7

CHANNEL CAPACITY AND CODING

In Chapter 5, we considered the problem of digital modulation by means of $M = 2^k$ signal waveforms, where each waveform conveys k bits of information. We observed that some modulation methods provide better performance than others. In particular, we demonstrated that orthogonal signaling waveforms allow us to make the probability of error arbitrarily small by letting the number of waveforms $M \to \infty$, provided that the SNR per bit $\gamma_b \geq -1.6$ dB. Thus, we can operate at the capacity of the additive, white gaussian noise channel in the limit as the bandwidth expansion factor $B_e = W/R \to \infty$. This is a heavy price to pay, because B_e grows exponentially with the block length k. Such inefficient use of channel bandwidth is highly undesirable.

In this and the following chapter, we consider signal waveforms generated from either binary or nonbinary sequences. The resulting waveforms are generally characterized by a bandwidth expansion factor that grows only linearly with k. Consequently, coded waveforms offer the potential for greater bandwidth efficiency than orthogonal M-ary waveforms. We shall observe that, in general, coded waveforms offer performance advantages not only in power-limited applications where $R/W < 1$, but also in bandwidth-limited systems where $R/W > 1$.

We begin by establishing several channel models that will be used to evaluate the benefits of channel coding, and we shall introduce the concept of channel capacity for the various channel models. Then, we treat the subject of code design for efficient communications.

7-1 CHANNEL MODELS AND CHANNEL CAPACITY

In the model of a digital communication system described in Section 1-1, we recall that the transmitter building blocks consist of the discrete-input, discrete-output channel encoder followed by the modulator. The function of the discrete channel encoder is to introduce, in a controlled manner, some redundancy in the binary information sequence, which can be used at the receiver to overcome the effects of noise and interference encountered in the transmission of the signal through the channel. The encoding process generally involves taking k information bits at a time and mapping each k-bit sequence into a unique n-bit sequence, called a *code word.* The amount of redundancy introduced by the encoding of the data in this manner is measured by the ratio n/k. The reciprocal of this ratio, namely k/n, is called the *code rate.*

The binary sequence at the output of the channel encoder is fed to the modulator, which serves as the interface to the communication channel. As we have discussed, the modulator may simply map each binary digit into one of two possible waveforms, i.e., a 0 is mapped into $s_1(t)$ and a 1 is mapped into $s_2(t)$. Alternatively, the modulator may transmit q-bit blocks at a time by using $M = 2^q$ possible waveforms.

At the receiving end of the digital communication system, the demodulator processes the channel-corrupted waveform and reduces each waveform to a scalar or a vector that represents an estimate of the transmitted data symbol (binary or M-ary). The detector, which follows the demodulator, may decide on whether the transmitted bit is a 0 or a 1. In such a case, the detector has made a *hard decision.* If we view the decision process at the detector as a form of quantization, we observe that a hard decision corresponds to binary quantization of the demodulator output. More generally, we may consider a detector that quantizes to $Q > 2$ levels, i.e., a Q-ary detector. If M-ary signals are used then $Q \geq M$. In the extreme case when no quantization is performed, $Q = \infty$. In the case where $Q > M$, we say that the detector has made a *soft decision.*

The quantized output from the detector is then fed to the channel decoder, which exploits the available redundancy to correct for channel disturbances.

In the following sections, we describe three channel models that will be used to establish the maximum achievable bit rate for the channel.

7-1-1 Channel Models

In this section we describe channel models that will be useful in the design of codes. The simplest is the *binary symmetric channel* (BSC), which corresponds to the case with $M = 2$ and hard decisions at the detector.

Binary Symmetric Channel Let us consider an additive noise channel and let the modulator and the demodulator/detector be included as parts of the

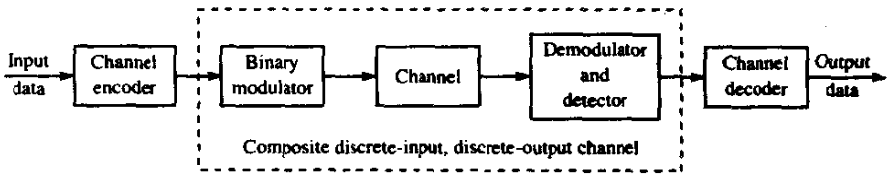

FIGURE 7-1-1 A composite discrete-input, discrete-output channel formed by including the modulator and the demodulator/detector as part of the channel.

channel. If the modulator employs binary waveforms and the detector makes hard decisions, then the composite channel, shown in Fig. 7-1-1, has a discrete-time binary input sequence and a discrete-time binary output sequence. Such a composite channel is characterized by the set $X = \{0, 1\}$ of possible inputs, the set of $Y = \{0, 1\}$ of possible outputs, and a set of conditional probabilities that relate the possible outputs to the possible inputs. If the channel noise and other disturbances cause statistically independent errors in the transmitted binary sequence with average probability p then

$$\begin{aligned} P(Y=0 \mid X=1) &= P(Y=1 \mid X=0) = p \\ P(Y=1 \mid X=1) &= P(Y=0 \mid X=0) = 1-p \end{aligned} \tag{7-1-1}$$

Thus, we have reduced the cascade of the binary modulator, the waveform channel, and the binary demodulator and detector into an equivalent discrete-time channel which is represented by the diagram shown in Fig. 7-1-2. This binary-input, binary-output, symmetric channel is simply called a *binary symmetric channel* (BSC). Since each output bit from the channel depends only on the corresponding input bit, we say that the channel is memoryless.

Discrete Memoryless Channels The BSC is a special case of a more general discreté-input, discrete-output channel. Suppose that the output from the channel encoder are q-ary symbols, i.e., $X = \{x_0, x_1, \ldots, x_{q-1}\}$ and the output of the detector consists of Q-ary symbols, where $Q \geq M = 2^q$. If the

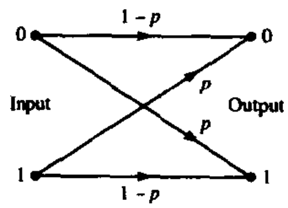

FIGURE 7-1-2 Binary symmetric channel.

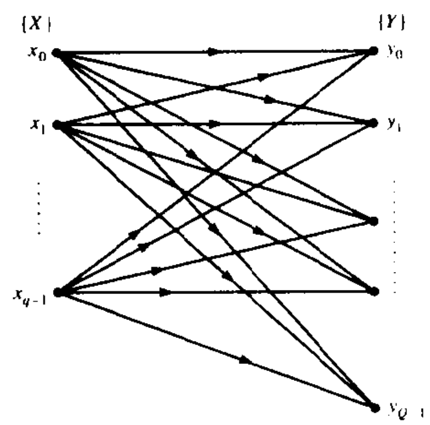

FIGURE 7-1-3 Discrete q-ary input, Q-ary output channel.

channel and the modulation are memoryless, then the input–output characteristics of the composite channel, shown in Fig. 7-1-1, are described by a set of qQ conditional probabilities

$$P(Y = y_i \mid X = x_j) \equiv P(y_i \mid x_j) \tag{7-1-2}$$

where $i = 0, 1, \ldots, Q-1$ and $j = 0, 1, \ldots, q-1$. Such a channel is called a *dicrete memoryless channel* (DMC), and its graphical representation is shown in Fig. 7-1-3. Hence, if the input to a DMC is a sequence of n symbols $u_1, u_2, \ldots, u_n$ selected from the alphabet X and the corresponding output is the sequence $v_1, v_2, \ldots, v_n$ of symbols from the alphabet Y, the joint conditional probability is

$$\begin{aligned} P(Y_1 = v_1, Y_2 = v_2, \ldots, Y_n = v_n \mid X = u_1, \ldots, X = u_n) \\ = \prod_{k=1}^{n} P(Y = v_k \mid X = u_k) \end{aligned} \tag{7-1-3}$$

This expression is simply a mathematical statement of the memoryless condition.

In general, the conditional probabilities $\{P(y_i \mid x_j)\}$ that characterize a DMC can be arranged in the matrix form $\mathbf{P} = [p_{ji}]$, where, by definition, $p_{ji} \equiv P(y_i \mid x_j)$. $\mathbf{P}$ is called the *probability transition matrix* for the channel.

Discrete-Input, Continuous-Output Channel Now, suppose that the input to the modulator comprises symbols selected from a finite and discrete input alphabet $X = \{x_0, x_1, \ldots, x_{q-1}\}$ and the output of the detector is unquantized ($Q = \infty$). Then, the input to the channel decoder can assume any value on the real line, i.e., $Y = \{-\infty, \infty\}$. This leads us to define a composite discrete-time memoryless channel that is characterized by the discrete input X, the continuous output Y, and the set of conditional probability density functions

$$p(y \mid X = x_k), \qquad k = 0, 1, \ldots, q-1$$

The most important channel of this type is the additive white gaussian noise channel (AWGN), for which

$$Y = X + G \tag{7-1-4}$$

where G is a zero-mean gaussian random variable with variance σ^2 and $X = x_k$, $k = 0, 1, \ldots, q-1$. For a given X, it follows that Y is gaussian with mean x_k and variance σ^2. That is,

$$p(y \mid X = x_k) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(y-x_k)^2/2\sigma^2} \tag{7-1-5}$$

For any given input sequence, X_i, $i = 1, 2, \ldots, n$, there is a corresponding output sequence

$$Y_i = X_i + G_i, \qquad i = 1, 2, \ldots, n \tag{7-1-6}$$

The condition that the channel is memoryless may be expressed as

$$p(y_1, y_2, \ldots, y_n \mid X_1 = u_1, X_2 = u_2, \ldots, X_n = u_n) = \prod_{i=1}^{n} p(y_i \mid X_i = u_i) \tag{7-1-7}$$

Waveform Channels We may separate the modulator and demodulator from the physical channel, and consider a channel model in which the inputs are waveforms and the outputs are waveforms. Let us assume that such a channel has a given bandwidth W, with ideal frequency response $C(f) = 1$ within the bandwidth W, and the signal at its output is corrupted by additive white gaussian noise. Suppose, that $x(t)$ is a band-limited input to such a channel and $y(t)$ is the corresponding output. Then,

$$y(t) = x(t) + n(t) \tag{7-1-8}$$

where $n(t)$ represents a sample function of the additive noise process. A suitable method for defining a set of probabilities that characterize the channel is to expand $x(t)$, $y(t)$, and $n(t)$ into a complete set of orthonormal functions. That is, we express $x(t)$, $y(t)$, and $n(t)$ in the form

$$\begin{aligned} y(t) &= \sum_i y_i f_i(t) \\ x(t) &= \sum_i x_i f_i(t) \\ n(t) &= \sum_i n_i f_i(t) \end{aligned} \tag{7-1-9}$$

where $\{y_i\}$, $\{x_i\}$, and $\{n_i\}$ are the sets of coefficients in the corresponding expansions, e.g.,

$$\begin{aligned} y_i &= \int_0^T y(t) f_i^*(t)\, dt \\ &= \int_0^T [x(t) + n(t)] f_i^*(t)\, dt \\ &= x_i + n_i \end{aligned} \tag{7-1-10}$$

The functions $\{f_i(t)\}$ form a complete orthonormal set over the interval $(0, T)$, i.e.,

$$\int_0^T f_i(t) f_j^*(t)\, dt = \delta_{ij} = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases} \tag{7-1-11}$$

where δ_{ij} is the Kronecker delta function. Since the gaussian noise is white, any complete set of orthonormal functions may be used in the expansions (7-1-9).

We may now use the coefficients in the expansion for characterizing the channel. Since

$$y_i = x_i + n_i$$

where n_i is gaussian, it follows that

$$p(y_i \mid x_i) = \frac{1}{\sqrt{2\pi}\, \sigma_i} e^{-(y_i - x_i)^2/2\sigma_i^2}, \qquad i = 1, 2, \ldots \tag{7-1-12}$$

Since the functions $\{f_i(t)\}$ in the expansion are orthonormal, it follows that the $\{n_i\}$ are uncorrelated. Since they are gaussian, they are also statistically independent. Hence,

$$p(y_1, y_2, \ldots, y_N \mid x_1, x_2, \ldots, x_N) = \prod_{i=1}^{N} p(y_i \mid x_i) \tag{7-1-13}$$

for any N. In this manner, the waveform channel is reduced to an equvalent discrete-time channel characterized by the conditional pdf given in (7-1-12).

When the additive noise is white and gaussian with spectral density $\frac{1}{2}N_0$, the variances $\sigma_i^2 = \frac{1}{2}N_0$ for all i in (7-1-12). In this case, samples of $x(t)$ and $y(t)$ may be taken at the Nyquist rate of $2W$ samples/s, so that $x_i = x(i/2W)$ and $y_i = y(i/2W)$. Since the noise is white, the noise samples are statistically independent. Thus, (7-1-12) and (7-1-13) describe the statistics of the sampled signal. We note that in a time interval of length T, there are $N = 2WT$ samples. This parameter is used below in obtaining the capacity of the band-limited AWGN waveform channel.

The choice of which channel model to use at any one time depends on our objectives. If we are interested in the design and analysis of the performance of the discrete channel encoder and decoder, it is appropriate to consider channel models in which the modulator and demodulator are a part of the composite channel. On the other hand, if our intent is to design and analyze the performance of the digital modulator and digital demodulator, we use a channel model for the waveform channel.

7-1-2 Channel Capacity

Now let us consider a DMC having an input alphabet $X = \{x_0, x_1, \ldots, x_{q-1}\}$, an output alphabet $Y = \{y_0, y_1, \ldots, y_{Q-1}\}$, and the set of transition probabilities $P(y_i \mid x_j)$ as defined in (7-1-2). Suppose that the symbol x_j is transmitted and the symbol y_i is received. The mutual information provided about the event $X = x_j$ by the occurrence of the event $Y = y_i$ is $\log [P(y_i \mid x_j)/P(y_i)]$, where

$$P(y_i) \equiv P(Y = y_i) = \sum_{k=0}^{q-1} P(x_k)P(y_i \mid x_k) \tag{7-1-14}$$

Hence, the average mutual information provided by the output Y about the input X is

$$I(X; Y) = \sum_{j=0}^{q-1} \sum_{i=0}^{Q-1} P(x_j)P(y_i \mid x_j) \log \frac{P(y_i \mid x_j)}{P(y_i)} \tag{7-1-15}$$

The channel characteristics determine the transition probabilities $P(y_i \mid x_j)$, but the probabilities of the input symbols are under the control of the discrete channel encoder. The value of $I(X; Y)$ maximized over the set of input symbol probabilities $P(x_j)$ is a quantity that depends only on the characteristics of the DMC through the conditional probabilities $P(y_i \mid x_j)$. This quantity is called the *capacity* of the channel and is denoted by C. That is, the capacity of a DMC is defined as

$$\begin{aligned} C &= \max_{P(x_j)} I(X; Y) \\ &= \max_{P(x_j)} \sum_{j=0}^{q-1} \sum_{i=0}^{Q-1} P(x_j)P(y_i \mid x_j) \log \frac{P(y_i \mid x_j)}{P(y_i)} \end{aligned} \tag{7-1-16}$$

The maximization of $I(X; Y)$ is performed under the constraints that

$$P(x_j) \geqslant 0$$

$$\sum_{j=0}^{q-1} P(x_j) = 1$$

The units of C are bits per input symbol into the channel (bits/channel use)

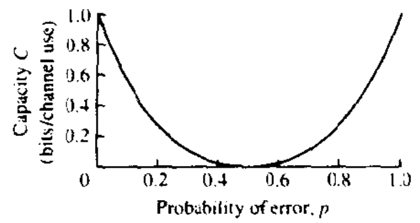

FIGURE 7-1-4 The capacity of a BSC as a function of the error probability p.

when the logarithm is base 2, and nats/input symbol when the natural logarithm (base e) is used. If a symbol enters the channel every τ_s seconds, the channel capacity in bits/s or nats/s is C/τ_s.

Example 7-1-1

For the BSC with transition probabilities

$$P(0 \mid 1) = P(1 \mid 0) = p$$

the average mutual information is maximized when the input probabilities $P(0) = P(1) = \frac{1}{2}$. Thus, the capacity of the BSC is

$$C = p \log 2p + (1-p) \log 2(1-p) = 1 - H(p) \qquad \text{(7-1-17)}$$

where $H(p)$ is the binary entropy function. A plot of C versus p is illustrated in Fig. 7-1-4. Note that for $p = 0$, the capacity is 1 bit/channel use. On the other hand, for $p = \frac{1}{2}$, the mutual information between input and output is zero. Hence, the channel capacity is zero. For $\frac{1}{2} < p \leq 1$, we may reverse the position of 0 and 1 at the output of the BSC, so that C becomes symmetric with respect to the point $p = \frac{1}{2}$. In our treatment of binary modulation and demodulation given in Chapter 5, we showed that p is a monotonic function of the signal-to-noise ratio (SNR) as illustrated in Fig. 7-1-5(a). Consequently when C is plotted as a function of the SNR, it increases monotonically as the SNR increases. This characteristic behavior of C versus SNR is illustrated in Fig. 7-1-5(b).

Next let us consider the discrete-time AWGN memoryless channel described by the transition probability density functions defined by (7-1-5). The

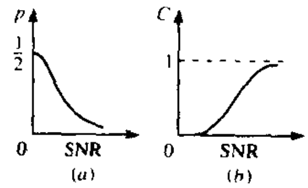

FIGURE 7-1-5 General behavior of error probability and channel capacity as a function of SNR.

average mutual information between the discrete input $X = \{x_0, x_1, \ldots, x_{q-1}\}$ and the output $Y = \{-\infty, \infty\}$ is given by the capacity of this channel in bits/channel use is

$$C = \max_{P(x_i)} \sum_{i=0}^{q-1} \int_{-\infty}^{\infty} p(y \mid x_i) P(x_i) \log_2 \frac{p(y \mid x_i)}{p(y)} \, dy \tag{7-1-18}$$

where

$$p(y) = \sum_{k=0}^{q-1} p(y \mid x_k) P(x_k) \tag{7-1-19}$$

Example 7-1-2

Let us consider a binary-input AWGN memoryless channel with possible inputs $X = A$ and $X = -A$. The average mutual information $I(X; Y)$ is maximized when the input probabilities are $P(X = A) = P(X = -A) = \frac{1}{2}$. Hence, the capacity of this channel in bits/channel use is

$$C = \tfrac{1}{2} \int_{-\infty}^{\infty} p(y \mid A) \log_2 \frac{p(y \mid A)}{p(y)} \, dy + \tfrac{1}{2} \int_{-\infty}^{\infty} p(y \mid -A) \log_2 \frac{p(y \mid -A)}{p(y)} \, dy \tag{7-1-20}$$

Figure 7-1-6 illustrates C as a function of the ratio $A^2/2\sigma^2$. Note that C increases monotonically from 0 to 1 bit/symbol as this ratio increases.

It is interesting to note that in the two channel models described above, the choice of equally probable input symbols maximizes the average mutual information. Thus, the capacity of the channel is obtained when the input symbols are equally probable. This is not always the solution for the capacity formulas given in (7-1-16) and (7-1-18), however. Nothing can be said in general about the input probability assignment that maximizes the average mutual information. However, in the two channel models considered above,

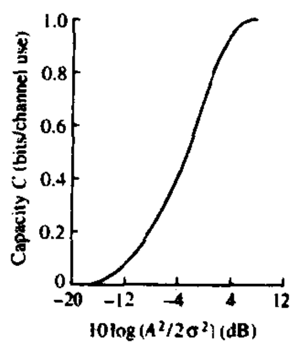

FIGURE 7-1-6 Channel capacity as a function of $A^2/2\sigma^2$ for binary-input AWGN memoryless channel.

the channel transition probabilities exhibit a form of symmetry that results in the maximum of $I(X; Y)$ being obtained when the input symbols are equally probable. The symmetry condition can be expressed in terms of the elements of the probability transition matrix $\mathbf{P}$ of the channel. When each row of this matrix is a permutation of any other row and each column is a permutation of any other column, the probability transition matrix is symmetric and input symbols with equal probability maximize $I(X; Y)$.

In general, necessary and sufficient conditions for the set of input probabilities $\{P(x_j)\}$ to maximize $I(X; Y)$ and, thus, to achieve capacity on a DMC are that (Problem 7-1)

$$\begin{aligned} I(x_j; Y) &= C \quad \text{for all } j \text{ with } P(x_j) > 0 \\ I(x_j; Y) &\leqslant C \quad \text{for all } j \text{ with } P(x_j) = 0 \end{aligned} \tag{7-1-21}$$

where C is the capacity of the channel and

$$I(x_j; Y) = \sum_{i=0}^{Q-1} P(y_i \mid x_j) \log \frac{P(y_i \mid x_j)}{P(y_i)} \tag{7-1-22}$$

Usually, it is relatively easy to check if the equally probable set of input symbols satisfy the conditions (7-1-21). If they do not, then one must determine the set of unequal probabilities $\{P(x_j)\}$ that satisfy (7-1-21).

Now let us consider a band-limited waveform channel with additive white gaussian noise. Formally, the capacity of the channel per unit time has been defined by Shannon (1948b) as

$$C = \lim_{T \to \infty} \max_{p(x)} \frac{1}{T} I(X; Y) \tag{7-1-23}$$

where the average mutual information $I(X; Y)$ is given in (3-2-17). Alternatively, we may use the samples or the coefficients $\{y_i\}$, $\{x_i\}$, and $\{n_i\}$ in the series expansions of $y(t)$, $x(t)$, and $n(t)$, respectively, to determine the average mutual information between $\mathbf{x}_N = [x_1 \; x_2 \; \dots \; x_N]$ and $\mathbf{y}_N = [y_1 \; y_2 \; \dots \; y_N]$, where $N = 2WT$, $y_i = x_i + n_i$, and $p(y_i \mid x_i)$ is given by (7-1-12). The average mutual information between $\mathbf{x}_N$ and $\mathbf{y}_N$ for the AWGN channel is

$$\begin{aligned} I(\mathbf{X}_N; \mathbf{Y}_N) &= \int_{\mathbf{x}_N} \cdots \int \int_{\mathbf{y}_N} \cdots \int p(\mathbf{y}_N \mid \mathbf{x}_N) p(\mathbf{x}_N) \log \frac{p(\mathbf{y}_N \mid \mathbf{x}_N)}{p(\mathbf{y}_N)} \, d\mathbf{x}_N \, d\mathbf{y}_N \\ &= \sum_{i=1}^{N} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(y_i \mid x_i) p(x_i) \log \frac{p(y_i \mid x_i)}{p(y_i)} \, dy_i \, dx_i \end{aligned} \tag{7-1-24}$$

where

$$p(y_i \mid x_i) = \frac{1}{\sqrt{\pi N_0}} e^{-(y_i - x_i)^2/N_0} \tag{7-1-25}$$

The maximum of $I(X; Y)$ over the input pdfs $p(x_i)$ is obtained when the $\{x_i\}$ are statistically independent zero-mean gaussian random variables, i.e.,

$$p(x_i) = \frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-x_i^2/2\sigma_x^2} \tag{7-1-26}$$

where σ_x^2 is the variance of each x_i. Then, it follows from (7-1-24) that

$$\begin{aligned}\max_{p(x)} I(\mathbf{X}_N; \mathbf{Y}_N) &= \sum_{i=1}^{N} \tfrac{1}{2} \log\left(1 + \frac{2\sigma_x^2}{N_0}\right) \\ &= \tfrac{1}{2} N \log\left(1 + \frac{2\sigma_x^2}{N_0}\right) \\ &= WT \log\left(1 + \frac{2\sigma_x^2}{N_0}\right)\end{aligned} \tag{7-1-27}$$

Suppose that we put a constraint on the average power in $x(t)$. That is,

$$\begin{aligned}P_{\text{av}} &= \frac{1}{T}\int_0^T E[x^2(t)]\,dt \\ &= \frac{1}{T}\sum_{i=1}^{N} E(x_i^2) \\ &= \frac{N\sigma_x^2}{T}\end{aligned} \tag{7-1-28}$$

Hence,

$$\begin{aligned}\sigma_x^2 &= \frac{TP_{\text{av}}}{N} \\ &= \frac{P_{\text{av}}}{2W}\end{aligned} \tag{7-1-29}$$

Substitution of this result into (7-1-27) for σ_x^2 yields

$$\max_{p(x)} I(\mathbf{X}_N; \mathbf{Y}_N) = WT \log\left(1 + \frac{P_{\text{av}}}{WN_0}\right) \tag{7-1-30}$$

Finally, the channel capacity per unit time is obtained by dividing the result in (7-1-30) by T. Thus

$$C = W \log\left(1 + \frac{P_{\text{av}}}{WN_0}\right) \tag{7-1-31}$$

This is the basic formula for the capacity of the band-limited AWGN

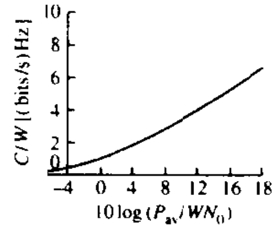

FIGURE 7-1-7 Normalized channel capacity as a function of SNR for band-limited AWGN channel.

waveform channel with a band-limited and average power-limited input. It was originally derived by Shannon (1948b).

A plot of the capacity in bits/s normalized by the bandwidth W is plotted in Fig. 7-1-7 as a function of the ratio of signal power P_{av} to noise power WN_0. Note that the capacity increases monotonically with increasing SNR. Thus, for a fixed bandwidth, the capacity of the waveform channel increases with an increase in the transmitted signal power. On the other hand, if P_{av} is fixed, the capacity can be increased by increasing the bandwidth W. Figure 7-1-8 illustrates a graph of C versus W. Note that as W approaches infinity, the capacity of the channel approaches the asymptotic value

$$C_\infty = \frac{P_{av}}{N_0}\log_2 e = \frac{P_{av}}{N_0 \ln 2}\ \text{bits/s} \tag{7-1-32}$$

It is instructive to express the normalized channel capacity C/W as a function of the SNR per bit. Since P_{av} represents the average transmitted power and C is the ratio in bits/s, it follows that

$$P_{av} = C\mathscr{E}_b \tag{7-1-33}$$

where $\mathscr{E}_b$ is the energy per bit. Hence, (7-1-31) may be expressed as

$$\frac{C}{W} = \log_2\left(1 + \frac{C}{W}\frac{\mathscr{E}_b}{N_0}\right) \tag{7-1-34}$$

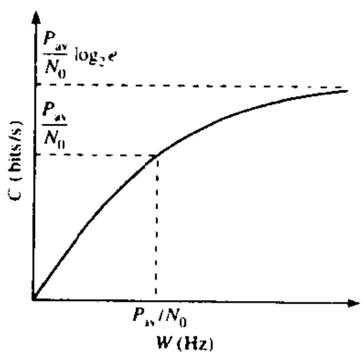

FIGURE 7-1-8 Channel capacity as a function of bandwidth with a fixed transmitted average power.

Consequently,

$$\frac{\mathscr{E}_b}{N_0} = \frac{2^{C/W} - 1}{C/W} \tag{7-1-35}$$

When $C/W = 1$, $\mathscr{E}_b/N_0 = 1$ (0 dB). As $C/W \to \infty$,

$$\begin{aligned}\frac{\mathscr{E}_b}{N_0} &\approx \frac{2^{C/W}}{C/W} \\ &\approx \exp\left(\frac{C}{W}\ln 2 - \ln\frac{C}{W}\right)\end{aligned} \tag{7-1-36}$$

Thus, $\mathscr{E}_b/N_0$ increases exponentially as $C/W \to \infty$. On the other hand, as $C/W \to 0$,

$$\frac{\mathscr{E}_b}{N_0} = \lim_{C/W \to 0} \frac{2^{C/W} - 1}{C/W} = \ln 2 \tag{7-1-37}$$

which is -1.6 dB. A plot of C/W versus $\mathscr{E}_b/N_0$ is shown in Fig. 5-2-17.

Thus, we have derived the channel capacities of three important channel models that are considered in this book. The first is the discrete-input, discrete-output channel, of which the BSC is a special case. The second is a discrete-input, continuous-output memoryless additive white gaussian noise channel. From these two channel models, we can obtain benchmarks for the coded performance with hard- and soft-decision decoding in digital communications systems.

The third channel model focuses on the capacity in bits/s of a waveform channel. In this case, we assumed that we have a bandwidth limitation on the channel, an additive gaussian noise that corrupts the signal, and an average power constraint at the transmitter. Under these conditions, we derived the result given in (7-1-31).

The major significance of the channel capacity formulas given above is that they serve as upper limits on the transmission rate for reliable communication over a noisy channel. The fundamental rate that the channel capacity plays is given by the *noisy channel coding theorem* due to Shannon (1948a).

Noisy Channel Coding Theorem

There exist channel codes (and decoders) that make it possible to achieve reliable communication, with as small an error probability as desired, if the transmission rate $R < C$, where C is the channel capacity. If $R > C$, it is not possible to make the probability of error tend toward zero with any code.

In the following section, we explore the benefits of coding for the additive noise channel models described above, and use the channel capacity as the benchmark for accessing code performance.

7-1-3 Achieving Channel Capacity with Orthogonal Signals

In Section 5-2, we used a simple union bound to show that, for orthogonal signals, the probability of error can be made as small as desired by increasing the number M of waveforms, provided that $\mathscr{E}_b/N_0 > 2 \ln 2$. We indicated that the simple union bound does not produce the smallest lower bound on the SNR per bit. The problem is that the upper bound used on $Q(x)$ is very loose for small x.

An alternative approach is to use two different upper bounds for $Q(x)$, depending on the value of x. Beginning with (5-2-21), we observe that

$$1 - [1 - Q(y)]^{M-1} \leq (M-1)Q(y) < Me^{-y^2/2} \tag{7-1-38}$$

This is just the union bound, which is tight when y is large, i.e., for $y > y_0$, where y_0 depends on M. When y is small, the union bound exceeds unity for large M. Since

$$1 - [1 - Q(y)]^{M-1} \leq 1 \tag{7-1-39}$$

for all y, we may use this bound for $y < y_0$ because it is tighter than the union bound. Thus (5-2-21) may be upper-bounded as

$$P_M < \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y_0} e^{-(y-\sqrt{2\gamma})^2/2}\, dy + \frac{M}{\sqrt{2\pi}} \int_{y_0}^{\infty} e^{-y^2/2} e^{-(y-\sqrt{2\gamma})^2/2}\, dy \tag{7-1-40}$$

The value of y_0 that minmizes this upper bound is found by differentiating the right-hand side of (7-1-40) and setting the derivative equal to zero. It is easily verified that the solution is

$$e^{y_0^2/2} = M \tag{7-1-41}$$

or, equivalently,

$$\begin{aligned} y_0 &= \sqrt{2 \ln M} = \sqrt{2 \ln 2 \log_2 M} \\ &= \sqrt{2k \ln 2} \end{aligned} \tag{7-1-42}$$

Having determined y_0, let us now compute simple exponential upper bounds for the integrals in (7-1-40). For the first integral, we have

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y_0} e^{-(y-\sqrt{2\gamma})^2/2}\, dy &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{-(\sqrt{2\gamma}-y_0)/\sqrt{2}} e^{-x^2}\, dx \\ &= Q(\sqrt{2\gamma} - y_0), \qquad y_0 \leq \sqrt{2\gamma} \\ &< e^{-(\sqrt{2\gamma}-y_0)^2/2}, \qquad y_0 \leq \sqrt{2\gamma} \end{aligned} \tag{7-1-43}$$

The second integral is upper-bounded as follows:

$$\frac{M}{\sqrt{2\pi}}\int_{y_0}^{\infty} e^{-y^2/2}e^{-(y-\sqrt{2\gamma})^2/2}\,dy = \frac{M}{\sqrt{2\pi}}e^{-\gamma/2}\int_{y_0-\sqrt{\gamma/2}}^{\infty} e^{-x^2}\,dx$$
$$< \begin{cases} Me^{-\gamma/2} & (y_0 \le \sqrt{\tfrac{1}{2}\gamma}) \\ Me^{-\gamma/2}e^{-(y_0-\sqrt{\gamma/2})^2} & (y_0 \ge \sqrt{\tfrac{1}{2}\gamma}) \end{cases} \tag{7-1-44}$$

Combining the bounds for the two integrals and substituting $e^{y_0^2/2}$ for M, we obtain

$$P_M < \begin{cases} e^{-(\sqrt{2\gamma}-y_0)^2/2} + e^{(y_0^2-\gamma)/2} & (0 \le y_0 \le \sqrt{\tfrac{1}{2}\gamma}) \\ e^{-(\sqrt{2\gamma}-y_0)^2/2} + e^{(y_0^2-\gamma)/2}e^{-(y_0-\sqrt{\gamma/2})^2} & (\sqrt{\tfrac{1}{2}\gamma} \le y_0 \le \sqrt{2\gamma}) \end{cases} \tag{7-1-45}$$

In the range $0 \le y_0 \le \sqrt{\frac{1}{2}\gamma}$, the bound may be expressed as

$$P_M < e^{(y_0^2-\gamma)/2}(1 + e^{-(y_0-\sqrt{\gamma/2})^2})$$
$$< 2e^{(y_0^2-\gamma)/2}, \quad 0 \le y_0 \le \sqrt{\tfrac{1}{2}\gamma} \tag{7-1-46}$$

In the range $\sqrt{\frac{1}{2}\gamma} \le y_0 \le \sqrt{2\gamma}$, the two terms in (7-1-45) are identical. Hence,

$$P_M < 2e^{-(\sqrt{2\gamma}-y_0)^2/2}, \quad \sqrt{\tfrac{1}{2}\gamma} \le y_0 \le \sqrt{2\gamma} \tag{7-1-47}$$

Now we substitute for y_0 and γ. Since $y_0 = \sqrt{2\ln M} = \sqrt{2k\ln 2}$ and $\gamma = k\gamma_b$, the bounds in (7-1-46) and (7-1-47) may be expresed as

$$P_M < \begin{cases} 2e^{-k(\gamma_b - 2\ln 2)/2} & (\ln M \le \tfrac{1}{4}\gamma) \\ 2e^{-k(\sqrt{\gamma_b}-\sqrt{\ln 2})^2} & (\tfrac{1}{4}\gamma \le \ln M \le \gamma) \end{cases} \tag{7-1-48}$$

The first upper bound coincides with the union bound presented earlier, but it is loose for large values of M. The second upper bound is better for large values of M. We note that $P_M \to 0$ as $k \to \infty$ $(M \to \infty)$ provided that $\gamma_b > \ln 2$. But, $\ln 2$ is the limiting value of the SNR per bit required for reliable transmission when signaling at a rate equal to the capacity of the infinite-bandwidth AWGN channel as shown in Section 7-1-2. In fact, when the substitutions

$$y_0 = \sqrt{2k\ln 2} = \sqrt{2RT\ln 2}$$
$$\gamma = \frac{TP_{av}}{N_0} = TC_\infty \ln 2 \tag{7-1-49}$$

are made into the two upper bounds given in (7-1-46) and (7-1-47), where $C_\infty = P_{av}/(N_0 \ln 2)$ is the capacity of the infinite-bandwidth AWGN channel, the result is

$$P_M < \begin{cases} 2 \cdot 2^{-T(\frac{1}{2}C_\infty - R)} & (0 \le R \le \tfrac{1}{4}C_\infty) \\ 2 \cdot 2^{-T(\sqrt{C_\infty}-\sqrt{R})^2} & (\tfrac{1}{4}C_\infty \le R \le C_\infty) \end{cases} \tag{7-1-50}$$

Thus we have expressed the bounds in terms of C_∞ and the bit rate in the channel. The first upper bound is appropriate for rates below $\frac{1}{4}C_\infty$, while the second is tighter than the first for rates between $\frac{1}{4}C_\infty$ and C_∞. Clearly, the probability of error can be made arbitrarily small by making $T \to \infty$ ($M \to \infty$ for fixed R), provided that $R < C_\infty = P_{av}/(N_0 \ln 2)$. Furthermore, we observe that the set of orthogonal waveforms achieves the channel capacity bound as $M \to \infty$, when the rate $R < C_\infty$.

7-1-4 Channel Reliability Functions

The exponential bounds on the error probability for M-ary orthogonal signals on an infinite-bandwidth AWGN channel given by (7-1-50) may be expressed as

$$P_M < 2 \cdot 2^{-TE(R)} \tag{7-1-51}$$

The exponential factor

$$E(R) = \begin{cases} \frac{1}{2}C_\infty - R & (0 \le R \le \frac{1}{4}C_\infty) \\ (\sqrt{C_\infty} - \sqrt{R})^2 & (\frac{1}{4}C_\infty \le R \le C_\infty) \end{cases} \tag{7-1-52}$$

in (7-1-51) is called the *channel reliability function* for the infinite-bandwidth AWGN channel. A plot of $E(R)/C_\infty$ is shown in Fig. 7-1-9. Also shown is the exponential factor for the union bound on P_M, given by (5-2-27), which may be expressed as

$$P_M \le \tfrac{1}{2} \cdot 2^{-T(\frac{1}{2}C_\infty - R)}, \qquad 0 \le R \le \tfrac{1}{2}C_\infty \tag{7-1-53}$$

Clearly, the exponential factor in (7-1-53) is not as tight as $E(R)$, due to the looseness of the union bound.

The bound given by (7-1-51) and (7-1-52) has been shown by Gallager (1965) to be *exponentially tight.* This means that there does not exist another reliability function, say $E_1(R)$, satisfying the condition $E_1(R) > E(R)$ for any R. Consequently, the error probability is bounded from above and below as

$$K_l 2^{-TE(R)} \le P_e \le K_u 2^{-TE(R)} \tag{7-1-54}$$

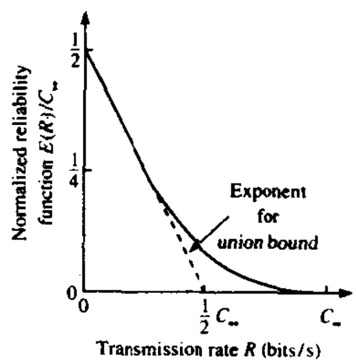

FIGURE 7-1-9 Channel reliability function for the infinite-bandwidth AWGN channel.

where the constants have only a weak dependence on T, i.e., they vary slowly with T.

Since orthogonal signals provide essentially the same performance as the optimum simplex signals for large M, the lower bound in (7-1-54) applies for any signal set. Hence, the reliability function $E(R)$ given by (7-1-52) determines the exponential characteristics of the error probability for digital signaling over the infinite-bandwidth AWGN channel.

Although the error probability can be made arbitrarily small by increasing the number of either orthogonal, biorthogonal, or simplex signals, with $R < C_\infty$, for a relatively modest number of signals, there is a large gap between the actual performance and the best achievable performance given by the channel capacity formula. For example, from Fig. 5-2-17, we observe that a set of $M = 16$ orthogonal signals detected coherently requires a SNR per bit of approximately 7.5 dB to achieve a bit error rate of $P_e = 10^{-5}$. In contrast, the channel capacity formula indicates that for a $C/W = 0.5$, reliable transmission is possible with a SNR of -0.8 dB. This represents a rather large difference of 8.3 dB/bit and serves as a motivation for searching for more efficient signaling waveforms. In this chapter and in Chapter 8, we demonstrate that coded waveforms can reduce this gap considerably.

Similar gaps in performance also exist in the bandwidth-limited region of Fig. 5-2-17, where $R/W > 1$. In this region, however, we must be more clever in how we use coding to improve performance, because we cannot expand the bandwidth as in the power-limited region. The use of coding techniques for bandwidth-efficient communication is also treated in Chapter 8.

7-2 RANDOM SELECTION OF CODES

The design of coded modulation for efficient transmission of information may be divided into two basic approaches. One is the algebraic approach, which is primarily concerned with the design of coding and decoding techniques for specific classes of codes, such as cyclic block codes and convolutional codes. The second is the probabilistic approach, which is concerned with the analysis of the performance of a general class of coded signals. This approach yields bounds on the probability of error that can be attained for communication over a channel having some specified characteristic.

In this section, we adopt the probabilistic approach to coded modulation. The algebraic approach, based on block codes and on convolutional codes, is treated in Chapter 8.

7-2-1 Random Coding Based on M-ary Binary Coded Signals

Let us consider a set of M coded signal waveforms constructed from a set of n-dimensional binary code words of the form

$$\mathbf{C}_i = [c_{i1} c_{i2} \dots c_{in}], \qquad i = 1, 2, \dots, M \tag{7-2-1}$$

where $c_{ij} = 0$ or 1. Each bit in the code word is mapped into a binary PSK waveform, so that the signal waveform corresponding to the code word $\mathbf{C}_i$ may be expressed as

$$s_i(t) = \sum_{j=1}^{n} s_{ij} f_j(t), \qquad i = 1, 2, \ldots, M \tag{7-2-2}$$

where

$$s_{ij} = \begin{cases} \sqrt{\mathscr{E}_c} & \text{when } c_{ij} = 1 \\ -\sqrt{\mathscr{E}_c} & \text{when } c_{ij} = 0 \end{cases} \tag{7-2-3}$$

and $\mathscr{E}_c$ is the energy per code bit. Thus, the waveforms $s_i(t)$ are equivalent to the n-dimensional vectors

$$\mathbf{s}_i = [s_{i1} \quad s_{i2} \quad \ldots \quad s_{in}], \qquad i = 1, 2, \ldots, M \tag{7-2-4}$$

which correspond to the vertices of a hypercube in n-dimensional space.

Now, suppose that the information rate into the encoder is R bits/s and we encode blocks of k bits at a time into one of the M waveforms. Hence, $k = RT$ and $M = 2^k = 2^{RT}$ signals are required. It is convenient to define a parameter D as

$$D = \frac{n}{T} \text{dimensions/s} \tag{7-2-5}$$

Thus, $n = DT$ is the dimensionality of the signal space.

The hypercube has $2^n = 2^{DT}$ vertices, of which $M = 2^{RT}$ may be used to transmit the information. If we impose the condition that $D > R$, the fraction of the vertices that we use as signal points is

$$F = \frac{2^k}{2^n} = \frac{2^{RT}}{2^{DT}} = 2^{-(D-R)T} \tag{7-2-6}$$

Clearly, if $D > R$, we have $F \rightarrow 0$ as $T \rightarrow \infty$.

The question that we wish to pose is the following. Can we choose a subset $M = 2^{RT}$ vertices out of the $2^n = 2^{DT}$ available vertices such that the probability of error $P \rightarrow 0$ as $T \rightarrow \infty$ or, equivalently, as $n \rightarrow \infty$? Since the fraction F of vertices used approaches zero as $T \rightarrow \infty$, it should be possible to select M signal waveforms having a minimum distance that increases as $T \rightarrow \infty$ and, thus, $P_e \rightarrow 0$.

Instead of attempting to find a single set of M coded waveforms for which we compute the error probability, let us consider the ensemble of $(2^n)^M$ distinct ways in which we can select M vertices from the 2^n available vertices of the hypercube. Associated with each of the 2^{nM} selections, there is a communication system, consisting of a modulator, a channel, and a demodulator, that is optimum for the selected set of M waveforms. Thus, there are 2^{nM}

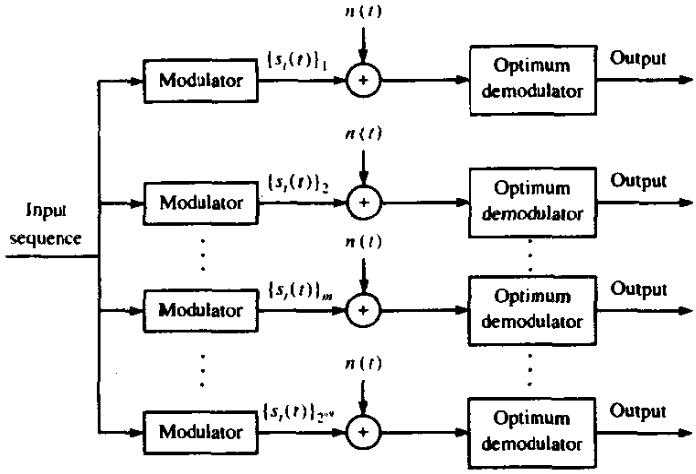

FIGURE 7-2-1 An ensemble at 2^{nM} communications system. Each system employs a different set of M signals from the set of 2^{nM} possible choices.

communication systems, one for each choice of the M coded waveforms, as illustrated in Fig. 7-2-1. Each communication system is characterized by its probability of error.

Suppose that our choice of M coded waveforms is based on random selection from the set of 2^{nM} possible sets of codes. Thus, the random selection of the mth code, denoted by $\{\mathbf{s}_i\}_m$, occurs with probability

$$P(\{\mathbf{s}_i\}_m) = 2^{-nM} \tag{7-2-7}$$

and the corresponding conditional probability of error for this choice of coded signals is $P_e(\{\mathbf{s}_i\}_m)$. Then, the average probability of error over the ensemble of codes is

$$\begin{aligned} \bar{P}_e &= \sum_{m=1}^{2^{nM}} P_e(\{\mathbf{s}_i\}_m) P(\{\mathbf{s}_i\}_m) \\ &= 2^{-nM} \sum_{m=1}^{2^{nM}} P_e(\{\mathbf{s}_i\}_m) \end{aligned} \tag{7-2-8}$$

where the overbar on P_e denotes an average over the ensemble of codes.

It is clear that some choices of codes will result in large probability of error. For example, the code that assigns all M k-bit sequences to the same vertex of the hypercube will result in a large probability of error. In such a case, $P_e(\{\mathbf{s}_i\}_m) > \bar{P}_e$. However, there will also be choices of codes for which $P_e(\{\mathbf{s}_i\}_m) < \bar{P}_e$. Consequently, if we obtain an upper bound on $\bar{P}_e$, this bound will also hold for those codes for which $P_e(\{\mathbf{s}_i\}_m) < \bar{P}_e$. Furthermore, if $\bar{P}_e \to 0$ as $T \to \infty$ then we conclude that, for these codes, $P(\{\mathbf{s}_i\}_m) \to 0$ as $T \to \infty$.

In order to determine an upper bound on $\bar{P}_e$, we consider the transmission of a k-bit message $\mathbf{X}_k \equiv [x_1 x_2 x_3 \ldots x_k]$, where $x_j = 0$ or 1 for $j = 1, 2, \ldots, k$. The conditional probability of error averaged over the ensemble of codes is

$$\overline{P_e(\mathbf{X}_k)} = \sum_{\text{all codes}} P_e(\mathbf{X}_k, \{\mathbf{s}_i\}_m) P(\{\mathbf{s}_i\}_m) \tag{7-2-9}$$

where $P_e(\mathbf{X}_k, \{\mathbf{s}_i\}_m)$ is the conditional probability of error for a given k-bit message $\mathbf{X}_k$, which is transmitted by use of the code $\{\mathbf{s}_i\}_m$. For the mth code, the probability of error $P_e(\mathbf{X}_k, \{\mathbf{s}_i\}_m)$ is upper-bounded as

$$P_e(\mathbf{X}_k, \{\mathbf{s}_i\}_m) \leq \sum_{\substack{l=1 \\ l \neq k}}^{M} P_{2m}(\mathbf{s}_l, \mathbf{s}_k) \tag{7-2-10}$$

where $P_{2m}(\mathbf{s}_l, \mathbf{s}_k)$ is the probability of error for a binary communication system that employs the signal vectors $\mathbf{s}_l$ and $\mathbf{s}_k$ to communicate one of two equally likely k-bit messages. Hence,

$$\overline{P_e(\mathbf{X}_k)} \leq \sum_{\text{all codes}} P_e(\{\mathbf{s}_i\}_m) \sum_{\substack{l=1 \\ l \neq k}}^{M} P_{2m}(\mathbf{s}_l, \mathbf{s}_k) \tag{7-2-11}$$

If we interchange the order of the summations in (7-2-11) we obtain

$$\begin{aligned} \overline{P_e(\mathbf{X}_k)} &\leq \sum_{\substack{l=1 \\ l \neq k}}^{M} \left[\sum_{\text{all codes}} P_e(\{\mathbf{s}_i\}_m) P_{2m}(\mathbf{s}_l, \mathbf{s}_k) \right] \\ &\leq \sum_{\substack{l=1 \\ l \neq k}}^{M} \overline{P_2(\mathbf{s}_l, \mathbf{s}_k)} \end{aligned} \tag{7-2-12}$$

where $\overline{P_2(\mathbf{s}_l, \mathbf{s}_k)}$ represents the ensemble average of $P_{2m}(\mathbf{s}_l, \mathbf{s}_k)$ over the 2^{nM} codes or the 2^{nM} communication systems.

For the additive white gaussian noise channel, the binary error probability $P_{2m}(\mathbf{s}_l, \mathbf{s}_k)$ is

$$P_{2m}(\mathbf{s}_l, \mathbf{s}_k) = Q\left(\sqrt{\frac{d_{lk}^2}{2N_0}}\right) \tag{7-2-13}$$

where $d_{lk}^2 = |\mathbf{s}_l - \mathbf{s}_k|^2$. If $\mathbf{s}_l$ and $\mathbf{s}_k$ differ in d coordinates,

$$d_{lk}^2 = |\mathbf{s}_l - \mathbf{s}_k|^2 = \sum_{j=1}^{n} (s_{lj} - s_{kj})^2 = d(2\sqrt{\mathscr{E}_c})^2 = 4d\mathscr{E}_c \tag{7-2-14}$$

Hence,

$$P_{2m}(\mathbf{s}_l, \mathbf{s}_k) = Q\left(\sqrt{\frac{2d\mathscr{E}_c}{N_0}}\right) \tag{7-2-15}$$

Now, we can average $P_{2m}(\mathbf{s}_l, \mathbf{s}_k)$ over the ensemble of codes. Since all the codes are equally probable, the signal vector $\mathbf{s}_l$ is equally likely to be any of the 2^n possible vertices of the hypercube and it is statistically independent of the signal vector $\mathbf{s}_k$. Therefore, $P(s_{li} = s_{ki}) = \frac{1}{2}$ and $P(s_{li} \neq s_{ki}) = \frac{1}{2}$, independently for all $i = 1, 2, \ldots, n$. Consequently, the probability that $\mathbf{s}_l$ and $\mathbf{s}_k$ differ in d positions is simply

$$P(d) = (\tfrac{1}{2})^n \binom{n}{d} \tag{7-2-16}$$

Hence, the expected value of $P_{2m}(\mathbf{s}_l, \mathbf{s}_k)$ over the ensemble of codes may be expressed as

$$\begin{aligned} \overline{P_2(\mathbf{s}_l, \mathbf{s}_k)} &= \sum_{d=0}^{n} P(d) Q\left(\sqrt{\frac{2d\mathscr{E}_c}{N_0}}\right) \\ &= \frac{1}{2^n} \sum_{d=0}^{n} \binom{n}{d} Q\left(\sqrt{\frac{2d\mathscr{E}_c}{N_0}}\right) \end{aligned} \tag{7-2-17}$$

The result (7-2-17) can be simplified if we upper-bound the Q-function as

$$Q\left(\sqrt{\frac{2d\mathscr{E}_c}{N_0}}\right) < e^{-d\mathscr{E}_c/N_0}$$

Thus,

$$\begin{aligned} \overline{P_2(\mathbf{s}_l, \mathbf{s}_k)} &\leq 2^{-n} \sum_{d=0}^{n} \binom{n}{d} e^{-d\mathscr{E}_c/N_0} \\ &\leq 2^{-n}(1 + e^{-\mathscr{E}_c/N_0})^n \\ &\leq [\tfrac{1}{2}(1 + e^{-\mathscr{E}_c/N_0})]^n \end{aligned} \tag{7-2-18}$$

We observe that the right-hand side of (7-2-18) is independent of the indices l and k. Hence, when we substitute the bound (7-2-18) into (7-2-12), we obtain

$$\begin{aligned} \overline{P_e(\mathbf{X}_k)} &\leq \sum_{\substack{l=1 \\ l \neq k}}^{M} \overline{P_2(\mathbf{s}_l, \mathbf{s}_k)} = (M - 1)[\tfrac{1}{2}(1 + e^{-\mathscr{E}_c/N_0})]^n \\ &< M[\tfrac{1}{2}(1 + e^{-\mathscr{E}_c/N_0})]^n \end{aligned}$$

Finally, the unconditional average error probability $\bar{P}_e$ is obtained by averaging $\overline{P_e(X_k)}$ over all possible k-bit information sequences. Thus,

$$\begin{aligned} \bar{P}_e &= \sum_k \overline{P_e(\mathbf{X}_k)} P(\mathbf{X}_k) < M[\tfrac{1}{2}(1 + e^{-\mathscr{E}_c/N_0})]^n \sum_k P(\mathbf{X}_k) \\ &< M[\tfrac{1}{2}(1 + e^{-\mathscr{E}_c/N_0})]^n \end{aligned} \tag{7-2-19}$$

This result can be expressed in a more convenient form by first defining a parameter R_0, which is called the *cutoff rate* and has units of bits/dimension, as

$$\begin{aligned} R_0 &= \log_2 \frac{2}{1 + e^{-\mathscr{E}_c/N_0}} \\ &= 1 - \log_2 (1 + e^{-\mathscr{E}_c/N_0}), \qquad \text{antipodal signaling} \end{aligned} \tag{7-2-20}$$

Then, (7-2-19) becomes

$$\bar{P}_e < M 2^{-nR_0} = 2^{RT} 2^{-nR_0} \tag{7-2-21}$$

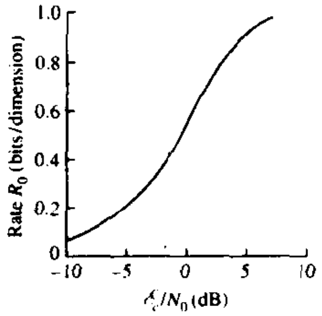

FIGURE 7-2-2 The cutoff rate R_0 as a function of the SNR per dimension in decibels.

Since $n = DT$, (7-2-21) may be expressed as

$$\bar{P}_e < 2^{-T(DR_0 - R)} \tag{7-2-22}$$

The parameter R_0 is plotted as a function of $\mathscr{E}_c/N_0$ in Fig. 7-2-2. We observe that $0 \le R_0 \le 1$. Consequently, $\bar{P}_e \to 0$ as $T \to \infty$, provided that the information rate $R < DR_0$.

Alternatively, (7-2-21) may be expressed as

$$\bar{P}_e < 2^{-n(R_0 - R/D)} \tag{7-2-23}$$

The ratio R/D also has units of bits/dimension and may be defined as

$$R_c = \frac{R}{D} = \frac{R}{n/T} = \frac{RT}{n} = \frac{k}{n} \tag{7-2-24}$$

Hence, R_c is the code rate and

$$\bar{P}_e < 2^{-n(R_0 - R_c)} \tag{7-2-25}$$

We conclude that when $R_c < R_0$, the average probability of error $\bar{P}_e \to 0$ as the code block length $n \to \infty$. Since the average value of the probability error can be made arbitrarily small as $n \to \infty$, it follows that there exist codes in the ensemble of 2^{nM} codes that have a probability of error no larger than $\bar{P}_e$.

From the derivation of the average error probability given above, we conclude that good codes exist. Although we do not normally select codes at random, it is interesting to consider the question of whether or not a randomly selected code is likely to be a good code. In fact, we can easily show that there are many good codes in the ensemble. First, we note that $\bar{P}_e$ is an ensemble average of error probabilities over all codes and that all these probabilities are obviously positive quantities. If a code is selected at random, the probability that its error probability $P_e > \alpha \bar{P}_e$ is less than $1/\alpha$. Consequently, no more than 10% of the codes have an error probability that exceeds $10\bar{P}_e$ and no more than 1% of the codes have an error probability that exceeds $100\bar{P}_e$.

We should emphasize that codes with error probabilities exceeding $\bar{P}_e$ are not necessarily poor codes. For example, suppose that an average error rate of $\bar{P}_e < 10^{-10}$ can be attained by using codes with dimensionality n_0 when $R_0 > R_c$. Then, if we select a code with error probability $1000\bar{P}_e = 10^{-7}$, we may compensate for this reduction in error probability by increasing n from n_0 to $n = 10n_0/7$. Thus, by a modest increase in dimensionality, we have a code with $\bar{P}_e < 10^{-10}$. In summary, good codes are abundant and, hence, they are easily found even by random selection.

It is also interesting to express the average error probability in (7-2-25) in terms of the SNR per bit, γ_b. To accomplish this, we express the energy per signal waveform as

$$\mathscr{E} = n\mathscr{E}_c = k\mathscr{E}_b \tag{7-2-26}$$

Hence, $n = k\mathscr{E}_b/\mathscr{E}_c$. We also note that $R_c\mathscr{E}_b/\mathscr{E}_c = 1$. Therefore, (7-2-25) may be expressed as

$$\bar{P}_e < 2^{-k(\gamma_b/\gamma_0 - 1)} \tag{7-2-27}$$

where γ_0 is a normalized SNR parameter, defined as

$$\begin{aligned} \gamma_0 &= \frac{R_c}{R_0}\gamma_b \\ &= \frac{R_c\gamma_b}{1 - \log_2(1 + e^{-R_c\gamma_b})} \end{aligned} \tag{7-2-28}$$

Now, we note that $\bar{P}_e \to 0$ as $k \to \infty$, provided that the SNR per bit, $\gamma_b > \gamma_0$.

The parameter γ_0 is plotted in Fig. 7-2-3 as a function of $R_c\gamma_b$. Note that as $R_c\gamma_b \to 0$, $\gamma_0 \to 2\ln 2$. Consequently, the error probability for M-ary binary coded signals is equivalent to the error probability obtained from the union bound for M-ary orthogonal signals, provided that the signal dimensionality is sufficiently large so that $\gamma_0 \approx 2\ln 2$.

The dimensionality parameter D that we introduced in (7-2-5) is proportional to the channel bandwidth required to transmit the signals. Recall from the sampling theorem that a signal of bandwidth W may be represented by samples taken at a rate of $2W$ samples/s. Thus, in the time interval of length T

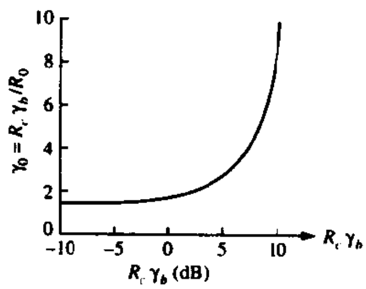

FIGURE 7-2-3 Lower bound on SNR per bit, γ_b, for binary antipodal signals.

there are $n = 2WT$ samples or, equivalently, n degrees of freedom (dimensions). Consequently, D may be equated with $2W$.

Finally, we note that the binary coded signals considered in this section are appropriate when the SNR per dimension is small, e.g., $\mathscr{E}_c/N_0 < 10$. However, when $\mathscr{E}_c/N_0 > 10$, R_0 saturates at 1 bit/dimension. Since the code rate is restricted to be less than R_0, binary coded signals become inefficient at $\mathscr{E}_c/N_0 > 10$. In such a case, we may use nonbinary-coded signals to achieve an increase in the number of bits per dimension. For example, multiple-amplitude coded signal sets can be constructed from nonbinary codes by mapping each code element into one of q possible amplitude levels (as in PAM). Such codes are considered below.

7-2-2 Random Coding Based on M-ary Multiamplitude Signals

Instead of constructing binary-coded signals, suppose we employ nonbinary codes with code words of the form given by (7-2-1), where the code elements c_{ij} are selected from the set $\{0, 1, \ldots, q-1\}$. Each code element is mapped into one of q possible amplitude levels. Thus, we construct signals corresponding to n-dimensional vectors $\{\mathbf{s}_i\}$ as in (7-2-4), where the components $\{s_{ij}\}$ are selected from a multiamplitude set of q possible values. Now, we have q^n possible signals, from which we select $M = 2^{RT}$ signals to transmit k-bit blocks of information. The q amplitudes corresponding to the code elements $\{0, 1, \ldots, q-1\}$ may be denoted by $\{a_1, a_2, \ldots, a_q\}$, and they are assumed to be selected according to some specified probabilities $\{p_i\}$. The amplitude levels are assumed to be equally spaced over the interval $[-\sqrt{\mathscr{E}_c}, \sqrt{\mathscr{E}_c}]$. For example, Fig. 7-2-4 illustrates the amplitude values for $q = 4$. In general, adjacent amplitude levels are separated by $2\sqrt{\mathscr{E}_c}/(q-1)$. This assignment guarantees not only that each component s_{ij} is peak-energy-limited to $\sqrt{\mathscr{E}_c}$, but, also, each code word is constrained in average energy to satisfy the condition

$$|\mathbf{s}_i|^2 < n\mathscr{E}_c \tag{7-2-29}$$

By repeating the derivation given above for random selection of codes in an AWGN channel, we find that the average probability of error is upper-bounded as

$$\bar{P}_e < M2^{-nR_0} = 2^{RT}2^{-nT_0} = 2^{-n(R_0 - R/D)} \tag{7-2-30}$$

where R_0 is defined as

$$R_0 = -\log_2\left(\sum_{l=1}^{q}\sum_{m=1}^{q} p_l p_m e^{-d_{lm}^2/4N_0}\right) \tag{7-2-31}$$

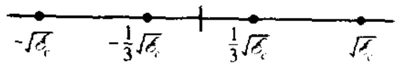

FIGURE 7-2-4 Signal alphabet consisting of four amplitude levels.

and

$$d_{lm} = |a_l - a_m|, \qquad l, m = 1, 2, \ldots, q \tag{7-2-32}$$

In the special case where all the amplitude levels are equally likely, $p_l = p_m = 1/q$ and (7-2-31) reduces to

$$R_0 = -\log_2 \left(\frac{1}{q^2} \sum_{l=1}^{q} \sum_{m=1}^{q} e^{-d_{lm}^2/4N_0} \right) \tag{7-2-33}$$

For example, where $q = 2$ and $a_1 = -\sqrt{\mathscr{E}_c}$, $a_2 = \sqrt{\mathscr{E}_c}$, we have $d_{11} = d_{22} = 0$, $d_{12} = d_{21} = 2\sqrt{\mathscr{E}_c}$, and, hence,

$$R_0 = \log_2 \frac{2}{1 + e^{-\mathscr{E}_c/N_0}}, \qquad q = 2$$

which agrees with our previous result. When $q = 4$, $a_1 = -\sqrt{\mathscr{E}_c}$, $a_2 = -\sqrt{\mathscr{E}_c}/3$, $a_3 = \sqrt{\mathscr{E}_c}/3$, and $a_4 = \sqrt{\mathscr{E}_c}$, we have $d_{mm} = 0$ for $m = 1, 2, 3, 4$, $d_{12} = d_{23} = d_{34} = d_{21} = d_{32} = d_{43} = 2\sqrt{\mathscr{E}_c}/3$, $d_{13} = d_{31} = d_{24} = d_{42} = 4\sqrt{\mathscr{E}_c}/3$, and $d_{14} = d_{41} = 2\sqrt{\mathscr{E}_c}$. Hence,

$$R_0 = \log_2 \frac{8}{2 + 3e^{-\mathscr{E}_c/9N_0} + 2e^{-4\mathscr{E}_c/9N_0} + e^{-\mathscr{E}_c/N_0}}, \qquad q = 4 \tag{7-2-34}$$

Clearly, R_0 now saturates at 2 bits/dimension as $\mathscr{E}_c/N_0$ increases.

The graphs of R_0 as a function of $\mathscr{E}_c/N_0$ for equally spaced and equally likely amplitude levels are shown in Fig. 7-2-5 for $q = 2$, 3, 4, 8, 16, 32, and 64. Note that the saturation level now occurs at $\log_2 q$ bits/dimension. Consequently, for high SNR, $\bar{P}_e \to 0$ as $n \to \infty$, provided that $R < DR_0 = 2WR_0$ bits/s.

If we remove the peak energy constraint on each of the elements, but retain

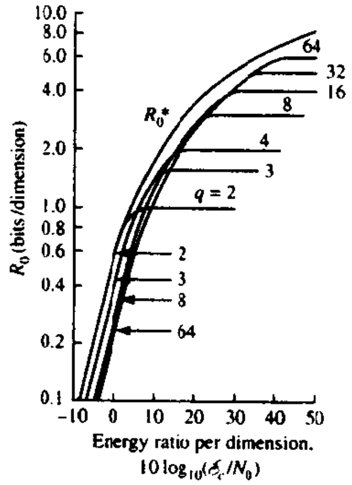

FIGURE 7-2-5 Cutoff rate R_0 for equally spaced q-level amplitude modulation with equal probabilities $p_c = 1/q$. [*From* Principles of Communication Engineering, *by J. M. Wozencraft and I. M. Jacobs,© 1965 by John Wiley and Sons, Inc. Reprinted with permission of the publisher.*]

the average energy constraint per code word as given by (7-2-29) it is possible to obtain a larger upper bound on the number of bits per dimension. For this case, the result obtained by Shannon (1959b) is

$$R_0^* = \frac{1}{2}\left[1 + \frac{\mathscr{E}_c}{N_0} - \sqrt{1 + \left(\frac{\mathscr{E}_c}{N_0}\right)^2}\right]\log_2 e$$

$$+ \tfrac{1}{2}\log_2\left[\frac{1}{2}\left(1 + \sqrt{1 + \left(\frac{\mathscr{E}_c}{N_0}\right)^2}\right)\right] \qquad (7\text{-}2\text{-}35)$$

The graph of R_0^* as a function of the SNR per dimension, $\mathscr{E}_c/N_0$, is also shown in Fig. 7-2-5. It is clear that our selection of the equally spaced, equally likely amplitude levels that result in R_0 is suboptimum. However, these coded signals are easily generated and implemented in practice. This is an important advantage that justifies their use.

7-2-3 Comparison of R_0^* with the Capacity of the AWGN Channel

The channel capacity of the band-limited additive white gaussian noise channel with an average power constraint on the input signal was derived in Section 7-1-2, and is given by

$$C = W\log_2\left(1 + \frac{P_{av}}{WN_0}\right) \text{ bits/s} \qquad (7\text{-}2\text{-}36)$$

where P_{av} is the average power of the input signal and W is the channel bandwidth. It is interesting to express the capacity of this channel in terms of bits/dimension and the average power in terms of energy/dimension. With $D = 2W$ and

$$\mathscr{E}_c = \frac{\mathscr{E}}{n} = \frac{P_{av}T}{n}$$

we have

$$P_{av} = \frac{n}{T}\mathscr{E}_c = D\mathscr{E}_c \qquad (7\text{-}2\text{-}37)$$

By defining $C_n = C/2W = C/D$ and substituting for W and P_{av}, (7-2-36) may be expressed as

$$C_n = \tfrac{1}{2}\log_2\left(1 + 2\frac{\mathscr{E}_c}{N_0}\right)$$

$$= \tfrac{1}{2}\log_2(1 + 2R_c\gamma_b) \text{ bits/dimension} \qquad (7\text{-}2\text{-}38)$$

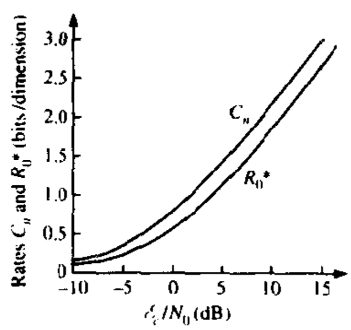

FIGURE 7-2-6 Comparison of cutoff rate R_0^* with the channel capacity for an AWGN channel.

This expression for the normalized capacity may be compared with R_0^*, as shown in Fig. 7-2-6. Since C_n is the ultimate upper limit on the transmission rate R/D, $R_0^* < C_n$ as expected. We also observe that for small values of $\mathscr{E}_c/N_0$, the difference between R_0^* and C_n is approximately 3 dB. Therefore, the use of randomly selected, optimum average power-limited, multiamplitude signals yields a rate function R_0^* that is within 3 dB of the channel capacity. More elaborate bounding techniques are required to show that the probability of error can be made arbitrarily small when $R < DC_n = 2WC_n = C$.

7-3 COMMUNICATION SYSTEM DESIGN BASED ON THE CUTOFF RATE

In the foregoing discussion, we characterized coding and modulation performance in terms of the error probability, which is certainly a meaningful criterion for system design. However, in many cases, the computation of the error probability is extremely difficult, especially if nonlinear operations such as signal quantization are performed in processing the signal at the receiver, or if the additive noise is nongaussian.

Instead of attempting to compute the exact probability of error for specific codes, we may use the ensemble average probability of error for randomly selected code words. The channel is assumed to have q input symbols $\{0, 1, \ldots, q-1\}$ and Q output symbols $\{0, 1, \ldots, Q-1\}$, and to be characterized by the transition probabilities $P(i \mid j)$, where $j = 0, 1, \ldots, q-1$ and $i = 0, 1, \ldots, Q-1$, with $Q \geq q$. The input symbols occur with probabilities $\{p_j\}$ and are assumed to be statistically independent. In addition, the noise on the channel is assumed to be statistically independent in time, so that there is no dependence among successive received symbols. Under these conditions, the ensemble average probability of error for random selected code words may be derived by applying the Chernoff bound (see Viterbi and Omura, 1979).

The general result that is obtained for the discrete memoryless channel is

$$\bar{P}_e < 2^{-n(R_Q - R/D)} \tag{7-3-1}$$

where n is the block length of the code, R is the information rate in bits/s, D is

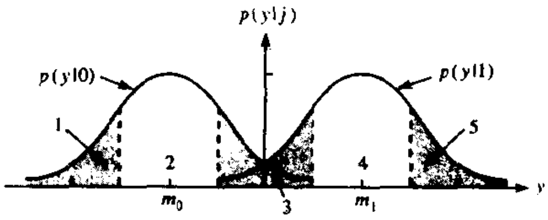

FIGURE 7-3-1 Example of quantization of the demodulator output into five levels.

the number of dimensions per second, and R_Q is the cutoff rate for a quantizer with Q levels, defined as

$$R_Q = \max_{\{p_j\}} \left\{ -\log_2 \sum_{i=0}^{Q-1} \left[\sum_{j=0}^{q-1} p_j \sqrt{P(i \mid j)} \right]^2 \right\} \tag{7-3-2}$$

From the viewpoint of code design, the combination of modulator, waveform channel, and demodulator constitutes a discrete-time channel with q inputs and Q outputs. The transition probabilities $\{P(i \mid j)\}$ depend on the channel noise characteristics, the number of quantization levels, and the type of quantizer, e.g., uniform or nonuniform. For example, in the binary-input AWGN channel, the output of the correlator at the sampling instant may be expressed as

$$p(y \mid j) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(y-m_j)^2/2\sigma^2}, \quad j = 0, 1 \tag{7-3-3}$$

where $m_0 = -\sqrt{\mathscr{E}_c}$, $m_1 = \sqrt{\mathscr{E}_c}$, and $\sigma^2 = \frac{1}{2}N_0$. These two pdfs are shown in Fig. 7-3-1. Also illustrated in the figure is a quantization scheme that subdivides the real line into five regions. From such a subdivision, we may compute the transition probabilities and optimally select the thresholds that subdivide the regions in a way that maximizes R_Q for any given Q. Thus,

$$P(i \mid j) = \int_{r_i} p(y \mid j)\, dy \tag{7-3-4}$$

where the integral of $p(y \mid j)$ is evaluated over the region r_i that corresponds to the transition probability $P(i \mid j)$.

The value of the rate R_Q in the limit as $Q \to \infty$ yields the cutoff rate for the unquantized decoder. It is relatively straightforward to show that as $Q \to \infty$, the first summation (sum from $i = 0$ to $Q - 1$) in (7-3-2) becomes an integral and the transition probabilities are replaced by the corresponding pdfs. Thus, when the channel consists of q discrete inputs and one continuous output y, which represents the unquantized output from a matched filter or a cross-correlator in a system that employs either PSK or a multiamplitude (PAM) modulation, the cutoff rate is given by

$$R_0 = \max_{\{p_j\}} \left\{ -\log_2 \int_{-\infty}^{\infty} dy \left[\sum_{j=0}^{q-1} p_j \sqrt{p(y \mid j)} \right]^2 \right\} \tag{7-3-5}$$

where p_j, $0 \leq j \leq q - 1$, is the probability of transmitting the jth symbol and

$p(y \mid j)$ is the conditional probability density function of the output y from the matched filter or cross-correlator when the jth signal is transmitted. This is the desired expression for unquantized (soft-decision) decoding.

We observe that when the input signal is binary PSK with $p_0 = p_1 = \frac{1}{2}$ and the noise is additive, white, and gaussian, (7-3-5) reduces to the familiar result given previously in (7-2-20).

The general expressions in (7-3-5) and (7-3-2) allow us to compare the performance of various receiver implementations based on a different number of quantization levels.

Example 7-3-1

Let us compare the performance of a binary PSK input signal in an AWGN channel when the receiver quantizes the output to $Q = 2$, 4, and 8 levels. To simplify the optimization problem for the quantization of the signal at the output of the demodulator, the quantization levels are placed at 0, $\pm\tau_h$, $\pm 2\tau_h, \ldots, \pm(2^{b-1} - 1)\tau_h$, where τ_h is the *quantizer step-size parameter,* which is to be selected, and b is the number of bits of the quantizer. A good strategy for the selection of τ_h is to choose it to minimize the SNR per bit γ_b that is required for operation at a code rate R_0. This implies that the step-size parameter must be optimized for every SNR, which in a practical implementation of the receiver means that the SNR must be measured. Fortunately, τ_h does not exhibit high sensitivity to small changes in SNR, so that it is possible to optimize τ_h for one SNR and obtain good performance for a wide range of SNRs about this nominal value by using a fixed τ_h.

Based on this approach, the expression for R_Q given by (7-3-2) was evaluated for $b = 1$ (hard-decision decoding), 2, and 3 bits, corresponding to $Q = 2$, 4, and 8 levels of quantization. The results are plotted in Fig. 7-3-2. The value of R_0 for unquantized soft-decision decoding, obtained by evaluating (7-3-5) is also shown in Fig. 7-3-2. We observe that two-bit quantization with $\tau_h = 1.0$ gains about 1.4 dB over hard-decision decoding, and three-bit quantization with $\tau_h = 0.5$ yields an additional 0.4 dB improvement. Thus, with a three-bit quantizer, we are within 0.2 dB of the

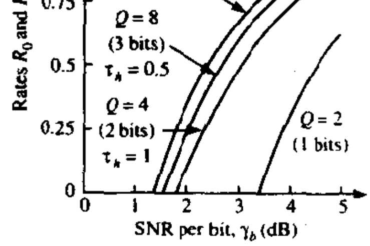

FIGURE 7-3-2 Effect of quantization on the performance of a coded communications system operating at a rate $R = R_0$ or $R = R_Q$, with binary PSK modulation on an AWGN channel.

unquantized soft-decision decoding limit. Clearly, there is little to be gained by increasing the precision any further.

When a nonbinary code is used in conjunction with M-ary ($M = q$) signaling, the received signal at the output of the M matched filters may be represented by the vector $\mathbf{y} = [y_1 \ y_2 \ \cdots \ y_M]$. The cutoff rate for this M-input, M-output (unquantized) channel is

$$R_0 = \max_{\{p_j\}} \left\{ -\log_2 \sum_{j=0}^{M-1} \sum_{i=0}^{M-1} p_j p_i \int_{-\infty}^{\infty} \sqrt{p(\mathbf{y} \mid j) p(\mathbf{y} \mid i)}\, d\mathbf{y} \right\} \tag{7-3-6}$$

where $p(\mathbf{y} \mid j)$ is the conditional probability density function of the output vector $\mathbf{y}$ from the demodulator given that the jth signal was transmitted. Note that (7-3-6) is similar in form to (7-3-5) except that we now have an M-fold integral to perform because there are M outputs from the demodulator.

Let us assume that the M signals are orthogonal so that the M outputs conditioned on a particular input signal are statistically independent. As a consequence,

$$p(\mathbf{y} \mid j) = p_{s+n}(y_j) \prod_{\substack{i=0 \\ i \neq j}}^{M-1} p_n(y_i) \tag{7-3-7}$$

where $p_{s+n}(y_j)$ is the pdf of the matched filter output corresponding to the transmitted signal and $\{p_n(y_i)\}$ corresponds to the noise-only outputs from the other $M-1$ matched filters. When (7-3-7) is incorporated into (7-3-6) we obtain

$$R_0 = \max_{\{p_j\}} \left\{ -\log_2 \left[\sum_{j=0}^{M-1} p_j^2 + \sum_{j=0}^{M-1} \sum_{\substack{i=0 \\ i \neq j}}^{M-1} p_i p_j \left(\int_{-\infty}^{\infty} dy \sqrt{p_{s+n}(y) p_n(y)} \right)^2 \right] \right\} \tag{7-3-8}$$

The maximization of R_0 over the set of input probabilities yields $p_j = 1/M$ for $1 \leq j \leq M$. Consequently, (7-3-8) reduces to

$$\begin{aligned} R_0 &= \log_2 \left\{ \frac{M}{1 + (M-1)[\int_{-\infty}^{\infty} \sqrt{p_{s+n}(y) p_n(y)}\, dy]^2} \right\} \\ &= \log_2 M - \log_2 \left\{ 1 + (M-1) \left[\int_{-\infty}^{\infty} \sqrt{p_{s+n}(y) p_n(y)}\, dy \right]^2 \right\} \end{aligned} \tag{7-3-9}$$

This is the desired result for the cutoff rate of an M-ary input, M-ary vector output unquantized channel.

For phase coherent detection of the M-ary orthogonal signals the appropriate pdfs are

$$\begin{aligned} P_{s+n}(y) &= \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(y-m)^2/2\sigma^2} \\ p_n(y) &= \frac{1}{\sqrt{2\pi}\,\sigma} e^{-y^2/2\sigma^2} \end{aligned} \tag{7-3-10}$$

where $m = \sqrt{\mathscr{E}}$ and $\sigma^2 = \frac{1}{2}N_0$. Substituting these relations into (7-3-9) and evaluating the integral yields

$$R_0 = \log_2 \left[\frac{M}{1 + (M-1)e^{-\mathscr{E}/2N_0}} \right]$$

$$= \log_2 \left[\frac{M}{1 + (M-1)e^{-R_w \gamma_b/2}} \right] \tag{7-3-11}$$

where $\mathscr{E}$ is the received energy per waveform, R_w is the information rate in bits/waveform, and $\gamma_b = \mathscr{E}_b/N_0$ is the SNR per bit.

We should emphasize that the rate parameter R_w has imbedded in it the code rate R_c. For example, if $M = 2$ and the code is binary then $R_w = R_c$. More generally, if the code is binary and $M = 2^\nu$ then each M-ary waveform conveys $R_w = \nu R_c$ bits of information. It is also interesting to note that if the code is binary and $M = 2$ then (7-3-11) reduces to

$$R_0 = \log_2 \left(\frac{2}{1 + e^{-R_c \gamma_b/2}} \right), \quad M = 2 \text{ orthogonal signals} \tag{7-3-12}$$

which is 3 dB worse than the cutoff rate for antipodal signals. If we set $R_w = R_0$ in (7-3-11) and solve for γ_b, we obtain

$$\gamma_b = \frac{2}{R_0} \ln \left(\frac{M-1}{2^{-R_0}M - 1} \right) \tag{7-3-13}$$

Graphs of R_0 versus γ_b for several values of M are illustrated in Fig. 7-3-3. Note that the curve for any value of M saturates at $R_0 = \log_2 M$.

It is also interesting to consider the limiting form of (7-3-11) in the limit as $M \to \infty$. This yields

$$\lim_{M \to \infty} R_0 = \frac{\mathscr{E}}{2 N_0 \ln 2} \text{ bits/waveform} \tag{7-3-14}$$

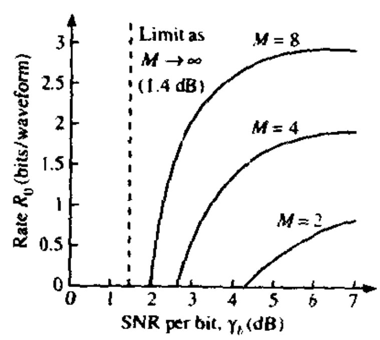

FIGURE 7-3-3 SNR per bit required to operate at a rate R_0 with M-ary orthogonal signals detected coherently in an AWGN channel.

Since $\mathcal{E} = P_{av}T$, where T is the time interval per waveform, it follows that

$$\lim_{M\to\infty} \frac{R_0}{T} = \frac{P_{av}}{2\,N_0 \ln 2} = \tfrac{1}{2}C_\infty \tag{7-3-15}$$

Hence, in the limit as $M \to \infty$, the cutoff rate is one-half of the capacity for the infinite bandwidth AWGN channel. Alternatively, the substitution of $\mathcal{E} = R_0\mathcal{E}_b$ into (7-3-14) yields $\gamma_b = 2\ln 2$ (1.4 dB), which is the minimum SNR required to operate at R_0 (as $M \to \infty$). Hence, signaling at a rate R_0 requires 3 dB more power than the Shannon limit.

The value of R_0 given in (7-3-11) is based on the use of M-ary orthogonal signals, which are clearly suboptimal when M is small. If we attempt to maximize R_0 by selecting the best set of M waveforms, we should not be surprised to find that the simplex set of waveforms is optimum. In fact, R_0 for these optimum waveforms is simply given as

$$R_0 = \log_2\left[\frac{M}{1 + (M-1)e^{-M\mathcal{E}/2(M-1)N_0}}\right] \tag{7-3-16}$$

If we compare this expression with (7-3-11) we observe that R_0 in (7-3-16) simply reflects the fact that the simplex set is more energy-efficient by a factor $M/(M-1)$.

In the case of noncoherent detection, the probability density functions corresponding to signal-plus-noise and noise alone may be expressed as

$$\begin{aligned} P_{s+n}(y) &= ye^{-(y^2+a^2)/2}I_0(ay), \qquad y \geq 0 \\ p_n(y) &= ye^{-y^2/2}, \qquad y \geq 0 \end{aligned} \tag{7-3-17}$$

where, by definition, $a = \sqrt{2\mathcal{E}/N_0}$. The computation of R_0 given by (7-3-9) does not yield a closed-form solution. Instead, the integral in (7-3-9) must be evaluated numerically. Results for this case have been given by Jordan (1966) and Bucher (1980). For example, the (normalized) cutoff rate R_0 for M-ary orthogonal signals with noncoherent detection is shown in Fig. 7-3-4 for

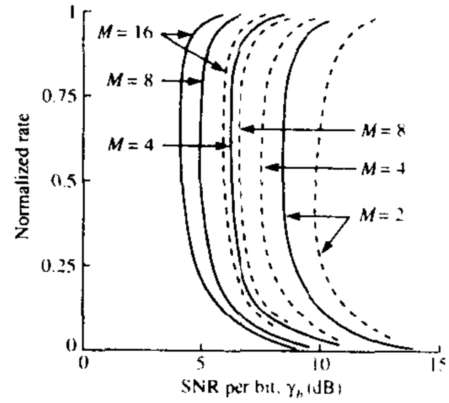

FIGURE 7-3-4 SNR per bit required to operate at a rate R_0 with M-ary orthogonal signals detected noncoherently in an AWGN channel.

$M = 2, 4, 8$, and 16. For purposes of comparison we also plot the cutoff rate for hard-decision decoding ($Q = M$) of the M-ary symbols. In this case, we have

$$R_Q = \log_2 \left\{ \frac{M}{[\sqrt{(1-P_M)} + \sqrt{(M-1)P_M}]^2} \right\}, \qquad Q = M \qquad (7\text{-}3\text{-}18)$$

where P_M is the probability of a symbol error. For a relatively broad range of rates, the difference between soft- and hard-decision decoding is approximately 2 dB.

The most striking characteristic of the performance curves in Fig. 7-3-4 is that there is an optimum code rate for any given M. Unlike the case of coherent detection, where the SNR per bit decreases monotonically with a decrease in code rate, the SNR per bit for noncoherent detection reaches a minimum in the vicinity of a normalized rate of 0.5, and increases for both high and low rates. The minimum is rather broad, so there is really a range of rates from 0.2 to 0.9 where the SNR per bit is within 1 dB of the minimum. This characteristic behavior in the performance with noncoherent detection is attributed to the nonlinear characteristic of the detector.

7-4 BIBLIOGRAPHICAL NOTES AND REFERENCES

The pioneering work on channel characterization in terms of channel capacity and random coding was done by Shannon (1948a, b, 1949). Additional contributions were subsequently made by Gilbert (1952), Elias (1955), Gallager (1965), Wyner (1965), Shannon *et al.* (1967), Forney (1968) and Viterbi (1969). All of these early publications are contained in the IEEE Press book entitled *Key Papers in the Development of Information Theory*, edited by Slepian (1974).

The use of the cutoff rate parameter as a design criterion was proposed and developed by Wozencraft and Kennedy (1966) and by Wozencraft and Jacobs (1965). It was used by Jordan (1966) in the design of coded waveforms for M-ary orthogonal signals with coherent and noncoherent detection. Following these pioneering works, the cutoff rate has been widely used as a design criterion for coded signals in a variety of different channel conditions.

PROBLEMS

7-1 Show that the following two relations are necessary and sufficient conditions for the set of input probabilities $\{P(x_j)\}$ to maximize $I(X; Y)$ and, thus, to achieve capacity for a DMC:

$$I(x_j; Y) = C \quad \text{for all } j \text{ with } P(x_j) > 0$$
$$I(x_j; Y) \leq C \quad \text{for all } j \text{ with } P(x_j) = 0$$

where C is the capacity of the channel and

$$I(x_j; Y) = \sum_{i=0}^{Q-1} P(y_i \mid x_j) \log \frac{P(y_i \mid x_j)}{P(y_i)}$$

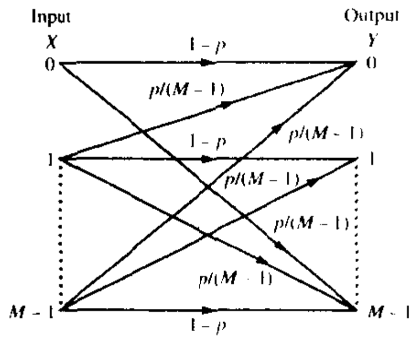

FIGURE P7-2

7-2 Figure P7-2 illustrates an M-ary symmetric DMC with transition probabilities $P(y \mid x) = 1 - p$ when $x = y = k$ for $k = 0, 1, \ldots, M - 1$, and $P(y \mid x) = p/(M - 1)$ when $x \neq y$.

a Show that this channel satisfies the condition given in Problem 7-1 when $P(x_k) = 1/M$.

b Determine and plot the channel capacity as a function of p.

7-3 Determine the capacities of the channels shown in Fig. P7-3.

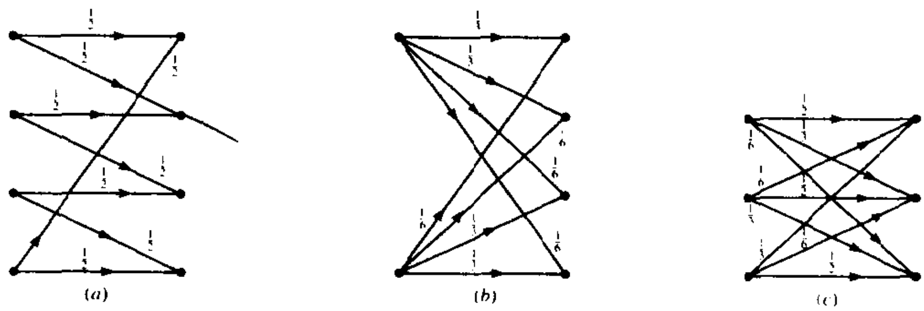

FIGURE P7-3

7-4 Consider the two channels with the transition probabilities as shown in Fig. P7-4. Determine if equally probable input symbols maximize the information rate through the channel.

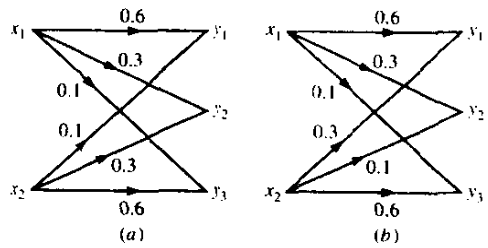

FIGURE P7-4

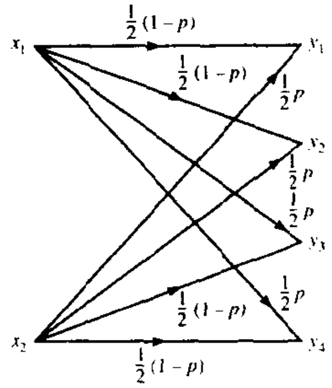

FIGURE P7-6

7-5 A telephone channel has a bandwidth $W = 3000$ Hz and a signal-to-noise power ratio of 400 (26 dB). Suppose we characterize the channel as a band-limited AWGN waveform channel with $P_{av}/WN_0 = 400$.

a Determine the capacity of the channel in bits/s.

b Is the capacity of the channel sufficient to support the transmission of a speech signal that has been sampled and encoded by means of logarithmic PCM?

c Usually, channel impairments other than additive noise limit the transmission rate over the telephone channel to less than the channel capacity of the equivalent band-limited AWGN channel considered in (a). Suppose that a transmission rate of $0.7C$ is achievable in practice without channel encoding. Which of the speech source encoding methods described in Section 3-5 provide sufficient compression to fit the bandwidth restrictions of the telephone channel?

7-6 Consider the binary-input, quaternary-output DMC shown in Fig. P7-6.

a Determine the capacity of the channel.

b Show that this channel is equivalent to a BSC.

7-7 Determine the channel capacity for the channel shown in Fig. P7-7.

8-8 Consider a BSC with crossover probability of error p. Suppose that R is the number of bits in a source code word that represents one of 2^R possible levels at the output of a quantizer. Determine

a the probability that a code word transmitted over the BSC is received correctly;

b the probability of having at least one bit error in a code word transmitted over the BSC;

c the probability of having n_e or less bit errors in a code word;

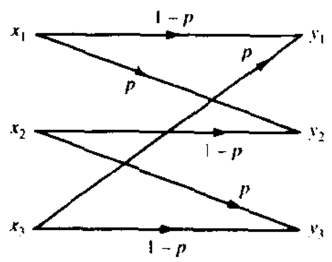

FIGURE P7-7

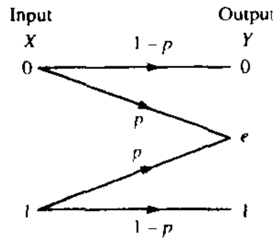

FIGURE P7-10

d Evaluate the probability in (a), (b), and (c) for $R = 5$, $p = 0.01$, and $n_e = 5$.

7-9 Show that, for a DMC, the average mutual information between a sequence $X_1 X_2 \cdots X_n$ of channel inputs and the corresponding channel outputs satisfy the condition

$$I(X_1 X_2 \cdots X_n; Y_1, Y_2, \ldots, Y_n) \leq \sum_{i=1}^{n} I(X_i; Y_i)$$

with equality if and only if the set of input symbols is statistically independent.

7-10 Figure P7-10 illustrates a binary erasure channel with transition probabilities $P(0 \mid 0) = P(1 \mid 1) = 1 - p$ and $P(e \mid 0) = P(e \mid 1) = p$. The probabilities for the input symbols are $P(X = 0) = \alpha$ and $P(X = 1) = 1 - \alpha$.

a Determine the average mutual information $I(X; Y)$ in bits.

b Determine the value of α that maximizes $I(X; Y)$, i.e., the channel capacity C in bits/channel use, and plot C as a function of p for the optimum value of α.

c For the value of α found in (b), determine the mutual information $I(x; y) = I(0; 0)$, $I(1; 1)$, $I(0; e)$, and $I(1; e)$.

7-11 Consider the binary-input, ternary-output channel with transition probabilities shown in Fig. P7-11, where e denotes an erasure. For the AWGN channel, α and p are defined as

$$\alpha = \frac{1}{\sqrt{\pi N_0}} \int_{-\beta}^{\beta} e^{-(x+\sqrt{\mathcal{E}_c})^2/N_0}\, dx$$

$$p = \frac{1}{\sqrt{\pi N_0}} \int_{\beta}^{\infty} e^{-(x+\sqrt{\mathcal{E}_c})^2/N_0}\, dx$$

a Determine R_Q for $Q = 3$ as a function of the probabilities α and p.

b The rate parameter R_Q depends on the choice of the threshold β through the probabilities α and p. For any $\mathcal{E}_c/N_0$, the value of β that maximizes R_Q can be determined by trial and error. For example, it can be shown that for $\mathcal{E}_c/N_0$ below 0 dB, $\beta_{opt} = 0.65\sqrt{\frac{1}{2}N_0}$; for $1 \leq \mathcal{E}_c/N_0 \leq 10$, β_{opt} varies approximately

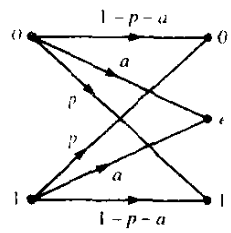

FIGURE P7-11

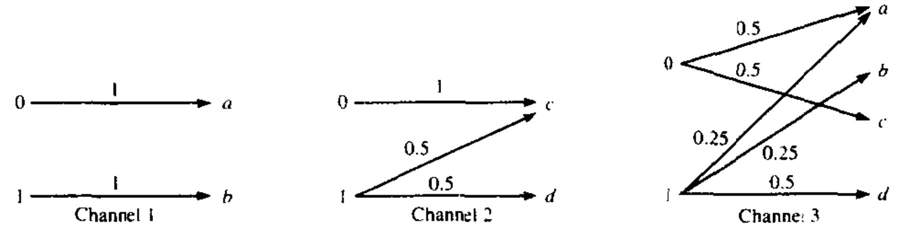

FIGURE P7-13

linearly between $0.65\sqrt{\frac{1}{2}N_0}$ and $1.0\sqrt{\frac{1}{2}N_0}$. By using $\beta = 0.65\sqrt{\frac{1}{2}N_0}$ for the entire range of $\mathscr{E}_c/N_0$, plot R_Q versus $\mathscr{E}_c/N_0$ and compare this result with R_Q ($Q = \infty$).

7-12 Find the capacity of the cascade connection of n binary-symmetric channels with the same crossover probability ϵ. What is the capacity when the number of channels goes to infinity?

7-13 Channels 1, 2, and 3 are shown in Fig. P7-13.
  a Find the capacity of channel 1. What input distribution achieves capacity?
  b Find the capacity of channel 2. What input distribution achieves capacity?
  c Let C denote the capacity of the third channel and C_1 and C_2 represent the capacities of the first and second channel. Which of the following relations holds true and why?

$$C < \tfrac{1}{2}(C_1 + C_2) \tag{i}$$
$$C = \tfrac{1}{2}(C_1 + C_2) \tag{ii}$$
$$C > \tfrac{1}{2}(C_1 + C_2) \tag{iii}$$

7-14 Let C denote the capacity of a discrete memoryless channel with input alphabet $\mathscr{X} = \{x_1, x_2, \ldots, x_N\}$ and output alphabet $\mathscr{Y} = \{y_1, y_2, \ldots, y_M\}$. Show that $C \leq \min\{\log M, \log N\}$.

7-15 The channel C (known as the Z channel) is shown in Fig. P7-15.
  a Find the input probability distribution that achieves capacity.
  b What is the input distribution and capacity for the special cases $\epsilon = 0$, $\epsilon = 1$, and $\epsilon = 0.5$?
  c Show that if n such channels are cascaded, the resulting channel will be equivalent to a Z channel with $\epsilon_1 = \epsilon^n$.
  d What is the capacity of the equivalent Z channel when $n \to \infty$.

7-16 Find the capacity of an additive white Gaussian noise channel with a bandwidth 1 MHz, power 10 W, and noise power spectral density $\frac{1}{2}N_0 = 10^{-9}$ W/Hz.

7-17 Channel C_1 is an additive white gaussian noise channel with a bandwidth W, average transmitter power P, and noise power spectral density $\frac{1}{2}N_0$. Channel C_2 is

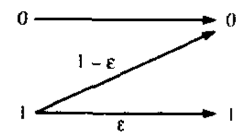

FIGURE P7-15

an additive gaussian noise channel with the same bandwidth and power as channel C_1 but with noise power spectral density $\Phi_n(f)$. It is further assumed that the total noise power for both channels is the same; that is,

$$\int_{-W}^{W} \Phi_n(f)\, df = \int_{-W}^{W} \tfrac{1}{2}N_0\, df = N_0 W$$

Which channel do you think has a larger capacity? Give an intuitive reasoning.

7-18 A discrete-time memoryless gaussian source with mean 0 and variance σ^2 is to be transmitted over a binary-symmetric channel with crossover probability p.

a What is the minimum value of the distortion attainable at the destination (distortion is measured in mean-squared error)?

b If the channel is a discrete-time memoryless additive gaussian noise channel with input power P and noise power P_n, what is the minimum attainable distortion?

c Now assume that the source has the same basic properties but is not memoryless. Do you expect the distortion in transmission over the binary-symmetric channel to be decreased or increased? Why?

7-19 X is a binary memoryless source with $p(X = 0) = 0.3$. This source is transmitted over a binary-symmetric channel with crossover probability $p = 0.1$.

a Assume that the source is directly connected to the channel, i.e., no coding is employed. What is the error probability at the destination?

b If coding is allowed, what is the minimum possible error probability in the reconstruction of the source.

c For what values of p is reliable transmission possible (with coding, of course)?

7-20 Plot the capacity of an AWGN channel that employs binary antipodal signaling, with optimal bit-by-bit detection at the receiver, as a function of $\mathscr{E}_b/N_0$. On the same axis, plot the capacity of the same channel when binary orthogonal signaling is employed.

7-21 In a coded communication system, M messages $1, 2, \ldots, M = 2^k$ are transmitted by M *baseband* signals $x_1(t), x_2(t), \ldots, x_M(t)$, each of duration nT. The general form of $x_i(t)$ is given by

$$x_i(t) = \sum_{j=0}^{n-1} f_{ij}(t - jT)$$

where $f_{ij}(t)$ can be either of the two signals $f_1(t)$ or $f_2(t)$, where $f_1(t) = f_2(t) \equiv 0$ for all $t \notin [0, T]$. We further assume that $f_1(t)$ and $f_2(t)$ have equal energy $\mathscr{E}$ and the channel is ideal (no attenuation) with additive white gaussian noise of power spectral density $\frac{1}{2}N_0$. This means that the received signal is $r(t) = x(t) + n(t)$, where $x(t)$ is one of the $x_i(t)$ and $n(t)$ represents the noise.

a With $f_1(t) = -f_2(t)$, show that N, the dimensionality of the signal space, satisfies $N \leq n$.

b Show that, in general, $N \leq 2n$.

c With $M = 2$, show that, for general $f_1(t)$ and $f_2(t)$,

$$p(\text{error} \mid x_1(t) \text{ sent}) \leq \int \cdots \int_{R^N} \sqrt{p(\mathbf{r} \mid \mathbf{x}_1) p(\mathbf{r} \mid \mathbf{x}_2)}\, d\mathbf{r}$$

where $\mathbf{r}$, $\mathbf{x}_1$, and $\mathbf{x}_2$ are the vector representations of $r(t)$, $x_1(t)$, and $x_2(t)$ in the N-dimensional space.

d Using the result of (c), show that, for general M,

$$p(\text{error} \mid x_m(t) \text{ sent}) \leq \sum_{\substack{1 \leq m' \leq M \\ m' \neq m}} \int_{R^N} \cdots \int \sqrt{p(\mathbf{r} \mid \mathbf{x}_m) p(\mathbf{r} \mid \mathbf{x}_{m'})}\, d\mathbf{r}$$

e Show that

$$\int_{R^N} \cdots \int \sqrt{p(\mathbf{r} \mid \mathbf{x}_m) p(\mathbf{r} \mid \mathbf{x}_{m'})}\, d\mathbf{r} = \exp\left(-\frac{|\mathbf{x}_m - \mathbf{x}_{m'}|^2}{4N_0}\right)$$

and, therefore,

$$p(\text{error} \mid x_m(t) \text{ sent}) \leq \sum_{\substack{1 \leq m' \leq M \\ m' \neq m}} \exp\left(-\frac{|\mathbf{x}_m - \mathbf{x}_{m'}|^2}{4N_0}\right)$$