclic Code Generation

One of the advantages of cyclic codes is that coding and decoding can be implemented using such simple elements as shift registers and modulo-2 adders. A systematic code generation is described in Eqs. (16.16). This involves division of $x^{n-k}d(x)$ by $g(x)$ and can be implemented by a dividing circuit, which is a shift register with feedback connections according to the generator polynomial* $g(x) = x^{n-k} + g_1x^{n-k-1} + \dots + g_{n-k-1}x + 1$. The gain g_k 's are either 0 or 1. An encoding circuit with $n - k$ shift registers is shown in Fig. 16.2. An understanding of this dividing circuit requires some background in linear sequential networks. An explanation of its functioning can be found in Peterson and Weldon.¹ The k data digits are shifted in one at a time at the input with the switch s held at position p_1 . The symbol D represents a one-digit delay. As the data digits move through the encoder, they are also shifted out onto the output line, because the first k digits of the code word are the data digits themselves. As soon as the last (or k th) data digit clears the last [($n - k$)th] register, all the registers contain the parity-check digits. The switch s is now thrown to position p_2 , and the $n - k$ parity-check digits are shifted out one at a time onto the line.

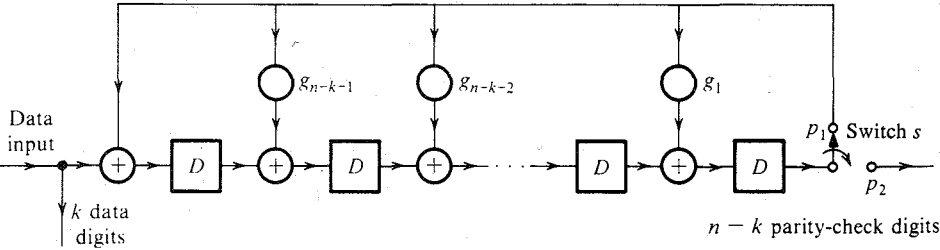


Figure 16.2 Encoder for systematic cyclic code.

* It can be shown that for cyclic codes, the generator polynomial must be of this form.

Decoding

Every valid code polynomial $c(x)$ is a multiple of $g(x)$. If an error occurs during the transmission, the received word polynomial $r(x)$ will not be a multiple* of $g(x)$. Thus,

$$\frac{r(x)}{g(x)} = m_1(x) + \frac{s(x)}{g(x)} \tag{16.18}$$

and

$$s(x) = \text{Rem} \frac{r(x)}{g(x)} \tag{16.19}$$

where the syndrome polynomial $s(x)$ has a degree $n - k - 1$ or less.

If $e(x)$ is the error polynomial, then

$$r(x) = c(x) + e(x)$$

Remembering that $c(x)$ is a multiple of $g(x)$,

$$\begin{aligned} s(x) &= \text{Rem} \frac{r(x)}{g(x)} \\ &= \text{Rem} \frac{c(x) + e(x)}{g(x)} \\ &= \text{Rem} \frac{e(x)}{g(x)} \end{aligned} \tag{16.20}$$

Again, as before, a received word r could result from any one of the 2^k code words and a suitable error. For example, for the code in Table 16.5, if $r = 0110010$, this could mean $c = 1110010$ and $e = 1000000$, or $c = 1101000$ and $e = 1011010$, or 14 more such combinations. As seen earlier, the most likely error pattern is the one with the minimum weight (or minimum number of 1's). Hence, here $c = 1110010$ and $e = 1000000$ is the correct decision.

It is convenient to prepare a decoding table, that is, to list the syndromes for all correctable errors. For any r , we compute the syndrome from Eq. (16.19), and from the table we find the corresponding correctable error e . Then we determine $c = r \oplus e$. Note that computation of $s(x)$ [see Eq. (16.18)] involves exactly the same operation as that required to compute $\rho(x)$ in coding [see Eq. (16.17a)]. Hence, the circuit in Fig. 16.2 can also be used to compute $s(x)$.

EXAMPLE 16.5 Construct the decoding table for the single-error correcting (7, 4) code in Table 16.5. Determine the data vectors transmitted for the following received vectors r : (a) 1101101; (b) 0101000; (c) 0001100.

The first step is to construct the decoding table. Because $n - k - 1 = 2$, the syndrome polynomial is of the second order, and there are seven possible nonzero syndromes. There are also seven possible correctable single-error patterns because $n = 7$. Using Eq. (16.20), we compute the syndrome for each of the seven correctable error patterns. For example, for $e = 1000000$, $e(x) = x^6$. Because $g(x) = x^3 + x^2 + 1$ for this code (see Example 16.4), we have

* This assumes that the number of errors in r is correctable.

$$\begin{array}{r} x^3 + x^2 + x \\ x^3 + x^2 + 1 \overline{) x^6} \\ \underline{x^6 + x^5 + x^3} \\ x^5 + x^3 \\ \underline{x^5 + x^4 + x^2} \\ x^4 + x^3 + x^2 \\ \underline{x^4 + x^3 + x} \\ x^2 + x \leftarrow s(x) \end{array}$$

Hence,

$s = 110$

In a similar way, we compute the syndromes for the remaining error patterns (see Table 16.6).

Table 16.6

<i>e</i>	<i>s</i>
1000000	110
0100000	011
0010000	111
0001000	101
0000100	100
0000010	010
0000001	001

When the received word *r* is 1101101,

$r(x) = x^6 + x^5 + x^3 + x^2 + 1$

We now compute *s*(*x*) according to Eq. (16.19):

$$\begin{array}{r} x^3 \\ x^3 + x^2 + 1 \overline{) x^6 + x^5 + x^3 + x^2 + 1} \\ \underline{x^6 + x^5 + x^3} \\ x^2 + 1 \end{array}$$

Hence, *s* = 101. From Table 16.6, this gives *e* = 0001000, and

$c = r \oplus e = 1101101 \oplus 0001000 = 1100101$

Hence, from Table 16.5 we have

$d = 1100$

In a similar way, we determine for *r* = 0101000, *s* = 110 and *e* = 1000000; hence *c* = *r* ⊕ *e* = 1101000, and *d* = 1101. For *r* = 0001100, *s* = 001 and *e* = 0000001; hence *c* = *r* ⊕ *e* = 0001101, and *d* = 0001.

Bose-Chaudhuri-Hocquenghen (BCH) Codes

The BCH codes are perhaps the most powerful of the random-error correcting cyclic codes. Moreover, their decoding procedure can be implemented simply. Hamming code is a special case of BCH codes. These codes are described as follows: For any positive integers *m*' and *t* (*t* < 2^{*m*'} − 1), there exists a *t*-error correcting (*n*, *k*) code with *n* = 2^{*m*'} − 1 and *n* − *k* ≤ *m*'*t*. The minimum distance *d*_{min} between code words is related by the inequality 2*t* + 1 ≤ *d*_{min} ≤ 2*t* + 2.

The detailed treatment of BCH codes requires extensive use of modern algebra and is beyond our scope. For further discussion of BCH codes, the reader is referred to Lin and Costello² or Peterson and Weldon.¹

16.4 BURST-ERROR DETECTING AND CORRECTING CODES

Thus far we have considered detecting or correcting errors that occur independently, or randomly, in digit positions. On some channels, disturbances can wipe out an entire block of digits. For instance, a stroke of lightning or a human-made electrical disturbance can affect several adjacent transmitted digits. On magnetic storage systems, magnetic tape defects usually affect more than one digit. Burst errors are those errors that wipe out some or all of a sequential set of digits. In general, random-error correcting codes are not efficient for correcting burst errors. Hence, special **burst-error correcting codes** are used for this purpose.

A burst of length *b* is defined as a sequence of digits in which the first digit and the *b*th digit are in error, with the *b* − 2 digits in between either in error or received correctly. For example, an error vector *e* = 0010010100 has a burst length of 6.

It can be shown that for detecting all burst errors of length *b* or less with a linear block code of length *n*, *b* parity-check bits are necessary and sufficient.¹ We shall prove the sufficiency part of this theorem by constructing a code of length *n* with *b* parity-check digits that will detect a burst of length *b*.

To construct such a code, let us group *k* data digits into segments of *b* digits in length (Fig. 16.3). To this we add a last segment of *b* parity-check digits, which are determined as follows. The modulo-2 sum of the *i*th digit in each segment (including the parity-check segment) must be zero. For example, the first digits in the five data segments are 1, 0, 1, 1, and 1. Hence, we must have 0 as the first parity-check digit in order to obtain a modulo-2 sum zero. We continue in this way with the second digit, the third digit, and so on, to the *b*th digit. Because parity-check digits are a linear combination of data digits, this is a linear block code. Moreover, it is a systematic code.

It is easy to see that if a digit sequence of length *b* or less is in error, parity will be violated and the error will be detected (but not corrected), and the receiver can request retransmission of the digits lost. One of the interesting properties of this code is that *b*, the number of parity-check digits, is independent of *k* (or *n*), which makes it a very useful code for such systems as

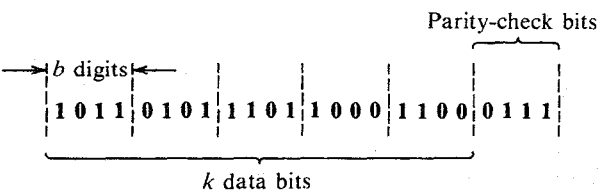


Figure 16.3 Burst-error detection.

packet switching, where the data digits may vary from packet to packet. It can be shown that a linear code with b parity bits detects not only all bursts of length b or less, but also a high percentage of longer bursts.¹

If we are interested in correcting rather than detecting burst errors, we require twice as many parity-check digits. A theorem says: In order to correct all burst errors of length b or less, a linear block code must have at least $2b$ parity-check digits.¹

16.5 INTERLACED CODES FOR BURST- AND RANDOM-ERROR CORRECTION

In general, random-error correcting codes are not efficient for burst-error correcting, and burst-error correcting codes are not efficient for random-error correcting. Unfortunately, in most practical systems, we have errors of both kinds. Out of the several methods proposed to simultaneously correct random and burst errors, the interlaced code is the most effective.

For an (n, k) code, if we interlace λ code words, we have what is known as a $(\lambda n, \lambda k)$ **interlaced code**. Instead of transmitting code words one by one, we group λ code words and interlace them. Consider, for example, the case of $\lambda = 3$ and a two-error correcting $(15, 8)$ code. Each code word has 15 digits. We group code words to be transmitted in groups of three. Suppose the first three code words to be transmitted are $x(x_1, x_2, \dots, x_{15})$, $y(y_1, y_2, \dots, y_{15})$, and $z(z_1, z_2, \dots, z_{15})$, respectively. Then instead of transmitting xyz in sequence as $x_1, x_2, \dots, x_{15}, y_1, y_2, \dots, y_{15}, z_1, z_2, \dots, z_{15}$, we transmit $x_1, y_1, z_1, x_2, y_2, z_2, x_3, y_3, z_3, \dots, x_{15}, y_{15}, z_{15}$. This can be explained graphically by Fig. 16.4, where λ code words (three in this case) are arranged in rows. In usual transmission, we transmit one row after another. In the interlaced case, we transmit columns (of λ elements) in sequence. When all the 15 (n) columns are transmitted, we repeat the procedure for the next λ code words to be transmitted.

To explain the error correcting capabilities of this code, we observe that the decoder will first remove the interlacing and regroup the received digits as $x_1, x_2, \dots, x_{15}, y_1, y_2, \dots, y_{15}, z_1, z_2, \dots, z_{15}$. Suppose the shaded digits in Fig. 16.4 were in error. Because the code is a two-error correcting code, two or less errors in each row will be corrected. Hence, all the errors in Fig. 16.4 are correctable. We see that there are two random, or independent, errors and one burst of length 4 in all the 45 digits transmitted. In general, if the original (n, k) code is t -error correcting, the interlaced code can correct any combination of t bursts of length λ or less.

x_1	x_2	x_3	$\dots$	x_{14}	x_{15}
y_1	y_2	y_3	$\dots$	y_{14}	y_{15}
z_1	z_2	z_3	$\dots$	z_{14}	z_{15}

Figure 16.4 Random- and burst-error correction.

16.6 CONVOLUTIONAL CODES

Convolutional (or recurrent) codes, first introduced in 1955 by Elias,³ differ from block codes as follows. In a block code, the block of n code digits generated by the encoder in any particular time unit depends only on the block of k input data digits within that time unit. In a convolutional code, on the other hand, the block of n code digits generated by the encoder in a particular time unit depends not only on the block of k message digits within that time unit but also on the block of data digits within a previous span of $N - 1$ time units ($N > 1$). For convolutional codes, k and n are usually small. Convolutional codes can be devised for correcting random errors, burst errors, or both. Encoding is easily implemented by shift registers. As a class, convolutional codes invariably outperform block codes of the same order of complexity.

A convolutional coder with constraint length N consists of an N -stage shift register and v modulo-2 adders. Figure 16.5 shows such a coder for the case of $N = 3$ and $v = 2$. The message digits are applied at the input of the shift register. The coded digit stream is obtained at the commutator output. The commutator samples the v modulo-2 adders in sequence, once during each input-bit interval. We shall explain this operation with reference to the input digits **11010**. Initially, all the stages of the register are clear; that is, they are in a **0** state. When the first data digit **1** enters the register, the stage s_1 shows **1** and all the other stages (s_2 and s_3) are unchanged; that is, they are in a **0** state. The two modulo-2 adders show $v_1 = 1$ and $v_2 = 1$. The commutator samples this output. Hence, the coder output is **11**. When the second message bit **1** enters the register, it enters the stage s_1 , and the previous **1** in s_1 is shifted to s_2 . Hence, s_1 and s_2 both show **1**, and s_3 is still unchanged; that is, it is in a **0** state. The modulo-2 adders now show $v_1 = 0$ and $v_2 = 1$. Hence, the decoder output is **01**. In the same way, when the third message digit **0** enters the register, we have $s_1 = 0$, $s_2 = 1$, and $s_3 = 1$, and the decoder output is **01**. Observe that each data digit influences N groups of v digits in the output (in this case three groups of two digits). The process continues until the last data digit enters the stage s_1 .^{*} We cannot stop here, however. We add $N - 1$ number of **0**'s to the input stream (dummy or augmented data) to make sure that the last data digit (**0** in this case) proceeds all the way through the shift register in order to influence the N groups of v digits. Hence, when the input digits are **11010**, we actually apply **1101000** (the digits augmented by $N - 1$ zeros) to the

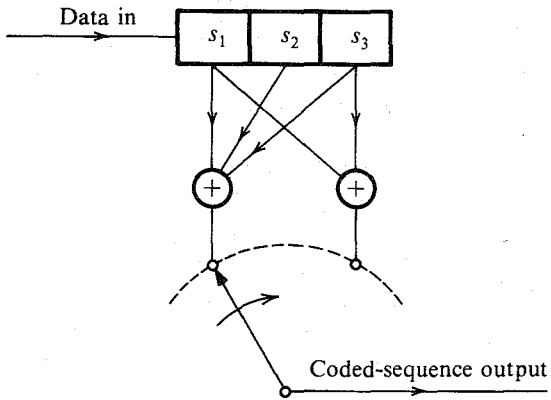


Figure 16.5 Convolutional coder.

* For a systematic code, one of the output digits must be the data digit itself.

input of the shift register. It can be seen that when the last digit of the augmented message stream enters s_1 , the last digit of the message stream has passed through all the N stages of the register. The reader can verify that the coder output is given by **11 01 01 00 10 11 00**. Thus, there are in all $n = (N + k - 1)v$ digits in the coded output for every k data digits. In practice, $k \gg N$, and, hence, there are approximately kv coded output digits for every k data digits, giving an efficiency $\eta \simeq 1/v$.*

It can be seen that unlike the block coder, the convolutional coder operates on a continuous basis, and each data digit influences N groups of v digits in the output.

Code Tree

Coding and decoding is considerably facilitated by what is known as the **code tree**, which shows the coded output for any possible sequence of data digits. The code tree for the coder in Fig. 16.5 with $k = 5$ is shown in Fig. 16.6. When the first digit is **0**, the coder output is **00**, and when it is **1**, the output is **11**. This is shown by the two tree branches that start at the initial node. The upper branch represents **0**, and the lower branch represents **1**. This convention will be followed throughout. At the terminal node of each of the two branches, we follow a similar procedure, corresponding to the second data digit. Hence, two branches initiate from each node, the upper one for **0** and the lower one for **1**. This continues until the k th data digit. From there on, all the input digits are **0** (augmented digit), and we have only one branch until the end. Hence, in all there are 32 (or 2^k) outputs corresponding to 2^k possible data vectors. The coded output for input **11010** can be easily read from this tree (the path shown dashed in Fig. 16.6).

Figure 16.6 shows that the code tree becomes repetitive after the third branch. This can be seen from the fact that the two blocks enclosed inside the dashed lines are identical. It means that the output from the fourth input digit is the same whether the first digit was **1** or **0**. This is not surprising in view of the fact that when the fourth input digit enters the shift register, the first input digit is shifted out of the register, and it ceases to influence the output digits. In other words, the data vector $1x_1x_2x_3x_4 \dots$ and the data vector $0x_1x_2x_3x_4 \dots$ generate the same output after the third group of output digits. It is convenient to label the four third-level nodes (the nodes appearing at the beginning of the third branch) as nodes a, b, c , and d (Fig. 16.6). The repetitive structure begins at the fourth-level nodes and continues at the fifth-level nodes, whose behavior is similar to that of nodes a, b, c , and d at the third level. Hence, we label the fourth- and fifth-level nodes also as either a, b, c , or d . What this means is that at the fifth-level nodes, the first two data digits have become irrelevant; that is, any of the four combinations (**11**, **10**, **01**, or **00**) for the first two data digits will give the same output after the fifth node.

This behavior can be seen from another point of view. When a data bit enters the shift register (in stage s_1), the output bits are determined not only by the data bit in s_1 , but by the two previous data bits already in stages s_3 and s_2 . There are four possible combinations of the two previous bits (in s_3 and s_2): **00**, **01**, **10**, and **11**. We shall label these four states as a, b, c , and d , respectively, as shown in Fig. 16.7a. Thus, when the previous two bits are **01** ($s_3 = 0, s_2 = 1$), the state is b , and so on. The number of states is equal to 2^{N-1} .

A data bit **0** or **1** generates four different outputs, depending on the encoder state. If the data bit is **0**, the encoder output is **00**, **10**, **11**, or **01**, depending on whether the encoder state is

* In general, instead of shifting one digit at a time, b digits may be shifted at a time. In this case $\eta \simeq b/v$.

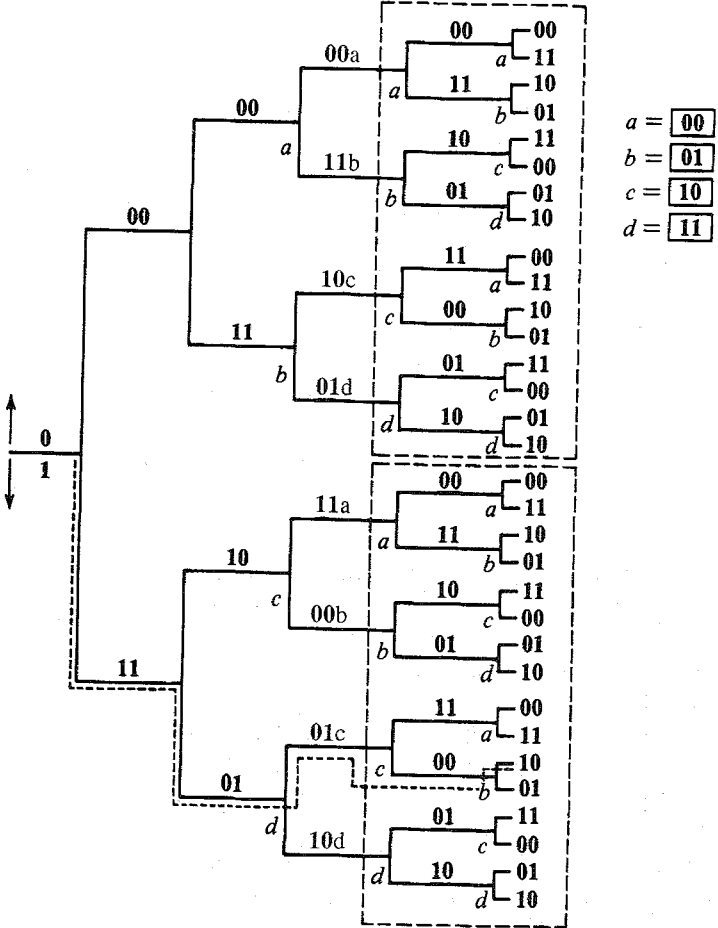


Figure 16.6 Code tree for the coder in Fig. 16.5.

a, b, c , or d . Similarly if the data bit is **1**, the encoder output is **11**, **01**, **00**, or **10**, depending on whether the encoder state is a, b, c , or d . This entire behavior can be concisely expressed by the state diagram shown in Fig. 16.7b. This is a four-state directed graph used to uniquely represent the input-output relation of this encoder. We use solid lines when the input bit is **0**, and dashed lines when the input bit is **1**. For instance, when the encoder is in state a , and we input **1**, the encoder output is **11** (dashed line). The encoder now goes to state b for the next data bit because at this point the previous two bits become $s_3 = 0$ and $s_2 = 1$. Similarly, when the encoder is in state a and the input is **0**, the output is **00** (solid line), and the encoder remains in state a . Note that the encoder cannot go directly from state a to states c or d . From any given state, the encoder can go to only two states directly by inputting a single data bit. This is an extremely important observation, which will be used later. The encoder goes from state a to state b (when the input is **1**), or to state a (when the input is **0**), and so on. The encoder cannot go from a to c in one step. It must go from a to b to c , or from a to b to d to c , and so on. We can also verify these facts from the code tree. Figure 16.7b contains the complete information of the code tree.

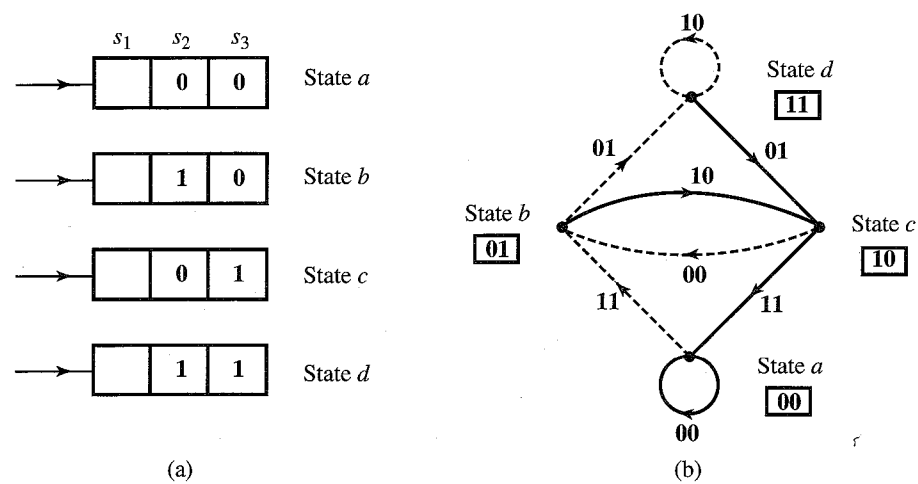


Figure 16.7 State diagram of the coder in Fig. 16.5.

Another useful way of representing the code tree is the trellis diagram in Fig. 16.8. The diagram starts from scratch (all 0's in the shift register, that is, state *a*) and makes transitions corresponding to each input data digit. These transitions are denoted by a solid line for the next data digit 0 and by a dashed line for the next data digit 1. Thus, when the first input digit is 0, the encoder output is 00 (solid line), and when the input digit is 1, the encoder output is 11 (dashed line). This is readily seen from Fig. 16.7b. We continue this way with the second input digit. After the first two input digits, the encoder is in one of the four states *a*, *b*, *c*, or *d*, as shown in Fig. 16.8. If the encoder is in state *a* (previous two data digits 00), it goes to state *b* if the next input bit is 1 or remains in state *a* if the next input bit is 0. In so doing, the encoder output is 11 (*a* to *b*) or 00 (*a* to *a*). Note that the structure of the trellis diagram is completely repetitive, as expected, and can be readily drawn using the state diagram in Fig. 16.7b.

Decoding

We shall consider two important techniques: (1) maximum-likelihood decoding (Viterbi's algorithm) and (2) sequential decoding.

Maximum-Likelihood Decoding—Viterbi's Algorithm: Among various decoding methods for convolutional codes, Viterbi's maximum-likelihood algorithm⁴ is one of the best techniques yet evolved for digital communications where energy efficiency dominates in importance. It permits major equipment simplification while obtaining the full performance benefits of maximum-likelihood decoding. The decoder structure is relatively simple for short constraint lengths, making decoding feasible at relatively high rates of up to 100 Mbit/s.

The maximum-likelihood receiver implies selecting a code word closest to the received word. Because there are 2^k code words, the maximum-likelihood decision involves storage of 2^k words and their comparison with the received word. This calculation is extremely difficult for large *k* and would result in an overly complex decoder.

A major simplification was made by Viterbi in the maximum-likelihood calculation by noting that each of the four nodes (*a*, *b*, *c*, and *d*) has only two predecessors; that is, each node

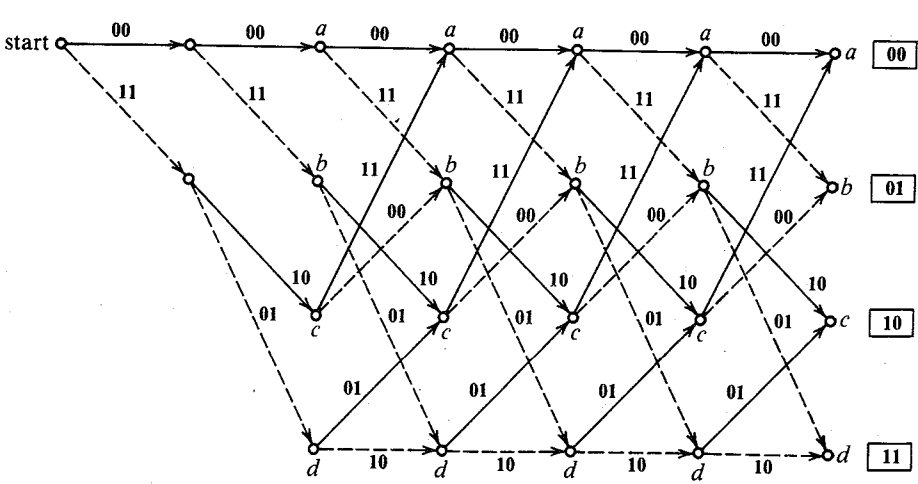


Figure 16.8 Trellis diagram for the coder in Fig. 16.5.

can be reached through two nodes only (see Fig. 16.7 or 16.8), and only the path that agrees most with the received sequence (the minimum-distance path) need be retained for each node. To understand this, consider the trellis diagram in Fig. 16.8. Our problem is as follows: Given a received sequence of bits, we need to find a path in the trellis diagram with the output digit sequence that agrees most with the received sequence.

Suppose the first six received digits are 01 00 01. We shall consider two paths of three branches (for six digits) leading to each of the nodes *a*, *b*, *c*, and *d*. Out of the two paths reaching into each node, we shall retain only the one that agrees most with the received sequence 01 00 01 (the minimum-distance path). The retained path is called the **survivor** at that node. There are two paths (00 00 00 and 11 10 11) that arrive at the third-level node *a*. These paths are at distances of 2 and 3, respectively, from the received sequence 01 00 01. Hence, the survivor at the third-level node *a* is 00 00 00. We repeat the procedure for nodes *b*, *c*, and *d*. For example, the two paths reaching the third-level node *c* (the node after three branches) are 00 11 10 and 11 01 01, at distances of 5 and 2, respectively, from the received sequence 01 00 01. Hence, the survivor at the third-level node *c* is 11 01 01. Similarly, we find survivors at the third-level nodes *b* and *d*. With four paths eliminated, the four survivor paths are the only contenders. The reason behind eliminating the other four paths is as follows. The two paths merging at the third-level node *a*, for example, imply that the previous two data digits are identical (viz., 00). Hence, regardless of what the future data digits are, both paths must merge at this node *a* and follow a common path in the future. Clearly, the survivor path is the minimum-distance path between the two, regardless of future data digits. What we need to remember is the four survivor paths and their distances from the received sequence. In general, the number of survivor paths is equal to the number of states, that is, 2^{N-1} .

Once we have survivors at all the third-level nodes, we look at the next two received digits. Suppose these are 00 (i.e., the received sequence is 01 00 01 00). We now compare the two survivors that merge into the fourth-level node *a*. These are the survivors at nodes *a* and *c* of the third level, with paths 00 00 00 00 and 11 01 01 11, respectively, and distances

of 2 and 4 from the received sequence **01 00 01 00**. Hence, the path **00 00 00 00** is the survivor at the fourth-level node *a*. We repeat this procedure for nodes *b*, *c*, and *d* and continue in this manner until the end. Note that only two paths merge in a node and there are only four contending paths (the four survivors at nodes *a*, *b*, *c*, and *d*) until the end. The only remaining problem is how to truncate the algorithm and ultimately decide on one path rather than four. This is done by forcing the last two data digits to be **00** (dummy or augmented data). When the first dummy **0** enters the register, we consider the survivors only at nodes *a* and *c*. The survivors at nodes *b* and *d* are discarded because these nodes can be reached only when the input bit is **1**, as seen from the state or trellis diagram. When the second dummy **0** enters the register, we consider only the survivor at node *a*. We discard the survivor at node *c* because the last two dummy data bits **00** lead the encoder to state *a*. In terms of the trellis diagram, this means that the number of states is reduced from four to two (*a* and *c*) by insertion of the first zero and to a single state (*a*) by insertion of the second zero.

With the Viterbi algorithm, storage and computational complexity are proportional to 2^N and are very attractive for constraint length $N < 10$. To achieve very low error probabilities, longer constraint lengths are required, and sequential decoding (to be discussed next) becomes attractive.

Sequential Decoding: In this technique, proposed by Wozencraft, the decoder complexity increases linearly rather than exponentially. To explain this technique, let us consider a coder with $N = 4$ and $v = 3$ (Fig. 16.9). The code tree for this coder is shown in Fig. 16.10. Each data digit generates three ($v = 3$) output digits but affects four groups of three digits (12 digits) in all.

In this decoding scheme, we observe only three (or v) digits at a time to make a tentative decision, with readiness to change our decision if it creates difficulties later. A sequential detector acts much like a driver who occasionally makes a wrong choice at a fork in the road, but quickly discovers the error (because of road signs), goes back, and tries the other path.

Applying this insight to our decoding problem, the analogous procedure would be as follows. We look at the first three received digits. There are only two paths of three digits from

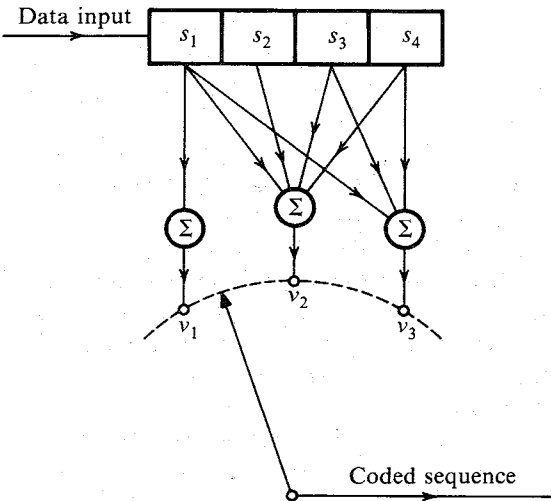


Figure 16.9 Convolution coder.

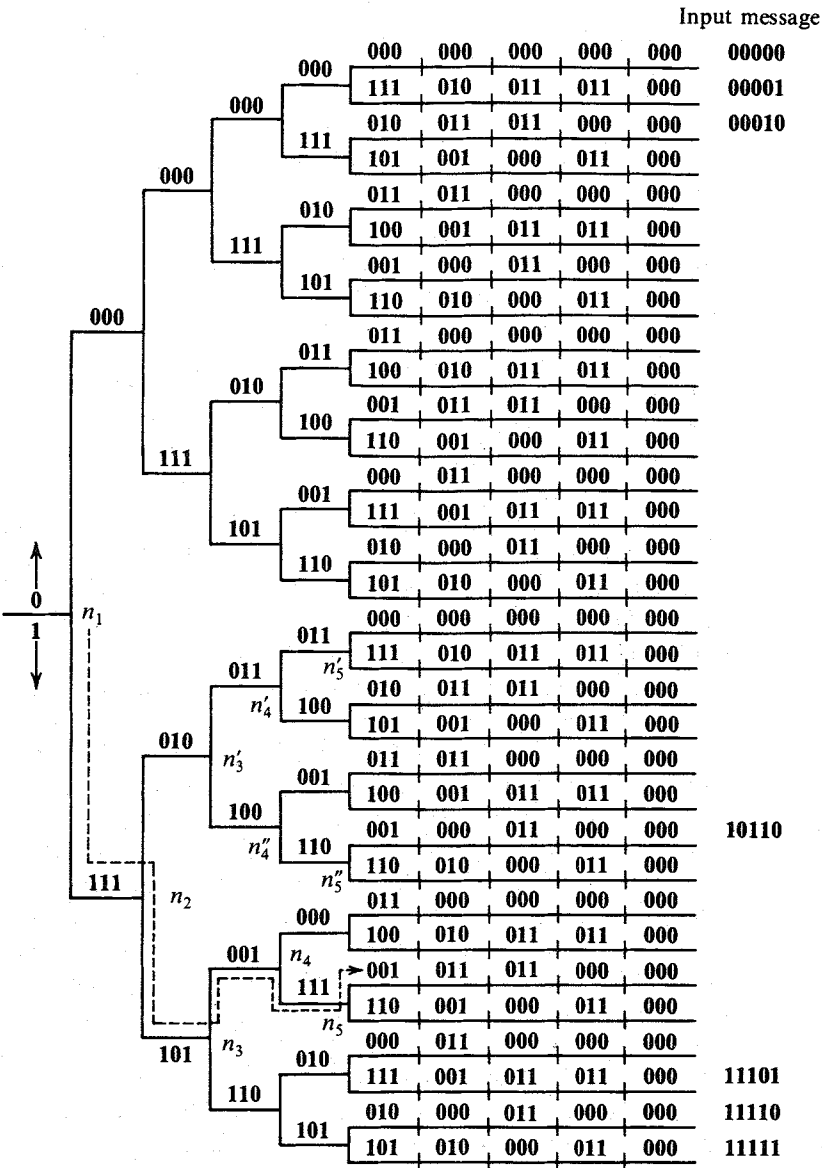


Figure 16.10 Code tree for the coder in Fig. 16.9.

the initial node n_1 . We choose that path whose sequence is at the shortest Hamming distance from the first three received digits. We thus progress to the most likely node. From this node there are two paths of three digits. We look at the second group of the three received digits and choose that path whose sequence is closest to these received digits. We progress this way until the fourth node. If we were unlucky enough to have a large number of errors in a certain

received group of v digits, we will take a wrong turn, and from there on we will find it more difficult to match the received digits with those along the paths available from the wrong node. This is the clue to the realization that an error has been made. Let us explain this by an example.

Suppose a data sequence **11010** is encoded by the coder in Fig. 16.9. Because $N = 4$, we add three dummy **0**'s to this sequence so that the augmented data sequence is **11010000**. The coded sequence will be (see the code tree in Fig. 16.10) **111 101 001 111 001 011 011 000**. Let the received sequence be **101 011 001 111 001 011 011 000** (three errors: one in the first group and two in the second group). We start at the initial node n_1 . The first received group **101** (one error) being closer to **111**, we make a correct decision to go to node n_2 . But the second group **001** (two errors) is closer to **010** than to **101** and will lead us to the wrong node n'_3 rather than to n_3 . From here on we are on the wrong track, and, hence, the received digits will not match any path starting from n'_3 . The third received group is **001** and does not match any sequence starting at n'_3 (viz., **001** and **100**). But it is closer to **011**. Hence, we go to node n'_4 . Here again the fourth received group **111** does not match any group starting at n'_4 (viz., **011** and **100**). But it is closer to **011**. This takes us to node n'_5 . It can be seen that the Hamming distance between the sequence of 12 digits along the path $n_1 n_2 n'_3 n'_4 n'_5$ and the first 12 received digits is 4, indicating four errors in 12 digits (if our path is correct). Such a high number of errors should immediately make us suspicious. If P_e is the digit error probability, then the expected number of errors n_e in d digits is $P_e d$. Because P_e is on the order of 10^{-4} to 10^{-6} , four errors in 12 digits is unreasonable. Hence, we go back to node n'_3 and try the lower branch, leading to n''_5 . This path $n_1 n_2 n'_3 n''_4 n''_5$ is even worse than the previous one, because it gives five errors in 12 digits. Hence, we go back even farther to node n_2 and try the path leading to n_3 and farther. We find the path $n_1 n_2 n_3 n_4 n_5$, giving three errors. If we go back still farther to n_1 and try alternate paths, we find that none yields less than five errors. Thus, the correct path is taken as $n_1 n_2 n_3 n_4 n_5$, giving three errors. If we go back still farther to n_1 and try alternate paths, we find that none yields less than five errors. Thus, the correct path is taken as $n_1 n_2 n_3 n_4 n_5$. This enables us to decode the first transmitted digit as **1**. Next, we start at node n_2 , discard the first three received digits, and repeat the procedure to decode the second transmitted digit. We repeat this until all the digits are decoded.

The next important question concerns the criterion for deciding when the wrong path is chosen. The plot of the expected number of errors n_e as a function of the number of decoded digits d is a straight line ($n_e = P_e d$) with slope P_e , as shown in Fig. 16.11. The actual number of errors along the path is also plotted. If the errors remain within a limit (the discard level), the decoding continues. If at some point it is found that the errors exceed the discard level, we go back to the nearest decision node and try an alternate path. If errors still increase beyond the discard level, we then go back one more node along the path and try an alternate path. The process continues until the errors are within the set limit. By making the discard level very stringent (close to the expected error curve), we reduce the average number of computations. On the other hand, if the discard level is made too stringent, the decoder will discard all possible paths in some extremely rare cases where the noise may cause an unusually large number of errors. This difficulty is usually resolved by starting with a stringent discard level. If on rare occasions the decoder rejects all paths, the discard level can be relaxed bit by bit until one of the paths is acceptable.

It can be shown that the error probability in this scheme decreases exponentially as N , whereas the system complexity grows only linearly with k . The efficiency is $\eta \simeq 1/v$. It can be

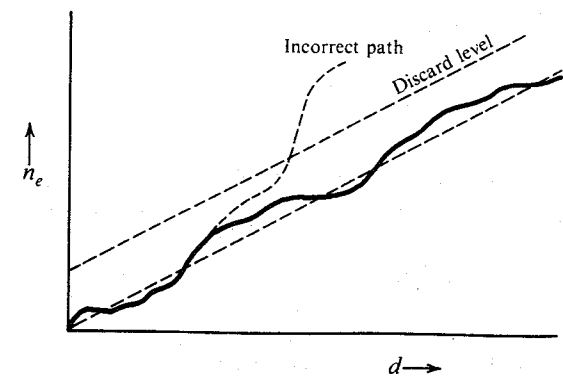


Figure 16.11 Setting the threshold in sequential decoding.

shown that for $\eta < \eta_o$, the average number of incorrect branches searched per decoded digit is bounded, whereas for $\eta > \eta_o$ it is not; hence η_o is called the **computational cutoff rate**.

There are several disadvantages to sequential decoding:

1. The number of incorrect path branches, and consequently the computation complexity, is a random variable depending on the channel noise.
2. To make storage requirements easier, the decoding speed has to be maintained at 10 to 20 times faster than the incoming data rate. This limits the maximum data rate capability.
3. The average number of branches can occasionally become very large and may result in a storage overflow, causing relatively long sequences to be erased.

A third technique for decoding convolutional codes is **feedback decoding**, with threshold decoding⁵ as a subclass. Threshold decoders are easily implemented. Their performance, however, does not compare favorably with the previous two methods.

16.7 COMPARISON OF CODED AND UNCODED SYSTEMS

It is instructive to compare the error probability of coded and uncoded schemes under similar constraints of power and information rate.

Let us consider a t -error correcting (n, k) code. In this case, k information digits are coded into n digits. For a proper comparison, we shall assume that k information digits are transmitted in the same time interval over both systems and that the transmitted power S_i is also maintained the same for both systems. Because only k digits are required to be transmitted over the uncoded system (versus n over the coded one), the bit rate R_b is lower for the uncoded system by a factor of k/n as compared to the coded system. Hence, for a given power, E_b is higher for the uncoded case as compared to the coded case by a factor n/k . This tends to reduce the bit error probability for the uncoded case. Let P_{eu} and P_{ec} represent the digit error probabilities in the uncoded and coded cases, respectively.

For the uncoded case, a word of k digits will be received wrong if any one of the k digits is in error. If P_{Eu} and P_{Ec} represent the word error probabilities of the uncoded and coded systems, respectively, then

$$P_{Eu} = 1 - P(\text{all } k \text{ digits received correct})$$

$$= 1 - (1 - P_{eu})^k \quad (16.21a)$$

$$\simeq kP_{eu} \quad P_{eu} \ll 1 \quad (16.21b)$$

For a t -error correcting (n, k) code, the received word will be in error if more than t errors occur in n digits. If $P(j, n)$ is the probability of j errors in n digits, then

$$P_{Ec} = \sum_{j=t+1}^n P(j, n)$$

Because there are $\binom{n}{j}$ ways in which j errors can occur in n digits (Example 10.6),

$$P(j, n) = \binom{n}{j} (P_{ec})^j (1 - P_{ec})^{n-j}$$

and

$$P_{Ec} = \sum_{j=t+1}^n \binom{n}{j} (P_{ec})^j (1 - P_{ec})^{n-j} \quad (16.22)$$

For $P_{Ec} \ll 1$, the first term in the summation in Eq. (16.22) dominates all the other terms, and we are justified in ignoring all but the first term. Hence,

$$P_{Ec} = \binom{n}{t+1} (P_{ec})^{t+1} (1 - P_{ec})^{n-(t+1)} \quad (16.23a)$$

$$\simeq \binom{n}{t+1} (P_{ec})^{t+1} \quad P_{ec} \ll 1 \quad (16.23b)$$

For further comparison, we must assume some specific transmission scheme. Let us consider a coherent PSK scheme. In this case for an AWGN channel,

$$P_{eu} = Q\left(\sqrt{\frac{2E_b}{\mathcal{N}}}\right) \quad (16.24a)$$

and because E_b for the coded case is k/n times that for the uncoded case,

$$P_{ec} = Q\left(\sqrt{\frac{2kE_b}{n\mathcal{N}}}\right) \quad (16.24b)$$

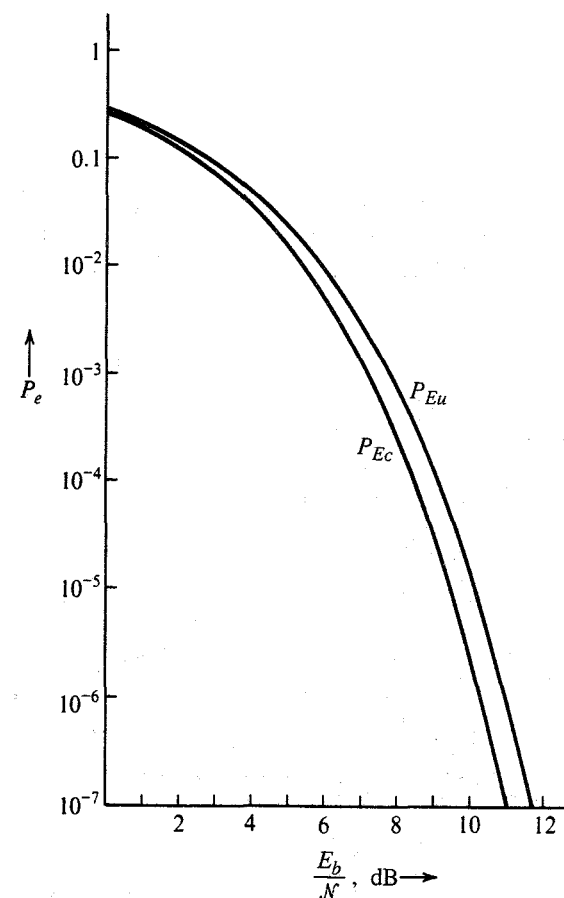
Hence,

$$P_{Eu} = kQ\left(\sqrt{\frac{2E_b}{\mathcal{N}}}\right) \quad (16.25a)$$

$$P_{Ec} = \binom{n}{t+1} \left[Q\left(\sqrt{\frac{2kE_b}{n\mathcal{N}}}\right) \right]^{t+1} \quad P_{ec} \ll 1 \quad (16.25b)$$

To compare coded and uncoded systems, we could plot P_{Eu} and P_{Ec} as functions of $E_b/\mathcal{N}$. Because Eqs. (16.25) involve parameters t , n , and k , a proper comparison requires families

Figure 16.12 Comparison of coded and uncoded systems.



of plots. For the case of a $(7, 4)$ single-error correcting code ($t = 1$, $n = 7$, and $k = 4$), P_{Ec} and P_{Eu} in Eqs. (16.25) are plotted in Fig. 16.12 as a function of $E_b/\mathcal{N}$. Observe that the coded scheme is superior to the uncoded scheme, but the improvement (about 1 dB) is not too significant. For large n and k , however, the coded scheme can become significantly superior to the uncoded one. For practical channels plagued by fading and impulse noise, coding can yield substantial gains.

EXAMPLE 16.6 Compare the performance of an AWGN BSC using a single-error correcting $(15, 11)$ code with that of the same system using uncoded transmission, given that $E_b/\mathcal{N} = 9.12$ for the uncoded scheme and coherent PSK is used to transmit the data.

From Eq. (16.25a),

$$P_{Eu} = 11Q(\sqrt{18.24}) = 1.1 \times 10^{-4}$$

and from Eq. (16.25b),

$$\begin{aligned}
 P_{Ec} &= \binom{15}{2} \left[Q \left(\sqrt{\frac{11}{15}} (18.24) \right) \right]^2 \\
 &= 105 (1.37 \times 10^{-4})^2 = 1.96 \times 10^{-6}
 \end{aligned}$$

Note that the word error probability of the coded system is reduced by a factor of 56. On the other hand, if we wish to achieve the error probability of the coded transmission (1.96×10^{-6}) using the uncoded system, we must increase the transmitted power. If E'_b is the new value of E_b to achieve $P_{Eu} = 1.96 \times 10^{-6}$,

$$P_{Eu} = 11 Q \left(\sqrt{\frac{2E'_b}{\mathcal{N}}} \right) = 1.96 \times 10^{-6}$$

This gives $E'_b/\mathcal{N} = 10.7$. This is an increase over the old value of 9.12 by a factor of 1.17, or 0.7 dB.

REFERENCES

- W. W. Peterson and E. J. Weldon, Jr., *Error Correcting Codes*, 2nd ed., Wiley, New York, 1972.
- S. Lin and D. Costello, *Error Control Coding: Fundamentals and Applications*, Prentice-Hall, Englewood-Cliffs, NJ, 1983.
- P. Elias, "Coding for Noisy Channels," *IRE Nat. Conv. Rec.*, vol. 3, part 4, pp. 37–46, 1955.
- A. J. Viterbi, "Convolutional Codes and Their Performance in Communications Systems," *IEEE Trans. Commun. Technol.*, vol. CT-19, pp. 751–771, Oct. 1971.
- J. L. Massey, *Threshold Decoding*, M.I.T. Press, Cambridge, MA, 1963.

16.1-1 Golay's (23, 12) codes are three-error correcting codes. Verify that $n = 23$ and $k = 12$ satisfies the Hamming bound exactly for $t = 3$.

16.1-2 (a) Determine the Hamming bound for a ternary code.
 (b) A ternary (11, 6) code exists that can correct up to two errors. Verify that this code satisfies the Hamming bound exactly.

16.1-3 Confirm the possibility of a (18, 7) binary code that can correct up to three errors. Can this code correct up to four errors?

16.2-1 If G and H are the generator and parity-check matrices, respectively, then show that

$$GH^T = 0$$

16.2-2 Given a generator matrix

$$G = [1 \quad 1 \quad 1]$$

construct a (3, 1) code. How many errors can this code correct? Find the code word for data vectors $d = 0$ and $d = 1$. Comment.

16.2-3 Repeat Prob. 16.2-2 for

$$G = [1 \quad 1 \quad 1 \quad 1 \quad 1]$$

This gives a (5, 1) code.

16.2-4 Suppose we wish to increase reliability by repeating a message three times, e.g., by transmitting data 0 by 000 and 1 by 111. This is a (3, 1) code.

- (a) Is this a systematic code?
- (b) If so, find the generating matrix G .

16.2-5 Consider the following $(k + 1, k)$ systematic linear block code with the parity-check digit c_{k+1} given by

$$c_{k+1} = d_1 + d_2 + \cdots + d_k$$

- (a) Construct the appropriate generator matrix for this code.
- (b) Construct the code generated by this matrix for $k = 3$.
- (c) Determine the error detecting or correcting capabilities of this code.
- (d) Show that

$$cH^T = 0$$

and

$$rH^T = \begin{cases} 0 & \text{if no error occurs} \\ 1 & \text{if single error occurs} \end{cases}$$

16.2-6 Consider a generator matrix G for a nonsystematic (6, 3) code:

$$G = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Construct the code for this G , and show that $d_{\min}$, the minimum distance between code words, is 3. Consequently, this code can correct at least one error.

16.2-7 Repeat Prob. 16.2-6 if

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

16.2-8 Find a generator matrix G for a (15, 11) single-error correcting linear block code. Find the code word for the data vector 10111010101.

16.2-9 For a (6, 3) systematic linear block code, the three parity-check digits c_4, c_5 , and c_6 are

$$\begin{aligned}
 c_4 &= d_1 + d_2 + d_3 \\
 c_5 &= d_1 + d_2 \\
 c_6 &= d_1 + d_3
 \end{aligned}$$

- (a) Construct the appropriate generator matrix for this code.
- (b) Construct the code generated by this matrix.
- (c) Determine the error correcting capabilities of this code.
- (d) Prepare a suitable decoding table.
- (e) Decode the following received words: **101100**, **000110**, **101010**.

- 16.2-10** (a) Construct a code table for the (6, 3) code generated by the matrix G in Prob. 16.2-6.
 (b) Prepare a suitable decoding table.

- 16.2-11** Construct a single-error correcting (7, 4) linear block code (Hamming code) and the corresponding decoding table.

- 16.2-12** For the (6, 3) code in Example 16.1, the decoding table is Table 16.3. Show that if we use this decoding table, and a two-error pattern **010100** or **001001** occurs, it will not be corrected. If it is desired to correct a single two-error pattern **010100** (along with six single-error patterns), construct the appropriate decoding table and verify that it does indeed correct one two-error pattern **010100** and that it cannot correct any other two-error patterns.

- 16.2-13** (a) Given $k = 8$, find the minimum value of n for a code that can correct at least one error.
 (b) Choose a generator matrix G for this code.
 (c) How many double errors can this code correct?
 (d) Construct a decoding table (syndromes and corresponding correctable error patterns).

- 16.2-14** Consider a (6, 2) code generated by the matrix

$$\begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix}$$

- (a) Construct the code table for this code and determine the minimum distance between code words.
- (b) Prepare a suitable decoding table. *Hint:* This code can correct all single-error patterns, seven double-error patterns, and two triple-error patterns. Choose the desired seven double-error patterns and the two triple-error patterns.

- 16.3-1** (a) Construct a systematic (7, 4) cyclic code using the generator polynomial $g(x) = x^3 + x + 1$.
 (b) What are the error correcting capabilities of this code?
 (c) Construct the decoding table.
 (d) If the received word is **1101100**, determine the transmitted data word.

- 16.3-2** A three-error correcting (23, 12) Golay code is a cyclic code with a generator polynomial

$$g(x) = x^{11} + x^9 + x^7 + x^6 + x^5 + x + 1$$

Determine the code words for the data vectors **000011110000**, **101010101010**, and **1100010101110**.

- 16.3-3** Factorize the polynomial

$$x^3 + x^2 + x + 1$$

Hint: A third-order polynomial must have one factor of first order. The only first-order polynomials that are prime (not factorable) are x and $x + 1$. Since x is not a factor of the given polynomial, try $x + 1$. Divide $x^3 + x^2 + x + 1$ by $x + 1$.

- 16.3-4** The concept explained in Prob. 16.3-3 can be extended to factorize any higher order polynomial. Using this technique, factorize

$$x^5 + x^4 + x^2 + 1$$

Hint: There must be at least one first-order factor. Try dividing by the two first-order prime polynomials x and $x + 1$. The given fifth-order polynomial can now be expressed as $\phi_1(x)\phi_4(x)$, where $\phi_1(x)$ is a first-order polynomial and $\phi_4(x)$ is a fourth-order polynomial which may or may not contain a first-order factor. Try dividing $\phi_4(x)$ by x and $x + 1$. If it does not work, it must have two second-order polynomials both of which are prime. The possible second-order polynomials are x^2 , $x^2 + 1$, $x^2 + x$, and $x^2 + x + 1$. Determine which of these are prime (not divisible by x or $x + 1$). Now try dividing $\phi_4(x)$ by these prime polynomials of the second order. If neither divides, $\phi_4(x)$ must be a prime polynomial of the fourth order and the factors are $\phi_1(x)$ and $\phi_4(x)$.

- 16.3-5** Use the concept explained in Prob. 16.3-4 to factorize a seventh-order polynomial $x^7 + 1$.
Hint: Determine prime factors of first-, second-, and third-order polynomials. The possible third-order polynomials are x^3 , $x^3 + 1$, $x^3 + x$, $x^3 + x + 1$, $x^3 + x^2$, $x^3 + x^2 + 1$, $x^3 + x^2 + x$, and $x^3 + x^2 + x + 1$. See hint in Prob. 16.3-4.

- 16.3-6** Equation (16.15) suggests a method of constructing a generator matrix G' for a cyclic code,

$$G' = \begin{bmatrix} x^{k-1}g(x) \\ x^{k-2}g(x) \\ \vdots \\ g(x) \end{bmatrix} = \begin{bmatrix} g_1 & g_2 & \cdots & g_{n-k+1} & 0 & 0 & \cdots & 0 \\ 0 & g_1 & g_2 & \cdots & g_{n-k+1} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & g_1 & g_2 & \cdots & g_{n-k+1} \end{bmatrix}$$

where $g(x) = g_1x^{n-k} + g_2x^{n-k-1} + \cdots + g_{n-k+1}$ is the generator polynomial. This is, in general, a nonsystematic cyclic code.

- (a) For a single-error correcting (7, 4) cyclic code with a generator polynomial $g(x) = x^3 + x^2 + 1$, find G' and construct the code.
- (b) Verify that this code is identical to that derived in Example 16.3 (Table 16.4).

- 16.3-7** The generator matrix G for a systematic cyclic code (see Prob. 16.3-6) can be obtained by realizing that adding any row of a generator matrix to any other row yields another valid generator matrix, because the code word is formed by linear combinations of data digits. Also, a generator matrix for a systematic code must have an identity matrix I_k in the first k columns. Such a matrix is formed step by step as follows. Observe that each row in G' in Prob. 16.3-6 is a left shift of the row below it, with the last row being $g(x)$. Start with the k th (last) row $g(x)$. Because $g(x)$ is of the order $n - k$, this row has the element 1 in the k th column, as required. For the $(k - 1)$ th row, use the last row with one left shift. We require a 0 in the k th column of the $(k - 1)$ th row to form I_k . If there is a 0 in the k th column of this $(k - 1)$ th row, we accept it as a valid $(k - 1)$ th row. If not, then we add the k th row to the $(k - 1)$ th row to obtain 0 in its k th column. The resulting row is the final $(k - 1)$ th row. This row with a single left shift serves as the $(k - 2)$ th row. But if this newly formed $(k - 2)$ th row does not have a 0 in its k th column, we add the k th (last) row to it to get the desired 0. We continue this way until all k rows are formed. This gives the generator matrix for a systematic (n, k) cyclic code.

- (a) For a single-error correcting (7, 4) systematic cyclic code with a generator polynomial $g(x) = x^3 + x^2 + 1$, find G and construct the code.
- (b) Verify that this code is identical to that in Table 16.5 (Example 16.4).

16.3-8 (a) Find the generator matrix G' for a nonsystematic (7, 4) cyclic code using the generator polynomial $g(x) = x^3 + x + 1$.

(b) Find the code generated by this matrix G' .

(c) Determine the error correcting capabilities of this code.

16.3-9 Find the generator matrix G for a systematic (7, 4) cyclic code using the generator polynomial $g(x) = x^3 + x + 1$ (see Prob. 16.3-7).

16.4-1 The simple burst-error detecting code in Fig. 16.3 can also be used as a single-error correcting code with a slight modification. The k data digits are divided into groups of b digits in length, as in Fig. 16.3. To each group we add one parity-check digit, so that each segment now has $b + 1$ digits (b data digits and one parity-check digit). The parity-check digit is chosen to ensure that the total number of 1's in each segment of $b + 1$ digits is even. Now we consider these digits as our new data and augment them with the last segment of $b + 1$ parity-check digits, as was done in Fig. 16.3. The data in Fig. 16.3 will be transmitted thus:

10111 01010 11011 10001 11000 01111

Show that this (30, 20) code is capable of single error correction as well as the detection of a single burst of length 5.

- 16.5-1** Discuss the error correcting capabilities of an interlaced $(\lambda n, \lambda k)$ cyclic code with $\lambda = 10$ and using a three-error correcting (31, 16) BCH code.
- 16.6-1** For the convolutional encoder in Fig. 16.5, the received bits are 01 00 01 00 10 11 11 00. Decode this sequence using Viterbi's algorithm and the trellis diagram in Fig. 16.8.

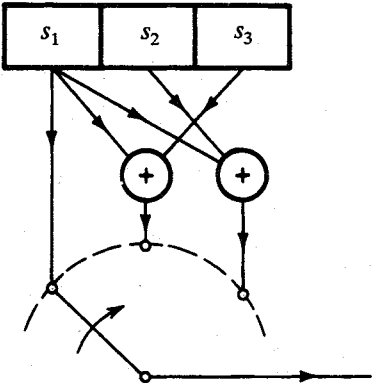


Figure P16.6-2

- 16.6-2** For the convolutional encoder shown in Fig. P16.6-2:
- (a) Draw the state and trellis diagrams and determine the output digit sequence for the data digits 11010100.

- (b) Use Viterbi's algorithm to decode the following received sequences:
- (i) 100 110 111 101 001 101 001 010
 - (ii) 010 110 111 101 101 101 001 010
 - (iii) 111 110 111 111 001 101 001 101

16.7-1 Uncoded data is transmitted using PSK over an AWGN channel with $E_b/\mathcal{N} = 9.12$. This data is now coded using a three-error correcting (23, 12) Golay code (see Prob. 16.1-1) and transmitted over the same channel at the same data rate and with the same transmitted power.

(a) Determine the word error probabilities P_{Eu} and P_{Ec} for the coded and the uncoded systems.

(b) If it is decided to achieve the error probability P_{Ec} computed in part (a) using the uncoded system by increasing the transmitted power, determine the required value of $E_b/\mathcal{N}$.

Appendix A

ORTHOGONALITY OF SOME SIGNAL SETS

A.1 Orthogonality of the Trigonometric and Exponential Signal Set

Consider an integral I defined by

$$I = \int_{T_0} \cos n\omega_0 t \cos m\omega_0 t dt \quad (\text{A.1a})$$

where $\int_{T_0}$ stands for integration over any contiguous interval of $T_0 = 2\pi/\omega_0$ seconds. By using a trigonometric identity (see Appendix D), Eq. (A.1a) can be expressed as

$$I = \frac{1}{2} \left[\int_{T_0} \cos(n+m)\omega_0 t dt + \int_{T_0} \cos(n-m)\omega_0 t dt \right] \quad (\text{A.1b})$$

Since $\cos \omega_0 t$ executes one complete cycle during any interval of T_0 seconds, $\cos(n+m)\omega_0 t$ executes $(n+m)$ complete cycles during any interval of duration T_0 . Therefore, the first integral in Eq. (A.1b), which represents the area under $(n+m)$ complete cycles of a sinusoid, equals zero. The same argument shows that the second integral in Eq. (A.1b) is also zero, except when $n = m$. Hence, I in Eq. (A.1b) is zero for all $n \neq m$. When $n = m$, the first integral in Eq. (A.1b) is still zero, but the second integral yields

$$I = \frac{1}{2} \int_{T_0} dt = \frac{T_0}{2}$$

Thus,

$$\int_{T_0} \cos n\omega_0 t \cos m\omega_0 t dt = \begin{cases} 0 & n \neq m \\ \frac{T_0}{2} & m = n \neq 0 \end{cases} \quad (\text{A.2a})$$

Using similar arguments, we can show that

$$\int_{T_0} \sin n\omega_0 t \sin m\omega_0 t dt = \begin{cases} 0 & n \neq m \\ \frac{T_0}{2} & n = m \neq 0 \end{cases} \quad (\text{A.2b})$$

and

$$\int_{T_0} \sin n\omega_0 t \cos m\omega_0 t dt = 0 \quad \text{all } n \text{ and } m \quad (\text{A.2c})$$

A.2 Orthogonality of the Exponential Signal Set

The set of exponentials $e^{jn\omega_0 t}$ ($n = 0, \pm 1, \pm 2, \dots$) is orthogonal over any interval of duration T_0 , that is,

$$\int_{T_0} e^{jm\omega_0 t} (e^{jn\omega_0 t})^* dt = \int_{T_0} e^{j(m-n)\omega_0 t} dt = \begin{cases} 0 & m \neq n \\ T_0 & m = n \end{cases} \quad (\text{A.3})$$

Let the integral on the left-hand side of Eq. (A.3) be I ,

$$\begin{aligned} I &= \int_{T_0} e^{jm\omega_0 t} (e^{jn\omega_0 t})^* dt \\ &= \int_{T_0} e^{j(m-n)\omega_0 t} dt \end{aligned} \quad (\text{A.4})$$

The case $m = n$ is trivial. In this case the integrand is unity, and $I = T_0$. When $m \neq n$,

$$\begin{aligned} I &= \frac{1}{j(m-n)\omega_0} e^{j(m-n)\omega_0 t} \Big|_{t_1}^{t_1+T_0} \\ &= \frac{1}{j(m-n)\omega_0} e^{j(m-n)\omega_0 t_1} [e^{j(m-n)\omega_0 T_0} - 1] = 0 \end{aligned}$$

The last result follows from the fact that $\omega_0 T_0 = 2\pi$, and $e^{j2\pi k} = 1$ for all integral values of k .

Appendix B

SCHWARZ INEQUALITY

Prove the following Schwarz inequality for a pair of real finite energy signals $f(t)$ and $g(t)$:

$$\left[\int_a^b f(t)g(t) dt \right]^2 \leq \left[\int_a^b f^2(t) dt \right] \left[\int_a^b g^2(t) dt \right] \quad (\text{B.1})$$

with equality only if $g(t) = cf(t)$, where c is an arbitrary constant.

The Schwarz inequality for finite-energy, complex-valued functions $X(\omega)$ and $Y(\omega)$ is given by

$$\left| \int_{-\infty}^{\infty} X(\omega)Y(\omega) d\omega \right|^2 \leq \int_{-\infty}^{\infty} |X(\omega)|^2 d\omega \int_{-\infty}^{\infty} |Y(\omega)|^2 d\omega \quad (\text{B.2})$$

with equality only if $Y(\omega) = cX^*(\omega)$, where c is an arbitrary constant.

We can prove Eq. (B.1) as follows: For any real value of λ , we know that

$$\int_a^b [\lambda f(t) - g(t)]^2 dt \geq 0$$

or

$$\lambda^2 \int_a^b f^2(t) dt - 2\lambda \int_a^b f(t)g(t) dt + \int_a^b g^2(t) dt \geq 0$$

Because this quadratic in λ is nonnegative for any value of λ , its discriminant must be nonpositive, and Eq. (B.1) follows. If the discriminant equals zero, then for some value of $\lambda = c$, the quadratic equals zero. This is possible only if $cf(t) - g(t) = 0$, and the result follows.

To prove Eq. (B.2), we observe that $|X(\omega)|$ and $|Y(\omega)|$ are real functions and inequality B.1 applies. Hence,

$$\left[\int_a^b |X(\omega)Y(\omega)| d\omega \right]^2 \leq \int_a^b |X(\omega)|^2 d\omega \int_a^b |Y(\omega)|^2 d\omega \quad (\text{B.3})$$

with equality only if $|Y(\omega)| = c|X(\omega)|$, where c is an arbitrary constant. Now recall that

$$\left| \int_a^b X(\omega)Y(\omega) d\omega \right| \leq \int_a^b |X(\omega)||Y(\omega)| d\omega = \int_a^b |X(\omega)Y(\omega)| d\omega \quad (\text{B.4})$$

with equality if and only if $Y(\omega) = cX^*(\omega)$, where c is an arbitrary constant. Equation (B.2) immediately follows from Eqs. (B.3) and (B.4).

Appendix C

GRAM-SCHMIDT ORTHOGONALIZATION OF A VECTOR SET

We have defined the dimensionality of a vector space as equal to the maximum number of independent vectors in the space. Thus in an N -dimensional space, there can be no more than N vectors that are independent. Alternatively, it is always possible to find a set of N vectors that are independent. Once such a set is chosen, any vector in this space can be expressed in terms of (as a linear combination of) the vectors in this set. This set forms what we commonly refer to as a basis set, which forms the coordinate system. This set of N independent vectors is by no means unique. The reader is familiar with this fact in the physical space of three dimensions, where one can find an infinite number of independent sets of three vectors. This is clear from the fact that we have an infinite number of possible coordinate systems. An orthogonal set, however, is of special interest because it is easier to deal with as compared to nonorthogonal set. If we are given a set of N independent vectors, it is possible to obtain from this set another set of N independent vectors that is orthogonal. This is done by the Gram-Schmidt process of orthogonalization.

In the following derivation, we use the result [derived in Eq. (2.26)] that the projection (or component) of a vector x_2 upon another vector x_1 (see Fig. C.1) is $c_{12}x_1$, where

$$c_{12} = \frac{x_1 \cdot x_2}{|x_1|^2} y_1 \quad (C.1a)$$

The error in this approximation is the vector $x_2 - c_{12}x_1$, that is,

$$\text{Error vector} = x_2 - \frac{x_1 \cdot x_2}{|x_1|^2} x_1 \quad (C.1b)$$

The error vector, shown dashed in Fig. C.1 is orthogonal to vector x_1 .

In order to get a physical insight into this procedure, we shall consider a simple case of 2-dimensional space. Let x_1 and x_2 be two independent vectors in a 2-dimensional space (Fig. C.1). We wish to generate a new set of two orthogonal vectors y_1 and y_2 from x_1 and x_2 . For convenience, we shall choose

$$y_1 = x_1 \quad (C.2)$$

We now find another vector y_2 that is orthogonal to y_1 (and x_1). Figure C.1 shows that the error

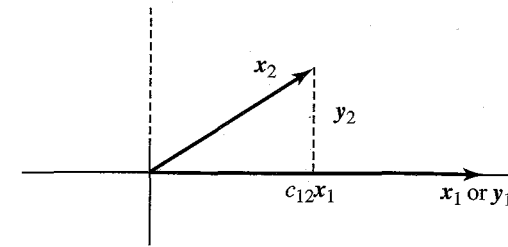


Figure C.1 Gram-Schmidt process for a 2-dimensional case.

vector in approximation of x_2 by y_1 (shown by dashed lines) is orthogonal to y_1 , and can be taken as y_2 . Hence,

$$\begin{aligned} y_2 &= x_2 - \frac{x_1 \cdot x_2}{|x_1|^2} x_1 \\ &= x_2 - \frac{y_1 \cdot x_2}{|y_1|^2} y_1 \end{aligned} \quad (C.3)$$

Equations (C.2) and (C.3) yield the desired orthogonal set. Note that this set is not unique. There is an infinite number of possible orthogonal vector sets (y_1, y_2) that can be generated from (x_1, x_2). In our derivation, we could as well have started with $y = x_2$ instead of $y_1 = x_1$. This starting point would have yielded an entirely different set.

The reader can extend these results to a 3-dimensional case. If vectors x_1, x_2, x_3 form an independent set in this space, then we form vectors y_1 and y_2 as in Eqs (C.2) and (C.3). To determine y_3 , we approximate x_3 in terms of vectors y_1 and y_2 . The error in this approximation must be orthogonal to both y_1 and y_2 and, hence, can be taken as the third orthogonal vector y_3 . Hence,

$$\begin{aligned} y_3 &= x_3 - \text{sum of projections of } x_3 \text{ on } y_1 \text{ and } y_2 \\ &= x_3 - \frac{y_1 \cdot x_3}{|y_1|^2} y_1 - \frac{y_2 \cdot x_3}{|y_2|^2} y_2 \end{aligned} \quad (C.4)$$

These results can be extended to an N -dimensional space. In general, if we are given N independent vectors $x_1, x_2, \dots, x_N$, then proceeding along similar lines, one can obtain an orthogonal set $y_1, y_2, \dots, y_N$, where

$$y_1 = x_1$$

and

$$y_j = x_j - \sum_{k=1}^{j-1} \frac{y_k \cdot x_j}{|y_k|^2} y_k \quad j = 2, 3, \dots, N \quad (C.5)$$

Note that this is one of the infinitely many orthogonal sets that can be formed from the set $x_1, x_2, \dots, x_N$. Moreover, this set is not an orthonormal set. The orthonormal set $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_N$ can be obtained by normalizing the lengths of the respective vectors,

$$\hat{y}_k = \frac{y_k}{|y_k|}$$

We can apply these concepts to signal space, because one-to-one correspondence exists between signals and vectors. If we have N independent signals $x_1(t)$, $x_2(t)$, ..., $x_N(t)$, we can form a set of N orthogonal signals $y_1(t)$, $y_2(t)$, ..., $y_N(t)$ as

$$\begin{aligned} y_1(t) &= x(t) \\ y_j(t) &= x_j(t) - \sum_{k=1}^{j-1} c_{kj} y_k(t) \quad j = 2, 3, \dots, N \end{aligned} \quad (\text{C.6})$$

where

$$c_{kj} = \frac{\int y_k(t) x_j(t) dt}{\int y_k^2(t) dt} \quad (\text{C.7})$$

Note that this is one of the infinitely many possible orthogonal sets that can be formed from the set $x_1(t)$, $x_2(t)$, ..., $x_N(t)$. The set can be normalized by dividing each signal $y_j(t)$ by its energy.

EXAMPLE C.1

The exponential signals

$$\begin{aligned} g_1(t) &= e^{-pt} u(t) \\ g_2(t) &= e^{-2pt} u(t) \\ &\vdots \\ g_N(t) &= e^{-Npt} u(t) \end{aligned}$$

form an independent set of signals in N -dimensional space, where N may be any integer. This set, however, is not orthogonal. We can use the Gram-Schmidt process to obtain an orthogonal set for this space. If $y_1(t)$, $y_2(t)$, ..., $y_N(t)$ is the desired orthogonal basis set, we choose

$$y_1(t) = g_1(t) = e^{-pt} u(t)$$

From Eqs. (C.6) and (C.7) we have

$$y_2(t) = x_2(t) - c_{12} y_1(t)$$

where

$$\begin{aligned} c_{12} &= \frac{\int_{-\infty}^{\infty} y_1(t) x_2(t) dt}{\int_{-\infty}^{\infty} y_1^2(t) dt} \\ &= \frac{\int_0^{\infty} e^{-pt} e^{-2pt} dt}{\int_0^{\infty} e^{-2pt} dt} \\ &= \frac{2}{3} \end{aligned}$$

Hence,

$$y_2(t) = (e^{-2pt} - \frac{2}{3} e^{-pt}) u(t)$$

Similarly, we can proceed to find the remaining functions $y_3(t)$, ..., $y_N(t)$, and so on. The reader can verify that all this represents a mutually orthogonal set.

Appendix D

MISCELLANEOUS

D.1 L'Hôpital's Rule

If $\lim f(x)/g(x)$ results in the indeterministic form $0/0$ or ∞/∞ , then

$$\lim \frac{f(x)}{g(x)} = \lim \frac{\dot{f}(x)}{\dot{g}(x)}$$

D.2 Taylor and Maclaurin Series

$$f(x) = f(a) + \frac{(x-a)}{1!} \dot{f}(a) + \frac{(x-a)^2}{2!} \ddot{f}(a) + \dots$$

$$f(x) = f(0) + \frac{x}{1!} \dot{f}(0) + \frac{x^2}{2!} \ddot{f}(0) + \dots$$

D.3 Power Series

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots + \frac{x^n}{n!} + \dots$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - \dots$$

$$\tan x = x + \frac{x^3}{3} + \frac{2x^5}{15} + \frac{17x^7}{315} + \dots \quad x^2 < \frac{\pi^2}{4}$$

$$Q(x) = \frac{e^{-x^2/2}}{x\sqrt{2\pi}} \left(1 - \frac{1}{x^2} + \frac{1 \cdot 3}{x^4} - \frac{1 \cdot 3 \cdot 5}{x^6} + \dots \right)$$

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{2!}x^2 + \frac{n(n-1)(n-2)}{3!}x^3 + \cdots + \binom{n}{k}x^k + \cdots + x^n$$

$$\approx 1 + nx \quad |x| \ll 1$$

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots \quad |x| < 1$$

D.4 Sums

$$\sum_{m=0}^k r^m = \frac{r^{k+1} - 1}{r - 1} \quad r \neq 1$$

$$\sum_{m=M}^N r^m = \frac{r^{N+1} - r^M}{r - 1} \quad r \neq 1$$

$$\sum_{m=0}^k \left(\frac{a}{b}\right)^m = \frac{a^{k+1} - b^{k+1}}{b^k(a - b)} \quad a \neq b$$

D.5 Complex Numbers

$$e^{\pm j\pi/2} = \pm j$$

$$e^{\pm jn\pi} = \begin{cases} 1 & n \text{ even} \\ -1 & n \text{ odd} \end{cases}$$

$$e^{\pm j\theta} = \cos \theta \pm j \sin \theta$$

$$a + jb = re^{j\theta} \quad r = \sqrt{a^2 + b^2}, \quad \theta = \tan^{-1}\left(\frac{b}{a}\right)$$

$$(re^{j\theta})^k = r^k e^{jk\theta}$$

$$(r_1 e^{j\theta_1})(r_2 e^{j\theta_2}) = r_1 r_2 e^{j(\theta_1 + \theta_2)}$$

D.6 Trigonometric Identities

$$e^{\pm jx} = \cos x \pm j \sin x$$

$$\cos x = \frac{1}{2}(e^{jx} + e^{-jx})$$

$$\sin x = \frac{1}{2j}(e^{jx} - e^{-jx})$$

$$\cos\left(x \pm \frac{\pi}{2}\right) = \mp \sin x$$

$$\sin\left(x \pm \frac{\pi}{2}\right) = \pm \cos x$$

$$2 \sin x \cos x = \sin 2x$$

$$\sin^2 x + \cos^2 x = 1$$

$$\cos^2 x - \sin^2 x = \cos 2x$$

$$\cos^2 x = \frac{1}{2}(1 + \cos 2x)$$

$$\sin^2 x = \frac{1}{2}(1 - \cos 2x)$$

$$\cos^3 x = \frac{1}{4}(3 \cos x + \cos 3x)$$

$$\sin^3 x = \frac{1}{4}(3 \sin x - \sin 3x)$$

$$\sin(x \pm y) = \sin x \cos y \pm \cos x \sin y$$

$$\cos(x \pm y) = \cos x \cos y \mp \sin x \sin y$$

$$\tan(x \pm y) = \frac{\tan x \pm \tan y}{1 \mp \tan x \tan y}$$

$$\sin x \sin y = \frac{1}{2}[\cos(x - y) - \cos(x + y)]$$

$$\cos x \cos y = \frac{1}{2}[\cos(x - y) + \cos(x + y)]$$

$$\sin x \cos y = \frac{1}{2}[\sin(x - y) + \sin(x + y)]$$

$$a \cos x + b \sin x = C \cos(x + \theta)$$

in which $C = \sqrt{a^2 + b^2}$ and $\theta = \tan^{-1}\left(\frac{-b}{a}\right)$

D.7 Indefinite Integrals

$$\int u dv = uv - \int v du$$

$$\int f(x) \dot{g}(x) dx = f(x)g(x) - \int \dot{f}(x)g(x) dx$$

$$\int \sin ax dx = -\frac{1}{a} \cos ax \quad \int \cos ax dx = \frac{1}{a} \sin ax$$

$$\int \sin^2 ax dx = \frac{x}{2} - \frac{\sin 2ax}{4a} \quad \int \cos^2 ax dx = \frac{x}{2} + \frac{\sin 2ax}{4a}$$

$$\int x \sin ax dx = \frac{1}{a^2}(\sin ax - ax \cos ax)$$

$$\int x \cos ax \, dx = \frac{1}{a^2}(\cos ax + ax \sin ax)$$
$$\int x^2 \sin ax \, dx = \frac{1}{a^3}(2ax \sin ax + 2 \cos ax - a^2 x^2 \cos ax)$$
$$\int x^2 \cos ax \, dx = \frac{1}{a^3}(2ax \cos ax - 2 \sin ax + a^2 x^2 \sin ax)$$
$$\int \sin ax \sin bx \, dx = \frac{\sin (a-b)x}{2(a-b)} - \frac{\sin (a+b)x}{2(a+b)} \quad a^2 \neq b^2$$
$$\int \sin ax \cos bx \, dx = -\left[\frac{\cos (a-b)x}{2(a-b)} + \frac{\cos (a+b)x}{2(a+b)} \right] \quad a^2 \neq b^2$$
$$\int \cos ax \cos bx \, dx = \frac{\sin (a-b)x}{2(a-b)} + \frac{\sin (a+b)x}{2(a+b)} \quad a^2 \neq b^2$$
$$\int e^{ax} \, dx = \frac{1}{a}e^{ax}$$
$$\int x e^{ax} \, dx = \frac{e^{ax}}{a^2}(ax - 1)$$
$$\int x^2 e^{ax} \, dx = \frac{e^{ax}}{a^3}(a^2 x^2 - 2ax + 2)$$
$$\int e^{ax} \sin bx \, dx = \frac{e^{ax}}{a^2 + b^2}(a \sin bx - b \cos bx)$$
$$\int e^{ax} \cos bx \, dx = \frac{e^{ax}}{a^2 + b^2}(a \cos bx + b \sin bx)$$
$$\int \frac{1}{x^2 + a^2} \, dx = \frac{1}{a} \tan^{-1} \frac{x}{a}$$
$$\int \frac{x}{x^2 + a^2} \, dx = \frac{1}{2} \ln (x^2 + a^2)$$

INDEX

- 2-binary 1-quaternary (2B1Q), 371
- Aliasing error (spectral folding), 60, 133, 257
- Alternate mark inversion (AMI), 295, 306, 359
- American national standards institute (ANSI), 432
- Amplitude modulation (AM), 162–170, 537–541, 718
- Amplitude phase shift keying (APK), 341
- Amplitude shift keying, 337–339, 591, 595
- Amplitude spectrum, 49, 74
- Analog-to-digital conversion, 5, 265
- Analog message, 3
- Analog signals, 20
 - conversion to digital, 5, 265–270
- Angle modulation, 208, 209, 541
- Asymmetric digital subscriber line (ADSL), 393
- Asynchronous transfer mode (ATM), 383
- Autocorrelation function, 40, 125, 127, 490
- Avalanche photodiode (APD), 423
- Bandpass limiter, 234
- Bandpass signals, 95, 514
 - distortionless transmission, 102
- Bandwidth, 8, 81
 - and duration of signal, 89
 - essential, 117
 - of product of signals, 89
 - of rect function, 81
- Baseband signal, 1, 151
- Basic rate interface (BRI), 369
- Bayes' receiver, 670
- Bayes' rule, 439
- Bearer channel (B channel), 369
- Bipolar return to zero (BPRZ), 359
- Binary
 - with 8 zero substitution (B8ZS), 309, 360
 - with 6 zero substitution (B6ZS), 309
 - Binary 8 zero suppression (B8ZS), 309, 360
- Binary message, 3, 20
- Binary symmetric channel (BSC) 447
 - channel capacity of, 696
- Bipolar (pseudoternary) signaling, 295, 306
- Bit (binary digit), 263
- Bit error rate (BER), 584, 609, 613, 617
- Bit stuffing, 345
- Block codes, 728, 731
- Bose-Chaudhuri-Hocquenghen (BCH) codes, 745
- Broadband ISDN (BISDN), 390
- Butterworth filters, 108
- Burst-error-detection and -correction codes, 745
- Capture effect, 242
- Capture range of PLL, 186
- Carrier, 10
- Carrier acquisition, 183
- Carrier communication, 151
- Carson's rule, 219
- Causal systems, 107
- CCITT, 275
- Cellular and digital packet data (CDPD), 412
- Cellular telephone, 404, 412
- Central-limit theorem, 472
- Channel, 1
- Channel bank, 358
- Channel capacity, 693, 709

- Channel matrix, 694
- Channel service unit (CSU), 363
- Chebyshev's inequality, 471
- Chip, 407
- Chip rate, 407
- Circuit, switched, 368
- Code-division multiple access (CDMA), 406, 412, 600
- Code efficiency, 687, 728
- Code polynomial, 738
- Code rate, 728
- Code tree, 748
- Coefficient of correlation, 36, 476, 657
- Coherent detection, 154, 590
- Communication systems, 1
 - analog, 532
 - digital, 294, 577
- Compact codes, 685
- Companding, 268, 563
- Compressors, 268, 563
- Conditional probability, 439
- Conjugate symmetry property, 75
- Consultative Committee on International Telegraphy and Telephony (CCITT), 275
- Continuous-time signals, 20
- Convolutional codes, 728, 747
- Correlation, 35, 473-476
- Correlation coefficient, 36
- Correlation functions, 39
- Correlator detector, 581
- Costas loop, 187
- Cross-correlation function, 40, 509
- Cross-power spectral density, 510
- Cumulative distribution function (CDF), 449
- Cyclic codes, 737-746
- Cyclostationary random process, 504n

- D4 framing (DF), 355
- Data channel (D channel), 369
- Decision regions, 646
- Deemphasis, 243
- Delta modulation, 281-288
 - adaptive, 285
 - comparison with PCM, 287
 - double integration, 285
 - output SNR, 286
 - overloading in, 284-285
 - threshold of coding, 284
- Demodulation, 10, 154, 161, 167, 177, 233
- Detection error probability, 329, 586-589
- Deviation ratio, 219
- Differential coding, 318, 600
- Differential GPS, 416

- Differential pulse code modulation (DPCM), 278-281
- Differentially coherent phase-shift keying (DPSK), 599
- Digit interleaving, 343
- Digital carrier systems, 337, 590
- Digital communication systems, 4, 294, 577
 - advantages of, 263
- Digital filters, 109
- Digital hierarchy, 346-348
- Digital multiplexing, 342
- Digital service unit (DSU), 363
- Digital signal cross connect, level 1 (DSX1), 363
- Digital signal at level one (DS1), 355
- Digital signal at level zero (DS0), 354
- Digital signals, 3, 20
 - noise immunity of, 4, 263
- Direchlet conditions, 49
 - strong, 49
 - weak, 49
- Direct sequence spread spectrum (DS/SS), 407, 602
- Discrete cosine transform (DCT), 396
- Discrete Fourier transform, 60, 130
- Discrete multitone (DMT), 394
- Discrete-time signals, 20
- Dispersion
 - material, 425
 - multimode, 425
 - profile, 425
 - waveguide, 425
- Distortion, 110
 - in audio and video signals, 104
 - due to multipaths, 114
 - linear, 110
 - nonlinear, 111
- Distortionless transmission, 102
- Double sideband modulation (DSB), 152, 534
- Duality, time frequency, 86
- Duobinary signaling, 295, 317, 527
 - modified, 296n

- Elastic store, 345
- Energy, 15
 - of complex signals, 15
 - of modulated signals, 120
- Energy spectral density (ESD), 116, 121
- Ensemble of a random process, 487
- Entropy of a source, 682
- Envelope detection, 168, 233, 538
- Envelope (group) delay, 104n
- Equalizer, 323-325
 - automatic and adaptive, 326

- least mean square error (LMSE), 326
- zero-forcing, 323
- Equivalent signal set, 662
- Equivocation, 694
- Error-free communication, 690, 699, 712
- Essential bandwidth, 117, 134
- Estimation, linear mean square, 476
- Events, 435
 - compliment of, 435
 - disjoint, 436
 - independent, 439
 - intersection of, 436
 - joint, 436
 - mutually exclusive, 436
 - union of, 435
- Extended superframe (ESF), 277, 356

- Fading, 115
 - selective, 115
- Fast Fourier transform, 61, 135
- Feedback decoding, 755
- Fiber-to-the-curb (FTTC), 393
- Fiber-optic link around the globe (FLAG), 424
- First-order hold circuit, 134, 255
- Folding frequency, 257
- Fortune* magazine, 226
- Fourier series, 43
 - convergence at jump discontinuities, 49
 - effect of symmetry, 53
 - exponential, 53
 - generalized, 43
 - periodicity of, 47
 - trigonometric, 44
 - trigonometric, compact, 45
- Fourier transform, 74
 - conjugate symmetry property of, 75
 - direct, 74
 - discrete, 60
 - duality (symmetry) property, 86
 - existence of, 76
 - fast, 61
 - frequency-convolution property, 98
 - frequency-shifting property, 92
 - inverse, 74
 - linearity of, 76
 - scaling property, 88
 - symmetry (duality) property, 86
 - time-convolution property, 98
 - time-differentiation property, 99
 - time-integration property, 99
 - time-shifting property, 90

- Frame, 355
- Frame relay, 372
- Frequency

- folding, 257
- fundamental, 44
- instantaneous, 208, 209
- n th harmonic, 44
- Frequency compression feedback (FCF), 555
- Frequency converter (mixer), 160, 189
- Frequency-convolution property, 98
- Frequency discriminators, 236
- Frequency-division multiple access (FDMA), 410
- Frequency-division mutiplexing, 12, 189
- Frequency domain description of a signal, 49
- Frequency-hopping spread spectrum system (FH/SS), 411, 606
 - fast hopping, 411
 - slow hopping, 411
- Frequency modulation (FM), 10, 211
- Frequency-shift keying (FSK), 337, 592, 598
- Full cosine roll-off characteristic, 314

- Gate function, 78
- Gaussian probability density, 331, 452
- Gaussian random process, 632
 - bandpass, 519
 - transmission through linear systems, 634
- Generalized angle, 209
- Generalized function, 29
- Generator matrix, 731
- Generator polynomial, 738
- Global positioning system (GPS), 413
- Global system for mobile communications (GSM), 412
- Graded index fiber, 425
- Gram-Schmidt orthogonalization process, 632, 768
- Ground reflected wave, 419
- Ground wave, 419

- Hamming bound, 729
- Hamming codes, 730
- Hamming distance, 691, 729
- Hamming sphere, 729
- High definition television (HDTV), 400
- High-bit rate digital subscriber line (HDSL), 364
- High-density bipolar (HDB) signaling, 308
- Hilbert transform, 173
- Hold-in range of PLL, 186
- Hybrid circuit, 427
- Hybrid fiber coax (HFC), 393

- Ideal filters, 106
- Image station, 190

- Impulse function, 28
- Industrial, scientific, and medical frequency band (ISM), 412
- Information measure, 680
- Instantaneous frequency, 208–210
- Integrated services digital network (ISDN), 369
- Interference
 - in AM systems, 242n
 - in angle-modulated systems, 241
- Interleaving effect, 247
- Intermediate frequency (IF), 189–190, 245
- Interpolation (sinc) function, 80
- Intersymbol interference (ISI), 310

- Jitter, 202, 328

- Laplace probability density function, 565
- Lasers, 423
- Light emitting diodes (LED), 423
- Line coding, 295, 297
- Linear mean-square estimation, 476
- Linear predictor (estimator), 276, 476
- Local access and transport areas (LATA), 432
- Local exchange companies (LEC), 432
- Lock range of PLL, 186
- Logarithmic units, 273
- Low earth orbit (LEO), 405

- Manchester (split-phase) signaling, 304
- M-ary communication, 334, 608
- M-ary message, 4, 334, 608
- Matched filter, 580
- Maximum a posteriori probability (MAP) detector, 643
- Maximum frequency deviation, 217
- Maximum length shift register sequences, 601
- Maximum-likelihood decoding, 734, 750
- Maximum-likelihood receiver, 671
- Mean square bandwidth, 549
- Minimax receiver, 672
- Minimum energy signal set, 664
- Minimum weight vector, 734
- Mobile radio, 404
- Mobile telephone switching office (MTSO), 404
- Modem, 342
- Modified duobinary, 296n
- Modulation, 10, 93
 - amplitude (AM), 93, 162, 537, 718
 - double sideband (DSB), 152, 534
 - frequency (FM), 10, 209
 - phase, (PM), 10, 209
 - single sideband (SSB), 171, 535
 - vestigial sideband (VSB), 179
- Modulation index, 164, 219
- Modulators
 - balanced, 157, 159
 - FM, 229
 - nonlinear, 156
 - switching, 157
- Moments of random variables, 466
- Motion picture experts group (MPEG), 396
- Multiamplitude signaling, 335, 608, 613
- Multipath effect, 114, 605
- Multiphase signaling (MPSK), 611
- Multiple access, 410, 603
- Multiple access interference (MAI), 603
- Multiplexing
 - asynchronous channels, 345
 - digital, 342
 - frequency division, 12, 189
 - time division, 12, 261
- Multitone signaling, 615
- Mutual information, 695

- Narrowband angle modulation, 216, 548
- Natural binary code (NBC), 262
- Near-far problem, 410, 604
- Negative frequency, 57
- Noise, 3
 - Noncoherent detection, 594, 618
 - Nonreturn to zero (NRZ), pulses, 296
 - North American digital hierarchy, 346, 365
 - Null event, 435
 - Nyquist criteria for zero ISI, 310
 - Nyquist interval, 253
 - Nyquist sampling rate, 253

- On-off keying (OOK), 337
- On-off signaling, 295, 304, 587
- Optical carrier (OC-n), 378
- Optimum filters, 522, 580
- Optimum preemphasis-deemphasis, 567
- Optimum receiver, 626
- Optical communication systems, 422
- Orthogonal set, 42
- Orthogonal signals, 33, 24
- Orthogonal signaling, 588, 615, 658
- Orthogonal vectors, 32
- Orthogonality
 - of exponential set, 765
 - of trigonometric set, 764
- Orthonormal set, 42

- Packet switched, 368

- Paley-Wiener criterion, 107, 107n
- Partial response signals, 316
- Parity-check digits, 731
- Parity-check matrix, 733
- Parseval's theorem, 59, 115
- PCM, 6, 262, 557
- PCM repeater, 442
- Permanent virtual circuit (PVC), 373
- Phase delay, 104n
- Phase-lock loop (PLL), 184, 236, 557
 - capture (pull-in) range, 186
 - hold-in (lock) range, 186
- Phase modulation (PM), 10, 209, 219
- Phase-shift keying (PSK), 337, 339, 591
- Phase spectrum, 49, 75
- Plesiochronous, 367
- Pointer, 381
- Polar signaling, 295, 302, 586
- Power
 - of complex signals, 15
 - of modulated signals, 130
 - of real signals, 15
- Power spectral density, 123, 496
 - interpretation of, 126
- Prediction coefficients, 276
- Preemphasis-deemphasis (PDE), 243, 567
- Preenvelope, 172n
- Primary rate interface (PRI), 369
- Probability density function (PDF), 331, 450
 - conditional, 439, 461
 - joint, 459
- Pseudonoise (PN) sequence, 407
 - generation of, 600
- Pseudoternary (bipolar) signaling, 295
- Public switched telephone network (PSTN), 430
- Pull-in range of PLL, 186
- Pulse-amplitude modulation (PAM), 260
- Pulse-code modulation (PCM), 6, 262, 557
 - companded, 268, 563
 - noise due to detection errors, 468–469
 - output SNR, 272, 559, 565
 - of quantization noise, 266, 468, 565
 - synchronizing and signaling, 276
- Pulse-position modulation (PPM), 260
- Pulse shaping, 310
- Pulse stuffing, 345
- Pulse-width (or duration) modulation (PWM), 260

- Quadrature-amplitude modulation (QAM), 170, 341, 613, 652
- Quadrature multiplexing, 171
- Qualcomm code-excited linear prediction (QCELP), 413

- Quantization, 5, 262
- Quantization noise, 265, 467, 565

- Raised cosine characteristics, 314
- Rake receiver, 410, 606
- Random processes, 487–495
 - bandpass, 514
 - classification of, 492
 - cross-power spectral density, 510
 - cyclostationary, 504n
 - ergodic, 494
 - gaussian, 519, 633
 - geometrical representation of, 635
 - incoherent, 509
 - independent, 509
 - mean square value (power) of, 499
 - orthogonal, 509
 - power spectral density (PSD), 499
 - stationary, 492
 - uncorrelated, 509
 - wide-sense (weakly) stationary, 493
- Random variables, 445
 - continuous, 446
 - correlation of, 473
 - discrete, 445
 - means, 463
- Rate of communication, 8, 696, 709, 711
- Ratio detector, 236
- Rayleigh probability density, 461, 595
- Realizable systems, 107
- Rectifier detector, 167
- Redundancy, 13, 687
- Regeneration, 358
- Regenerative repeater, 4, 263, 322
- Relative frequency, 436
- Return to zero (RZ) pulses, 296
- Rice probability density, 521, 595
- Robbed bit signaling, 277, 356
- Roll-off factor, 314

- Sample function, 487
- Sample point, 435
- Sample space, 435
- Sampling, instantaneous, 251
 - practical, 258
- Sampling (sifting) property, 28
- Sampling theorem, 5, 251
- Satellite communication systems, 421
- Scaling property of the Fourier transform, 88
- Scatter diagram, 474
- Schwartz inequality, 36, 568, 578, 766
- Scrambling, 319
- Sequential decoding, 752

Sidebands, 152
 Signal classification, 20
 aperiodic, 20, 21
 analog, 20
 continuous-time, 20
 deterministic, 24, 434
 digital, 20
 discrete-time, 20
 energy, 20, 23
 periodic, 20, 21
 power, 20, 23
 random, 20, 24, 434
 Signal comparison (correlation), 35
 Signal component, 32
 Signal distortion
 linear, 110
 multipath effect, 114
 nonlinear, 111
 Signal energy, 15, 23
 Signal power, 15, 23
 Signal size, 14
 Signal-to-noise ratio (SNR), 3, 8, 14
 in amplitude-modulated systems, 534–541
 in angle-modulated systems, 541–552
 in baseband systems, 532
 in DM, 286
 exchange with bandwidth, 12, 611, 619, 718
 in PCM, 272
 Signal space, 626, 642
 Signaling system 7 (SS7), 383
 Signals
 analog, 3, 20
 bandpass, 95
 digital, 3, 20
 energy, 15
 periodic, 21
 power, 15
 vector representation of, 628
 Sign (sgn) function, 84
 Simplex signal set, 668
 Sinc function, 80
 Single sideband modulation (SSB), 171–179, 535
 Sky wave, 419
 Slope detector, 236
 Source encoding, 684
 Spectrum
 amplitude, 49, 75
 continuous, 75
 discrete (or line), 49
 energy density, 116
 exponential Fourier, 55
 phase, 49, 75
 power density, 123
 Split-phase (twinned binary) signaling, 304

Spread spectrum systems, 406
 direct sequence (DS/SS), 407, 602
 frequency-hopping, (FH/SS), 411, 606
 Square-law detector, 205
 Standard deviation, 467
 Statistical time division multiplexing (STATDM), 383
 Stochastic processes, 487
 Superframe, 277, 355
 extended, 277, 357
 Superheterodyne receiver, 189
 Switched circuit, 368
 Switched multi-megabit data service (SMDS), 391
 Symmetry (duality) property, 86
 Synchronization, 276, 328, 622
 Synchronous detection or demodulation, 161
 Synchronous optical network (SONET), 379
 Synchronous payload envelope (SPE), 379
 Synchronous transport signal (STS-n), 378
 Synchronous satellite, 421
 Syndrome, 733
 Systematic codes, 731, 740

 T-1 carrier system, 274–278, 343, 346, 347, 355
 Television, 191–201
 Thermal noise, 511, 528
 Threshold detection, 39, 455, 579
 Threshold effect, 540, 553, 560
 Time assignment system interpolation, 430
 Time convolution property, 98
 Time correlation, 121, 125
 Time-differentiation property, 99
 Time-division multiple-access (TDMA) systems, 345, 410, 413
 Time-division multiplexing, 12, 261
 Time domain description of a signal, 49
 Time-integration property, 99
 Time-shifting property, 90
 Timing extraction, 328
 Timing jitter, 328, 329
 Tone modulation, 164, 220
 Transmission media, 416–427
 Transparency, 297
 Transversal filter, 115, 148, 279, 319
 Trellis diagram, 751
 Triangle function, 79
 Tropospheric scattering, 421
 Tropospheric wave, 419
 Trunk circuit, 432
 Twinned binary (split-phase or Manchester) signaling, 304

Unrealizable systems, 107

Variance, 467
 Vector space, 41, 626–628
 Very high data rate ADSL (VDSL), 395
 Vestigal sideband modulation, 179–182
 Vestigal spectrum, 314
 Video-on-demand (VOD), 392
 Virtual channel, (VC), 384
 Virtual path (VP), 384
 Virtual tributary (VT), 379
 Viterbi algorithm, 750

Voltage-controlled oscillator (VCO), 185, 236, 555

 Weaver's method of SSB generation, 176
 Wiener-Hopf filter, 522
 Wiener-Khinchine relation, 499

 x-DSL, 393

 Zero padding, 132
 Zero-forcing equalizer, 323

Modern Digital and Analog Communication Systems

THIRD EDITION

B.P.Lathi

T.Srinivasa Rao
Dept. of ECE
BEC, Bapatla.
hodibaaba@gmail.com