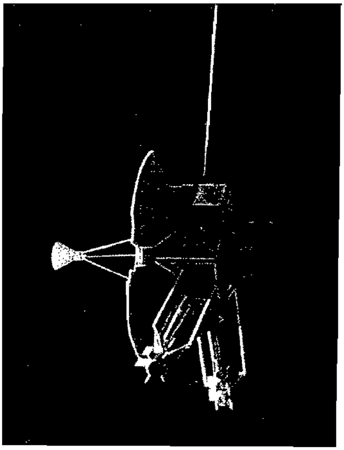

COMMUNICATION SYSTEMS

4th Edition

Simon Haykin

COMMUNICATION SYSTEMS

Haykin

4th Ed

TABLE 9.3 *Average probability of error for repetition code*

Code Rate, r = 1/n	*Average Probability of Error,* P_e
1	10^{-2}
$\frac{1}{3}$	3×10^{-4}
$\frac{1}{5}$	10^{-6}
$\frac{1}{7}$	4×10^{-7}
$\frac{1}{9}$	10^{-8}
$\frac{1}{11}$	5×10^{-10}

Consider first a binary symmetric channel with transition probability $p = 10^{-2}$. For this value of p, we find from Equation (9.60) that the channel capacity $C = 0.9192$. Hence, from the channel coding theorem, we may state that for any $\epsilon > 0$ and $r \leq 0.9192$, there exists a code of large enouigh length n and code rate r, and an appropriate decoding algorithm, such that when the coded bit stream is sent over the given channel, the average probability of channel decoding error is less than ϵ. This result is depicted in Figure 9.12 for the limiting value $\epsilon = 10^{-8}$.

To put the significance of this result in perspective, consider next a simple coding scheme that involves the use of a *repetition code*, in which each bit of the message is repeated several times. Let each bit (0 or 1) be repeated n times, where $n = 2m + 1$ is an odd integer. For example, for $n = 3$, we transmit 0 and 1 as 000 and 111, respectively. Intuitively, it would seem logical to use a *majority rule* for decoding, which operates as follows: *If in a block of n received bits (representing one bit of the message), the number of 0s exceeds the number of 1s, the decoder decides in favor of a 0. Otherwise, it decides in favor of a 1.* Hence, an error occurs when $m + 1$ or more bits out of $n = 2m + 1$ bits are received incorrectly. Because of the assumed symmetric nature of the channel, the *average probability of error* P_e is independent of the *a priori* probabilities of 0 and 1. Accordingly, we find that P_e is given by (see Problem 9.24)

$$P_e = \sum_{i=m+1}^{n} \binom{n}{i} p^i (1-p)^{n-i} \tag{9.65}$$

where p is the transition probability of the channel.

Table 9.3 gives the average probability of error P_e for a repetition code, which is calculated by using Equation (9.65) for different values of the code rate r. The values given here assume the use of a binary symmetric channel with transition probability $p = 10^{-2}$. The improvement in reliability displayed in Table 9.3 is achieved at the cost of decreasing code rate. The results of this table are also shown plotted as the curve labeled "repetition code" in Figure 9.12. This curve illustrates the *exchange of code rate for message reliability*, which is a characteristic of repetition codes.

This example highlights the unexpected result presented to us by the channel coding theorem. The result is that it is not necessary to have the code rate r approach zero (as in the case of repetition codes) so as to achieve more and more reliable operation of the communication link. The theorem merely requires that the code rate be less than the channel capacity C. ◀

9.9 *Differential Entropy and Mutual Information for Continuous Ensembles*

The sources and channels considered in our discussion of information-theoretic concepts thus far have involved ensembles of random variables that are *discrete* in amplitude. In this section, we extend some of these concepts to *continuous* random variables and random vectors. The motivation for doing so is to pave the way for the description of another fundamental limit in information theory, which we take up in Section 9.10.

Consider a continuous random variable X with the *probability density function* $f_X(x)$. By analogy with the entropy of a discrete random variable, we introduce the following definition:

$$h(X) = \int_{-\infty}^{\infty} f_X(x) \log_2\left[\frac{1}{f_X(x)}\right] dx \tag{9.66}$$

We refer to $h(X)$ as the *differential entropy of* X to distinguish it from the ordinary or *absolute entropy*. We do so in recognition of the fact that although $h(X)$ is a useful mathematical quantity to know, it is *not* in any sense a measure of the randomness of X. Nevertheless, we justify the use of Equation (9.66) in what follows. We begin by viewing the continuous random variable X as the limiting form of a discrete random variable that assumes the value $x_k = k\,\Delta x$, where $k = 0, \pm 1, \pm 2, \ldots$, and Δx approaches zero. By definition, the continuous random variable X assumes a value in the interval $[x_k, x_k + \Delta x]$ with probability $f_X(x_k)\,\Delta x$. Hence, permitting Δx to approach zero, the ordinary entropy of the continuous random variable X may be written in the limit as follows:

$$\begin{aligned}
H(X) &= \lim_{\Delta x \to 0} \sum_{k=-\infty}^{\infty} f_X(x_k)\,\Delta x \log_2\left(\frac{1}{f_X(x_k)\,\Delta x}\right) \\
&= \lim_{\Delta x \to 0} \left[\sum_{k=-\infty}^{\infty} f_X(x_k) \log_2\left(\frac{1}{f_X(x_k)}\right) \Delta x - \log_2 \Delta x \sum_{k=-\infty}^{\infty} f_X(x_k)\,\Delta x\right] \\
&= \int_{-\infty}^{\infty} f_X(x) \log_2\left(\frac{1}{f_X(x)}\right) dx - \lim_{\Delta x \to 0} \log_2 \Delta x \int_{-\infty}^{\infty} f_X(x)\,dx \\
&= h(X) - \lim_{\Delta x \to 0} \log_2 \Delta x
\end{aligned} \tag{9.67}$$

where, in the last line, we have made use of Equation (9.66) and the fact that the total area under the curve of the probability density function $f_X(x)$ is unity. In the limit as Δx approaches zero, $-\log_2 \Delta x$ approaches infinity. This means that the entropy of a continuous random variable is infinitely large. Intuitively, we would expect this to be true, because a continuous random variable may assume a value anywhere in the interval $(-\infty, \infty)$ and the uncertainty associated with the variable is on the order of infinity. We avoid the problem associated with the term $\log_2 \Delta x$ by adopting $h(X)$ as a differential entropy, with the term $-\log_2 \Delta x$ serving as reference. Moreover, since the information transmitted over a channel is actually the difference between two entropy terms that have a common reference, the information will be the same as the difference between the corresponding differential entropy terms. We are therefore perfectly justified in using the term $h(X)$, defined in Equation (9.66), as the differential entropy of the continuous random variable X.

When we have a continuous random vector $\mathbf{X}$ consisting of n random variables $X_1, X_2, \ldots, X_n$, we define the differential entropy of $\mathbf{X}$ as the *n-fold integral*

$$h(\mathbf{X}) = \int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x}) \log_2\left[\frac{1}{f_{\mathbf{X}}(\mathbf{x})}\right] d\mathbf{x} \tag{9.68}$$

where $f_{\mathbf{X}}(\mathbf{x})$ is the *joint probability density function* of $\mathbf{X}$.

▶ EXAMPLE 9.7 Uniform Distribution

Consider a random variable X uniformly distributed over the interval $(0, a)$. The probability density function of X is

$$f_X(x) = \begin{cases} \dfrac{1}{a}, & 0 < x < a \\ 0, & \text{otherwise} \end{cases}$$

Applying Equation (9.66) to this distribution, we get

$$\begin{aligned} h(X) &= \int_0^a \frac{1}{a} \log(a)\, dx \\ &= \log a \end{aligned} \tag{9.69}$$

Note that $\log a < 0$ for $a < 1$. Thus this example shows that, unlike a discrete random variable, the differential entropy of a continuous random variable can be negative. ◀

▶ EXAMPLE 9.8 Gaussian Distribution

Consider an arbitrary pair of random variables X and Y, whose probability density functions are respectively denoted by $f_Y(x)$ and $f_X(x)$ where x is merely a dummy variable. Adapting the fundamental inequality of Equation (9.12) to the situation at hand, we may write[9]

$$\int_{-\infty}^{\infty} f_Y(x) \log_2\left(\frac{f_X(x)}{f_Y(x)}\right) dx \le 0 \tag{9.70}$$

or, equivalently,

$$-\int_{-\infty}^{\infty} f_Y(x) \log_2 f_Y(x)\, dx \le -\int_{-\infty}^{\infty} f_Y(x) \log_2 f_X(x)\, dx \tag{9.71}$$

The quantity on the left-hand side of Equation (9.71) is the differential entropy of the random variable Y; hence,

$$h(Y) \le -\int_{-\infty}^{\infty} f_Y(x) \log_2 f_X(x)\, dx \tag{9.72}$$

Suppose now the random variables X and Y are described as follows:

- The random variables X and Y have the *same mean* μ and the *same variance* σ^2.
- The random variable X is *Gaussian distributed* as shown by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{9.73}$$

Hence, substituting Equation (9.73) into Equation (9.72), and changing the base of the logarithm from 2 to $e = 2.7183$, we get

$$h(Y) \le -\log_2 e \int_{-\infty}^{\infty} f_Y(x)\left(-\frac{(x-\mu)^2}{2\sigma^2} - \log(\sqrt{2\pi}\sigma)\right) dx \tag{9.74}$$

We now recognize the following properties of the random variable Y (given that its mean is μ and its variance is σ^2):

$$\begin{aligned} \int_{-\infty}^{\infty} f_Y(x)\, dx &= 1 \\ \int_{-\infty}^{\infty} (x-\mu)^2 f_Y(x)\, dx &= \sigma^2 \end{aligned}$$

We may therefore simplify Equation (9.74) as

$$h(Y) \le \tfrac{1}{2}\log_2(2\pi e\sigma^2) \tag{9.75}$$

The quantity on the right-hand side of Equation (9.75) is in fact the differential entropy of the Gaussian random variable X:

$$h(X) = \tfrac{1}{2}\log_2(2\pi e\sigma^2) \tag{9.76}$$

Finally, combining Equations (9.75) and (9.76), we may write

$$h(Y) \le h(X), \qquad \begin{cases} X\text{: Gaussian random variable} \\ Y\text{: another random variable} \end{cases} \tag{9.77}$$

where equality holds if, and only if, $Y = X$.

We may now summarize the results of this important example as two entropic properties of a Gaussian random variable:

1. For a finite variance σ^2, the Gaussian random variable has the largest differential entropy attainable by any random variable.
2. The entropy of a Gaussian random variable X is uniquely determined by the variance of X (i.e., it is independent of the mean of X).

Indeed, it is because of Property 1 that the Gaussian channel model is so widely used as a conservative model in the study of digital communication systems. ◀

■ MUTUAL INFORMATION

Consider next a pair of continuous random variables X and Y. By analogy with Equation (9.47), we define the *mutual information* between the random variables X and Y as follows:

$$I(X;Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\log_2\left[\frac{f_X(x|y)}{f_X(x)}\right] dx\, dy \tag{9.78}$$

where $f_{X,Y}(x, y)$ is the joint probability density function of X and Y, and $f_X(x|y)$ is the conditional probability density function of X, given that $Y = y$. Also, by analogy with Equations (9.45), (9.50), (9.43), and (9.44) we find that the mutual information $I(X; Y)$ has the following properties:

$$1.\ I(X;Y) = I(Y;X) \tag{9.79}$$

$$2.\ I(X;Y) \ge 0 \tag{9.80}$$

$$\begin{aligned} 3.\ I(X;Y) &= h(X) - h(X|Y) \\ &= h(Y) - h(Y|X) \end{aligned} \tag{9.81}$$

The parameter $h(X)$ is the differential entropy of X; likewise for $h(Y)$. The parameter $h(X|Y)$ is the *conditional differential entropy* of X, given Y; it is defined by the double integral (see Equation (9.41))

$$h(X|Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\log_2\left[\frac{1}{f_X(x|y)}\right] dx\, dy \tag{9.82}$$

The parameter $h(Y|X)$ is the conditional differential entropy of Y, given X; it is defined in a manner similar to $h(X|Y)$.

9.10 *Information Capacity Theorem*

In this section, we use the idea of mutual information to formulate the information capacity theorem for *band-limited, power-limited Gaussian channels*. To be specific, consider a zero-mean stationary process $X(t)$ that is band-limited to B hertz. Let X_k, $k = 1, 2, \ldots, K$, denote the continuous random variables obtained by uniform sampling of the process $X(t)$ at the Nyquist rate of $2B$ samples per second. These samples are transmitted in T seconds over a noisy channel, also band-limited to B hertz. Hence, the number of samples, K, is given by

$$K = 2BT \tag{9.83}$$

We refer to X_k as a sample of the *transmitted signal*. The channel output is perturbed by *additive white Gaussian noise* (AWGN) of zero mean and power spectral density $N_0/2$. The noise is band-limited to B hertz. Let the continuous random variables Y_k, $k = 1, 2, \ldots, K$ denote samples of the received signal, as shown by

$$Y_k = X_k + N_k, \qquad k = 1, 2, \ldots, K \tag{9.84}$$

The noise sample N_k is Gaussian with zero mean and variance given by

$$\sigma^2 = N_0B \tag{9.85}$$

We assume that the samples Y_k, $k = 1, 2, \ldots, K$ are statistically independent.

A channel for which the noise and the received signal are as described in Equations (9.84) and (9.85) is called a *discrete-time, memoryless Gaussian channel*. It is modeled as in Figure 9.13. To make meaningful statements about the channel, however, we have to assign a *cost* to each channel input. Typically, the transmitter is *power limited*; it is therefore reasonable to define the cost as

$$E[X_k^2] = P, \qquad k = 1, 2, \ldots, K \tag{9.86}$$

where P is the *average transmitted power*. The *power-limited Gaussian channel* described herein is of not only theoretical but also practical importance in that it models many communication channels, including line-of-sight radio and satellite links.

The *information capacity* of the channel is defined as the maximum of the mutual information between the channel input X_k and the channel output Y_k over all distributions on the input X_k that satisfy the power constraint of Equation (9.86). Let $I(X_k; Y_k)$ denote

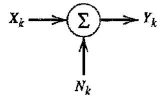

FIGURE 9.13 Model of discrete-time, memoryless Gaussian channel.

the mutual information between X_k and Y_k. We may then define the information capacity of the channel as

$$C = \max_{f_{X_k}(x)} \{I(X_k; Y_k): E[X_k^2] = P\} \tag{9.87}$$

where the maximization is performed with respect to $f_{X_k}(x)$, the probability density function of X_k.

The mutual information $I(X_k; Y_k)$ can be expressed in one of the two equivalent forms shown in Equation (9.81). For the purpose at hand, we use the second line of this equation and so write

$$I(X_k; Y_k) = h(Y_k) - h(Y_k | X_k) \tag{9.88}$$

Since X_k and N_k are independent random variables, and their sum equals Y_k, as in Equation (9.84), we find that the conditional differential entropy of Y_k, given X_k, is equal to the differential entropy of N_k (see Problem 9.28):

$$h(Y_k \mid X_k) = h(N_k) \tag{9.89}$$

Hence, we may rewrite Equation (9.88) as

$$I(X_k; Y_k) = h(Y_k) - h(N_k) \tag{9.90}$$

Since $h(N_k)$ is independent of the distribution of X_k, maximizing $I(X_k; Y_k)$ in accordance with Equation (9.87) requires maximizing $h(Y_k)$, the differential entropy of sample Y_k of the received signal. For $h(Y_k)$ to be maximum, Y_k has to be a Gaussian random variable (see Example 9.8). That is, the samples of the received signal represent a noiselike process. Next, we observe that since N_k is Gaussian by assumption, the sample X_k of the transmitted signal must be Gaussian too. We may therefore state that the maximization specified in Equation (9.87) is attained by choosing the samples of the transmitted signal from a noiselike process of average power P. Correspondingly, we may reformulate Equation (9.87) as

$$C = I(X_k; Y_k): X_k \text{ Gaussian}, \quad E[X_k^2] = P \tag{9.91}$$

where the mutual information $I(X_k; Y_k)$ is defined in accordance with Equation (9.90).

For the evaluation of the information capacity C, we proceed in three stages:

1. The variance of sample Y_k of the received signal equals $P + \sigma^2$. Hence, the use of Equation (9.76) yields the differential entropy of Y_k as

$$h(Y_k) = \tfrac{1}{2} \log_2[2\pi e(P + \sigma^2)] \tag{9.92}$$

2. The variance of the noise sample N_k equals σ^2. Hence, the use of Equation (9.76) yields the differential entropy of N_k as

$$h(N_k) = \tfrac{1}{2} \log_2(2\pi e\sigma^2) \tag{9.93}$$

3. Substituting Equations (9.92) and (9.93) into Equation (9.90) and recognizing the definition of information capacity given in Equation (9.91), we get the desired result:

$$C = \frac{1}{2} \log_2\left(1 + \frac{P}{\sigma^2}\right) \text{ bits per transmission} \tag{9.94}$$

With the channel used K times for the transmission of K samples of the process $X(t)$ in T seconds, we find that the information capacity per unit time is (K/T) times the result given in Equation (9.94). The number K equals $2BT$, as in Equation (9.83). Accordingly, we may express the information capacity in the equivalent form:

$$C = B \log_2\left(1 + \frac{P}{N_0 B}\right) \text{ bits per second} \tag{9.95}$$

where we have used Equation (9.85) for the noise variance σ^2.

Based on the formula of Equation (9.95), we may now state Shannon's third (and most famous) theorem, the *information capacity theorem*,[10] as follows:

The information capacity of a continuous channel of bandwidth B hertz, perturbed by additive white Gaussian noise of power spectral density $N_0/2$ and limited in bandwidth to B, is given by

$$C = B \log_2\left(1 + \frac{P}{N_0 B}\right) \text{ bits per second}$$

where P is the average transmitted power.

The information capacity theorem is one of the most remarkable results of information theory for, in a single formula, it highlights most vividly the interplay among three key system parameters: channel bandwidth, average transmitted power (or, equivalently, average received signal power), and noise power spectral density at the channel output. The dependence of information capacity C on channel bandwidth B is *linear*, whereas its dependence on signal-to-noise ratio P/N_0B is *logarithmic*. Accordingly, *it is easier to increase the information capacity of a communication channel by expanding its bandwidth than increasing the transmitted power for a prescribed noise variance.*

The theorem implies that, for given average transmitted power P and channel bandwidth B, we can transmit information at the rate of C bits per second, as defined in Equation (9.95), with arbitrarily small probability of error by employing sufficiently complex encoding systems. It is not possible to transmit at a rate higher than C bits per second by any encoding system without a definite probability of error. Hence, the channel capacity theorem defines the *fundamental limit* on the rate of error-free transmission for a power-limited, band-limited Gaussian channel. To approach this limit, however, the transmitted signal must have statistical properties approximating those of white Gaussian noise.

Sphere Packing[11]

To provide a plausible argument supporting the information capacity theorem, suppose that we use an encoding scheme that yields K code words, one for each sample of the transmitted signal. Let n denote the length (i.e., the number of bits) of each code word. It is presumed that the coding scheme is designed to produce an acceptably low probability of symbol error. Furthermore, the code words satisfy the power constraint; that is, the average power contained in the transmission of each code word with n bits is nP, where P is the average power per bit.

Suppose that any code word in the code is transmitted. The received vector of n bits is Gaussian distributed with mean equal to the transmitted code word and variance equal to $n\sigma^2$, where σ^2 is the noise variance. With high probability, the received vector lies inside a sphere of radius $\sqrt{n\sigma^2}$, centered on the transmitted code word. This sphere is itself contained in a larger sphere of radius $\sqrt{n(P + \sigma^2)}$, where $n(P + \sigma^2)$ is the average power of the received vector.

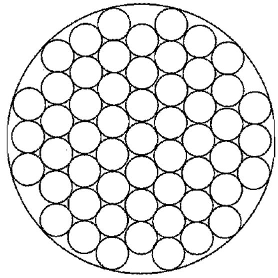

FIGURE 9.14 The sphere-packing problem.

We may thus visualize the picture portrayed in Figure 9.14. With everything inside a small sphere of radius $\sqrt{n\sigma^2}$ assigned to the code word on which it is centered, it is reasonable to say that when this particular code word is transmitted, the probability that the received vector will lie inside the correct "decoding" sphere is high. The key question is: How many decoding spheres can be packed inside the larger sphere of received vectors? In other words, how many code words can we in fact choose? To answer this question, we first recognize that the volume of an n-dimensional sphere of radius r may be written as $A_n r^n$, where A_n is a scaling factor. We may therefore make the following statements:

- The volume of the sphere of received vectors is $A_n[n(P + \sigma^2)]^{n/2}$.
- The volume of the decoding sphere is $A_n(n\sigma^2)^{n/2}$.

Accordingly, it follows that the maximum number of *nonintersecting* decoding spheres that can be packed inside the sphere of possible received vectors is

$$\begin{aligned}\frac{A_n[n(P + \sigma^2)]^{n/2}}{A_n(n\sigma^2)^{n/2}} &= \left(1 + \frac{P}{\sigma^2}\right)^{n/2} \\ &= 2^{(n/2)\log_2(1+P/\sigma^2)}\end{aligned} \tag{9.96}$$

Taking the logarithm of this result to base 2, we readily see that the maximum number of bits per transmission for a low probability of error is indeed as defined previously in Equation (9.94).

▶ EXAMPLE 9.9 Reconfiguration of Constellation for Reduced Power

To illustrate the idea of sphere packing, consider the 64-QAM square constellation of Figure 9.15*a*. The figure depicts two-dimensional nonintersecting decoding spheres centered on the message points in the constellation. In trying to pack the decoding spheres as tightly as possible while maintaining the same Euclidean distance between the message points as before, we obtain the alternative constellation shown in Figure 9.15*b*. With a common Euclidean distance between the message points, the two constellations of Figure 9.15 produce approximately the same bit error rate, assuming the use of a high enough signal-to-noise ratio over an AWGN channel; see, for example, Equation (5.95). However, comparing these two constellations, we find that the sum of squared Euclidean distances from the message points to the origin in Figure 9.15*b* is smaller than that in Figure 9.15*a*. It follows therefore that the tightly packed constellation of Figure 9.15*b* has an advantage over the square constellation

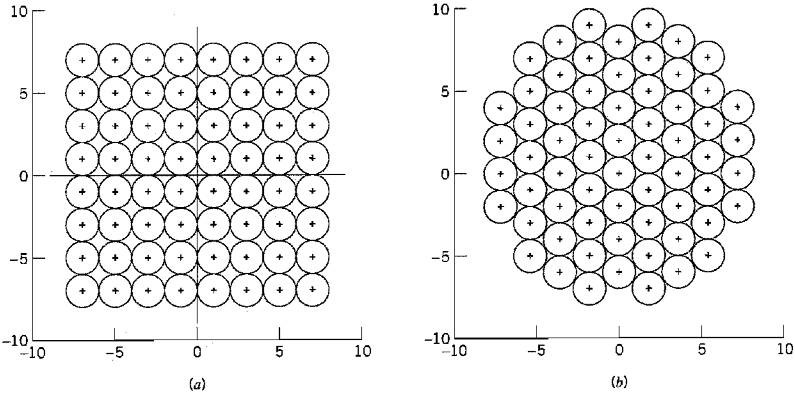

FIGURE 9.15 (*a*) Square 64-QAM constellation. (*b*) The most tightly coupled alternative to that of part *a*.

of Figure 9.15*a*: a smaller transmitted average signal energy per symbol for the same bit error rate on an AWGN channel. ◀

9.11 *Implications of the Information Capacity Theorem*

Now that we have an intuitive feel for the information capacity theorem, we may go on to discuss its implications in the context of a Gaussian channel that is limited in both power and bandwidth. For the discussion to be useful, however, we need an ideal framework against which the performance of a practical communication system can be assessed. To this end, we introduce the notion of an *ideal system* defined as one that transmits data at a bit rate R_b equal to the information capacity C. We may then express the average transmitted power as

$$P = E_b C \tag{9.97}$$

where E_b is the transmitted energy per bit. Accordingly, the ideal system is defined by the equation

$$\frac{C}{B} = \log_2\left(1 + \frac{E_b}{N_0}\frac{C}{B}\right) \tag{9.98}$$

Equivalently, we may define the *signal energy-per-bit to noise power spectral density ratio* E_b/N_0 in terms of the ratio C/B for the ideal system as

$$\frac{E_b}{N_0} = \frac{2^{C/B} - 1}{C/B} \tag{9.99}$$

A plot of bandwidth efficiency R_b/B versus E_b/N_0 is called the *bandwidth-efficiency diagram*. A generic form of this diagram is displayed in Figure 9.16, where the curve labeled "capacity boundary" corresponds to the ideal system for which $R_b = C$. Based on Figure 9.16, we can make the following observations:

1. For *infinite bandwidth*, the ratio E_b/N_0 approaches the limiting value

$$\begin{aligned}\left(\frac{E_b}{N_0}\right)_\infty &= \lim_{B\to\infty}\left(\frac{E_b}{N_0}\right) \\ &= \log 2 = 0.693\end{aligned} \tag{9.100}$$

This value is called the *Shannon limit* for an AWGN channel, assuming a code rate of zero. Expressed in decibels, it equals -1.6 dB. The corresponding limiting value of the channel capacity is obtained by letting the channel bandwidth B in Equation (9.95) approach infinity; we thus find that

$$\begin{aligned}C_\infty &= \lim_{B\to\infty} C \\ &= \frac{P}{N_0}\log_2 e\end{aligned} \tag{9.101}$$

where e is the base of the natural logarithm.

2. The *capacity boundary*, defined by the curve for the critical bit rate $R_b = C$, separates combinations of system parameters that have the potential for supporting error-free transmission ($R_b < C$) from those for which error-free transmission is not possible ($R_b > C$). The latter region is shown shaded in Figure 9.16.
3. The diagram highlights potential *trade-offs* among E_b/N_0, R_b/B, and probability of symbol error P_e. In particular, we may view movement of the operating point along

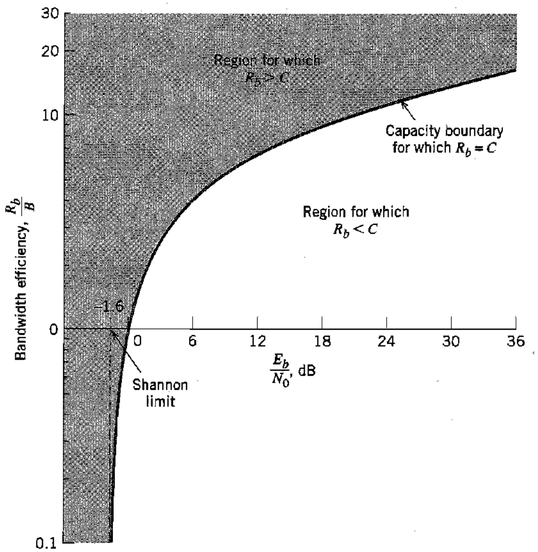

FIGURE 9.16 Bandwidth-efficiency diagram.

a horizontal line as trading P_e versus E_b/N_0 for a fixed R_b/B. On the other hand, we may view movement of the operating point along a vertical line as trading P_e versus R_b/B for a fixed E_b/N_0.

▶ EXAMPLE 9.10 *M*-ary PCM

In this example, we look at an M-ary PCM system in light of the channel capacity theorem under the assumption that the system operates above the error threshold. That is, the average probability of error due to channel noise is negligible.

We assume that the M-ary PCM system uses a code word consisting of n code elements, each having one of M possible discrete amplitude levels; hence the name "M-ary." From Chapter 3 we recall that for a PCM system to operate above the error threshold, there must be provision for a noise margin that is sufficiently large to maintain a negligible error rate due to channel noise. This, in turn, means there must be a certain separation between these M discrete amplitude levels. Call this separation $k\sigma$, where k is a constant and $\sigma^2 = N_0B$ is the noise variance measured in a channel bandwidth B. The number of amplitude levels M is usually an integer power of 2. The average transmitted power will be least if the amplitude range is symmetrical about zero. Then the discrete amplitude levels, normalized with respect to the separation $k\sigma$, will have the values $\pm 1/2$, $+3/2, \ldots, \pm (M-1)/2$. We assume that these M different amplitude levels are equally likely. Accordingly, we find that the average transmitted power is given by

$$\begin{aligned} P &= \frac{2}{M}\left[\left(\frac{1}{2}\right)^2 + \left(\frac{3}{2}\right)^2 + \cdots + \left(\frac{M-1}{2}\right)^2\right](k\sigma)^2 \\ &= k^2\sigma^2\left(\frac{M^2-1}{12}\right) \end{aligned} \tag{9.102}$$

Suppose that the M-ary PCM system described herein is used to transmit a message signal with its highest frequency component equal to W hertz. The signal is sampled at the Nyquist rate of $2W$ samples per second. We assume that the system uses a quantizer of the midrise type, with L equally likely representation levels. Hence, the probability of occurrence of any one of the L representation levels is $1/L$. Correspondingly, the amount of information carried by a single sample of the signal is $\log_2 L$ bits. With a maximum sampling rate of $2W$ samples per second, the maximum rate of information transmission of the PCM system, measured in bits per second, is given by

$$R_b = 2W \log_2 L \text{ bits per second} \tag{9.103}$$

Since the PCM system uses a code word consisting of n code elements, each having one of M possible discrete amplitude values, we have M^n different possible code words. For a unique encoding process, we require

$$L = M^n \tag{9.104}$$

Clearly, the rate of information transmission in the system is unaffected by the use of an encoding process. We may therefore eliminate L between Equations (9.103) and (9.104) to obtain

$$R_b = 2Wn \log_2 M \text{ bits per second} \tag{9.105}$$

Equation (9.102) defines the average transmitted power required to maintain an M-ary PCM system operating above the error threshold. Hence, solving this equation for the number of discrete amplitude levels, M, we get

$$M = \left(1 + \frac{12P}{k^2N_0B}\right)^{1/2} \tag{9.106}$$

where $\sigma^2 = N_0 B$ is the variance of the channel noise measured in a bandwidth B. Therefore, substituting Equation (9.106) into Equation (9.105), we obtain

$$R_b = Wn \log_2\left(1 + \frac{12P}{k^2 N_0 B}\right) \tag{9.107}$$

The channel bandwidth B required to transmit a rectangular pulse of duration $1/2nW$ (representing a code element in the code word) is given by (see Chapter 3)

$$B = \kappa n W$$

where κ is a constant with a value lying between 1 and 2. Using the minimum possible value $\kappa = 1$, we find that the channel bandwidth $B = nW$. We may thus rewrite Equation (9.107) as

$$R_b = B \log_2\left(1 + \frac{12P}{k^2 N_0 B}\right) \tag{9.108}$$

The *ideal system* is described by Shannon's channel capacity theorem, given in Equation (9.95). Hence, comparing Equation (9.108) with Equation (9.95), we see that they are identical if the average transmitted power in the PCM system is increased by the factor $k^2/12$, compared with the ideal system. Perhaps the most interesting point to note about Equation (9.108) is that the form of the equation is right: *Power and bandwidth in a PCM system are exchanged on a logarithmic basis, and the information capacity C is proportional to the channel bandwidth B.* ◀

▶ EXAMPLE 9.11 *M*-ary PSK and *M*-ary FSK

In this example, we compare the bandwidth-power exchange capabilities of M-ary PSK and M-ary FSK signals in light of Shannon's information capacity theorem. Consider first a coherent M-ary PSK system that employs a *nonorthogonal* set of M phase-shifted signals for the transmission of binary data. Each signal in the set represents a symbol with $\log_2 M$ bits. Using the definition of null-to-null bandwidth, we may express the bandwidth efficiency of M-ary PSK as follows [see Equation (6.51)]:

$$\frac{R_b}{B} = \frac{\log_2 M}{2}$$

In Figure 9.17*a*, we show the operating points for different numbers of phase levels $M = 2$, 4, 8, 16, 32, 64. Each point corresponds to an average probability of symbol error $P_e = 10^{-5}$. In the figure we have also included the capacity boundary for the ideal system. We observe from Figure 9.17*a* that as M is increased, the bandwidth efficiency is improved, but the value of E_b/N_0 required for error-free transmission moves away from the Shannon limit.

Consider next a coherent M-ary FSK system that uses an *orthogonal* set of M frequency-shifted signals for the transmission of binary data, with the separation between adjacent signal frequencies set at $1/2T$, where T is the symbol period. As with the M-ary PSK, each signal in the set represents a symbol with $\log_2 M$ bits. The bandwidth efficiency of M-ary FSK is as follows [see Equation (6.143)]:

$$\frac{R_b}{B} = \frac{2 \log_2 M}{M}$$

In Figure 9.17*b*, we show the operating points for different numbers of frequency levels $M =$ 2, 4, 8, 16, 32, 64 for an average probability of symbol error $P_e = 10^{-5}$. In the figure, we have also included the capacity boundary for the ideal system. We see that increasing M in (orthogonal) M-ary FSK has the opposite effect to that in (nonorthogonal) M-ary PSK. In particular, as M is increased, which is equivalent to increased bandwidth requirement, the operating point moves closer to the Shannon limit. ◀

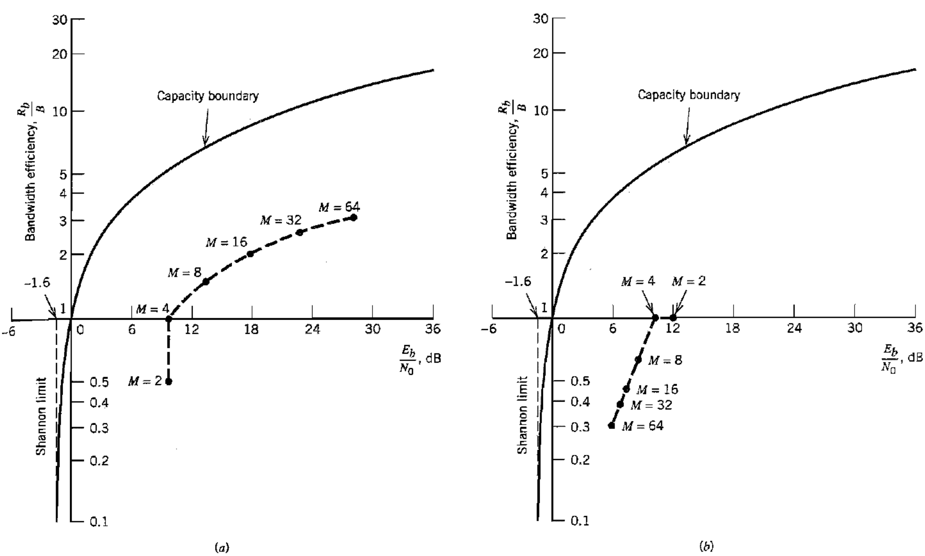

FIGURE 9.17 (*a*) Comparison of M-ary PSK against the ideal system for $P_e = 10^{-5}$ and increasing M. (*b*) Comparison of M-ary FSK against the ideal system for $P_e = 10^{-5}$ and increasing M.

▶ EXAMPLE 9.12 Capacity of Binary-Input AWGN Channel

In this example, we investigate the capacity of an AWGN channel using *encoded* binary antipodal signaling (i.e., levels -1 and $+1$ for binary symbols 0 and 1, respectively). In particular, we address the issue of determining the minimum achievable bit error rate as a function of E_b/N_0 for varying code rate r. It is assumed that the binary symbols 0 and 1 are equiprobable.

Let the random variables X and Y denote the channel input and channel output, respectively; X is a discrete variable, whereas Y is a continuous variable. In light of the second line of Equation (9.81), we may express the mutual information between the channel input and channel output as

$$I(X; Y) = h(Y) - h(Y|X)$$

The second term, $h(Y|X)$, is the conditional differential entropy of the channel output Y, given the channel input X. By virtue of Equations (9.89) and (9.93), this term is just the entropy of a Gaussian distribution. Hence, using σ^2 to denote the variance of the channel noise, we may write

$$h(Y|X) = \frac{1}{2}\log_2(2\pi e\sigma^2)$$

Next, the first term, $h(Y)$, is the differential entropy of the channel output Y. With the use of binary antipodal signaling, the probability density function of Y, given $X = x$, is a mixture of two Gaussian distributions with common variance σ^2 and mean values -1 and $+1$, as shown by

$$f_Y(y_i|x) = \frac{1}{2}\left[\frac{\exp(-(y_i+1)^2/2\sigma^2)}{\sqrt{2\pi}\sigma} + \frac{\exp(-(y_i-1)^2/2\sigma^2)}{\sqrt{2\pi}\sigma}\right] \qquad (9.109)$$

Hence, we may determine the differential entropy of Y using the formula

$$h(Y) = -\int_{-\infty}^{\infty} f_Y(y_i|x) \log_2[f_Y(y_i|x)]\, dy_i$$

where $f_Y(y_i|x)$ is defined by Equation (9.109). From the formulas of $h(Y|X)$ and $h(Y)$, it is clear that the mutual information is solely a function of the noise variance σ^2. Using $M(\sigma^2)$ to denote this functional dependence, we may thus write

$$I(X; Y) = M(\sigma^2)$$

Unfortunately, there is no closed formula that we can derive for $M(\sigma^2)$ because of the difficulty of determining $h(Y)$. Nevertheless, the differential entropy $h(Y)$ can be well approximated using *Monte Carlo integration*, which is straightforward to program on a digital computer; see Problem 9.36.

Because symbols 0 and 1 are equiprobable, it follows that the channel capacity C is equal to the mutual information between X and Y. Hence, for error-free data transmission over the AWGN channel, the code rate r must satisfy the condition

$$r < M(\sigma^2) \tag{9.110}$$

A robust measure of the ratio E_b/N_0 is

$$\frac{E_b}{N_0} = \frac{P}{N_0 r} = \frac{P}{2\sigma^2 r}$$

where P is the average transmitted power, and $N_0/2$ is the two-sided power spectral density of the channel noise. Without loss of generality, we may set $P = 1$. We may then express the noise variance as

$$\sigma^2 = \frac{N_0}{2E_b r} \tag{9.111}$$

Substituting Equation (9.111) into (9.110) and rearranging terms, we get the desired relation:

$$\frac{E_b}{N_0} = \frac{1}{2rM^{-1}(r)} \tag{9.112}$$

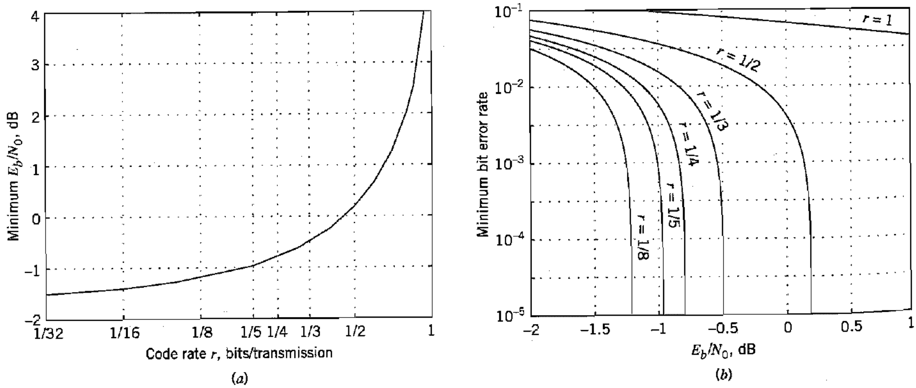

FIGURE 9.18 Binary antipodal signaling over an AWGN channel. (*a*) Minimum E_b/N_0 versus the code rate r. (*b*) Minimum bit error rate (BER) versus E_b/N_0 for varying code rate r.

where $M^{-1}(r)$ is the *inverse* of the mutual information between the channel input and output, expressed as a function of the code rate r.

Using the Monte Carlo method to estimate the differential entropy $h(Y)$ and therefore $M^{-1}(r)$, the plots of Figure 9.18 are computed.[12] Figure 9.18*a* plots the minimum E_b/N_0 versus the code rate r for error-free communication. Figure 9.18*b* plots the minimum achievable bit error rate versus E_b/N_0 with the code rate r as a running parameter. From Figure 9.18 we may draw the following conclusions:

- For uncoded binary signaling (i.e., $r = 1$), an infinite E_b/N_0 is required for error-free communication, which agrees with what we know about uncoded data transmission over an AWGN channel.
- The minimum E_b/N_0 decreases with decreasing code rate r, which is intuitively satisfying. For example, for $r = 1/2$, the minimum value of E_b/N_0 is slightly less than 0.2 dB.
- As r approaches zero, the minimum E_b/N_0 approaches the limiting value of -1.6 dB, which agrees with the Shannon limit derived earlier; see Equation (9.100).

9.12 *Information Capacity of Colored Noise Channel*[13]

The information capacity theorem as formulated in Equation (9.95) applies to a band-limited white noise channel. In this section, we extend Shannon's information capacity theorem to the more general case of a *nonwhite*, or *colored*, *noise channel*. To be specific, consider the channel model shown in Figure 9.19*a* where the transfer function of the channel is denoted by $H(f)$. The channel noise $n(t)$, which appears additively at the channel output, is modeled as the sample function of a stationary Gaussian process of zero mean and power spectral density $S_N(f)$. The requirement is twofold:

1. Find the input ensemble, described by the power spectral density $S_X(f)$, that maximizes the mutual information between the channel output $y(t)$ and the channel input $x(t)$, subject to the constraint that the average power of $x(t)$ is fixed at a constant value P.
2. Hence, determine the optimum information capacity of the channel.

This problem is a constrained optimization problem. To solve it, we proceed as follows:

- Because the channel is linear, we may replace the model of Figure 9.19*a* with the equivalent model shown in Figure 9.19*b*. From the viewpoint of the spectral characteristics of the signal plus noise measured at the channel output, the two models of Figure 9.19 are equivalent, provided that the power spectral density of the noise

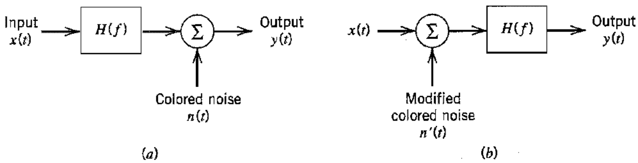

FIGURE 9.19 (*a*) Model of band-limited, power-limited noisy channel. (*b*) Equivalent model of the channel.