is a Poisson impulse train with random amplitudes $\{A_i\}$. Assume the $\{A_i\}$ are identically distributed random variables, with mean $\bar{A}$ and second moment $\overline{A^2}$, which are statistically independent of each other and of the $\{\tau_i\}$. *Hint.* Show that

$$\overline{y(t_1)\, y(t_2)} = \overline{A^2}\, \mathrm{E}\left[\sum_i h(t_1 - \tau_i)\, h(t_2 - \tau_i)\right] + \bar{A}^2\, \mathrm{E}\left[\sum_i \sum_{j \neq i} h(t_1 - \tau_i)\, h(t_2 - \tau_j)\right].$$

3.18 Property 4 on p. 156 states that every weighted linear sum of jointly Gaussian random variables is a Gaussian random variable.

a. Prove the converse statement that if

$$y = \sum_{i=1}^{k} a_i x_i,$$

is a Gaussian random variable for every (nonzero) constant vector $\mathbf{a} = (a_1, a_2, \ldots, a_k)$, the $\{x_i\}$ are jointly Gaussian. *Hint.* Calculate the joint characteristic function of the $\{x_i\}$ by noting that

$$M_{\mathbf{x}}(\mathbf{v}) = M_y(1)\,\big|_{\mathbf{a}=\mathbf{v}} = e^{-\frac{1}{2}\sigma_y^2} e^{j\bar{y}}$$

and compare with Eq. 3.76 after evaluating σ_y^2 and $\bar{y}$.

b. The converse statement may be taken as an alternate definition of jointly Gaussian random variables. Prove properties 2 and 4 (p. 156) directly from this definition without recourse to the multivariate characteristic function. Observe that with this alternate definition the multivariate central limit theorem can be reduced to a single-variable theorem.

4

Optimum Receiver Principles

The concepts and methods of random processes studied in Chapter 3, together with the a posteriori probability viewpoint of communication discussed in Chapter 2, provide the background necessary to treat the problem of optimum communication receiver design. In this chapter we apply this background to the particular communication system diagrammed in Fig. 4.1. Here one of a discrete set of specified waveforms

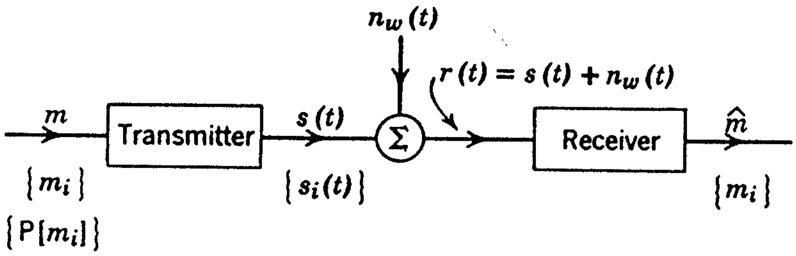


Figure 4.1 Communication over an additive white Gaussian noise channel.

$\{s_i(t)\}$, $i = 0, 1, \ldots, M-1$, is transmitted over a channel disturbed by the addition of white Gaussian noise, so that the received signal process is

$$r(t) = s(t) + n_w(t). \tag{4.1}$$

Which waveform is actually transmitted depends on the random message input, m; when $m = m_i$, the transmitted signal is $s_i(t)$. Thus the correspondence

$$m = m_i \Leftrightarrow s(t) = s_i(t) \tag{4.2}$$

defines the transmitter. The a priori probabilities $\{P[m_i]\}$ specify the input source.

The first part of this chapter is devoted to investigating how the received signal $r(t)$ should be processed in order to produce an estimate, $\hat{m}$, of the transmitter input m that is optimum in the sense that the probability of error

$$P[\mathcal{E}] \triangleq P[\hat{m} \neq m] \tag{4.3}$$

is minimum. The investigation results in the determination of the optimum receiver structure; that is, in the specification of what operations to perform on $r(t)$.

In formulating the optimum receiver design problem, we assume that the a priori probabilities $\{P[m_i]\}$ and signals $\{s_i(t)\}$ are known. The chapter concludes with a discussion of how the minimum achievable probability of error depends on the choice of these a priori data. In particular, certain signal sets of practical importance are evaluated and compared.

In Chapter 7 we extend the results of this chapter to the design and evaluation of optimum receivers for certain channels that disturb the transmitted signal in ways more complicated than by the simple addition of white Gaussian noise.

4.1 BASIC APPROACH TO OPTIMUM RECEIVER DESIGN

In Chapter 3 we have seen that the transmitted signal $s(t)$, the disturbing noise $n_w(t)$, and the received signal $r(t)$ in Fig. 4.1 are random processes. In addition, we have seen that a random process is specified in terms of the joint density functions that it implies. The key to analyzing communication situations such as that in Fig. 4.1 is to find some way to *replace all waveforms by finite dimensional vectors*, for which we can then calculate the joint density function. We show in Section 4.3 and Appendix 4A that this replacement is permissible. As a preliminary, however, it is convenient first to establish the operations performed by an optimum receiver under the assumption that the replacement of waveforms by vectors has already been accomplished.

4.2 VECTOR CHANNELS

The N-dimensional vector communication system diagrammed in Fig. 4.2 is a straightforward extension of the single random variable system discussed in Chapter 2 in connection with Fig. 2.34. The transmitter is defined by a set of M signal vectors, $\{\mathbf{s}_i\}$. When $m = m_i$, the vector $\mathbf{s}_i$ is transmitted,

$$\mathbf{s}_i = (s_{i1}, s_{i2}, \ldots, s_{iN}); \qquad i = 0, 1, \ldots, M - 1. \tag{4.4a}$$

The vector channel disturbs the transmission and emits a random vector

$$\mathbf{r} = (r_1, r_2, \ldots, r_N). \tag{4.4b}$$

We consider a vector channel to be defined mathematically if and only if the entire set of M conditional density functions $\{p_{\mathbf{r}}(\ \mid \mathbf{s} = \mathbf{s}_i)\}$ is known. For brevity, we follow the usage of Eq. 2.104c and denote this set by $p_{\mathbf{r}|\mathbf{s}}$.

For our vector communication system the optimum receiver is specified as follows: given that any particular vector, say $\mathbf{r} = \boldsymbol{\rho}$, is received, where

$$\boldsymbol{\rho} \triangleq (\rho_1, \rho_2, \ldots, \rho_N), \tag{4.5}$$

the optimum receiver must determine from its knowledge of $p_{\mathbf{r}|\mathbf{s}}$, $\{\mathbf{s}_i\}$, and $\{P[m_i]\}$ which one of the possible transmitter inputs $\{m_i\}$ has maximum a posteriori probability. More precisely, the optimum receiver sets $\hat{m} = m_k$ whenever

$$P[m_k \mid \mathbf{r} = \boldsymbol{\rho}] > P[m_i \mid \mathbf{r} = \boldsymbol{\rho}]; \qquad \text{for } i = 0, 1, \ldots, M-1, i \neq k. \tag{4.6}$$

Proof that such a *maximum a posteriori probability* receiver is in fact optimum follows from noting that when the receiver sets $\hat{m} = m_k$, the

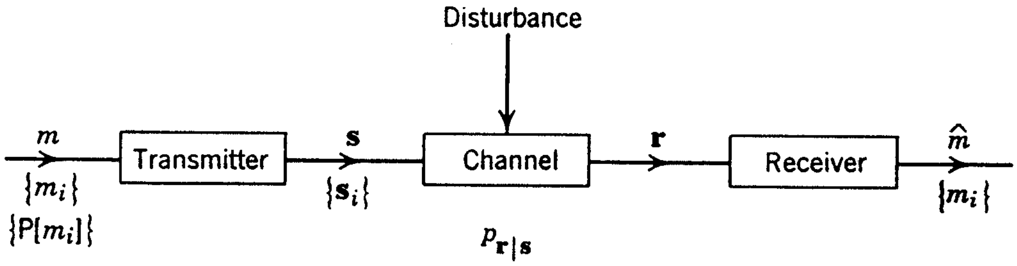


Figure 4.2 A vector communication system.

conditional probability of a correct decision, given that $\mathbf{r} = \boldsymbol{\rho}$, is

$$P[\mathcal{C} \mid \mathbf{r} = \boldsymbol{\rho}] = P[m_k \mid \mathbf{r} = \boldsymbol{\rho}]. \tag{4.7a}$$

The unconditional probability of correct decision can be written

$$P[\mathcal{C}] = \int_{-\infty}^{\infty} P[\mathcal{C} \mid \mathbf{r} = \boldsymbol{\rho}]\, p_{\mathbf{r}}(\boldsymbol{\rho})\, d\boldsymbol{\rho}. \tag{4.7b}$$

Since

$$p_{\mathbf{r}}(\boldsymbol{\rho}) \geqslant 0,$$

it is clear that $P[\mathcal{C}]$ is maximized by maximizing $P[\mathcal{C} \mid \mathbf{r} = \boldsymbol{\rho}]$ for each received vector $\boldsymbol{\rho}$. If two or more m_i yield the same a posteriori probability, the receiver may select $\hat{m}$ from among them in any arbitrary way—for instance, by choosing the one with the smallest index—without affecting the probability of error.

Determination of the a posteriori probabilities $\{P[m_i \mid \mathbf{r} = \boldsymbol{\rho}]\}$ follows from the mixed form of Bayes rule, Eq. 2.103a:

$$P[m_i \mid \mathbf{r} = \boldsymbol{\rho}] = \frac{P[m_i]\, p_{\mathbf{r}}(\boldsymbol{\rho} \mid m_i)}{p_{\mathbf{r}}(\boldsymbol{\rho})}. \tag{4.8a}$$

Since the event $m = m_i$ implies the event $\mathbf{s} = \mathbf{s}_i$, and conversely, we have

$$p_{\mathbf{r}}(\boldsymbol{\rho} \mid m_i) = p_{\mathbf{r}}(\boldsymbol{\rho} \mid \mathbf{s} = \mathbf{s}_i). \tag{4.8b}$$

Finally, since $p_{\mathbf{r}}(\boldsymbol{\rho})$ is independent of the index i, we conclude from Eqs. 4.6 and 4.8 that the optimum receiver, on observing $\mathbf{r} = \boldsymbol{\rho}$, sets $\hat{m} = m_k$ whenever the *decision function*

$$P[m_i]\, p_{\mathbf{r}}(\boldsymbol{\rho} \mid \mathbf{s} = \mathbf{s}_i); \qquad i = 0, 1, \ldots, M-1, \tag{4.9}$$

is maximum for $i = k$.

A receiver that determines $\hat{m}$ by maximizing only the factor $p_{\mathbf{r}}(\boldsymbol{\rho} \mid \mathbf{s} = \mathbf{s}_i)$ without regard to the factor $P[m_i]$ is called a *maximum-likelihood* receiver. Such a receiver is often used when the a priori probabilities $\{P[m_i]\}$ are not known. A maximum-likelihood receiver yields the minimum probability of error when the transmitter inputs are all equally likely.

Decision Regions

The nature of the optimum vector receiver may be clarified by considering the two-dimensional example shown in Fig. 4.3*a*, wherein the vectors are described in terms of coordinates φ_1 and φ_2. We assume three possible input messages, with known a priori probabilities $P[m_0]$, $P[m_1]$, and $P[m_2]$. The corresponding transmitted vectors are assumed to be

$$\mathbf{s}_0 = (1, 2),$$
$$\mathbf{s}_1 = (2, 1), \tag{4.10}$$
$$\mathbf{s}_2 = (1, -2).$$

If we now receive some point $\mathbf{r} = \boldsymbol{\rho}$, as shown, the receiver can calculate $P[m_i]\, p_{\mathbf{r}}(\boldsymbol{\rho} \mid \mathbf{s} = \mathbf{s}_i)$ from knowledge of the functions $p_{\mathbf{r}|\mathbf{s}}$ which define the channel and thereby determine $\hat{m}$ in accordance with the preceding discussion.

We note that this calculation can be carried out for every point $\boldsymbol{\rho}$ in the (φ_1, φ_2) plane and that each such point is thereby assigned to one and only one of the possible inputs $\{m_i\}$. Thus the decision rule of Eq. 4.9 implies a partitioning of the entire plane into disjoint regions, say $\{I_i\}$, $i = 0, 1, 2$, similar in general to those shown in Fig. 4.3*b*. Each region comprises all points such that whenever the received vector $\mathbf{r}$ is in I_k the optimum receiver sets $\hat{m}$ equal to m_k. The correspondence

$$\mathbf{r} \text{ in } I_k \Leftrightarrow \hat{m} = m_k \tag{4.11}$$

defines the optimum receiver.

The regions $\{I_i\}$ are called optimum *decision regions* and are a natural extension of the *decision intervals* considered in Fig. 2.35. We note for future reference that the optimum receiver makes an error when $m = m_k$ if and only if $\mathbf{r}$ falls outside I_k.

It is clear that the concept of decision regions, which for simplicity we have illustrated for a two-dimensional plane, extends directly to the case

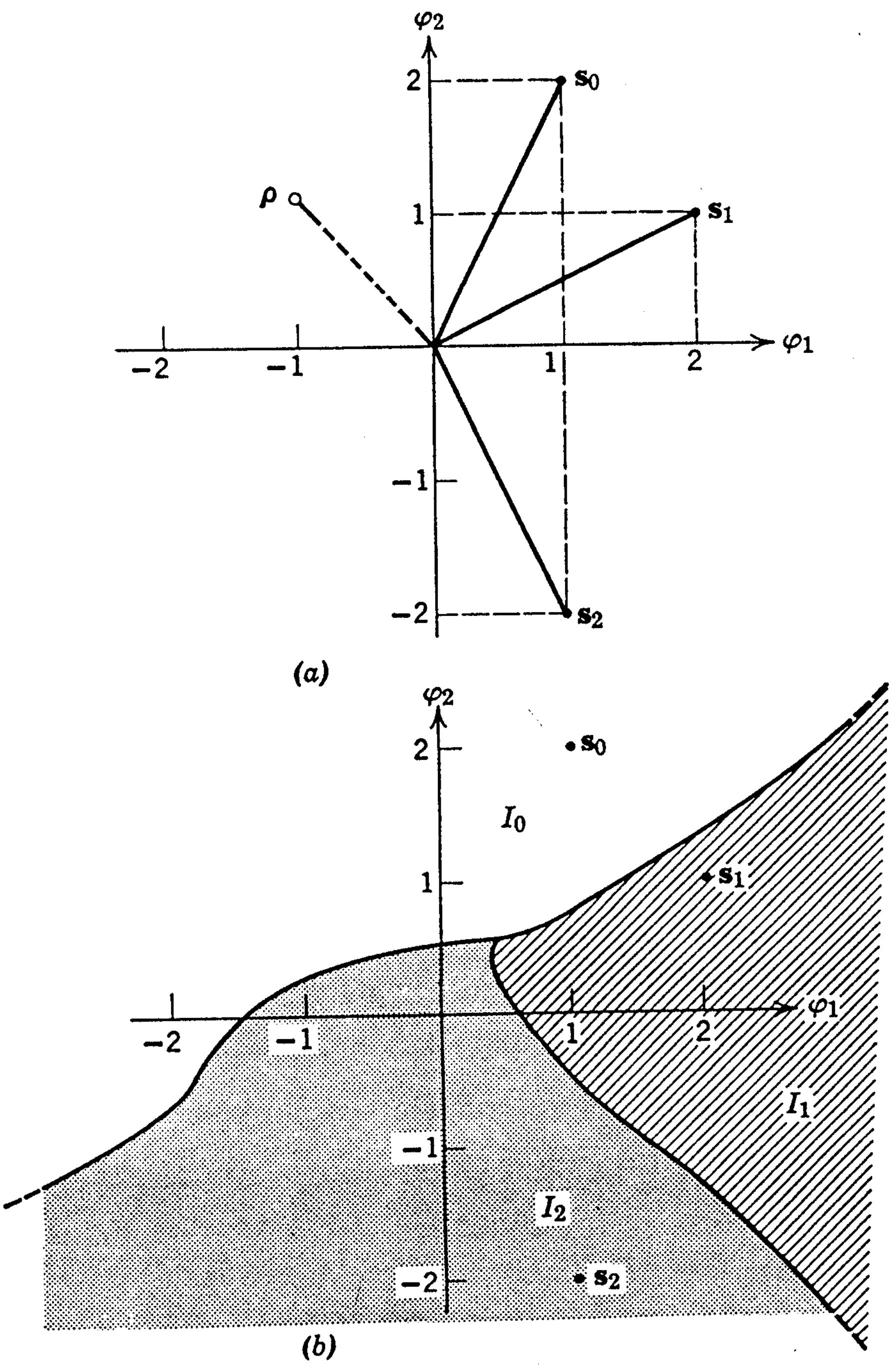


Figure 4.3 A three-signal vector communication problem: (*a*) three two-dimensional signal vectors and a possible received signal $\boldsymbol{\rho}$; (*b*) decision regions.

of an arbitrary number of possible inputs $\{m_i\}$ and to corresponding signals $\{\mathbf{s}_i\}$ that are defined on an arbitrary number of dimensions. The decision function of Eq. 4.9 then implies a partitioning of an N-dimensional received signal space into M disjoint N-dimensional decision regions $\{I_i\}$.

Additive Gaussian Noise

The actual boundaries of the decision regions in any particular case depend by Eq. 4.9 on the a priori probabilities $\{P[m_i]\}$, the signals $\{\mathbf{s}_i\}$, and the definition of the channel $p_{\mathbf{r}|\mathbf{s}}$. In some instances the calculation of these boundaries may be simple; in most it is exceedingly difficult. Fortunately, many situations of practical interest fall into the simple category.

To illustrate a relatively straightforward situation, consider the case in which the channel disturbs the signal vector (as shown in Fig. 4.4) simply

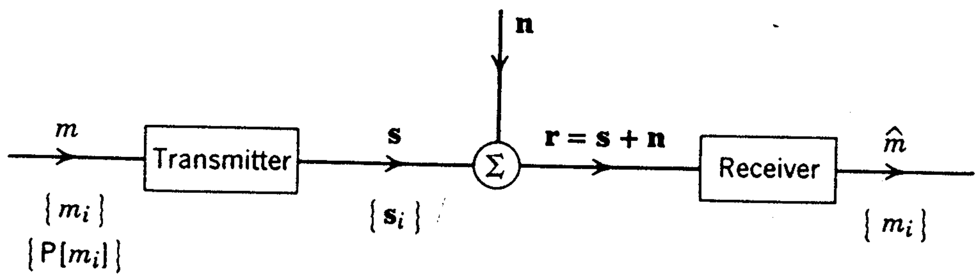


Figure 4.4 An N-dimensional vector communication system.

by adding to it a random noise vector

$$\mathbf{n} = (n_1, n_2, \ldots, n_N). \tag{4.12}$$

The random signal vector $\mathbf{s} = (s_1, s_2, \ldots, s_N)$ and received vector $\mathbf{r}$ are then related by

$$\mathbf{r} = \mathbf{s} + \mathbf{n} = (s_1 + n_1, s_2 + n_2, \ldots, s_N + n_N). \tag{4.13}$$

Since Eq. 4.13 implies that $\mathbf{r} = \boldsymbol{\rho}$ when $\mathbf{s} = \mathbf{s}_i$ if and only if $\mathbf{n} = \boldsymbol{\rho} - \mathbf{s}_i$, the conditional density functions $p_{\mathbf{r}|\mathbf{s}}$ are given by

$$p_{\mathbf{r}}(\boldsymbol{\rho} \mid \mathbf{s} = \mathbf{s}_i) = p_{\mathbf{n}}(\boldsymbol{\rho} - \mathbf{s}_i \mid \mathbf{s} = \mathbf{s}_i); \qquad i = 0, 1, \ldots, M - 1. \tag{4.14}$$

We now make the often-reasonable assumption that $\mathbf{n}$ and $\mathbf{s}$ are *statistically independent* (cf. Eq. 2.104):

$$p_{\mathbf{n}|\mathbf{s}} = p_{\mathbf{n}}. \tag{4.15a}$$

Hence

$$p_{\mathbf{n}}(\boldsymbol{\rho} - \mathbf{s}_i \mid \mathbf{s} = \mathbf{s}_i) = p_{\mathbf{n}}(\boldsymbol{\rho} - \mathbf{s}_i); \qquad \text{all } i. \tag{4.15b}$$

The decision function of Eq. 4.9 is therefore

$$P[m_i]\, p_{\mathbf{n}}(\boldsymbol{\rho} - \mathbf{s}_i). \tag{4.16}$$

In order to simplify the decision function still more, we must specify the noise density function $p_{\mathbf{n}}$. An especially simple and important case is that in which the N components of $\mathbf{n}$ are statistically independent, zero-mean, Gaussian random variables, each with variance σ^2. From

Eq. 3.57 we then have

$$p_{\mathbf{n}}(\boldsymbol{\alpha}) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{j=1}^{N}\alpha_j^2\right). \tag{4.17a}$$

The notation can be contracted by observing that the squared-length of any vector $\boldsymbol{\alpha}$ is defined to be the dot product of $\boldsymbol{\alpha}$ with itself. In the familiar case of $N = 2$ or 3 we have

$$|\boldsymbol{\alpha}|^2 = \boldsymbol{\alpha}\cdot\boldsymbol{\alpha} = \sum_{j=1}^{N}\alpha_j^2, \tag{4.17b}$$

where the $\{\alpha_j\}$ are the Cartesian coordinates of $\boldsymbol{\alpha}$. For larger N length is defined in the same way and Eq. 4.17b remains valid. Thus Eq. 4.17a can be written

$$p_{\mathbf{n}}(\boldsymbol{\alpha}) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-|\boldsymbol{\alpha}|^2/2\sigma^2}. \tag{4.17c}$$

Substituting Eq. 4.17c in Eq. 4.16, we see that for this $p_{\mathbf{n}}$ the optimum receiver sets $\hat{m} = m_k$ whenever

$$P[m_i]\, e^{-|\boldsymbol{\rho}-\mathbf{s}_i|^2/2\sigma^2} \tag{4.18}$$

is maximum for $i = k$. [The factor $(2\pi\sigma^2)^{-N/2}$ is independent of i and its discard entails no loss of optimality.] Finally, we note that maximizing the expression of Eq. 4.18 is equivalent to finding that value of i which *minimizes*

$$|\boldsymbol{\rho} - \mathbf{s}_i|^2 - 2\sigma^2 \ln P[m_i]. \tag{4.19}$$

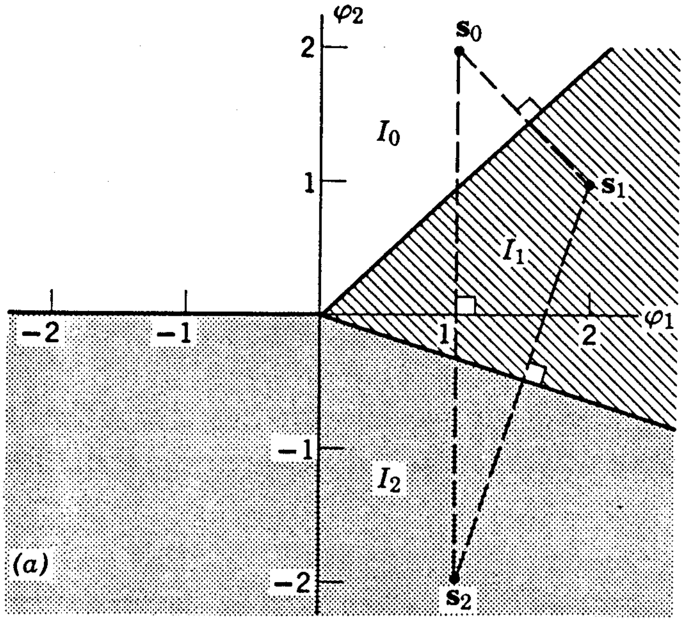
The decision function of Eq. 4.19 is easily visualized geometrically. We recognize that the term $|\boldsymbol{\rho} - \mathbf{s}_i|^2$ is the square of the Euclidean distance between the points $\boldsymbol{\rho}$ and $\mathbf{s}_i$:

$$|\boldsymbol{\rho} - \mathbf{s}_i|^2 = \sum_{j=1}^{N}(\rho_j - s_{ij})^2.$$

Whenever all m_i have equal a priori probability, the optimum decision rule is to assign a received point $\boldsymbol{\rho}$ to m_k if and only if $\boldsymbol{\rho}$ is *closer* to the point $\mathbf{s}_k$ than to any other possible signal. For example, consider the two-dimensional signal set of Eq. 4.10. If all three messages are equally probable, the decision regions are those shown in Fig. 4.5*a*; when the three messages have unequal a priori probabilities, the decision regions are modified in accordance with Eq. 4.19, as indicated in Fig. 4.5*b*.

Once the decision regions $\{I_i\}$ have been determined, an expression for the conditional probability of correct decision follows immediately:

$$P[\mathcal{C} \mid m_i] = P[\mathbf{r} \text{ in } I_i \mid m_i] = \int_{I_i} p_{\mathbf{r}}(\boldsymbol{\rho} \mid \mathbf{s} = \mathbf{s}_i)\, d\boldsymbol{\rho}. \tag{4.20a}$$



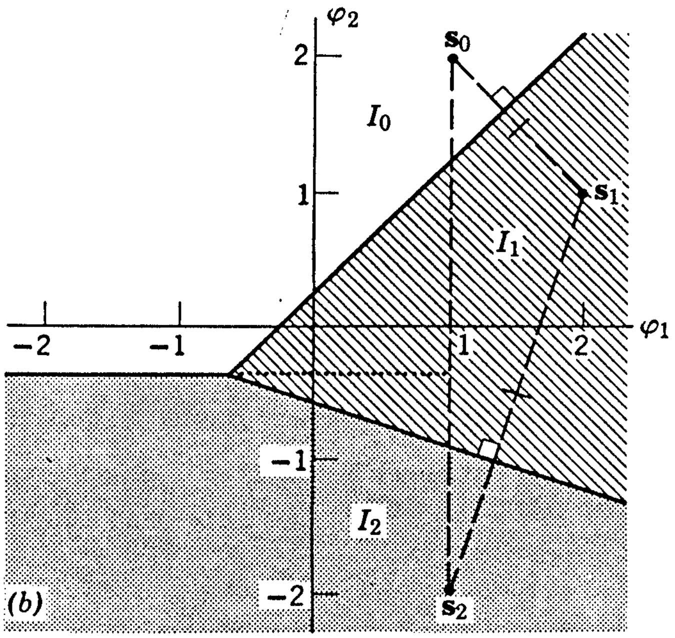


Figure 4.5 Optimum decision regions for additive Gaussian noise: (*a*) the boundaries of the $\{I_i\}$ are the perpendicular bisectors of the sides of the signal triangle whenever $P[m_0] = P[m_1] = P[m_2]$; (*b*) the boundaries of the $\{I_i\}$ are displaced when $P[m_1] > P[m_0] > P[m_2]$.

For additive equal-variance Gaussian noise this becomes

$$P[\mathcal{C} \mid m_i] = \int_{I_i} p_n(\boldsymbol{\rho} - \mathbf{s}_i)\, d\boldsymbol{\rho}$$

$$= \frac{1}{(2\pi\sigma^2)^{N/2}} \int_{I_i} e^{-|\boldsymbol{\rho}-\mathbf{s}_i|^2/2\sigma^2} d\boldsymbol{\rho}\,. \tag{4.20b}$$

The over-all probability of error is

$$P[\mathcal{E}] \triangleq 1 - P[\mathcal{C}] = 1 - \sum_{i=0}^{M-1} P[m_i]\, P[\mathcal{C} \mid m_i]. \tag{4.20c}$$

In Section 4.4 these expressions are evaluated for certain (important) situations in which the decision regions are such that the integrals can be easily calculated or approximated.

Multivector Channels

In the "diversity" communication system shown in Fig. 4.6, in which the transmitted vector **s** is applied at the input of two different channels and the receiver observes the output of both, it is natural to describe the

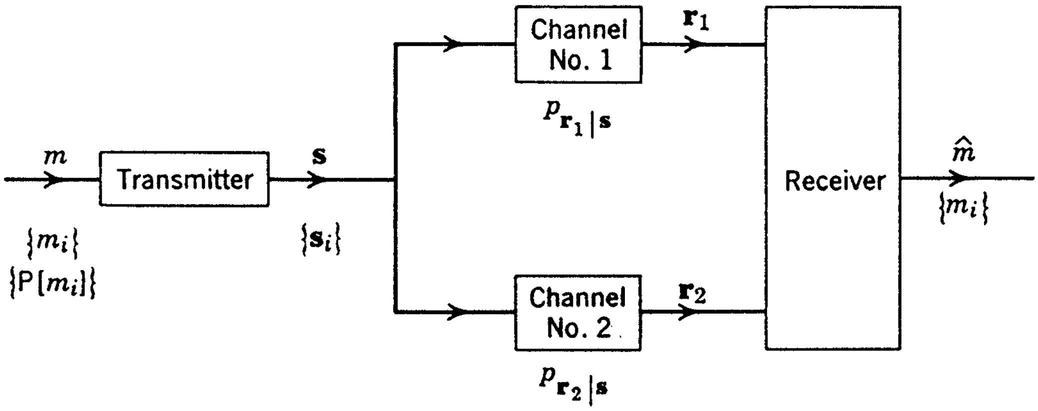


Figure 4.6 A "diversity" vector communication system. [In many situations the vectors **s**, $\mathbf{r}_1$, and $\mathbf{r}_2$ all have the same number of components, but this need not be so.]

total receiver input **r** in terms of vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ that are associated with each channel individually. Thus we write

$$\mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2) \triangleq (r_{11}, r_{12}, \ldots, r_{1k}, r_{21}, r_{22}, \ldots, r_{2l}), \tag{4.21a}$$

where

$$\mathbf{r}_1 = (r_{11}, r_{12}, \ldots, r_{1k}), \tag{4.21b}$$

$$\mathbf{r}_2 = (r_{21}, r_{22}, \ldots, r_{2l}). \tag{4.21c}$$

Given that vectors $\mathbf{r}_1 = \boldsymbol{\rho}_1$ and $\mathbf{r}_2 = \boldsymbol{\rho}_2$ are received, the a posteriori probability of the ith message is

$$P[m_i \mid \mathbf{r} = \boldsymbol{\rho}] = P[m_i \mid \mathbf{r}_1 = \boldsymbol{\rho}_1, \mathbf{r}_2 = \boldsymbol{\rho}_2], \tag{4.22a}$$

where $\boldsymbol{\rho} \triangleq (\boldsymbol{\rho}_1, \boldsymbol{\rho}_2)$. With this notation, the optimum decision rule of Eq. 4.9 is written: set $\hat{m} = m_k$ if and only if

$$P[m_i]\, p_{\mathbf{r}}(\boldsymbol{\rho} \mid \mathbf{s} = \mathbf{s}_i) = P[m_i]\, p_{\mathbf{r}_1,\mathbf{r}_2}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2 \mid \mathbf{s} = \mathbf{s}_i) \tag{4.22b}$$

is maximum for $i = k$.

The theorem of irrelevance. In many cases of practical importance a channel presents some data at its output which an optimum receiver can ignore. For instance, consider the arbitrary vector channel in Fig. 4.7, in which two inputs $\mathbf{r}_1$ and $\mathbf{r}_2$ are available to the receiver. Let us determine the conditions under which the receiver may disregard $\mathbf{r}_2$ without affecting the probability of error.

The optimum decision rule is again given by Eq. 4.22b. If we factor the right-hand side of this equation in accordance with Bayes rule (Eq. 2.103),

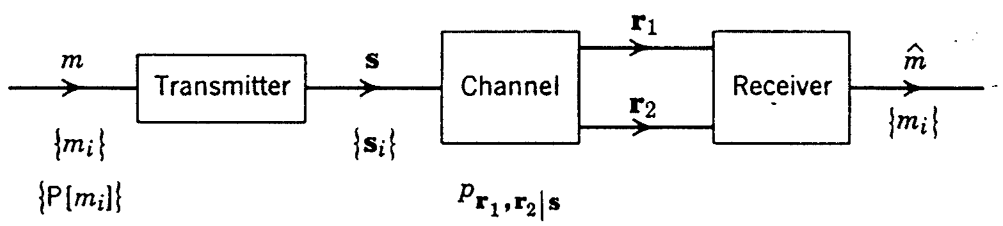


Figure 4.7 An arbitrary vector communication system described in terms of two output vectors.

we see that an optimum receiver sets $\hat{m} = m_k$ following the observation $\mathbf{r}_1 = \boldsymbol{\rho}_1$, $\mathbf{r}_2 = \boldsymbol{\rho}_2$ if and only if the decision function

$$P[m_i]\, p_{\mathbf{r}_1}(\boldsymbol{\rho}_1 \mid \mathbf{s} = \mathbf{s}_i)\, p_{\mathbf{r}_2}(\boldsymbol{\rho}_2 \mid \mathbf{s} = \mathbf{s}_i, \mathbf{r}_1 = \boldsymbol{\rho}_1) \tag{4.23}$$

is maximum for $i = k$. If $\mathbf{r}_2$ *when conditioned on* $\mathbf{r}_1$ is statistically independent of $\mathbf{s}$, then for every value of $\boldsymbol{\rho}_2$

$$\begin{aligned} p_{\mathbf{r}_2}(\boldsymbol{\rho}_2 \mid \mathbf{s} = \mathbf{s}_i, \mathbf{r}_1 = \boldsymbol{\rho}_1) &= p_{\mathbf{r}_2}(\boldsymbol{\rho}_2 \mid \mathbf{r}_1 = \boldsymbol{\rho}_1) \\ &= \text{a number independent of } i. \end{aligned} \tag{4.24}$$

When this is so, the knowledge that $\mathbf{r}_2 = \boldsymbol{\rho}_2$ can never enter into the determination of which value of i maximizes the expression of Eq. 4.23; an optimum receiver may therefore totally ignore $\mathbf{r}_2$. Thus we have the important *theorem of irrelevance: an optimum receiver may disregard a vector* $\mathbf{r}_2$ *if and only if*

$$p_{\mathbf{r}_2|\mathbf{r}_1,\mathbf{s}} = p_{\mathbf{r}_2|\mathbf{r}_1}. \tag{4.25a}$$

Equation 4.25a is a necessary and sufficient condition for ignoring $\mathbf{r}_2$. A sufficient condition is that

$$p_{\mathbf{r}_2|\mathbf{r}_1,\mathbf{s}} = p_{\mathbf{r}_2}. \tag{4.25b}$$

The meaning and utility of this theorem may be demonstrated by considering three examples, each of which involves two additive noise vectors $\mathbf{n}_1$ and $\mathbf{n}_2$ that are statistically independent of one another and of $\mathbf{s}$. The first example, shown in Fig. 4.8, illustrates a situation in which Eq. 4.25b is valid: the received vector $\mathbf{r}_2$ is just the noise $\mathbf{n}_2$, which is statistically independent of both $\mathbf{n}_1$ and $\mathbf{s}$, hence of $\mathbf{s}$ and $\mathbf{r}_1 = \mathbf{n}_1 + \mathbf{s}$.

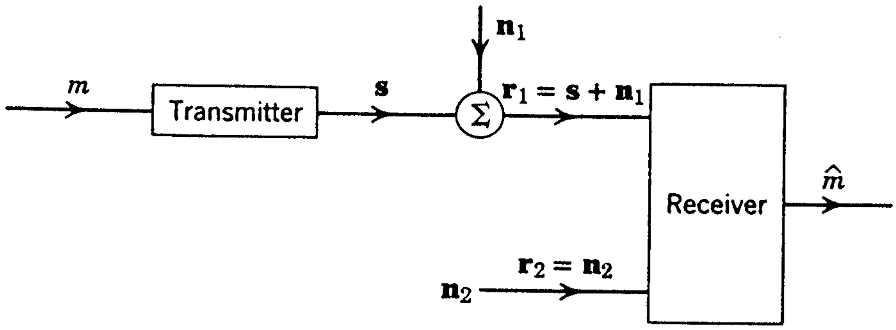


Figure 4.8 The vector r_2 is irrelevant because $p_{r_2|r_1,s} = p_{r_2}$.

Accordingly,

$$p_{r_2|r_1,s} = p_{r_2} \tag{4.26}$$

and r_2 is irrelevant, which is obviously sensible.

The second example, shown in Fig. 4.9, illustrates a situation in which Eq. 4.25a is valid but Eq. 4.25b is not. We have two vector channels in cascade and a receiver that has access to the intermediate output r_1 as well as to the final output r_2. Since r_2 is a corrupted version of r_1, hence

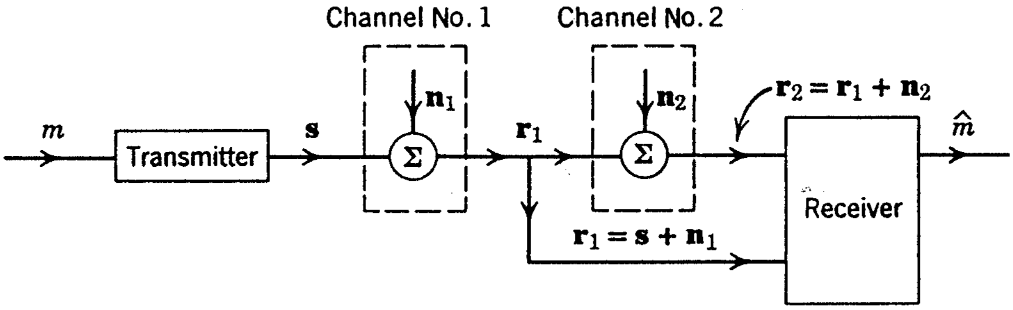


Figure 4.9 The vector r_2 is irrelevant because $p_{r_2|r_1,s} = p_{r_2|r_1}$.

depends on m only through r_1, we feel intuitively that r_2 can tell us nothing about s that is not already conveyed by r_1. We prove this formally by noting that, since $r_2 = r_1 + n_2$, when r_1 is known r_2 depends only on the noise n_2, which is independent of s. Thus for all ρ_2 and i

$$p_{r_2}(\rho_2 \mid r_1 = \rho_1, s = s_i) = p_{n_2}(\rho_2 - \rho_1) = p_{r_2}(\rho_2 \mid r_1 = \rho_1).$$

The condition of Eq. 4.25a is satisfied, and the theorem of irrelevance states that r_2 is of no value to an optimum receiver.

The third example, shown in Fig. 4.10, illustrates a situation in which r_2 *cannot* be discarded by an optimum receiver. We have

$$\begin{aligned} p_{r_2}(\rho_2 \mid r_1 = \rho_1, s = s_i) &= p_{r_2}(\rho_2 \mid n_1 = \rho_1 - s_i, s = s_i) \\ &= p_{n_2}(\rho_2 - \rho_1 + s_i \mid n_1 = \rho_1 - s_i, s = s_i) \\ &= p_{n_2}(\rho_2 - \rho_1 + s_i), \end{aligned}$$

which does depend explicitly on i. Thus Eq. 4.25 is not satisfied and r_2 is not irrelevant, even though r_2 and s are pairwise independent. This is

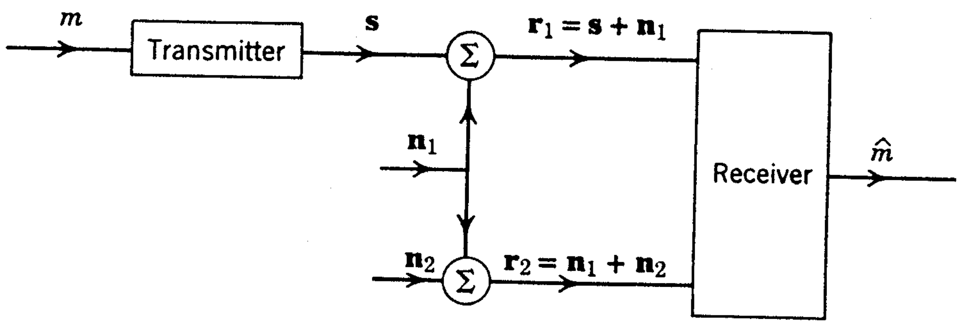


Figure 4.10 The vector $\mathbf{r}_2$ is not irrelevant.

clearly sensible, since (as an extreme case) knowledge of $\mathbf{r}_2$ provides a good estimate of $\mathbf{n}_1$, hence of s, when $p_{\mathbf{n}_2}$ is such that with high probability $\mathbf{n}_2$ is very small compared to $\mathbf{n}_1$.

The theorem of reversibility. An important corollary of the theorem of irrelevance is the *theorem of reversibility*, which states that *the minimum attainable probability of error is not affected by the introduction of a reversible operation at the output of a channel*, as in Fig. 4.11*a*. As indicated in Fig. 4.11*b*, an operation **G** is reversible if the input $\mathbf{r}_2$ can be exactly recovered from the output $\mathbf{r}_1$. In such a case it is obvious that

$$p_{\mathbf{r}_2|\mathbf{r}_1,\mathbf{s}} = p_{\mathbf{r}_2|\mathbf{r}_1},$$

so that Eq. 4.25a is satisfied, $\mathbf{r}_2$ may be discarded, and the theorem is proved. An alternative proof follows from noting that a receiver for $\mathbf{r}_1$ can be built which first recovers $\mathbf{r}_2$, as shown in Fig. 4.11*c*, and then operates on $\mathbf{r}_2$ to determine $\hat{m}$.

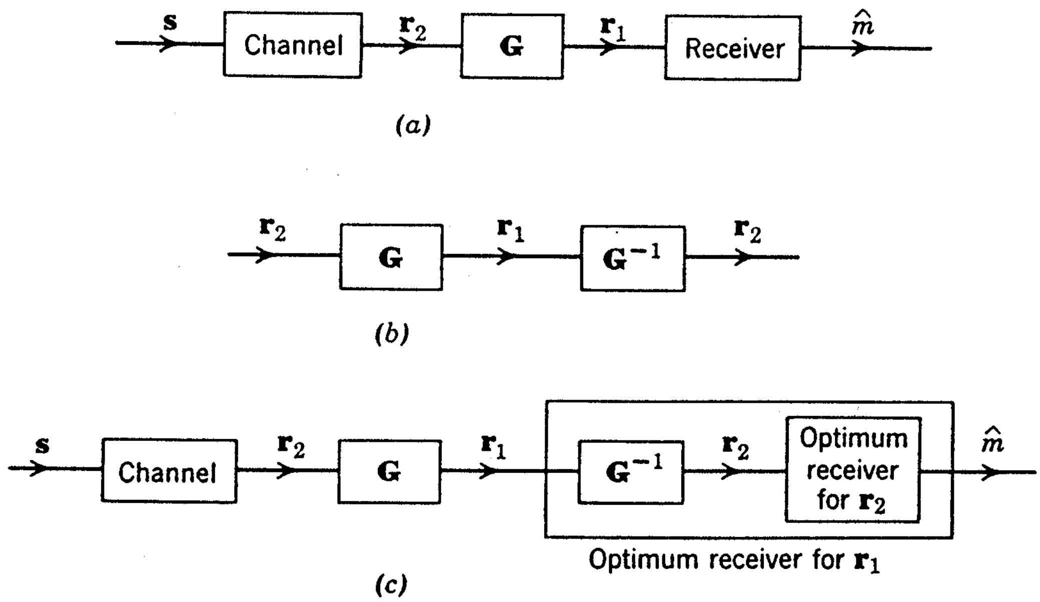


Figure 4.11 Insertion of a reversible operation, **G**, between channel and receiver. The operation inverse to **G** is denoted $\mathbf{G}^{-1}$. For example, **G** might be the addition, and $\mathbf{G}^{-1}$ the subtraction, of a fixed vector **a**.

4.3 WAVEFORM CHANNELS

The foregoing discussion of irrelevance provides the analytic tool that is required in order to replace the waveform communication problem of Fig. 4.1 by an equivalent vector communication problem. We therefore return to consideration of this figure, in which the received waveform $r(t)$ is given by

$$r(t) = s(t) + n_w(t) \tag{4.27}$$

and $n_w(t)$ is a zero-mean white Gaussian noise process with power density

$$\mathcal{S}_w(f) = \frac{\mathcal{N}_0}{2}; \qquad -\infty < f < \infty. \tag{4.28}$$

We first represent the signal process $s(t)$ in an equivalent vector form and then show that the relevant noise process may also be represented by a random vector.

Waveform Synthesis

A convenient way to synthesize the signal set $\{s_i(t)\}$ at the transmitter of Fig. 4.1 is shown in Fig. 4.12. A set of N filters is used, with the impulse response of the jth filter denoted by $\varphi_j(t)$. When the transmitter input is m_i, the first filter is excited by an impulse of value s_{i1}, the second filter by an impulse of value s_{i2}, and so on, with the Nth filter excited by an impulse of value s_{iN}. The filter outputs are summed to yield $s_i(t)$. Thus the transmitted waveform is one of the M signals

$$s_i(t) = \sum_{j=1}^{N} s_{ij}\, \varphi_j(t); \qquad i = 0, 1, \ldots, M-1. \tag{4.29}$$

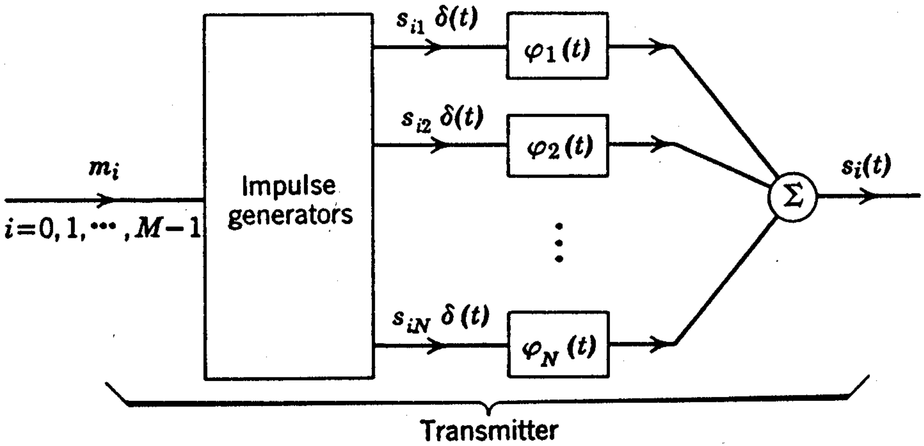


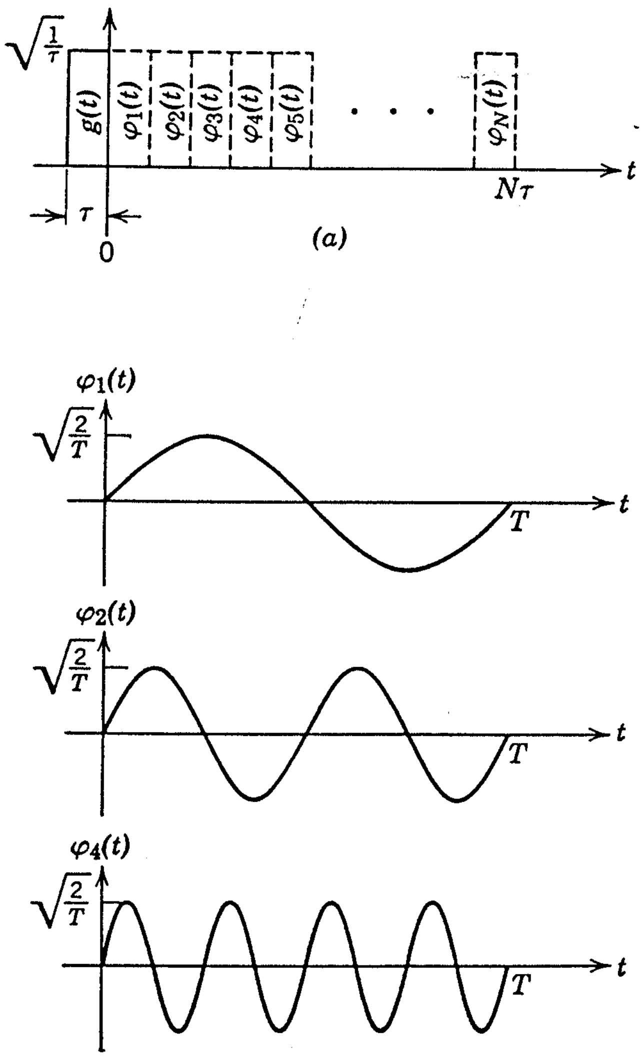
Figure 4.12 Signal synthesis. The output $s_i(t)$ depends on i through the choice of the impulse weighting coefficients $\{s_{ij}\}$.

For ease of analysis we assume that the N "building-block" waveforms $\{\varphi_j(t)\}$ are *orthonormal*, by which we mean

$$\int_{-\infty}^{\infty} \varphi_j(t)\, \varphi_l(t)\, dt = \begin{cases} 1; & j = l \\ 0; & j \neq l \end{cases} \tag{4.30}$$

for all j and l, $1 \leqslant j, l \leqslant N$.

We shall soon see that the error performance which can be achieved with signal sets generated in this way is completely independent of the



(b)

Figure 4.13 Examples of orthonormal waveforms: (*a*) orthonormal time-translated pulses; (*b*) orthonormal frequency-translated pulses.

actual waveshapes chosen for the $\{\varphi_j(t)\}$; only the coefficients $\{s_{ij}\}$ and the noise power density $\mathcal{N}_0/2$ affect the minimum attainable $P[\mathcal{E}]$. Thus the $\{\varphi_j(t)\}$ may be chosen for engineering convenience. In application, one frequently encounters the set of time-translated pulses

$$\varphi_j(t) = g(t - j\tau); \qquad j = 1, 2, \ldots, N \tag{4.31a}$$

shown in Fig. 4.13*a*, where $g(t)$ is the unit energy pulse

$$g(t) = \begin{cases} \sqrt{\dfrac{1}{\tau}}; & -\tau \leqslant t < 0, \\ 0; & \text{elsewhere.} \end{cases} \tag{4.31b}$$

A second common example is the set of frequency-translated pulses

$$\varphi_j(t) = \begin{cases} \sqrt{\dfrac{2}{T}} \sin 2\pi \dfrac{j}{T} t; & 0 \leqslant t < T, \\ 0; & \text{elsewhere,} \end{cases} \quad j = 1, 2, \ldots, N \tag{4.32}$$

shown in Fig. 4.13*b*. It may be readily verified that both sets of waveforms satisfy the orthonormality condition of Eq. 4.30. [The prefix "ortho" comes from "orthogonal," meaning that the integral of $\varphi_j(t)\,\varphi_l(t)$ is zero whenever $j \neq l$; the suffix "normal" means that the integral is unity whenever $j = l$.]

It may seem restrictive at first to consider only waveforms $\{s_i(t)\}$ that are constructed in accordance with Eq. 4.29. This is not so: *any set of M finite-energy waveforms can be synthesized in this way.* This and the fact that the *number of filters required to do so never exceeds M*, is proved in Appendix 4A. It follows that there is no loss of generality entailed in considering only transmitters that operate as shown in Fig. 4.12.

Geometric Interpretation of Signals

Once a convenient set of orthonormal functions $\{\varphi_j(t)\}$ has been adopted, each of the transmitter waveforms $\{s_i(t)\}$ is completely determined by the vector of its coefficients:

$$\mathbf{s}_i = (s_{i1}, s_{i2}, \ldots, s_{iN}); \qquad i = 0, 1, \ldots, M-1. \tag{4.33}$$

As usual, we visualize the M vectors $\{\mathbf{s}_i\}$ as defining M points in an N-dimensional geometric space, called the *signal space*, with N mutually perpendicular axes labeled $\varphi_1, \varphi_2, \ldots, \varphi_N$. If we let $\boldsymbol{\varphi}_j$ denote the unit vector along the *j*th-axis, $j = 1, 2, \ldots, N$, each N-tuple in Eq. 4.33 denotes the vector

$$\mathbf{s}_i = s_{i1}\boldsymbol{\varphi}_1 + s_{i2}\boldsymbol{\varphi}_2 + \cdots + s_{iN}\boldsymbol{\varphi}_N. \tag{4.34}$$

The idea of visualizing transmitter signals geometrically is of fundamental importance. For example, Fig. 4.3 (which we have already considered) represents a two-dimensional space with three signals: $N = 2$,

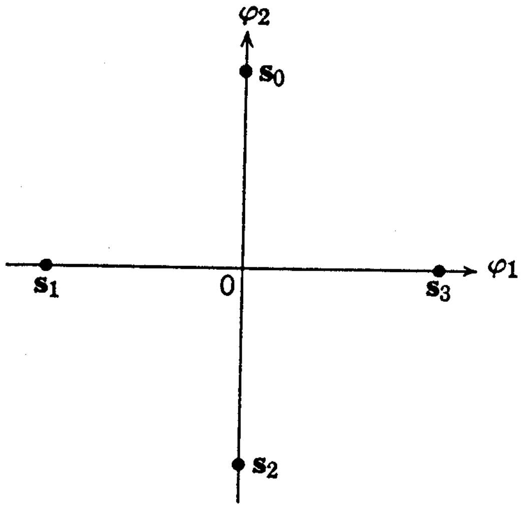


Figure 4.14 Four signals in a two-dimensional signal space. Each vector $\mathbf{s}_i$ is located a distance $\sqrt{E_s}$ from the origin.

$M = 3$. As another example, consider the set of two orthonormal functions

$$\varphi_1(t) = \begin{cases} \sqrt{\dfrac{2}{T}} \sin 2\pi f_0 t; & 0 \leqslant t < T \\ 0; & \text{elsewhere} \end{cases} \tag{4.35a}$$

$$\varphi_2(t) = \begin{cases} \sqrt{\dfrac{2}{T}} \cos 2\pi f_0 t; & 0 \leqslant t < T \\ 0; & \text{elsewhere,} \end{cases} \tag{4.35b}$$

where f_0 is an integral multiple of $1/T$. If we choose

$$\begin{aligned} \mathbf{s}_0 &= (0, \sqrt{E_s}) \\ \mathbf{s}_1 &= (-\sqrt{E_s}, 0) \\ \mathbf{s}_2 &= (0, -\sqrt{E_s}) \\ \mathbf{s}_3 &= (\sqrt{E_s}, 0), \end{aligned} \tag{4.36}$$

the vector diagram of Fig. 4.14 represents the set of four phase-modulated transmitter waveforms

$$s_i(t) = \begin{cases} \sqrt{\dfrac{2E_s}{T}} \cos 2\pi\left(f_0 t + \dfrac{i}{4}\right); & 0 \leqslant t < T \\ 0; & \text{elsewhere} \end{cases} \qquad i = 0, 1, 2, 3, \tag{4.37a}$$

where

$$E_s = \int_{-\infty}^{\infty} s_i^2(t)\, dt; \qquad i = 0, 1, 2, 3 \tag{4.37b}$$

is the energy dissipated if $s_i(t)$ is a voltage across a 1-ohm load. Similarly, if $\varphi_1(t)$ and $\varphi_2(t)$ are two nonoverlapping unit pulses, the vectors of Eq. 4.36 and the diagram of Fig. 4.14 represent the four entirely different waveforms shown in Fig. 4.15. The actual waveforms $\{s_i(t)\}$ depend on

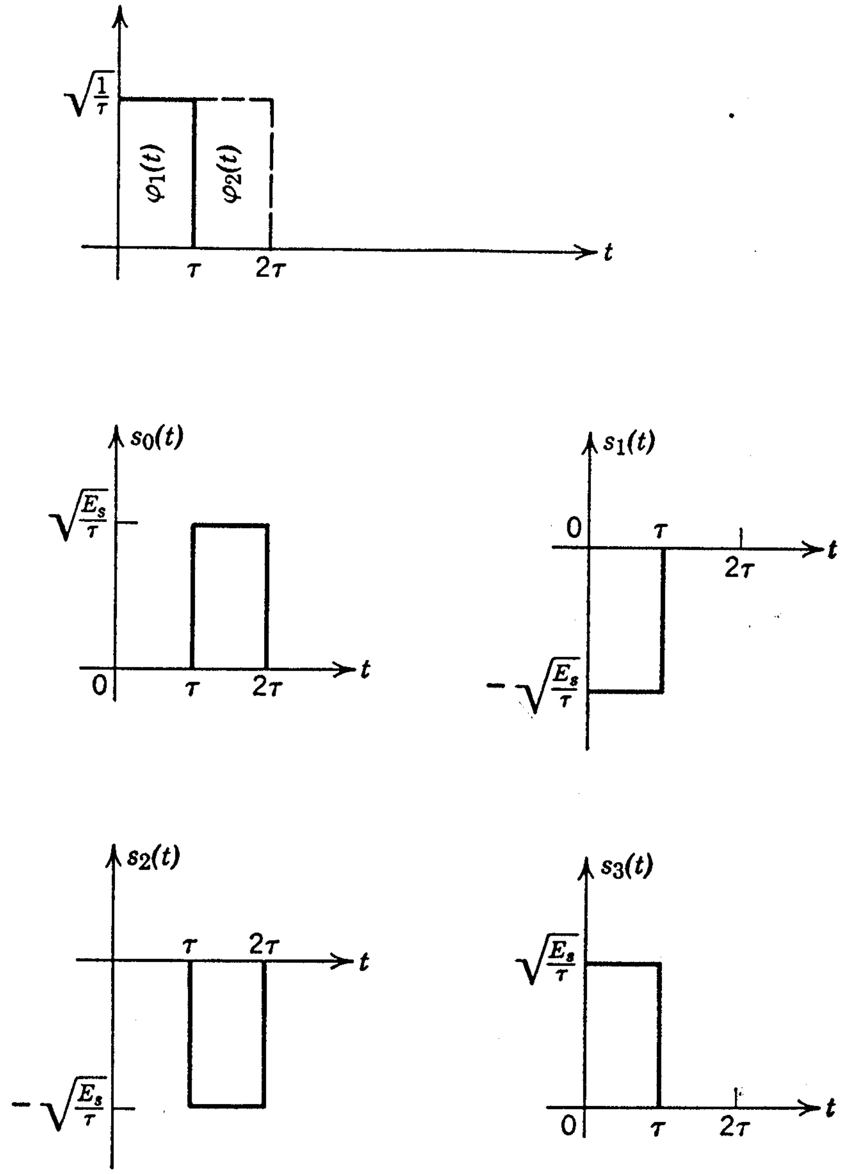


Figure 4.15 Another set of waveforms corresponding to the vector diagram of Fig. 4.14.

the choice of the $\{\varphi_j(t)\}$, but their geometric representation depends only on the $\{\mathbf{s}_i\}$.

Recovery of the Signal Vectors

So far we have considered the synthesis of the signal waveforms $\{s_i(t)\}$ from corresponding signal vectors $\{\mathbf{s}_i\}$. It is also straightforward to recover the vectors from the waveforms. We observe that by virtue of the orthonormality of the $\{\varphi_j(t)\}$

$$\int_{-\infty}^{\infty} s_i(t)\, \varphi_l(t)\, dt = \int_{-\infty}^{\infty} \left[\sum_{j=1}^{N} s_{ij}\, \varphi_j(t)\right] \varphi_l(t)\, dt$$
$$= \sum_{j=1}^{N} s_{ij} \int_{-\infty}^{\infty} \varphi_j(t)\, \varphi_l(t)\, dt$$
$$= \sum_{j=1}^{N} s_{ij}\, \delta_{jl} = s_{il}, \tag{4.38}$$

in which we use the Kronecker delta

$$\delta_{jl} \triangleq \begin{cases} 1; & l = j \\ 0; & l \neq j. \end{cases} \tag{4.39}$$

Carrying out the multiplication and integration for each $\varphi_l(t)$, $1 \leqslant l \leqslant N$, we obtain

$$\mathbf{s}_i = (s_{i1}, s_{i2}, \ldots, s_{iN}).$$

The procedure can be implemented as shown in the block diagram of Fig. 4.16. If $s(t)$ is applied at the input, the output is a vector

$$\mathbf{s} \triangleq (s_1, s_2, \ldots, s_N) \tag{4.40a}$$

with components

$$s_j \triangleq \int_{-\infty}^{\infty} s(t)\, \varphi_j(t)\, dt; \qquad j = 1, 2, \ldots, N. \tag{4.40b}$$

If $s(t) = s_i(t)$, then $\mathbf{s} = \mathbf{s}_i$.

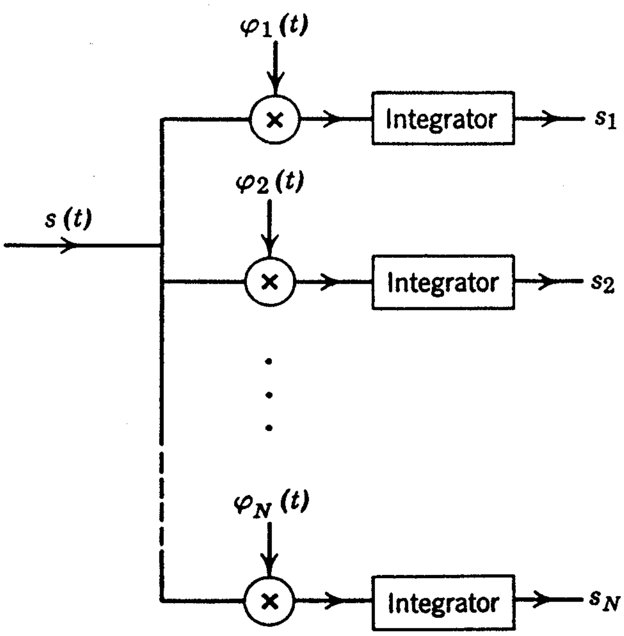


Figure 4.16 Extraction of $\mathbf{s} = (s_1, s_2, \ldots, s_N)$ from $s(t)$. Each of the integrations extends over the duration of the $\varphi_j(t)$ with which it is associated.

Irrelevant Data

Now suppose that the input to the bank of N multipliers and integrators in Fig. 4.16 is not $s(t)$, but rather the received random process $r(t)$ of Fig. 4.1. In this case the integrator outputs, say

$$r_j \triangleq \int_{-\infty}^{\infty} r(t)\, \varphi_j(t)\, dt; \qquad j = 1, 2, \ldots, N, \tag{4.41a}$$

are random variables† which together constitute a random vector

$$\mathbf{r}_1 \triangleq (r_1, r_2, \ldots, r_N). \tag{4.41b}$$

Since $r(t) = s(t) + n_w(t)$, we have

$$\mathbf{r}_1 = \mathbf{s} + \mathbf{n}, \tag{4.42}$$

where

$$\mathbf{n} = (n_1, n_2, \ldots, n_N) \tag{4.43a}$$

is the random vector with components

$$n_j \triangleq \int_{-\infty}^{\infty} n_w(t)\, \varphi_j(t)\, dt; \qquad j = 1, 2, \ldots, N. \tag{4.43b}$$

We assume that $n_w(t)$, hence $\mathbf{n}$, is statistically independent of $\mathbf{s}$.

Were it not for the noise vector $\mathbf{n}$, we have seen that $\mathbf{r}_1$ would coincide with whichever one of the $\{\mathbf{s}_i\}$ was actually transmitted. When the presence of $\mathbf{n}$ cannot be neglected, this, of course, is no longer true. What is true, however, is that the vector $\mathbf{r}_1$ in and by itself does contain *all data from $r(t)$ that is relevant to the optimum determination of the transmitted message.* The objective of this section is to prove this important fact.

The first step in the proof is to note that the waveform equation corresponding to the vector equality of Eq. 4.42 is

$$r_1(t) \triangleq \sum_{j=1}^{N} r_j\, \varphi_j(t) = s(t) + n(t), \tag{4.44a}$$

in which

$$s(t) = \sum_{j=1}^{N} s_j\, \varphi_j(t) \tag{4.44b}$$

$$n(t) \triangleq \sum_{j=1}^{N} n_j\, \varphi_j(t). \tag{4.44c}$$

† The value $r_j(\omega)$ assigned to any point ω of the sample space on which $r(t)$ is defined is $\int_{-\infty}^{\infty} r(\omega, t)\, \varphi_j(t)\, dt$.

The next step in the proof is to note that in terms of these random processes we may write

$$r(t) = r_1(t) + r_2(t), \tag{4.45a}$$

where

$$\begin{aligned} r_2(t) &\triangleq r(t) - r_1(t) \\ &= [s(t) + n_w(t)] - [s(t) + n(t)] \\ &= n_w(t) - n(t) \end{aligned} \tag{4.45b}$$

is a random process that is independent of the signal transmitted. The fact that $r_2(t)$ is not in general identically zero implies that the noise process $n_w(t)$ cannot be represented with complete fidelity by the finite orthonormal set $\{\varphi_j(t)\}$.

We have succeeded in Eq. 4.45a in decomposing the received waveform $r(t)$ into two waveforms, $r_1(t)$ and $r_2(t)$, the first entirely specified by the vector $\mathbf{r}_1$ and the second independent of the transmitted signal. We now show that the optimum receiver may disregard $r_2(t)$ and therefore base its decision solely upon the vector $\mathbf{r}_1 = \mathbf{s} + \mathbf{n}$.

Observe that any finite set of time samples taken from $r_2(t)$, say

$$\mathbf{r}_2 = (r_2(t_1), r_2(t_2), \ldots, r_2(t_q)), \tag{4.46}$$

depends only on $n_w(t)$. Since this is true also of $\mathbf{n}$, the vectors $\mathbf{r}_2$ and $\mathbf{n}$ are jointly independent of $\mathbf{s}$. As a preliminary to invoking the theorem of irrelevance (Eq. 4.25b and Fig. 4.8), we observe in consequence that

$$\begin{aligned} p_{\mathbf{r}_2|\mathbf{r}_1,\mathbf{s}} = p_{\mathbf{r}_2|\mathbf{n},\mathbf{s}} &= \frac{p_{\mathbf{r}_2,\mathbf{n},\mathbf{s}}}{p_{\mathbf{n},\mathbf{s}}} \\ &= \frac{p_{\mathbf{r}_2,\mathbf{n}}\, p_{\mathbf{s}}}{p_{\mathbf{n}}\, p_{\mathbf{s}}} = p_{\mathbf{r}_2|\mathbf{n}}. \end{aligned}$$

Thus $\mathbf{r}_2$ may be discarded by the optimum receiver provided that it is also independent of $\mathbf{n}$. Since a random process is completely described by the statistical behavior of finite sets of time samples, it follows that the entire process $r_2(t)$ may be discarded whenever the statistical independence of $\mathbf{r}_2$ and $\mathbf{n}$ holds true for every possible finite set of sampling instants $\{t_l\}$, $l = 1, 2, \ldots, q$. In other words, the random process $r_2(t)$ may be ignored if it is statistically independent of the process $n(t)$.

The required proof of statistical independence rests on the fact that both $n(t)$ and $r_2(t)$ result from linear operations—integration, addition, and subtraction—on the Gaussian process $n_w(t)$. Thus $n(t)$ and $r_2(t)$ are jointly Gaussian processes, so that by analogy with Eq. 3.130 any two random vectors obtained from $n(t)$ and $r_2(t)$, respectively, are statistically independent if the covariance

$$\mathrm{E}[n(s)\, r_2(t)] - \mathrm{E}[n(s)]\, \mathrm{E}[r_2(t)]$$

vanishes for all observation instants t and s. In particular, since $n_w(t)$, hence $n(s)$ and $r_2(t)$ as well, are zero mean, it suffices to show that

$$\mathrm{E}[n(s)\, r_2(t)] = 0; \qquad \text{for all } t \text{ and } s. \tag{4.47a}$$

From Eq. 4.44c we have

$$\mathrm{E}[n(s)\, r_2(t)] = \mathrm{E}\left[r_2(t) \sum_{j=1}^{N} n_j\, \varphi_j(s) \right]$$
$$= \sum_{j=1}^{N} \varphi_j(s)\, \mathrm{E}[n_j\, r_2(t)], \tag{4.47b}$$

so that we need prove only that

$$\overline{n_j r_2(t)} = 0; \qquad \text{for all } j \text{ and } t. \tag{4.47c}$$

In order to verify Eq. 4.47c, we note from the definitions of Eqs. 4.43 and 4.45 that

$$\overline{n_j r_2(t)} = \overline{n_j[n_w(t) - n(t)]} = \overline{n_j n_w(t)} - \overline{n_j n(t)}$$
$$= \int_{-\infty}^{\infty} \overline{n_w(t)\, n_w(\alpha)}\, \varphi_j(\alpha)\, d\alpha - \sum_{l=1}^{N} \overline{n_j n_l}\, \varphi_l(t). \tag{4.48a}$$

The integral can be evaluated with the help of Eq. 3.136b:

$$\int_{-\infty}^{\infty} \overline{n_w(t)\, n_w(\alpha)}\, \varphi_j(\alpha)\, d\alpha = \int_{-\infty}^{\infty} \mathcal{R}_w(t - \alpha)\, \varphi_j(\alpha)\, d\alpha$$
$$= \frac{\mathcal{N}_0}{2} \int_{-\infty}^{\infty} \delta(t - \alpha)\, \varphi_j(\alpha)\, d\alpha = \frac{\mathcal{N}_0}{2}\, \varphi_j(t). \tag{4.48b}$$

Evaluation of the sum follows from the fact that

$$\overline{n_l n_j} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \overline{n_w(\alpha)\, n_w(\beta)}\, \varphi_l(\alpha)\, \varphi_j(\beta)\, d\alpha\, d\beta$$
$$= \frac{\mathcal{N}_0}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(\alpha - \beta)\, \varphi_l(\alpha)\, \varphi_j(\beta)\, d\alpha\, d\beta$$
$$= \frac{\mathcal{N}_0}{2} \int_{-\infty}^{\infty} \varphi_l(\beta)\, \varphi_j(\beta)\, d\beta = \frac{\mathcal{N}_0}{2}\, \delta_{lj}. \tag{4.48c}$$

Thus

$$\sum_{l=1}^{N} \overline{n_l n_j}\, \varphi_l(t) = \frac{\mathcal{N}_0}{2}\, \varphi_j(t). \tag{4.48d}$$

Substituting Eqs. 4.48b and 4.48d into Eq. 4.48a, we have

$$\overline{n_j r_2(t)} = \frac{\mathcal{N}_0}{2}\,\varphi_j(t) - \frac{\mathcal{N}_0}{2}\,\varphi_j(t) = 0; \qquad \text{for all } j \text{ and } t, \tag{4.48e}$$

which was to be shown.

This completes the proof that the process $r_2(t)$ is statistically independent of $n(t)$. We conclude that the vector $\mathbf{r}_1$ defined by Eqs. 4.41 does in fact contain all data relevant to the optimum determination of $\hat{m}$ for the communication system of Fig. 4.1.

Joint Density Function of the Relevant Noise

In addition to the result that $r_2(t)$ is irrelevant, the foregoing analysis yields valuable information about $\mathbf{r}_1$. First, Eq. 4.42 establishes that the relevant effect of the additive white Gaussian noise $n_w(t)$ is to disturb the transmitted vector $\mathbf{s}$ by the addition of a random noise vector $\mathbf{n}$:

$$\mathbf{r}_1 = \mathbf{s} + \mathbf{n}. \tag{4.49a}$$

Second, the discussion leading to Eq. 4.47 implies that $\mathbf{n}$ is a set of N jointly Gaussian random variables, $\{n_j\}$, each of which has zero mean:

$$\overline{n_j} = 0; \qquad j = 1, 2, \ldots, N. \tag{4.49b}$$

Third, Eq. 4.48c establishes that the $\{n_j\}$ have zero covariance and equal variance:

$$\overline{n_l n_j} = \begin{cases} \dfrac{\mathcal{N}_0}{2}; & l = j \\ 0; & l \neq j. \end{cases} \tag{4.49c}$$

Thus the joint density function $p_{\mathbf{n}}$, in the notation of Eq. 4.17, is

$$p_{\mathbf{n}}(\boldsymbol{\alpha}) = \frac{1}{(\pi\mathcal{N}_0)^{N/2}}\, e^{-|\boldsymbol{\alpha}|^2/\mathcal{N}_0}, \tag{4.49d}$$

which implies that the $\{n_j\}$ are statistically independent. In particular, we note that $p_{\mathbf{n}}$ is *spherically symmetric*, that is, that $p_{\mathbf{n}}(\boldsymbol{\alpha})$ depends on the magnitude but not on the direction of the argument vector $\boldsymbol{\alpha}$.

Invariance of the Vector Channel to Choice of Orthonormal Base

Since a receiver need never consider the process $r_2(t)$ of Eq. 4.45, we shall henceforth disregard it and designate the relevant received vector simply by $\mathbf{r}$ rather than by $\mathbf{r}_1$.

Once provision is made for calculating the vector $\mathbf{r}$, the remaining receiver design problem is precisely the same as the vector receiver problem which we have already considered in connection with Fig. 4.2 and Eqs. 4.13 and 4.17, with the variance σ^2 set equal to $\mathcal{N}_0/2$. The relationship between the vector and waveform channels is illustrated in Fig. 4.17, in which we break both the transmitter and receiver into two parts. The "vector transmitter" accepts the input message m and generates the vector $\mathbf{s}_i$ whenever $m = m_i$; the "modulator" then constructs $s_i(t)$ from $\mathbf{s}_i$ and the waveforms $\{\varphi_j(t)\}$, which we call the *orthonormal base.* At the receiver the "detector" operates on the received waveform $r(t)$ and

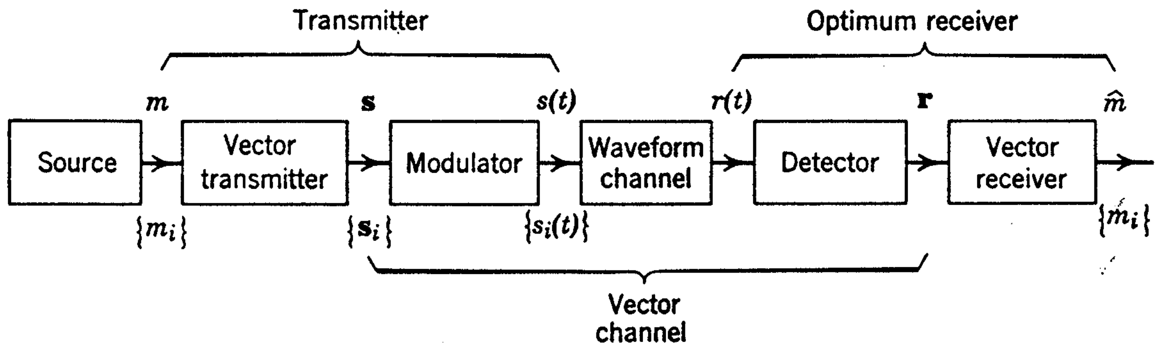


Figure 4.17 Reduction of waveform channel to vector channel. The modulator converts $\mathbf{s}$ to $s(t)$ by the mechanism of Fig. 4.12. The detector extracts the relevant received vector $\mathbf{r}$ from $r(t)$ by the mechanism of Fig. 4.16.

produces the relevant vector $\mathbf{r}$; the "vector receiver" then determines which message is most probable from observation of $\mathbf{r}$ and knowledge of the $\{\mathbf{s}_i\}$ and $\{\mathrm{P}[m_i]\}$.

We have already noted that a particular geometric configuration of the signal vectors $\{\mathbf{s}_i\}$ may be converted to many different sets of waveforms $\{s_i(t)\}$ by appropriate choice of the orthonormal base. In addition, we now note that the derivation of $p_{\mathbf{n}}$ relies only on the fact that the $\{\varphi_j(t)\}$ are orthonormal and depends in no way on the specific waveshapes of these functions. Thus, as claimed earlier, whenever their vector representations $\{\mathbf{s}_i\}$ are the same, systems with different sets of transmitter signals $\{s_i(t)\}$, $i = 0, 1, \ldots, M - 1$, reduce to the same vector channel and yield the same minimum probability of error, $\mathrm{P}[\mathcal{E}]$. The expression for $\mathrm{P}[\mathcal{E}]$ is given in Eq. 4.20, with σ^2 specialized to $\mathcal{N}_0/2$ in accordance with Eq. 4.49c.

4.4 RECEIVER IMPLEMENTATION

We have seen so far that the optimum receiver in Fig. 4.17 performs two functions: first, the receiver calculates the relevant data vector

$$\mathbf{r} = (r_1, r_2, \ldots, r_N), \tag{4.50a}$$

where

$$r_j = \int_{-\infty}^{\infty} r(t)\, \varphi_j(t)\, dt; \qquad j = 1, 2, \ldots, N. \tag{4.50b}$$

Then, in accordance with Eq. 4.19 (with $\sigma^2 = \mathcal{N}_0/2$), the receiver sets $\hat{m} = m_k$ if the decision function

$$|\mathbf{r} - \mathbf{s}_i|^2 - \mathcal{N}_0 \ln \mathrm{P}[m_i] \tag{4.51}$$

is *minimum* for $i = k$. In practice, squarers are avoided by recognizing that

$$|\mathbf{r} - \mathbf{s}_i|^2 = \sum_{j=1}^{N} (r_j - s_{ij})^2$$

$$= \sum_{j=1}^{N} (r_j^2 - 2r_j s_{ij} + s_{ij}^2) = |\mathbf{r}|^2 - 2\mathbf{r} \cdot \mathbf{s}_i + |\mathbf{s}_i|^2, \tag{4.52a}$$

in which

$$\mathbf{r} \cdot \mathbf{s}_i \triangleq \sum_{j=1}^{N} r_j s_{ij} \tag{4.52b}$$

is the dot product of the vectors $\mathbf{r}$ and $\mathbf{s}_i$. Since $|\mathbf{r}|^2$ is independent of i, a decision rule equivalent to Eq. 4.52 is to *maximize* the expression

$$(\mathbf{r} \cdot \mathbf{s}_i) + c_i, \tag{4.53a}$$

where

$$c_i \triangleq \tfrac{1}{2}(\mathcal{N}_0 \ln \mathrm{P}[m_i] - |\mathbf{s}_i|^2), \qquad i = 0, 1, \ldots, M - 1. \tag{4.53b}$$

Correlation Receiver

When the relevant received vector $\mathbf{r}$ is obtained from the received waveform by the bank of N multipliers and integrators shown in Fig. 4.16, the receiver is called a correlation receiver. When M is not large, the numbers

$$\mathbf{r} \cdot \mathbf{s}_i = \sum_{j=1}^{N} r_j s_{ij}; \qquad i = 0, 1, \ldots, M - 1$$

can be obtained from $\mathbf{r}$ and knowledge of the $\{\mathbf{s}_i\}$ by attaching a set of M resistor weighting networks (with weights proportional to the $\{s_{ij}\}$) to the integrator outputs or by other analog computer techniques. When M is very large, digital computation of the $\{\mathbf{r} \cdot \mathbf{s}_i\}$ becomes preferable. A block diagram of an optimum correlation receiver is shown in Fig. 4.18.

Matched Filter Receiver

If each member of the orthonormal base $\{\varphi_j(t)\}$ is identically zero outside some finite time interval, say $0 \leqslant t \leqslant T$, the use of the multipliers

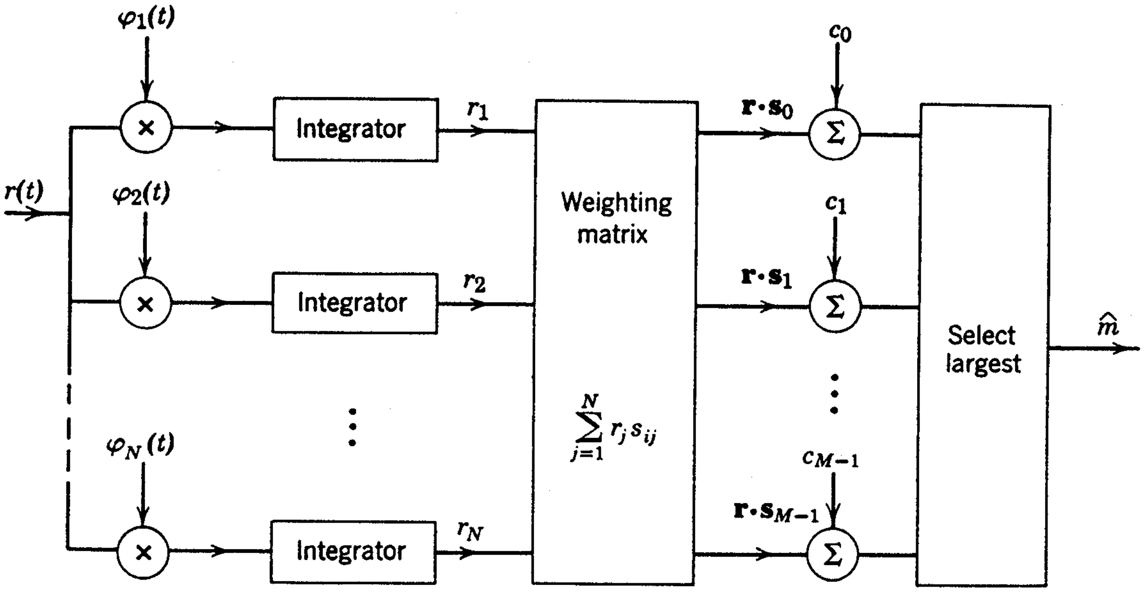


Figure 4.18 Diagram of the correlation receiver. The bias terms $\{c_i\}$ are given by Eq. 4.53b.

shown in Fig. 4.18 can be avoided. This is desirable, since accurate analog multipliers are hard to build. Consider, for instance, the output $u_j(t)$ of a linear filter with impulse response $h_j(t)$. When $r(t)$ is the filter input, we have

$$u_j(t) = \int_{-\infty}^{\infty} r(\alpha)\, h_j(t - \alpha)\, d\alpha. \tag{4.54a}$$

If we now set

$$h_j(t) = \varphi_j(T - t), \tag{4.54b}$$

the output is

$$u_j(t) = \int_{-\infty}^{\infty} r(\alpha)\, \varphi_j(T - t + \alpha)\, d\alpha. \tag{4.54c}$$

Finally, the output sampled at time $t = T$ is

$$u_j(T) = \int_{-\infty}^{\infty} r(\alpha)\, \varphi_j(\alpha)\, d\alpha \triangleq r_j, \tag{4.54d}$$

where the second equality follows from Eq. 4.50b. Thus the optimum decision rule of Eq. 4.53 can also be implemented by the receiver shown in Fig. 4.19.

A filter whose impulse response is a delayed, time-reversed version of a signal $\varphi_j(t)$ is called *matched* to $\varphi_j(t)$ and the optimum receiver realization of Fig. 4.19 is a *matched filter* receiver. The requirement that $\varphi_j(t)$ vanish for $t > T$ is necessary in order that the matched filter may be physically realizable, that is, in order that $h_j(t) \equiv 0$ for $t < 0$.

For both the correlation and matched filter optimum receiver realizations we note that the "bias" terms

$$c_i = \tfrac{1}{2}(\mathcal{N}_0 \ln P[m_i] - |\mathbf{s}_i|^2)$$

represent a priori data that are available to the receiver independent of the received signal $r(t)$. In the particular case in which the bias term is the same for every i (in particular, when $|\mathbf{s}_i|^2$ is constant and $P[m_i] = 1/M$ for all i), these bias terms do not affect the choice of index i that maximizes the decision function of Eq. 4.53 and may therefore be deleted from the receiver diagrams in Figs. 4.18 and 4.19 without loss of optimality.

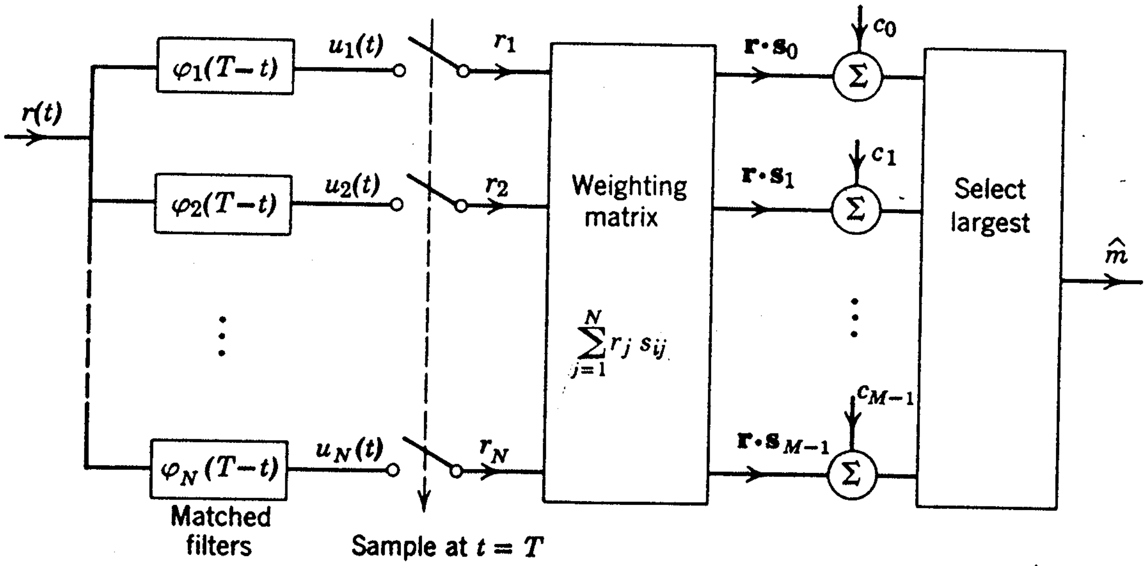


Figure 4.19 Diagram of the matched-filter receiver.

A simple example of a matched filter occurs when the signal to be matched is

$$\varphi_j(t) = \begin{cases} -\sqrt{\dfrac{2}{T}} \cos 2\pi f_j t; & 0 \leqslant t \leqslant T \\ 0; & \text{elsewhere,} \end{cases} \tag{4.55a}$$

where f_j is an integral multiple of $1/2T$. Then

$$h_j(t) = \varphi_j(T - t)$$
$$= \begin{cases} +\sqrt{\dfrac{2}{T}} \cos 2\pi f_j t; & 0 \leqslant t \leqslant T \\ 0; & \text{elsewhere,} \end{cases} \tag{4.55b}$$

as shown in Fig. 4.20*a*.

The voltage response of the infinite-Q parallel tuned circuit shown in Fig. 4.20*b* to a unit impulse of current is

$$h(t) = \frac{1}{C} \cos \frac{t}{\sqrt{LC}}; \qquad 0 \leqslant t < \infty,$$

where we have assumed that the initial energy storage in L and C at time $t = 0$ is zero. It is clear that when $1/\sqrt{LC} = 2\pi f_j$ and $1/C = \sqrt{2/T}$ the impulse response $h(t)$ coincides with $h_j(t)$ over the interval $0 \leqslant t \leqslant T$, although it does not do so for $t > T$. Thus the matched filtering operation for $\varphi_j(t)$ can be instrumented as shown in Fig. 4.20*c*. The parallel switch closing briefly at time $t = 0$ dumps any residual energy in the filter, ensuring that signal energy received earlier than $t = 0$ does not contribute to the output at time $t = T$. The series switch closing briefly at $t = T$ samples the filter output at the proper time. The entire cycle can be repeated during the interval $T \leqslant t \leqslant 2T$, although care must be taken to

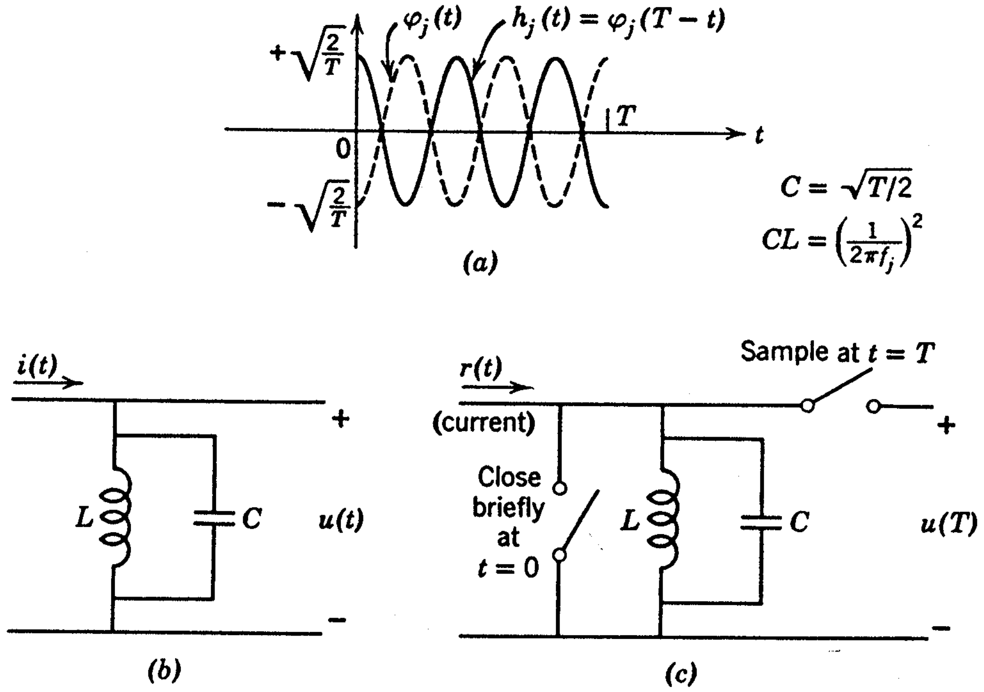


Figure 4.20 Integrate-and-dump filter. In application the resonant circuit may be lossy, so long as its time constant is much greater than T.

be sure that the desired output is always sampled just before the filter is dumped. A matched filter of this sort is called an *integrate-and-dump*[87] circuit. Such a filter is not time invariant, but it does give the desired impulse response as long as the timing of the switches is properly synchronized with respect to $\varphi_j(t)$.

Parseval relationships. The vector decision function of Eq. 4.53 can be interpreted directly in terms of time functions by means of the following Parseval relationship. Consider an orthonormal set $\{\varphi_j(t)\}$, $j = 1, 2, \ldots, N$, and any two waveforms defined by

$$f(t) \triangleq \sum_{j=1}^{N} f_j \, \varphi_j(t) \tag{4.56a}$$

$$g(t) \triangleq \sum_{j=1}^{N} g_j \, \varphi_j(t), \tag{4.56b}$$

with corresponding vector representations

$$\mathbf{f} = (f_1, f_2, \ldots, f_N) \tag{4.57a}$$

$$\mathbf{g} = (g_1, g_2, \ldots, g_N). \tag{4.57b}$$

Then

$$\int_{-\infty}^{\infty} f(t)\, g(t)\, dt = \int_{-\infty}^{\infty} \sum_{j=1}^{N} \sum_{l=1}^{N} f_j g_l\, \varphi_j(t)\, \varphi_l(t)\, dt$$

$$= \sum_{j=1}^{N} \sum_{l=1}^{N} f_j g_l \int_{-\infty}^{\infty} \varphi_j(t)\, \varphi_l(t)\, dt$$

$$= \sum_{j=1}^{N} \sum_{l=1}^{N} f_j g_l\, \delta_{jl} = \sum_{j=1}^{N} f_j g_j = \mathbf{f} \cdot \mathbf{g}.$$

Thus the well-known Parseval equation[62] from Fourier theory,

$$\int_{-\infty}^{\infty} f(t)\, g(t)\, dt = \int_{-\infty}^{\infty} F(f)\, G^*(f)\, df,$$

where $F(f)$ and $G(f)$ are the Fourier transforms of $f(t)$ and $g(t)$, can be extended to read

$$\int_{-\infty}^{\infty} f(t)\, g(t)\, dt = \int_{-\infty}^{\infty} F(f)\, G^*(f)\, df = \mathbf{f} \cdot \mathbf{g}. \tag{4.58a}$$

In particular, when $g(t) = f(t)$, we have

$$\int_{-\infty}^{\infty} f^2(t)\, dt = \int_{-\infty}^{\infty} |F(f)|^2\, df = |\mathbf{f}|^2. \tag{4.58b}$$

Equation 4.58a states that the "correlation" of $f(t)$ and $g(t)$, defined as the integral of their product, equals the dot product of the corresponding vectors. Equation 4.58b states that the "energy" of $f(t)$, normalized to a one-ohm load, equals the square of the length of the corresponding vector $\mathbf{f}$.

Equation 4.58b provides an immediate interpretation of the bias term c_i in the additive white Gaussian noise decision rule of Eq. 4.53. We have

$$c_i = \tfrac{1}{2}(\mathcal{N}_0 \ln \mathrm{P}[m_i] - E_i), \tag{4.59a}$$

where

$$E_i \triangleq \int_{-\infty}^{\infty} s_i^2(t)\, dt = \text{energy of the } i\text{th signal}. \tag{4.59b}$$

Moreover, from Eqs. 4.29 and 4.50 we also have

$$\int_{-\infty}^{\infty} r(t)\, s_i(t)\, dt = \int_{-\infty}^{\infty} r(t) \left[\sum_{j=1}^{N} s_{ij}\, \varphi_j(t)\right] dt$$

$$= \sum_{j=1}^{N} s_{ij} \int_{-\infty}^{\infty} r(t)\, \varphi_j(t)\, dt = \sum_{j=1}^{N} s_{ij} r_j = \mathbf{r} \cdot \mathbf{s}_i.$$

Thus, in terms of the complete received waveform $r(t)$, the optimum decision function of Eq. 4.53a is

$$\int_{-\infty}^{\infty} r(t)\, s_i(t)\, dt + c_i. \qquad (4.60)$$

In view of Eq. 4.60, the matched filter (or the correlation) receiver can be instrumented directly in terms of the $\{s_i(t)\}$, $i = 0, 1, \ldots, M-1$, as indicated in Fig. 4.21. At first glance this might appear to eliminate the need for the weighting and summing operations in Figs. 4.18 and 4.19.

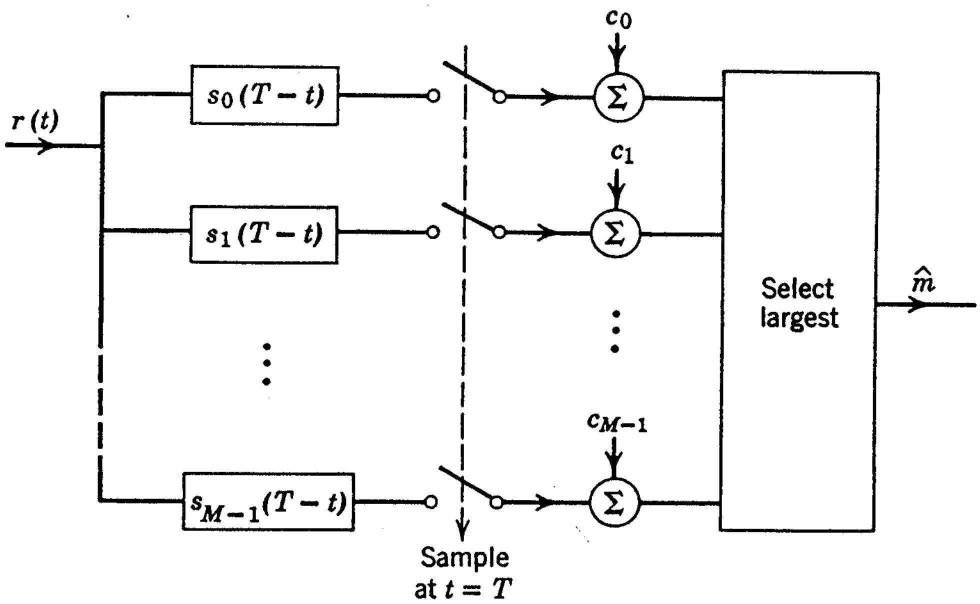


Figure 4.21 An optimum receiver with M filters matched directly to the signals $\{s_i(t)\}$, which are assumed to have duration T.

Actually, of course, these operations are still being performed but now occur within the M matched filters (or correlators). We have already remarked (and prove in Appendix 4A) that the number, N, of orthonormal functions required to express any set of M signals $\{s_i(t)\}$ in the form of Eq. 4.29 is always less than or equal to M. When $M \gg N$, a situation often encountered in practice, it is usually much less expensive to use N filters (or correlators) matched to the $\{\varphi_j(t)\}$, plus an analog or digital computer, than it is to use M filters (or correlators) matched directly to the $\{s_i(t)\}$.

Signal-to-noise ratio. We may gain insight into the optimality of the matched filtering operation by a signal-to-noise ratio analysis. Consider the situation illustrated in Fig. 4.22, in which $h(t)$ is an arbitrary linear filter, T is an arbitrary observation instant, and $\varphi(t)$ is any known signal. [In particular, we may choose $\varphi(t)$ to be one of the orthonormal base functions.] The sampled output r may be written

$$r = \bar{r} + n, \qquad (4.61)$$

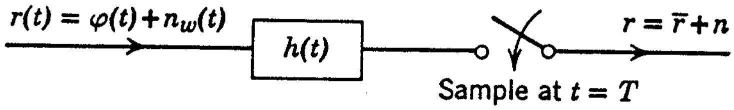


Figure 4.22 An arbitrary filter, the output of which is sampled at $t = T$.

where $\bar{r}$, the mean of r, depends on $\varphi(t)$ and the noise term n depends on $n_w(t)$. We now show that the maximum attainable signal-to-noise power ratio, defined as

$$\mathcal{S}/\mathcal{N} \triangleq \bar{r}^2/\overline{n^2}, \tag{4.62}$$

occurs when the filter is matched to $\varphi(t)$; that is, when

$$h(t) = \varphi(T - t). \tag{4.63}$$

In application, T is taken large enough that $h(t)$ is realizable.

We prove that this $h(t)$ maximizes $\mathcal{S}/\mathcal{N}$ by invoking the Schwarz inequality, one form of which states that for any pair of finite-energy waveforms $a(t)$ and $b(t)$,

$$\left[\int_{-\infty}^{\infty} a(t)\, b(t)\, dt\right]^2 \leqslant \left[\int_{-\infty}^{\infty} a^2(t)\, dt\right]\left[\int_{-\infty}^{\infty} b^2(t)\, dt\right]. \tag{4.64}$$

The equality obtains if and only if $b(t) = ca(t)$, where c is any constant.

The validity of Eq. 4.64 is evident if we make an orthonormal expansion of the waveforms $a(t)$ and $b(t)$ by means of the Gram-Schmidt procedure discussed in Appendix 4A. We then have

$$a(t) = a_1\, \psi_1(t) + a_2\, \psi_2(t)$$
$$b(t) = b_1\, \psi_1(t) + b_2\, \psi_2(t),$$

where

$$\int_{-\infty}^{\infty} \psi_i(t)\, \psi_j(t)\, dt = \delta_{ij}; \qquad i, j = 1, 2.$$

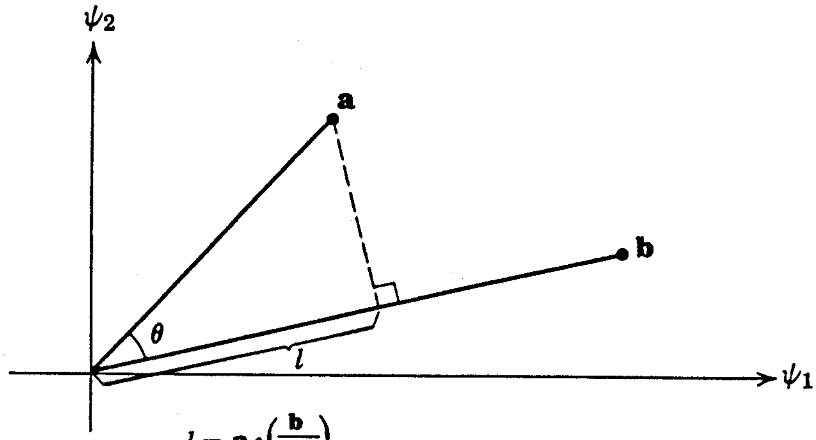
Figure 4.23 illustrates that the angle between the two vectors

$$\mathbf{a} \triangleq (a_1, a_2)$$
$$\mathbf{b} \triangleq (b_1, b_2) \tag{4.65a}$$

is given by

$$\cos\theta = \frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{a}|\,|\mathbf{b}|} = \frac{\int_{-\infty}^{\infty} a(t)\, b(t)\, dt}{\left[\int_{-\infty}^{\infty} a^2(t)\, dt \int_{-\infty}^{\infty} b^2(t)\, dt\right]^{1/2}}. \tag{4.65b}$$

The second equality above rests on the Parseval relations of Eq. 4.58.



$$l = \mathbf{a}\cdot\left(\frac{\mathbf{b}}{|\mathbf{b}|}\right)$$

$$\cos\theta = \frac{l}{|\mathbf{a}|} = \frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{a}|\ |\mathbf{b}|}$$

Figure 4.23 The angle between two vectors.

The Schwarz inequality (Eq. 4.64) results from recognizing that $|\cos\theta| \leqslant 1$. Furthermore, $|\cos\theta| = 1$ if and only if $\mathbf{b} = c\mathbf{a}$, that is, if and only if $a(t) = c\,b(t)$.

We now apply the Schwarz inequality to the maximization of $\mathcal{S}/\mathcal{N}$. For the random variable r of Fig. 4.22,

$$\bar{r} = \int_{-\infty}^{\infty} \varphi(T-\alpha)\,h(\alpha)\,d\alpha$$

and

$$\overline{n^2} = \mathrm{E}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} n_w(T-\alpha)\,n_w(T-\beta)\,h(\alpha)\,h(\beta)\,d\alpha\,d\beta\right]$$

$$= \frac{\mathcal{N}_0}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta(\beta-\alpha)\,h(\alpha)\,h(\beta)\,d\alpha\,d\beta$$

$$= \frac{\mathcal{N}_0}{2}\int_{-\infty}^{\infty} h^2(\alpha)\,d\alpha.$$

From Schwarz's inequality, for any $h(t)$ we have

$$\frac{\mathcal{S}}{\mathcal{N}} \triangleq \frac{\bar{r}^2}{\overline{n^2}} = \frac{\left[\int_{-\infty}^{\infty} \varphi(T-\alpha)\,h(\alpha)\,d\alpha\right]^2}{(\mathcal{N}_0/2)\int_{-\infty}^{\infty} h^2(\alpha)\,d\alpha}$$

$$\leqslant \frac{\int_{-\infty}^{\infty} \varphi^2(T-\alpha)\,d\alpha \int_{-\infty}^{\infty} h^2(\alpha)\,d\alpha}{(\mathcal{N}_0/2)\int_{-\infty}^{\infty} h^2(\alpha)\,d\alpha} = \frac{\int_{-\infty}^{\infty} \varphi^2(\alpha)\,d\alpha}{\mathcal{N}_0/2}. \qquad (4.66)$$

Since the substitution of $c\varphi(T-\alpha)$ for $h(\alpha)$ satisfies Eq. 4.66 with the equality, the ratio $\mathcal{S}/\mathcal{N}$ is indeed maximized when $h(t)$ is matched to $\varphi(t)$, as claimed.

The frequency-domain interpretation of this result is instructive. Since amplitude scaling affects the signal and noise in the same way, we need consider only $c = 1$. Then the transfer function of the matched filter is given by

$$H(f) = \int_{-\infty}^{\infty} \varphi(T-t)e^{-j2\pi ft}\,dt$$

$$= \int_{-\infty}^{\infty} \varphi(\alpha)e^{-j2\pi f(T-\alpha)}\,d\alpha$$

$$= e^{-j2\pi fT}\,\Phi^*(f), \tag{4.67a}$$

where the signal spectrum is

$$\Phi(f) = |\Phi(f)|\,e^{j\theta(f)} \triangleq \int_{-\infty}^{\infty} \varphi(t)e^{-j2\pi ft}\,dt. \tag{4.67b}$$

Thus

$$H(f) = |\Phi(f)|\,e^{-j[\theta(f)+2\pi fT]}. \tag{4.67c}$$

In accordance with the inverse Fourier transform,

$$\varphi(t) = \int_{-\infty}^{\infty} \Phi(f)e^{j2\pi ft}\,df, \tag{4.68}$$

we may interpret the filter input $\varphi(t)$ to be a composite of many small (complex) sinusoids: the sinusoid at frequency f_1 has amplitude $|\Phi(f_1)|\,df$ and phase $\theta(f_1)$. In passing through the filter this component is multiplied by $H(f_1)$, which changes its magnitude to $|\Phi(f_1)|^2$ and its phase to

$$\theta(f_1) - [\theta(f_1) + 2\pi f_1 T] = -2\pi f_1 T.$$

Thus the filter-output sinusoid at frequency f_1 is

$$|\Phi(f_1)|^2\,df\,e^{j2\pi f_1(t-T)},$$

which has a maximum at $t = T$. Since this is true for every f_1, all of the frequency components of $\varphi(t)$ are brought into phase coincidence and reinforce each other at $t = T$; as shown in Fig. 4.24, an output signal peak is produced at this instant.

Appreciation of the effect of the spectral-amplitude shaping caused by $|H(f)|$ can be gained by contrasting the matched filter with an *inverse filter*, which has the transfer function

$$\frac{e^{-j2\pi fT}}{\Phi(f)} = \frac{1}{|\Phi(f)|}\,e^{-j[\theta(f)+2\pi fT]}. \tag{4.69}$$

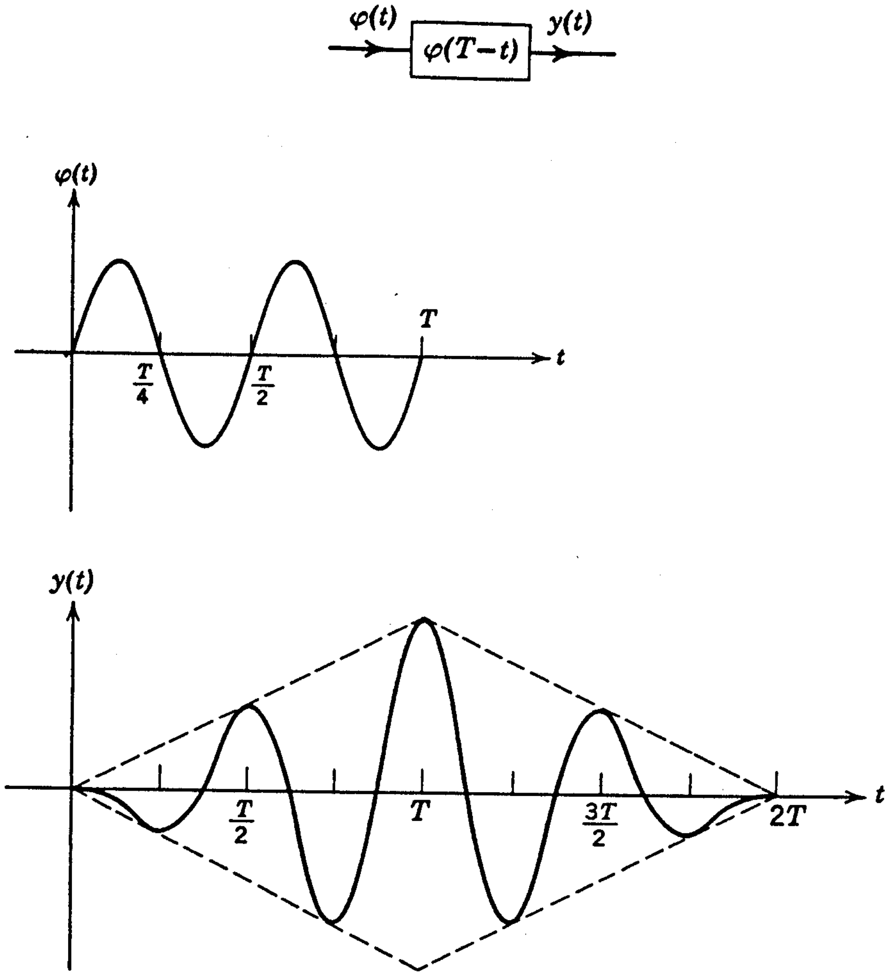


Figure 4.24 An example illustrating that the output of the matched filter is maximum at the instant $t = T$.

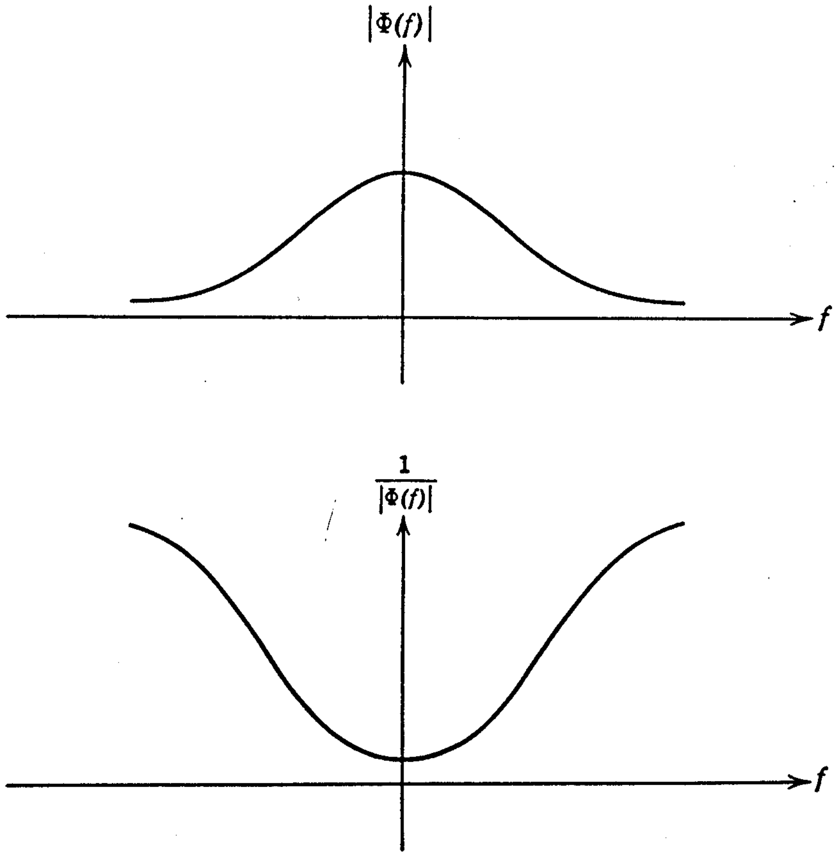


Figure 4.25 The inverse filter has high gain at frequencies for which $|\Phi(f)|$ is small, whereas the matched filter gain is proportional to $|\Phi(f)|$.

The inverse filter also brings all components of $\varphi(t)$ into phase coincidence. As shown in Fig. 4.25, however, the weaker components of $\varphi(t)$ are accentuated by the inverse filter, whereas they are suppressed by the matched filter. Since the noise spectrum $S_w(f)$ is flat over all frequencies, the inverse filter exalts the out-of-band noise and the matched filter subdues it.

Component Accuracy

So far we have presumed that the receiver knows exactly both the transmitter signal vectors $\{\mathbf{s}_i\}$ and the orthonormal base functions $\{\varphi_j(t)\}$. In practice, of course, limitations on component accuracy render this knowledge only approximate. Alternatively, in the interests of economy we might wish to settle for a system that is somewhat less than optimum.

In general, calculation of the precise trade-off between error performance and the precision of receiver instrumentation is both tedious and unrewarding. It is more instructive to visualize the nature and extent of the problem geometrically. For example, assume that there are two equally likely transmitter signals, say

$$s(t) = \pm s_1\varphi_1(t). \tag{4.70a}$$

The corresponding vector representation is illustrated by the black dots in Fig. 4.26. The receiver's approximations to these signals might be

$$\tilde{s}(t) = \pm[\tilde{s}_1\varphi_1(t) + \tilde{s}_2\varphi_2(t)]. \tag{4.70b}$$

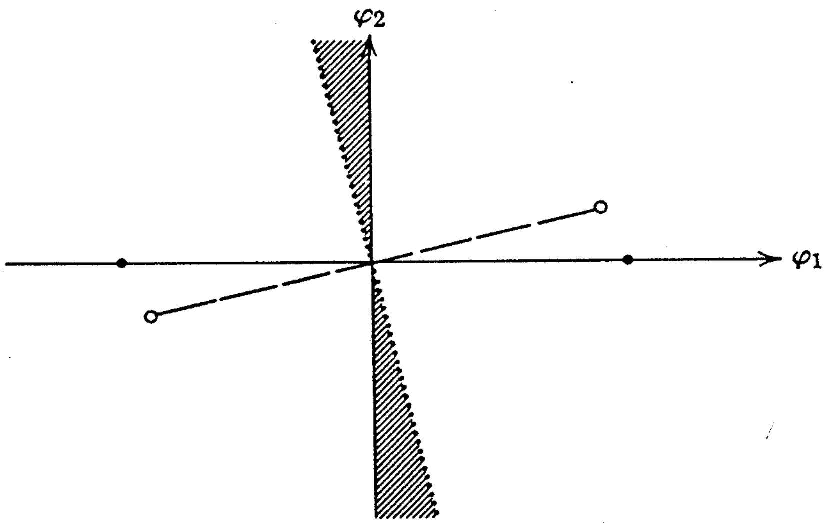


Figure 4.26 The effect of receiver approximation.

These approximations are represented vectorially by the open dots in the figure. The second orthonormal function $\varphi_2(t)$ is introduced to permit complete generality in representing the receiver's approximation of $\varphi_1(t)$.

A receiver matched to these approximate signals would employ the decision boundary indicated by the dotted line in Fig. 4.26, whereas an optimum receiver would use the φ_2-axis as the decision boundary. It is clear that the degradation in error performance is small as long as the receiver's approximations of the $\{s_i(t)\}$ are sufficiently accurate that the probability of the received vector $\mathbf{r}$ falling into the shaded area is small compared with the optimum P[$\mathcal{E}$]. This condition is met in general whenever

$$|\tilde{\mathbf{s}}_i - \mathbf{s}_i|^2 = \int_{-\infty}^{\infty} [\tilde{s}_i(t) - s_i(t)]^2 \, dt$$

is small compared with the square of each intersignal distance, $|\mathbf{s}_i - \mathbf{s}_k|^2$, for all i and $k \neq i$.

4.5 PROBABILITY OF ERROR

We have seen in Section 4.3 that the problem of communicating one of a set of M specified signals $\{s_i(t)\}$ over a channel disturbed only by additive white Gaussian noise always reduces to a corresponding vector communication problem. In particular, we recall that the transmitter signals are represented by M points $\{\mathbf{s}_i\}$ in an N-dimensional space and that the relevant noise disturbance is represented by an N-dimensional random vector, $\mathbf{n}$, with the spherically symmetric density function

$$p_{\mathbf{n}}(\boldsymbol{\alpha}) = \frac{1}{(\pi \mathcal{N}_0)^{N/2}} e^{-|\boldsymbol{\alpha}|^2/\mathcal{N}_0}. \tag{4.71a}$$

In accordance with the discussion leading to Eq. 4.19, the optimum receiver divides the signal space into a set of M disjoint decision regions $\{I_i\}$; any point $\boldsymbol{\rho}$ is assigned to I_k if and only if

$$|\boldsymbol{\rho} - \mathbf{s}_k|^2 - \mathcal{N}_0 \ln \mathrm{P}[m_k] < |\boldsymbol{\rho} - \mathbf{s}_i|^2 - \mathcal{N}_0 \ln \mathrm{P}[m_i]; \qquad \text{for all } i \neq k. \tag{4.71b}$$

The receiver output $\hat{m}$ is then set equal to m_k whenever the received vector

$$\mathbf{r} = \mathbf{s} + \mathbf{n} \tag{4.71c}$$

lies in I_k. Since the vector communication problem is invariant to the specific orthonormal base $\{\varphi_j(t)\}, j = 1, 2, \ldots, N$, that relates the $\{\mathbf{s}_i\}$ and the $\{s_i(t)\}$, the probability of error is independent of the waveshapes ascribed to the $\{\varphi_j(t)\}$.

In this section we evaluate the minimum attainable error probability (Eqs. 4.20, with $\sigma^2 = \mathcal{N}_0/2$) for certain important vector signal configurations. Except for $M = 2$, we assume that all M a priori probabilities

$\{P[m_i]\}$ are equal. The assumption is justified from an operational point of view in the discussion of "completely symmetric signals" at the end of this chapter.

Equivalent Signal Sets

In addition to signal sets that are equivalent by virtue of the fact that their geometrical configurations are identical, different geometrical configurations may also be equivalent insofar as error probability is concerned. Insight into this fact is gained by considering the geometry of the decision regions.

Rotation and translation of coordinates. In Fig. 4.27*a* we show a signal $\mathbf{s}_i$ and its decision region I_i. Whenever $\mathbf{s}_i$ is transmitted, a correct

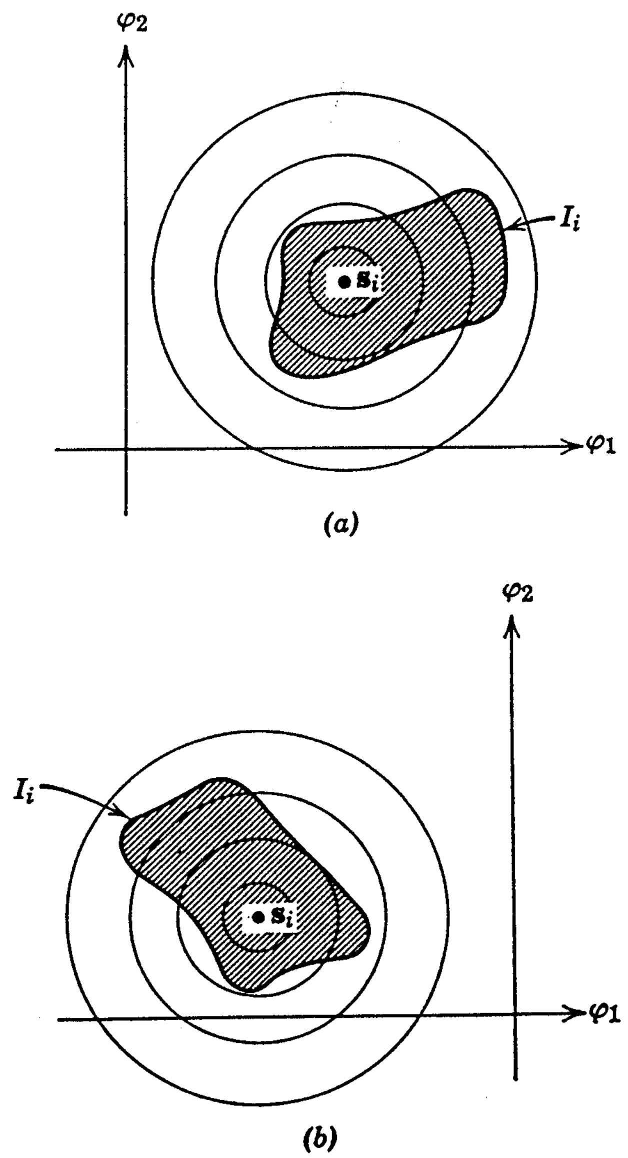


Figure 4.27 Equivalent decision regions. The concentric circles represent loci of constant $p_{\mathbf{n}}$.

decision results if $\mathbf{n} + \mathbf{s}_i$ falls within I_i. The probability of this event is unaffected if $\mathbf{s}_i$ and I_i are translated together through signal space. This follows, in accordance with Eqs. 4.71, from the fact that the noise $\mathbf{n}$ is additive and its density function $p_{\mathbf{n}}$ is independent of the signal. Moreover, since $p_{\mathbf{n}}$ is spherically symmetric, as indicated by the contours of constant probability density in the illustration, the probability that $\mathbf{n} + \mathbf{s}_i$ will fall in I_i is also unaffected by a rotation of I_i about $\mathbf{s}_i$. Thus $\mathbf{s}_i$ and I_i may be simultaneously translated and rotated, as in Fig. 4.27*b*, without affecting the conditional probability of a correct decision, $\mathrm{P}[\mathcal{C} \mid m_i]$.

Minimum-energy signals. Although the probability of a correct decision is invariant to translation, such a transformation does affect the energy required to transmit each signal: in general, $\mathbf{s}_i' = \mathbf{s}_i - \mathbf{a}$ implies

$$E_i \triangleq \int_{-\infty}^{\infty} s_i^2(t)\, dt = |\mathbf{s}_i|^2 \neq |\mathbf{s}_i - \mathbf{a}|^2 = |\mathbf{s}_i'|^2 \triangleq E_i'. \tag{4.72}$$

When there is a constraint, say E_s, on the peak energy permitted for any signal, the vectors $\{\mathbf{s}_i\}$ are constrained to lie within a sphere of radius $\sqrt{E_s}$, as indicated in Fig. 4.28. A somewhat weaker constraint is that the mean energy $\overline{E_m}$, defined as

$$\overline{E_m} \triangleq \sum_{i=0}^{M-1} \mathrm{P}[m_i]\, E_i = \sum_{i=0}^{M-1} \mathrm{P}[m_i]\, |\mathbf{s}_i|^2, \tag{4.73}$$

be less than some fixed value. For a given configuration of signal points the mean energy can be minimized, without affecting the probability of

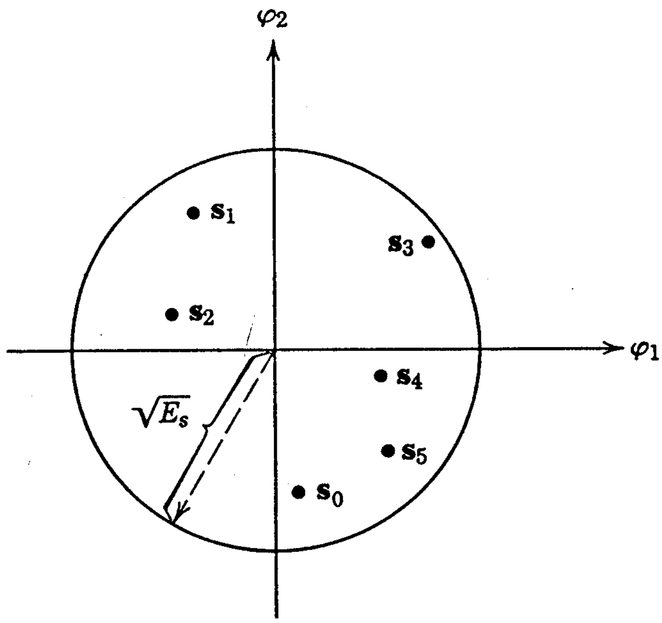


Figure 4.28 Peak energy constraint.

error, by subtracting from each signal $\mathbf{s}_i$ a constant vector $\mathbf{a}$ so chosen that

$$\sum_{i=0}^{M-1} \mathrm{P}[m_i]\, |\mathbf{s}_i - \mathbf{a}|^2$$

is minimum.

How to choose $\mathbf{a}$ is obvious once we have recognized that the expression for $\overline{E_m}$ is precisely the expression for the moment of inertia around the origin of a system of M point masses, where the mass of the ith point is $\mathrm{P}[m_i]$ and its position is $\mathbf{s}_i$. Since the moment of inertia is minimum when taken around the centroid (center of gravity) of a system, it follows that $\mathbf{a}$ should be chosen in such a way that the resulting centroid coincides with the origin. Given a set of probabilities $\{\mathrm{P}[m_i]\}$ and a set of signals $\{\mathbf{s}_i\}$, the appropriate choice of $\mathbf{a}$ is therefore

$$\mathbf{a} = \sum_{i=0}^{M-1} \mathrm{P}[m_i]\, \mathbf{s}_i \triangleq \mathrm{E}[\mathbf{s}]. \tag{4.74a}$$

As proof, we note that for any other translation, say $\mathbf{b}$, we have

$$\begin{aligned} \mathrm{E}[|\mathbf{s} - \mathbf{b}|^2] &= \mathrm{E}[|(\mathbf{s} - \mathbf{a}) + (\mathbf{a} - \mathbf{b})|^2] \\ &= \mathrm{E}[|\mathbf{s} - \mathbf{a}|^2] + 2(\mathbf{a} - \mathbf{b}) \cdot (\mathrm{E}[\mathbf{s}] - \mathbf{a}) + |\mathbf{a} - \mathbf{b}|^2 \qquad (4.74b) \\ &= \mathrm{E}[|\mathbf{s} - \mathbf{a}|^2] + |\mathbf{a} - \mathbf{b}|^2, \end{aligned}$$

where the last equality follows from Eq. 4.74a. The mean energy is increased when $\mathbf{b} \neq \mathbf{a}$. If the mean energy still exceeds the allowable maximum after the translation $\mathbf{a}$ is made, further reduction is possible only by transformations such as radial scaling that do affect the probability of error.

Rectangular Signal Sets

When the geometric configuration of M equally likely signal vectors is rectangular, the calculation of the error probability is especially easy. The simplest situation is that in which there are only two signals.

Binary signals. The general case of two signal vectors, each with probability $\frac{1}{2}$, is shown in Fig. 4.29*a*. From the standpoint of error probability, an equivalent signal set is that shown in Fig. 4.29*b*, in which the signal configuration has been rotated and translated in such a way that the centroid coincides with the origin and the vector $(\mathbf{s}_0 - \mathbf{s}_1)$ lies along the φ_1 axis.

The optimum decision regions for Fig. 4.29*b* are determined by the expression

$$\min_i \; \{|\boldsymbol{\rho} - \mathbf{s}_i|^2 - \mathcal{N}_0 \ln \mathrm{P}[m_i]\}. \tag{4.75}$$

For equal a priori probabilities, this decision rule is just

$$\min_i |\boldsymbol{\rho} - \mathbf{s}_i|^2.$$

It is clear from Fig. 4.29*b* that the locus of all points $\boldsymbol{\rho}$ equally distant from $\mathbf{s}_0$ and $\mathbf{s}_1$ is the φ_2 axis. Thus an error occurs when $\mathbf{s}_1$ is transmitted

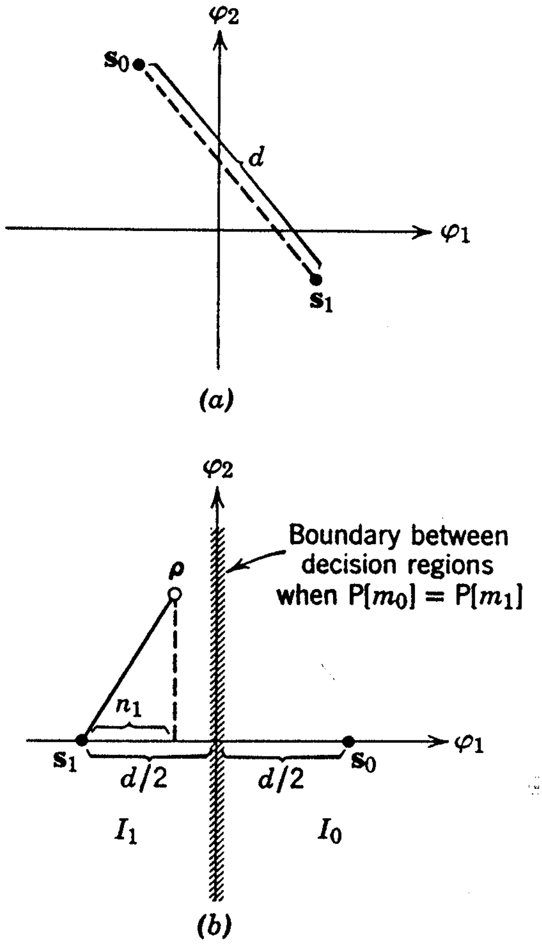


Figure 4.29 Binary signal sets for which P[ℰ] is the same. The signals in (*b*) are called "antipodal"; each has energy $E_s = (d/2)^2$.

if and only if the noise component n_1 exceeds $d/2$, where d is the distance between the two signals:

$$\mathrm{P}[\mathcal{E} \mid m_1] = \mathrm{P}[\boldsymbol{\rho} \text{ in } I_0 \mid m_1] = \mathrm{P}\left[n_1 > \frac{d}{2}\right],$$

where

$$d^2 \triangleq |\mathbf{s}_0 - \mathbf{s}_1|^2 = \int_{-\infty}^{\infty} [s_0(t) - s_1(t)]^2 \, dt. \tag{4.76a}$$

But n_1 is zero-mean Gaussian with variance $\mathcal{N}_0/2$, so that

$$\mathrm{P}[\mathcal{E} \mid m_1] = \int_{d/2}^{\infty} \frac{1}{\sqrt{\pi \mathcal{N}_0}} e^{-\alpha^2/\mathcal{N}_0} \, d\alpha.$$

Setting $\gamma = \alpha\sqrt{2/\mathcal{N}_0}$, we have

$$P[\mathcal{E} \mid m_1] = \int_{\frac{d/2}{\sqrt{\mathcal{N}_0/2}}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\gamma^2/2}\, d\gamma \triangleq Q\left(\frac{d}{\sqrt{2\mathcal{N}_0}}\right).$$

Since, by symmetry, the conditional probability of error is the same for either signal, we also have

$$P[\mathcal{E}] = \sum_{i=0}^{1} P[m_i]\, P[\mathcal{E} \mid m_i] = P[\mathcal{E} \mid m_1] = Q\left(\frac{d}{\sqrt{2\mathcal{N}_0}}\right). \qquad (4.76b)$$

The function $Q(\ )$ was defined in Eq. 2.50 and plotted in Fig. 2.36.

Equation 4.76b is the minimum error probability for any pair of equally likely signal vectors separated by a distance d, regardless of their actual location in signal space. When the signals have minimum energy and are therefore antipodal as in Fig. 4.29*b*, the length of each vector is $\sqrt{E_s}$, so that $d = 2\sqrt{E_s}$ and

$$P[\mathcal{E}] = Q(\sqrt{2E_s/\mathcal{N}_0});\ \text{equally likely antipodal signals.} \qquad (4.77)$$

On the other hand, when the signals are orthogonal, as in Fig. 4.30, we have $d = \sqrt{2E_s}$ and

$$P[\mathcal{E}] = Q(\sqrt{E_s/\mathcal{N}_0});\ \text{two equally likely orthogonal signals.} \qquad (4.78)$$

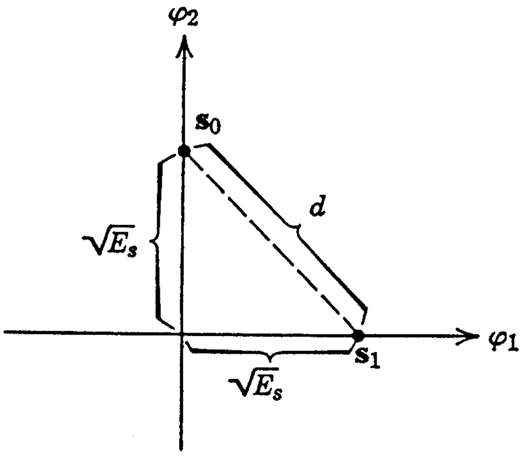


Figure 4.30 Two orthogonal signals.

It is common engineering practice to express energy ratios in units of decibels (db), where

$$\left.\frac{E_s}{\mathcal{N}_0}\right|_{\text{db}} \triangleq 10 \log_{10} \frac{E_s}{\mathcal{N}_0}.$$

For example

$\frac{E_s}{\mathcal{N}_0}$	$\left.\frac{E_s}{\mathcal{N}_0}\right\rvert_{db}$
0.1	−10 db
1.0	0 db
2.0	3 db
3.0	4.8 db
10.0	10 db
100.0	20 db

The probabilities of error for antipodal and orthogonal signaling are plotted in Fig. 4.31 with $E_s/\mathcal{N}_0$ in units of db. The figure illustrates that antipodal signaling is 3 db more efficient than orthogonal signaling in communicating one of two equally likely messages.

With binary signals, it is also easy to determine $P[\mathcal{E}]$ when the a priori probabilities are not equal. As shown in Fig. 4.32, the decision boundary is shifted from $\mathbf{s}_1$ toward $\mathbf{s}_0$ by an amount

$$\Delta = \frac{\mathcal{N}_0/2}{d} \ln \frac{P[m_1]}{P[m_0]}. \tag{4.79a}$$

Equation 4.79a is derived from the decision rule of Eq. 4.75 by solving the equation

$$|\boldsymbol{\rho} - \mathbf{s}_1|^2 - \mathcal{N}_0 \ln P[m_1] = |\boldsymbol{\rho} - \mathbf{s}_0|^2 - \mathcal{N}_0 \ln P[m_0]$$

for $\boldsymbol{\rho} = (\rho_1, \rho_2)$. For any value of ρ_2 we then have

$$\left(\rho_1 + \frac{d}{2}\right)^2 - \mathcal{N}_0 \ln P[m_1] = \left(\rho_1 - \frac{d}{2}\right)^2 - \mathcal{N}_0 \ln P[m_0].$$

Since Δ is the value of ρ_1 satisfying this equation,

$$2\Delta d = \mathcal{N}_0 \ln \frac{P[m_1]}{P[m_0]}.$$

The resulting error probability is

$$P[\mathcal{E}] = P[m_0]\, Q\left(\frac{d - 2\Delta}{\sqrt{2\mathcal{N}_0}}\right) + P[m_1]\, Q\left(\frac{d + 2\Delta}{\sqrt{2\mathcal{N}_0}}\right). \tag{4.79b}$$

Rectangular decision regions. The ease of calculating the error probability for binary signals is directly attributable to the fact that an error occurs if and only if *one* random variable exceeds a given magnitude. A situation that is only slightly more complicated exists whenever the decision region boundaries are rectangular. Consider, for example, the

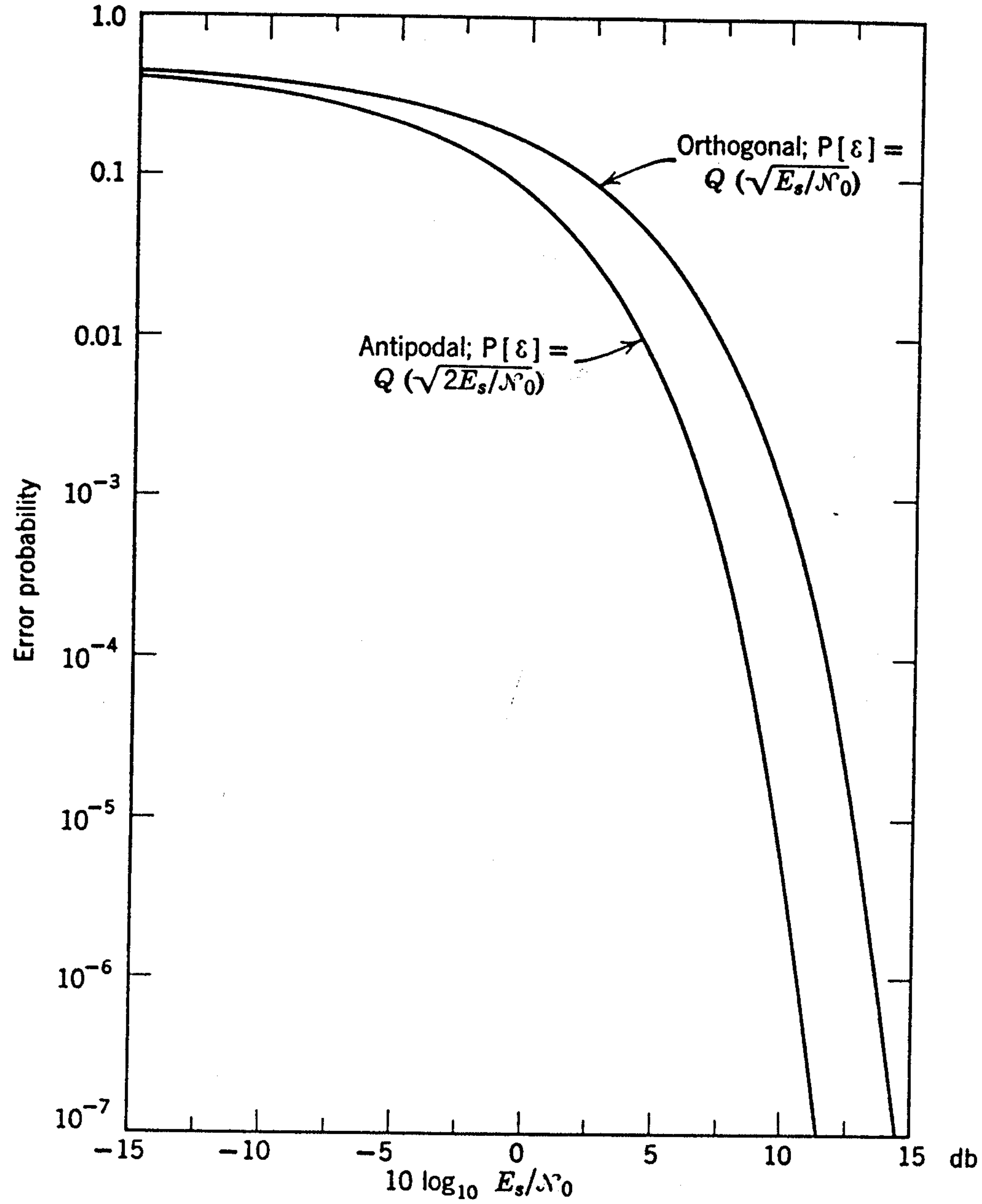


Figure 4.31 Probability of error for binary antipodal and binary orthogonal signaling with equally likely messages.

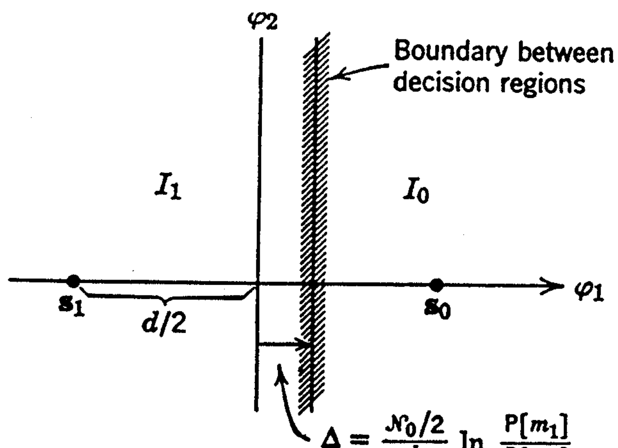


Figure 4.32 Decision regions for antipodal signals with distance d and unequal a priori probabilities.

signal $\mathbf{s}_i$ and decision region I_i shown in Fig. 4.33*a*. After translating $\mathbf{s}_i$ to the origin and rotating the configuration as shown in Fig. 4.33*b*, we see immediately that $\mathbf{s}_i + \mathbf{n}$ falls within I_i whenever, simultaneously,

$$(a_1 < n_1 < b_1) \quad \text{and} \quad (a_2 < n_2 < b_2). \tag{4.80a}$$

But n_1 and n_2 are statistically independent (cf. Eq. 4.49), and the density

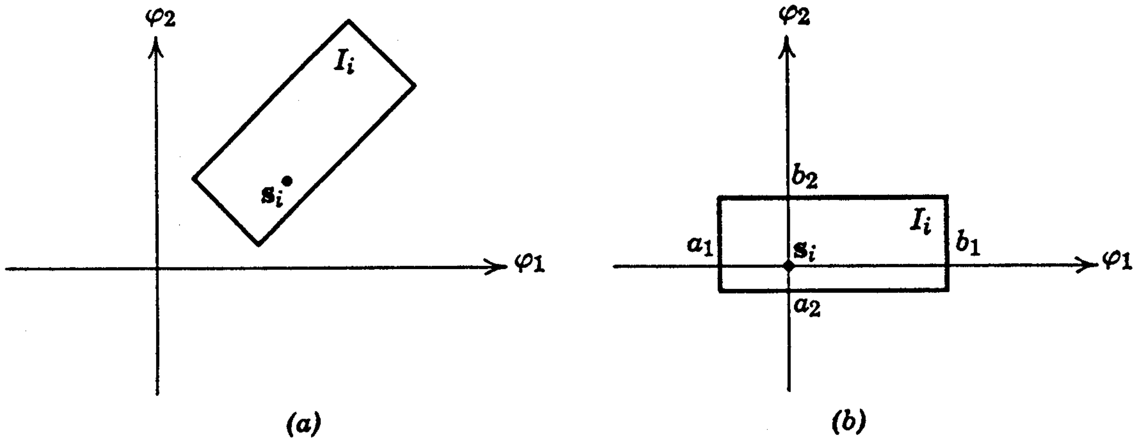


Figure 4.33 A single rectangular decision region.

function, say p_n, of each is the same:

$$p_{n_j}(\alpha) = p_n(\alpha) = \frac{1}{\sqrt{\pi \mathcal{N}_0}} e^{-\alpha^2/\mathcal{N}_0}; \qquad j = 1, 2. \tag{4.80b}$$

Thus

$$\begin{aligned} P[\mathcal{C} \mid m_i] &= P[a_1 < n_1 < b_1, a_2 < n_2 < b_2] \\ &= P[a_1 < n_1 < b_1]\, P[a_2 < n_2 < b_2] \\ &= \int_{a_1}^{b_1} p_n(\alpha)\, d\alpha \int_{a_2}^{b_2} p_n(\alpha)\, d\alpha. \end{aligned} \tag{4.80c}$$

The optimum decision boundaries are always rectangular when the signal vector configuration is rectangular and all signals are equally likely. A simple example is the rectangular configuration of six equally likely signals shown in Fig. 4.34. We have

$$P[\mathcal{C} \mid m_0] = \int_{-\infty}^{d/2} p_n(\alpha)\, d\alpha \int_{-\infty}^{d/2} p_n(\alpha)\, d\alpha = (1 - p)^2, \tag{4.81a}$$

where $p = Q(d/\sqrt{2\mathcal{N}_0})$ is the probability of error for two signals separated by a distance d. From symmetry,

$$P[\mathcal{C} \mid m_0] = P[\mathcal{C} \mid m_1] = P[\mathcal{C} \mid m_2] = P[\mathcal{C} \mid m_3]. \tag{4.81b}$$

Similarly,

$$P[\mathcal{C} \mid m_4] = P[\mathcal{C} \mid m_5] = \int_{-d/2}^{d/2} p_n(\alpha)\, d\alpha \int_{-d/2}^{\infty} p_n(\alpha)\, d\alpha$$
$$= (1 - 2p)(1 - p). \tag{4.81c}$$

Thus

$$P[\mathcal{C}] = \sum_{i=0}^{5} P[\mathcal{C} \mid m_i]\, P[m_i]$$
$$= \tfrac{4}{6}(1 - p)^2 + \tfrac{2}{6}(1 - 2p)(1 - p). \tag{4.81d}$$

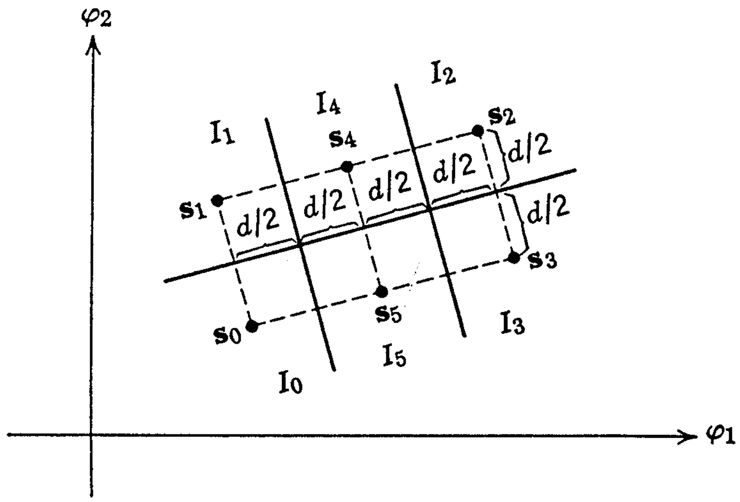


Figure 4.34 Rectangular decision regions.

Vertices of a hypercube. A special case of rectangular decision regions occurs when $M = 2^N$ equally likely messages are located on the vertices of an N-dimensional hypercube centered on the origin. This configuration is shown geometrically in Fig. 4.35 for $N = 2$ and 3. Analytically, we have

$$\mathbf{s}_i = (s_{i1}, s_{i2}, \ldots, s_{iN}); \quad i = 0, 1, \ldots, 2^N - 1, \tag{4.82a}$$

where

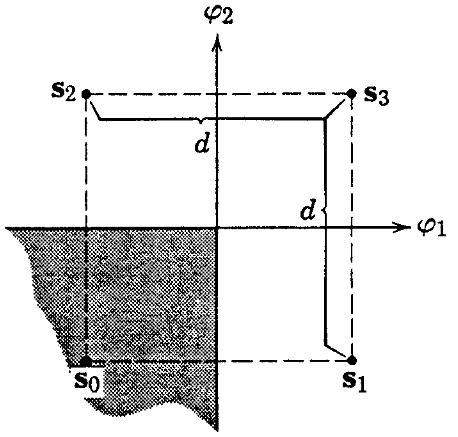
$$s_{ij} = \begin{cases} +\, d/2 \\ \text{or} & \text{for all } i, j. \\ -d/2 \end{cases} \tag{4.82b}$$

To evaluate the error probability, assume that the signal

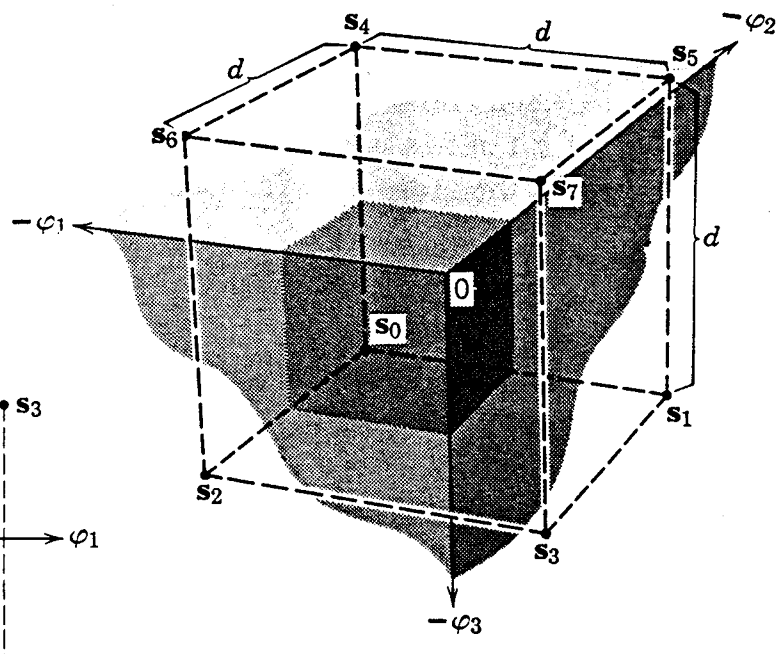
$$\mathbf{s}_0 \triangleq \left(-\frac{d}{2}, -\frac{d}{2}, \ldots, -\frac{d}{2}\right) \tag{4.83}$$

is transmitted. We first claim that no error is made if

$$n_j < \frac{d}{2}; \qquad \text{for all } j = 1, 2, \ldots, N. \tag{4.84a}$$



(a) $N = 2$



(b) $N = 3$

Figure 4.35 Signals on the vertices of two- and three-dimensional cubes: (a) $N = 2$; (b) $N = 3$. The decision regions I_0 are shaded.

The proof is immediate. When $\mathbf{r} = \boldsymbol{\rho}$ is received the jth component of $\boldsymbol{\rho} - \mathbf{s}_i$ is

$$(\rho_j - s_{ij}) = \begin{cases} n_j; & \text{if } s_{ij} = -\dfrac{d}{2} \\ n_j - d\,; & \text{if } s_{ij} = +\dfrac{d}{2}. \end{cases} \tag{4.84b}$$

Since Eq. 4.84a implies

$$d - n_j > n_j; \quad \text{all } j, \tag{4.84c}$$

it follows that

$$|\boldsymbol{\rho} - \mathbf{s}_i|^2 = \sum_{j=1}^{N} (\rho_j - s_{ij})^2 > \sum_{j=1}^{N} n_j^2 = |\boldsymbol{\rho} - \mathbf{s}_0|^2 \tag{4.84d}$$

for all $\mathbf{s}_i \neq \mathbf{s}_0$ whenever Eq. 4.84a is satisfied.

We next claim that an error is made if, for at least one j,

$$n_j > \frac{d}{2}. \tag{4.85}$$

This follows from the fact that $\boldsymbol{\rho}$ is closer to $\mathbf{s}_j$ than to $\mathbf{s}_0$ whenever Eq. 4.85 is satisfied, where $\mathbf{s}_j$ denotes that signal with components $+d/2$ in the jth direction and $-d/2$ in all other directions. (Of course, $\boldsymbol{\rho}$ may be still closer to some signal other than $\mathbf{s}_j$, but it cannot be closest to $\mathbf{s}_0$.)

Equations 4.84d and 4.85 together imply that a correct decision is made if and only if Eq. 4.84a is satisfied. The probability of this event, given that $m = m_0$, is therefore

$$\begin{aligned} P[\mathcal{C} \mid m_0] &= P\left[\text{all } n_j < \frac{d}{2};\quad j = 1, 2, \ldots, N\right] \\ &= \prod_{j=1}^{N} P\left[n_j < \frac{d}{2}\right] \\ &= \left(1 - \int_{d/2}^{\infty} p_n(\alpha)\, d\alpha\right)^N \\ &= (1 - p)^N; \end{aligned}$$

in which,

$$p = Q\left(\frac{d}{\sqrt{2\mathcal{N}_0}}\right) \tag{4.86}$$

is again the probability of error for two equally likely signals separated by distance d. Finally, from symmetry

$$P[\mathcal{C} \mid m_i] = P[\mathcal{C} \mid m_0]; \qquad \text{for all } i, \tag{4.87a}$$

hence

$$P[\mathcal{C}] = (1 - p)^N. \tag{4.87b}$$

In order to express this result in terms of signal energy, we again recognize that the distance squared from the origin to each signal $\mathbf{s}_i$ is the same. The transmitted energy is therefore independent of i, hence may be designated E_s. From Eqs. 4.58b and 4.82b we have

$$|\mathbf{s}_i|^2 = \sum_{j=1}^{N} s_{ij}^2 = N\frac{d^2}{4} = E_s, \tag{4.88a}$$

$$d = 2\sqrt{\frac{E_s}{N}}, \tag{4.88b}$$

and

$$p = Q\left(\sqrt{\frac{2E_s}{N\mathcal{N}_0}}\right). \tag{4.89}$$

The simple form of the result $P[\mathcal{C}] = (1 - p)^N$ suggests that a more immediate derivation may exist. Indeed one does. Note that the jth coordinate of the random signal $\mathbf{s}$ is a priori equally likely to be $+d/2$

or $-d/2$, independent of all other coordinates. Moreover, the noise n_j disturbing the jth coordinate is independent of the noise in all other coordinates. Hence, by the theorem on irrelevance, a decision may be made on the jth coordinate without examining any other coordinate. This single-coordinate decision corresponds to the problem of binary signals separated by distance d, for which the probability of correct decision is $1 - p$. Since in the original hypercube problem a correct decision is made if and only if a correct decision is made on every coordinate, and since these decisions are independent, it follows immediately that

$$P[\mathcal{C}] = (1 - p)^N. \tag{4.90}$$

Orthogonal and Related Signal Sets

Another class of equally likely signals for which the minimum attainable error probability is quite easy to calculate is the set of M equal-energy orthogonal vectors. Closely related to them are the simplex and bi-orthogonal signal sets. In treating these sets it is convenient to index the orthonormal axes $\{\varphi_j\}$ from $j = 0$ to $N - 1$ rather than from $j = 1$ to N, where N is the dimensionality of the signal space.

Orthogonal signals. When M equally likely and equal-energy signals are mutually orthogonal, so that $N = M$ and

$$\int_{-\infty}^{\infty} s_i(t)\, s_k(t)\, dt = \mathbf{s}_i \cdot \mathbf{s}_k = E_s\, \delta_{ik}; \qquad i, k = 0, 1, \ldots, M - 1, \tag{4.91}$$

the optimum decision region boundaries are no longer rectangular and are difficult to visualize. It is easier to proceed analytically. Letting $\boldsymbol{\varphi}_j$ denote the unit vector along the jth coordinate axis and

$$\mathbf{s}_j = \sqrt{E_s}\, \boldsymbol{\varphi}_j; \qquad j = 0, 1, \ldots, M - 1, \tag{4.92a}$$

we note that the squared distance from $\mathbf{s}_j$ to the received vector $\mathbf{r}$ is

$$\begin{aligned} |\mathbf{r} - \mathbf{s}_j|^2 &= |\mathbf{r}|^2 + |\mathbf{s}_j|^2 - 2\mathbf{r} \cdot (\sqrt{E_s}\, \boldsymbol{\varphi}_j) \\ &= |\mathbf{r}|^2 + E_s - 2r_j\sqrt{E_s}, \end{aligned} \tag{4.92b}$$

where r_j is the jth component of $\mathbf{r}$.

When $\mathbf{s}_k$ is transmitted, it follows that

$$|\mathbf{r} - \mathbf{s}_k|^2 < |\mathbf{r} - \mathbf{s}_i|^2; \qquad \text{all } i \neq k \tag{4.93a}$$

if and only if

$$-2r_k\sqrt{E_s} < -2r_i\sqrt{E_s},$$

i.e.

$$r_i < r_k; \qquad \text{all } i \neq k. \tag{4.93b}$$

As shown in Fig. 4.36, when $\mathbf{s}_0$ is transmitted we have

$$r_0 = n_0 + \sqrt{E_s} \tag{4.94a}$$

$$r_i = n_i; \qquad i > 0. \tag{4.94b}$$

Thus

$$\begin{aligned} P[\mathcal{C} \mid m_0, r_0 = \alpha] &= P[n_1 < \alpha, n_2 < \alpha, \ldots, n_{M-1} < \alpha] \\ &= (P[n_1 < \alpha])^{M-1}, \end{aligned} \tag{4.95a}$$

in which the last equality stems from the fact that all n_i are statistically

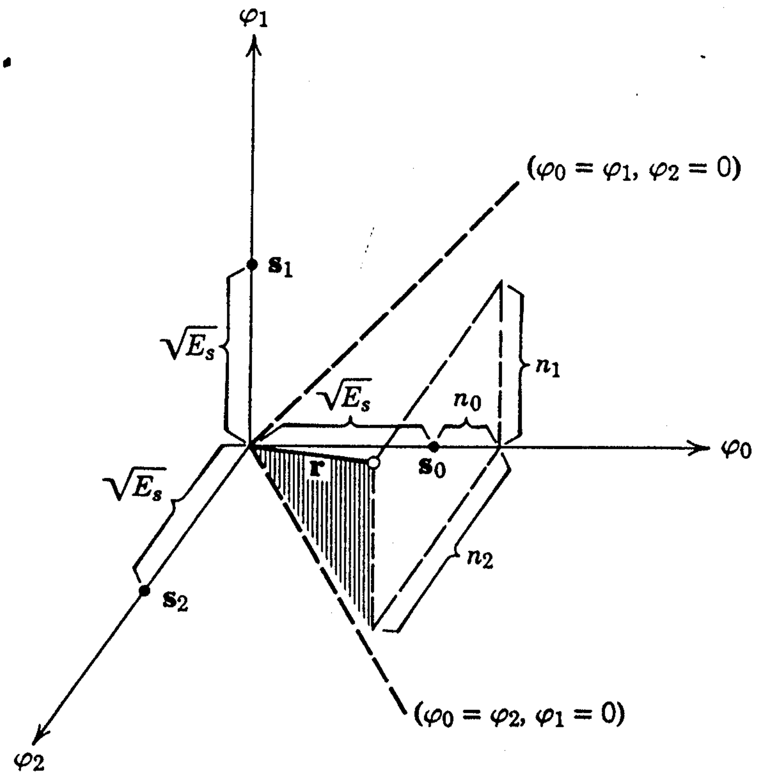


Figure 4.36 Three orthogonal signals. When $\mathbf{s}_0$ is transmitted, a correct decision is made if and only if n_1 and n_2 are both less than $\alpha = \sqrt{E_s} + n_0$. The heavy dashed lines are the intersections of the decision boundaries with the planes $\varphi_2 = 0$ and $\varphi_1 = 0$.

independent and identically distributed. Multiplying by

$$p_{r_0}(\alpha) = p_n(\alpha - \sqrt{E_s}) \tag{4.95b}$$

and integrating yields, for M equally likely equal-energy signals,

$$P[\mathcal{C} \mid m_0] = \int_{-\infty}^{\infty} p_n(\alpha - \sqrt{E_s})\, d\alpha \left[\int_{-\infty}^{\alpha} p_n(\beta)\, d\beta\right]^{M-1}, \tag{4.96a}$$

with

$$p_n(\alpha) \triangleq \frac{1}{\sqrt{\pi \mathcal{N}_0}}\, e^{-\alpha^2/\mathcal{N}_0} \tag{4.96b}$$

From symmetry,

$$P[\mathcal{C} \mid m_i] = P[\mathcal{C} \mid m_0] = P[\mathcal{C}], \tag{4.96c}$$

so that Eq. 4.96a is also the expression for the unconditional probability of a correct decision.

The integral in Eq. 4.96a cannot be simplified further but has been tabulated[36] as a function of M and $E_s/\mathcal{N}_0$; a plot of $P[\mathcal{E}] = 1 - P[\mathcal{C}]$ is provided in Fig. 4.37.

Simplex signals. A useful application of the energy minimization ideas discussed earlier is to M equally likely orthogonal signals. From Eqs.

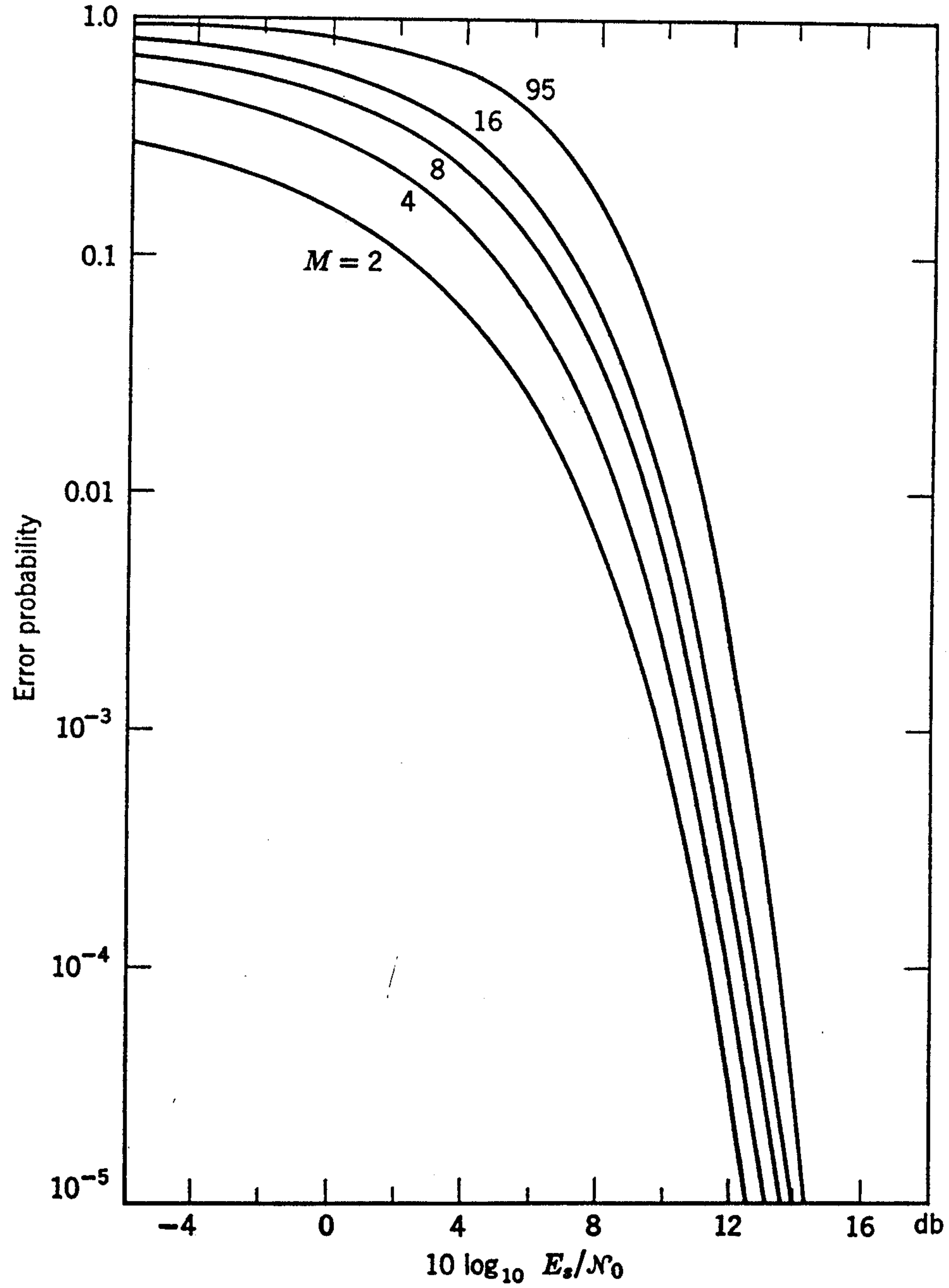


Figure 4.37 Error probability for M orthogonal signals.

4.74a and 4.92a the minimizing translation is

$$\mathbf{a} = \mathrm{E}[\mathbf{s}] = \frac{1}{M}\sum_{i=0}^{M-1} \mathbf{s}_i = \frac{\sqrt{E_s}}{M}\sum_{i=0}^{M-1} \boldsymbol{\varphi}_i. \tag{4.97a}$$

The resulting signal set

$$\{\mathbf{s}_i'\} = \{\mathbf{s}_i - \mathbf{a}\}; \qquad i = 0, 1, \ldots, M-1 \tag{4.97b}$$

is called a *simplex* and is the optimum[52] (minimum P[$\mathcal{E}$] set of M signals for use in white Gaussian noise when energy is constrained and $P[m_i] = 1/M$ for all i. The simplex signals for $M = 2$, 3, and 4 are shown in Fig. 4.38. Since

$$\sum_{i=0}^{M-1} \mathbf{s}_i' = \sum_{i=0}^{M-1} \mathbf{s}_i - M\mathbf{a} = \mathbf{0}, \tag{4.98}$$

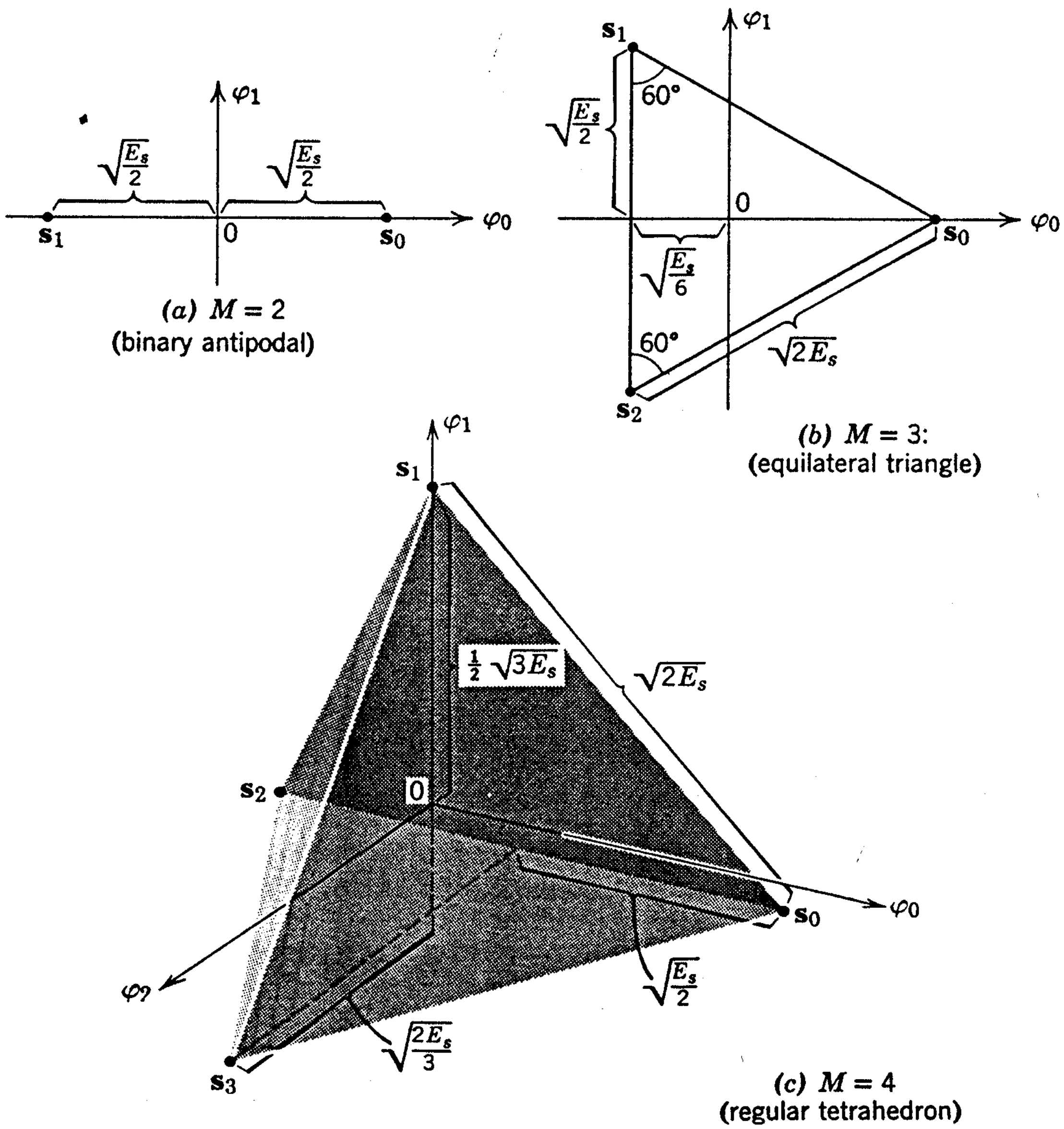


Figure 4.38 Simplex signals. All $\mathbf{s}_i$ are at distance $\sqrt{E_s(1 - 1/M)}$ from the origin.

any one of the $\{\mathbf{s}_i'\}$ can be expressed as a linear combination of the others. The M simplex signals therefore span a space of $N = M - 1$ dimensions. By virtue of the orthonormality of the $\{\boldsymbol{\varphi}_j\}$, for all i, k

$$
\begin{aligned}
\mathbf{s}_i' \cdot \mathbf{s}_k' &= (\mathbf{s}_i - \mathbf{a}) \cdot (\mathbf{s}_k - \mathbf{a}) \\
&= (\mathbf{s}_i \cdot \mathbf{s}_k) - \mathbf{a} \cdot (\mathbf{s}_k + \mathbf{s}_i) + |\mathbf{a}|^2 \\
&= E_s\,\delta_{ik} - 2\frac{E_s}{M} + \frac{E_s}{M} \\
&= \begin{cases} E_s\left(1 - \dfrac{1}{M}\right); & \text{for } i = k \\ -\dfrac{E_s}{M}; & \text{otherwise.} \end{cases}
\end{aligned}
\tag{4.99}
$$

We see that each signal in a simplex has the same energy, which is reduced by the factor $(1 - 1/M)$ from that required for the orthogonal signals, with *no change in error probability*. (Translations do not effect P[ℰ].) When $M = 2$, the saving is 3 db; for large M the saving is negligible.

Equation 4.99 may be used as the definition of a simplex. We note that a set of M vectors $\{\mathbf{s}_i'\}$ satisfying Eq. 4.99 may be transformed to a set of orthogonal vectors by adding a vector $\sqrt{E_s/M}\,\boldsymbol{\psi}$ to each $\mathbf{s}_i'$, where $\boldsymbol{\psi}$ is any unit vector orthogonal to all of the $\{\mathbf{s}_i'\}$.

Biorthogonal signals. The final specific signal configuration considered here is the biorthogonal set, illustrated for $N = 2$ and 3 in Fig. 4.39. This signal set can be obtained from an original orthogonal set of N signals by augmenting it with the negative of each signal. Obviously, for the biorthogonal set

$$M = 2N. \tag{4.100}$$

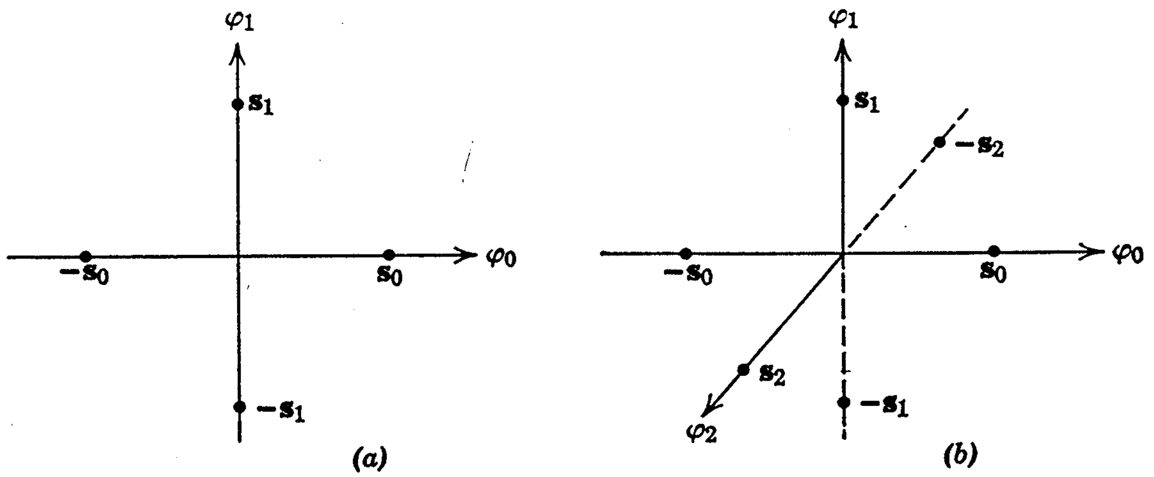


Figure 4.39 Biorthogonal signals, all at distance $\sqrt{E_s}$ from origin.

We denote the additional signals by $-\mathbf{s}_j$, $j = 0, 1, \ldots, N-1$, and assume each signal has energy E_s.

It is clear from Fig. 4.40 that the received message point is closer to $\mathbf{s}_0$ than to $-\mathbf{s}_0$ if and only if

$$r_0 > 0. \tag{4.101a}$$

Also, $\mathbf{r}$ is closer to $\mathbf{s}_0$ than to $\mathbf{s}_i$ if and only if

$$r_0 > r_i; \qquad i > 0, \tag{4.101b}$$

and $\mathbf{r}$ is closer to $\mathbf{s}_0$ than to $-\mathbf{s}_i$ if and only if

$$r_0 > -r_i; \qquad i > 0. \tag{4.101c}$$

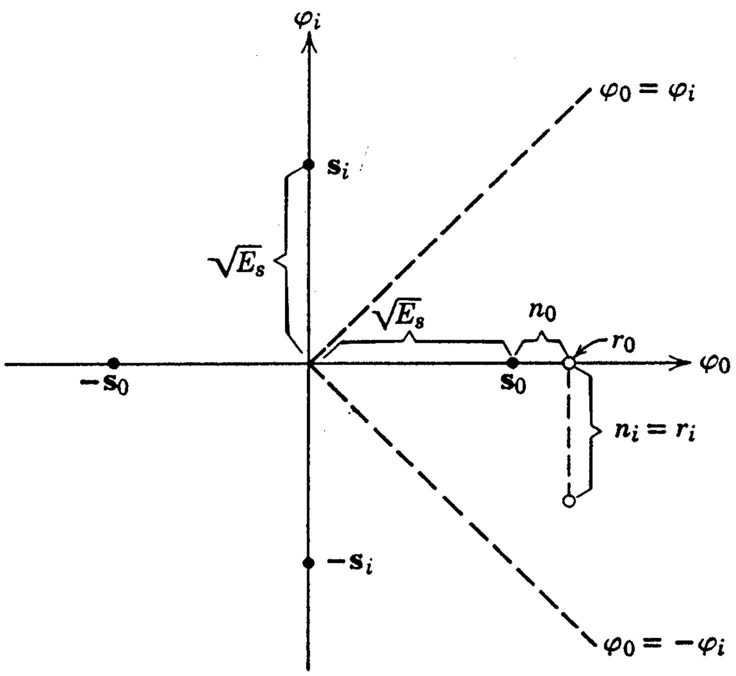


Figure 4.40 Biorthogonal signals. When $\mathbf{s}_0$ is transmitted, $\mathbf{r}$ is closer to $\pm\mathbf{s}_i$ than it is to $\mathbf{s}_0$ if and only if n_0 and n_i are such that one of the two heavy dashed lines is crossed.

It follows that the conditional probability of a correct decision for equally likely messages, given that $\mathbf{s}_0$ is transmitted and that

$$r_0 = n_0 + \sqrt{E_s} = \alpha > 0, \tag{4.102a}$$

is just

$$\begin{aligned} &\mathrm{P}[\mathcal{C} \mid m_0, r_0 = \alpha > 0] \\ &= \mathrm{P}[-\alpha < n_1 < \alpha, -\alpha < n_2 < \alpha, \ldots, -\alpha < n_{N-1} < \alpha] \\ &= \{\mathrm{P}[-\alpha < n < \alpha]\}^{N-1} \\ &= \left[\int_{-\alpha}^{\alpha} p_n(\beta)\, d\beta\right]^{N-1}. \end{aligned} \tag{4.102b}$$

The notation is that of Eq. 4.96b. Multiplying by $p_{r_0}(\alpha) = p_n(\alpha - \sqrt{E_s})$ and integrating over α from 0 to ∞ (because of the condition of Eq. 4.102a), we obtain

$$P[\mathcal{C} \mid m_0] = \int_0^\infty p_n(\alpha - \sqrt{E_s})\, d\alpha \left[\int_{-\alpha}^{\alpha} p_n(\beta)\, d\beta \right]^{N-1}. \tag{4.103}$$

Once again, by virtue of symmetry and the equal a priori probability of the $\{m_i\}$, Eq. 4.103 is also the expression for $P[\mathcal{C}]$. Noting that $N - 1 = (M/2) - 1$, we have†

$$P[\mathcal{C}] = \int_0^\infty p_n(\alpha - \sqrt{E_s})\, d\alpha \left[1 - 2\int_{\alpha}^{\infty} p_n(\beta)\, d\beta \right]^{\frac{M}{2} - 1}. \tag{4.104}$$

The difference in error performance for M biorthogonal and M orthogonal signals is negligible when M and E_s/N_0 are large, but the number of dimensions required is reduced by one half in the biorthogonal case.

Completely Symmetric Signal Sets and A Priori Knowledge

In almost all of the specific cases we have considered—in particular, the binary, orthogonal, simplex, biorthogonal, and vertices-of-a-hypercube signal sets—the error probability calculation is greatly simplified by the "complete symmetry" of the geometrical configuration of the $\{\mathbf{s}_i\}$. By *complete symmetry* we mean that any relabeling of the signal points can be undone by a rotation of coordinates, translation, and/or inversion of axes. As a counterexample, the signals of Fig. 4.34 are *not* completely symmetric.

Given complete symmetry, the condition

$$P[m_i] = \frac{1}{M}; \qquad \text{for all } i \tag{4.105}$$

leads to congruent decision regions $\{I_i\}$ and thus to a conditional probability of correct decision that is independent of the particular signal transmitted:

$$P[\mathcal{C} \mid m_i] = \text{a constant}; \qquad \text{for all } i. \tag{4.106a}$$

If such a congruent-decision-region receiver is used with message probabilities $\{P[m_i]\}$ that are not all the same, the resulting probability of correct decision is

$$P[\mathcal{C}] = \sum_{i=0}^{M-1} P[m_i]\, P[\mathcal{C} \mid m_i] = P[\mathcal{C} \mid m_0], \tag{4.106b}$$

which is unchanged from the equally likely message case. Thus the error performance of a congruent-decision-region receiver is invariant to the

† The integral of Eq. 4.104 is tabulated and plotted in reference 36.

actual source statistics $\{P[m_i]\}$. (Of course, if the source statistics are known in advance, the probability of correct decision can be increased by the use of a noncongruent-decision-region receiver designed in accordance with Eq. 4.71b.)

Invariance to message probabilities can be exploited by a communication system designer, who seldom knows in advance the exact input statistics of the source. If the transmitter is designed with completely symmetric signals and an optimum receiver is designed on the assumption that all messages are equally likely, Eqs. 4.106 will be satisfied and the error probability of the system can be specified independent of the message source to which it is connected. A receiver designed to be optimum under the assumption of equally likely messages is called a *maximum likelihood receiver*. (See also the discussion following Eq. 4.9.)

Minimax receivers. The foregoing discussion provides a powerful argument in support of a design assumption that all a priori message probabilities are equal. Even more cogently, with completely symmetric signals this assumption leads to a receiver design that is *minimax*, a term we now define.

For a fixed transmitter and channel, the probability of error depends only on the receiver and the message probabilities. For a given receiver (with transmitter and channel fixed) the probability of error depends only on the message source statistics and reaches a maximum value for some choice of these statistics. This maximum value of the $P[\mathcal{E}]$ is a useful criterion of goodness for the receiver in the absence of a priori knowledge of the $\{P[m_i]\}$: it represents a guaranteed minimum performance level beneath which the system will never operate, regardless of the statistics of the message source to which it may be connected. With this criterion, the receiver with the *smallest maximum* $P[\mathcal{E}]$ is most desirable. It is called the *minimax receiver*.

The argument that the maximum likelihood receiver is minimax when the $\{\mathbf{s}_i\}$ are completely symmetric is very simple. First, this receiver yields a probability of error that is independent of the actual $\{P[m_i]\}$ with which it may be used. Second, by the definition of optimum, any other receiver yields a greater probability of error when used with equally likely signals, hence must have a larger maximum. This concludes the proof.

Union Bound on the Probability of Error

An approximation to the $P[\mathcal{E} \mid m_i]$ for any set of M equally likely signals $\{\mathbf{s}_i\}$ in white Gaussian noise is obtained by noting that an error occurs when $\mathbf{s}_i$ is transmitted if and only if the received vector $\mathbf{r}$ is closer to at least one signal $\mathbf{s}_k$, $k \neq i$, than it is to $\mathbf{s}_i$. If $\mathcal{E}_{ik}$ is used to denote the event that $\mathbf{r}$ is closer to $\mathbf{s}_k$ than to $\mathbf{s}_i$ when $\mathbf{s}_i$ is transmitted, we have

$$P[\mathcal{E} \mid m_i] = P[\mathcal{E}_{i0} \cup \mathcal{E}_{i1} \cup \cdots \cup \mathcal{E}_{i,i-1} \cup \mathcal{E}_{i,i+1} \cup \cdots \cup \mathcal{E}_{i,M-1}]. \quad (4.107)$$

From Eq. 2.10 the probability of a finite union of events is bounded above by the sum of the probabilities of the constituent events, a result made geometrically evident in Fig. 4.41. Thus

$$P[\mathcal{E} \mid m_i] \leqslant \sum_{\substack{k=0 \\ (k \neq i)}}^{M-1} P[\mathcal{E}_{ik}]. \quad (4.108)$$

Note that $P[\mathcal{E}_{ik}]$ is not in general equal to $P[\hat{m} = m_k \mid m_i]$, because the latter is the probability that $\mathbf{r} = \mathbf{s}_i + \mathbf{n}$ is closer to $\mathbf{s}_k$ than to *every* other

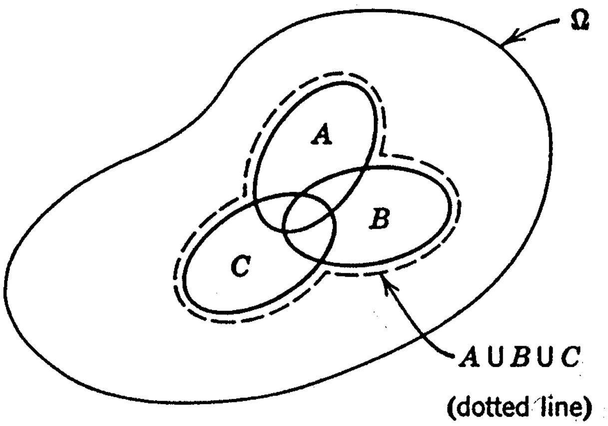


Figure 4.41 Venn diagram. It is apparent that $P[A \cup B \cup C] \leqslant P[A] + P[B] + P[C]$.

signal vector. To emphasize that $P[\mathcal{E}_{ik}]$ depends only on two vectors, $\mathbf{s}_i$ and $\mathbf{s}_k$, hereafter we write $P_2[\mathbf{s}_i, \mathbf{s}_k]$ in place of $P[\mathcal{E}_{ik}]$. Equation 4.108 then becomes

$$P[\mathcal{E} \mid m_i] \leqslant \sum_{\substack{k=0 \\ (k \neq i)}}^{M-1} P_2[\mathbf{s}_i, \mathbf{s}_k]. \quad (4.109)$$

We next observe that $P_2[\mathbf{s}_i, \mathbf{s}_k]$ is just the probability of error for a system that uses the vectors $\mathbf{s}_i$ and $\mathbf{s}_k$ as signals to communicate one of two equally likely messages. The bound of Eq. 4.109, and this interpretation of $P_2[\mathbf{s}_i, \mathbf{s}_k]$, holds for channels more general than that of additive Gaussian noise. For the Gaussian channel, however, the expression for $P_2[\mathbf{s}_i, \mathbf{s}_k]$ is particularly simple; from Eq. 4.76b, we have

$$P_2[\mathbf{s}_i, \mathbf{s}_k] = Q\left(\frac{|\mathbf{s}_i - \mathbf{s}_k|}{\sqrt{2\mathcal{N}_0}}\right). \quad (4.110)$$

The union bound of Eq. 4.109 is especially useful when the signal set $\{\mathbf{s}_i\}$ is completely symmetric, for in this case the unconditioned error probability P[$\mathcal{E}$] equals P[$\mathcal{E} \mid m_i$] and most of the terms $\{P_2[\mathbf{s}_i, \mathbf{s}_k]\}$ are identical. The following examples illustrate the application of the bound.

Orthogonal Signals:

$$P[\mathcal{E}] = P[\mathcal{E} \mid m_i] \leqslant (M - 1)Q(\sqrt{E_s/\mathcal{N}_0}). \tag{4.111}$$

Biorthogonal Signals:

$$P[\mathcal{E}] = P[\mathcal{E} \mid m_i] \leqslant (M - 2)Q(\sqrt{E_s/\mathcal{N}_0}) + Q(\sqrt{2E_s/\mathcal{N}_0}). \tag{4.112}$$

In many instances the union bound is a useful approximation to the actual P[$\mathcal{E}$]. It becomes increasingly tight for fixed M as $E_s/\mathcal{N}_0$ is increased.

APPENDIX 4A ORTHONORMAL EXPANSIONS AND VECTOR REPRESENTATIONS

When one of M signals $\{s_i(t)\}$ is communicated over an additive white Gaussian noise channel, the vector receiver to which the optimum waveform receiver reduces does not depend on the specific waveshapes of the N orthonormal base functions $\{\varphi_j(t)\}$. Only the vectors $\{\mathbf{s}_i\}$ are important; the particular set $\{\varphi_j(t)\}$ used to generate the signals $\{s_i(t)\}$ has no effect on the decision rule (Eq. 4.53), hence on the receiver error probability. In the design of communication systems for use in white Gaussian noise, the problem is to choose a good set of vectors $\{\mathbf{s}_i\}$ and a convenient set of functions $\{\varphi_j(t)\}$ that will propagate satisfactorily over the channel.

To prove that the transmitter structure of Fig. 4.12 and the correlation and matched filter receivers of Figs. 4.18 and 4.19 are completely general, we must show that *any* set of M finite-energy waveforms can always be expressed as

$$s_i(t) = \sum_{j=1}^{N} s_{ij}\, \varphi_j(t); \qquad i = 0, 1, \ldots, M - 1, \tag{4A.1a}$$

in which the waveforms $\{\varphi_j(t)\}$ are an appropriately chosen set of orthonormal functions:

$$\int_{-\infty}^{\infty} \varphi_l(t)\, \varphi_j(t)\, dt = \delta_{lj}; \qquad 1 \leqslant l, j \leqslant N. \tag{4A.1b}$$

In this appendix we prove the generality of Eq. 4A.1 and discuss some of its implications.

The Gram-Schmidt orthogonalization procedure. One convenient way in which an appropriate orthonormal set $\{\varphi_j(t)\}$ can be obtained from any given signal set $\{s_i(t)\}$ is by the Gram-Schmidt[43] orthogonalization procedure described in the following sequence of steps.

1. First consider $s_0(t)$. If $s_0(t) \equiv 0$ (has zero energy), renumber the signals. For $s_0(t) \not\equiv 0$, set

$$\varphi_1(t) = \frac{s_0(t)}{\sqrt{E_0}}, \tag{4A.2a}$$

where

$$E_0 \triangleq \int_{-\infty}^{\infty} s_0^2(t)\, dt. \tag{4A.2b}$$

Then $\varphi_1(t)$ is a waveform with unit energy. Since $s_0(t) = \sqrt{E_0}\, \varphi_1(t)$, the coefficient $s_{01} = \sqrt{E_0}$. The associated vector $\mathbf{s}_0$ is shown in Fig. 4A.1*a*.

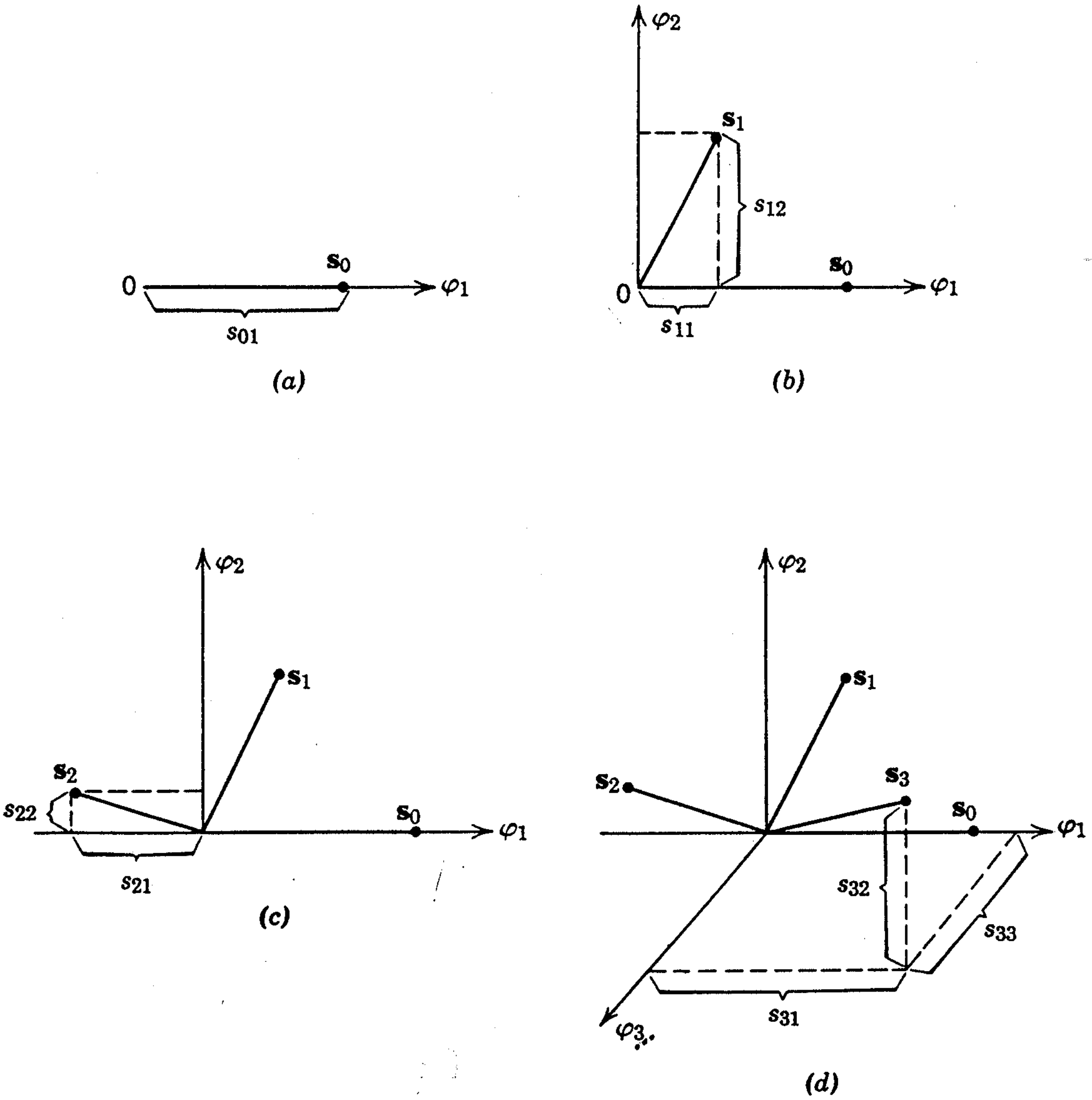


Figure 4A.1 Vectors obtained by the Gram-Schmidt procedure: $M = 4$, $N = 3$. Here $\mathbf{s}_2$ can be expressed as a linear combination of φ_1 and φ_2, so that $\theta_2(t) \equiv 0$.

2. Second, define the auxiliary function $\theta_1(t)$ as

$$\theta_1(t) = s_1(t) - s_{11}\,\varphi_1(t), \tag{4A.3a}$$

where

$$s_{11} \triangleq \int_{-\infty}^{\infty} s_1(t)\,\varphi_1(t)\,dt. \tag{4A.3b}$$

If $\theta_1(t) \not\equiv 0$, set

$$\varphi_2(t) = \frac{\theta_1(t)}{\sqrt{E_{\theta_1}}}, \tag{4A.3c}$$

where

$$E_{\theta_1} \triangleq \int_{-\infty}^{\infty} \theta_1^{\,2}(t)\,dt. \tag{4A.3d}$$

Then $\varphi_2(t)$ also has unit energy, and $s_{12} = \sqrt{E_{\theta_1}}$. Furthermore,

$$\int_{-\infty}^{\infty} \varphi_2(t)\,\varphi_1(t)\,dt = 0, \tag{4A.3e}$$

which follows from the equations

$$\begin{aligned}
\sqrt{E_{\theta_1}} \int_{-\infty}^{\infty} \varphi_2(t)\,\varphi_1(t)\,dt &= \int_{-\infty}^{\infty} \theta_1(t)\,\varphi_1(t)\,dt \\
&= \int_{-\infty}^{\infty} [s_1(t) - s_{11}\,\varphi_1(t)]\varphi_1(t)\,dt \\
&= \int_{-\infty}^{\infty} s_1(t)\,\varphi_1(t)\,dt - s_{11}\int_{-\infty}^{\infty} \varphi_1^{\,2}(t)\,dt \\
&= s_{11} - s_{11} = 0.
\end{aligned}$$

The vector $\mathbf{s}_1$ is shown in Fig. 4A.1*b* under the assumption that $\theta_1(t) \not\equiv 0$. If $\theta_1(t) \equiv 0$, proceed to (3).

3. The general step in the procedure is as follows. Assume that $(l - 1)$ orthonormal waveforms $\varphi_1(t), \varphi_2(t), \ldots, \varphi_{l-1}(t)$ have been defined through the use of $s_0(t), s_1(t), \ldots, s_{k-1}(t)$. It is clear that $(l - 1) \leqslant k$, since each new signal introduces at most one new orthonormal function. Now consider $s_k(t)$ and define the auxiliary function

$$\theta_k(t) = s_k(t) - \sum_{j=1}^{l-1} s_{kj}\,\varphi_j(t), \tag{4A.4a}$$

where

$$s_{kj} \triangleq \int_{-\infty}^{\infty} s_k(t)\,\varphi_j(t)\,dt; \qquad j = 1, 2, \ldots, l - 1. \tag{4A.4b}$$

If $\theta_k(t) \not\equiv 0$, set

$$\varphi_l(t) = \frac{\theta_k(t)}{\sqrt{E_{\theta_k}}}, \tag{4A.4c}$$

where

$$E_{\theta_k} \triangleq \int_{-\infty}^{\infty} \theta_k^2(t)\, dt. \tag{4A.4d}$$

Clearly, $\varphi_l(t)$ has unit energy, and $s_{kl} = \sqrt{E_{\theta_k}}$. Also,

$$\int_{-\infty}^{\infty} \varphi_l(t)\, \varphi_m(t)\, dt = 0; \qquad \text{for } 1 \leqslant m \leqslant l-1, \tag{4A.4e}$$

which follows from the equations

$$\begin{aligned}
\sqrt{E_{\theta_k}} \int_{-\infty}^{\infty} \varphi_k(t)\, \varphi_m(t)\, dt &= \int_{-\infty}^{\infty} \theta_k(t)\, \varphi_m(t)\, dt \\
&= \int_{-\infty}^{\infty} \left[s_k(t) - \sum_{j=1}^{l-1} s_{kj}\, \varphi_j(t) \right] \varphi_m(t)\, dt \\
&= \int_{-\infty}^{\infty} s_k(t)\, \varphi_m(t)\, dt - \sum_{j=1}^{l-1} s_{kj} \int_{-\infty}^{\infty} \varphi_j(t)\, \varphi_m(t)\, dt \\
&= s_{km} - \sum_{j=1}^{l-1} s_{kj}\, \delta_{jm} \\
&= s_{km} - s_{km} = 0; \qquad 1 \leqslant m \leqslant l-1.
\end{aligned}$$

The foregoing procedure can be continued until all M signals $\{s_i(t)\}$ have been exhausted, as shown in Figs. 4A.1*c*, *d*. There will then have been established $N \leqslant M$ orthonormal waveforms $\{\varphi_j(t)\}$ with the equality holding if and only if all M signals are *linearly independent*—that is, if and only if no one signal can be expressed as a linear combination of the others. The integer N is called the *dimensionality* of the signal space defined by the $\{s_i(t)\}$. By the nature of the construction, it is clear that each $s_i(t)$, $i = 0, 1, \ldots, M-1$, can indeed be expressed as a linear combination of the $\{\varphi_j(t)\}$ and thus that Eq. 4A.1 is satisfied.

A simple example of the Gram-Schmidt procedure is provided by the four waveforms shown in Fig. 4A.2. Starting with $s_0(t)$, we have

$$E_0 = 4 + 4 + 4 = 12,$$

and

$$\varphi_1(t) = \frac{s_0(t)}{\sqrt{E_0}} = \frac{s_0(t)}{\sqrt{12}}, \qquad s_{01} = \sqrt{12}.$$

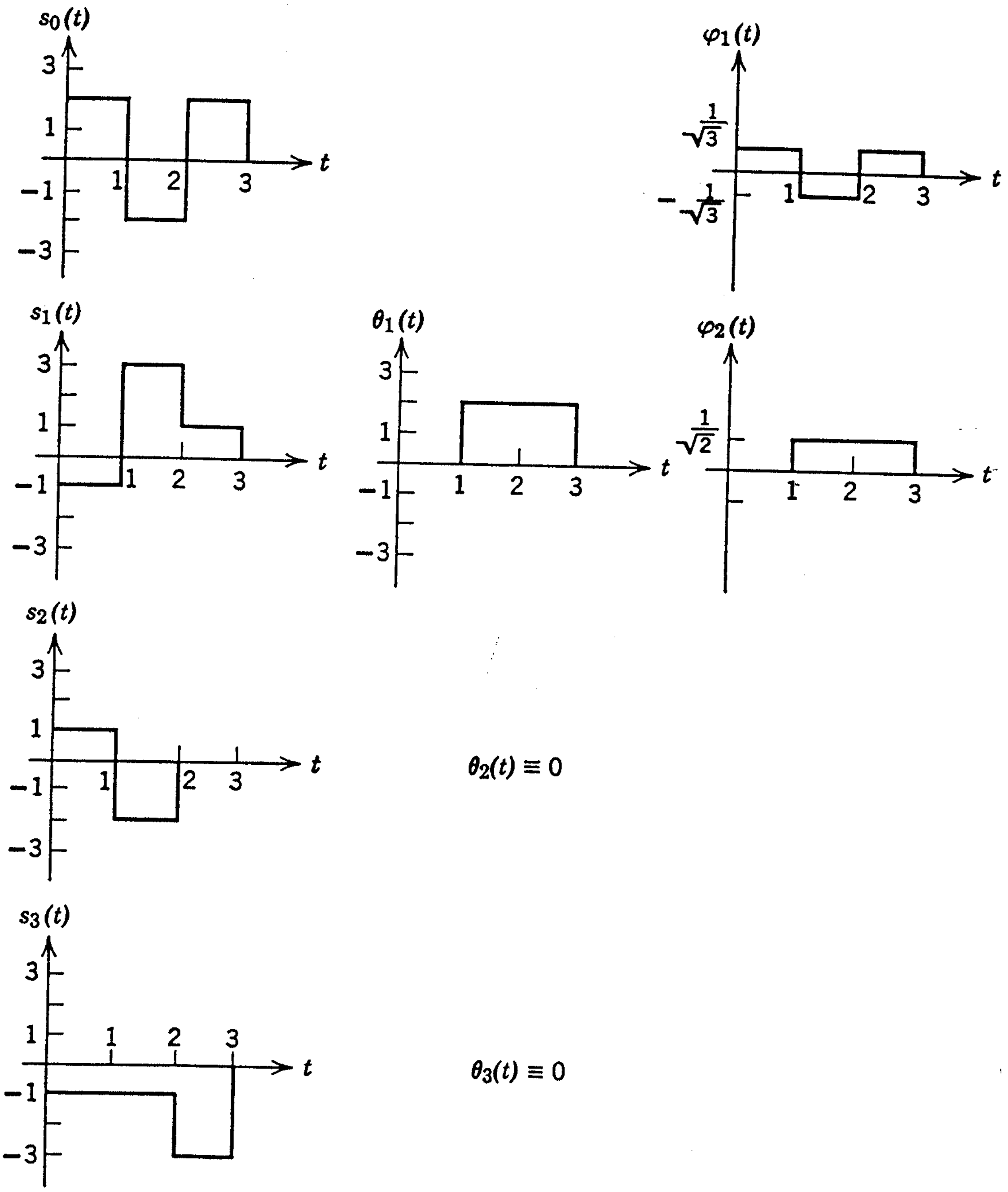


Figure 4A.2 An illustration of the Gram-Schmidt procedure.

Next, introducing $s_1(t)$, we have

$$s_{11} = \int_{-\infty}^{\infty} s_1(t)\, \varphi_1(t)\, dt = \sqrt{\tfrac{1}{3}}(-1 - 3 + 1) = -\sqrt{3}$$

$$\theta_1(t) = s_1(t) + \sqrt{3}\, \varphi_1(t)$$

$$E_{\theta_1} = 8, \qquad s_{12} = \sqrt{8}$$

$$\varphi_2(t) = \frac{1}{\sqrt{8}}\, \theta_1(t).$$

Introducing $s_2(t)$, we obtain

$$s_{21} = \int_{-\infty}^{\infty} s_2(t)\, \varphi_1(t)\, dt = \sqrt{3},$$

$$s_{22} = \int_{-\infty}^{\infty} s_2(t)\, \varphi_2(t)\, dt = -\sqrt{2},$$

$$\theta_2(t) = s_2(t) - \sqrt{3}\, \varphi_1(t) + \sqrt{2}\, \varphi_2(t) \equiv 0.$$

Finally, introducing $s_3(t)$, we have

$$s_{31} = \int_{-\infty}^{\infty} s_3(t)\, \varphi_1(t)\, dt = -\sqrt{3},$$

$$s_{32} = \int_{-\infty}^{\infty} s_3(t)\, \varphi_2(t)\, dt = -2\sqrt{2},$$

$$\theta_3(t) = s_3(t) + \sqrt{3}\, \varphi_1(t) + 2\sqrt{2}\, \varphi_2(t) \equiv 0.$$

Thus the four signals $\{s_i(t)\}$ span a space of two dimensions, and the vector representations are

$$\begin{aligned} s_0(t) &= \sqrt{12}\, \varphi_1(t) & \mathbf{s}_0 &= (\sqrt{12}, 0), \\ s_1(t) &= -\sqrt{3}\, \varphi_1(t) + \sqrt{8}\, \varphi_2(t) & \mathbf{s}_1 &= (-\sqrt{3}, \sqrt{8}), \\ s_2(t) &= +\sqrt{3}\, \varphi_1(t) - \sqrt{2}\, \varphi_2(t) & \mathbf{s}_2 &= (\sqrt{3}, -\sqrt{2}), \\ s_3(t) &= -\sqrt{3}\, \varphi_1(t) - \sqrt{8}\, \varphi_2(t) & \mathbf{s}_3 &= (-\sqrt{3}, -\sqrt{8}), \end{aligned} \tag{4A.5}$$

as shown in Fig. 4A.3.

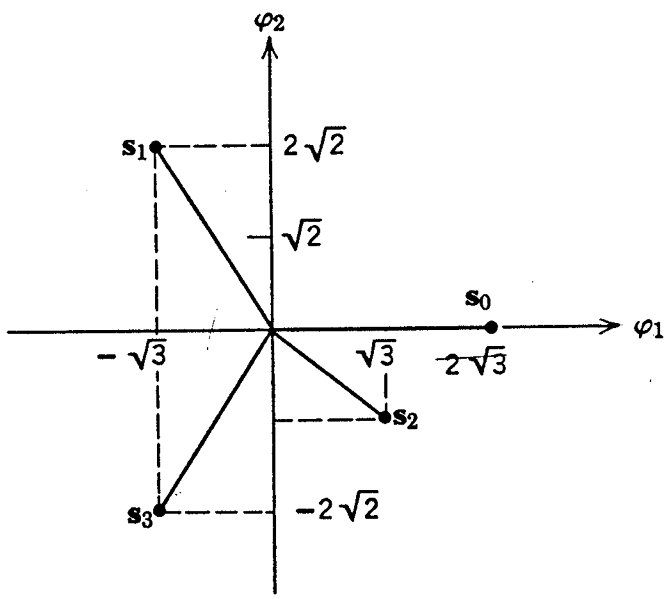


Figure 4A.3 A vector representation of the $\{s_i(t)\}$ of Fig. 4A.2.

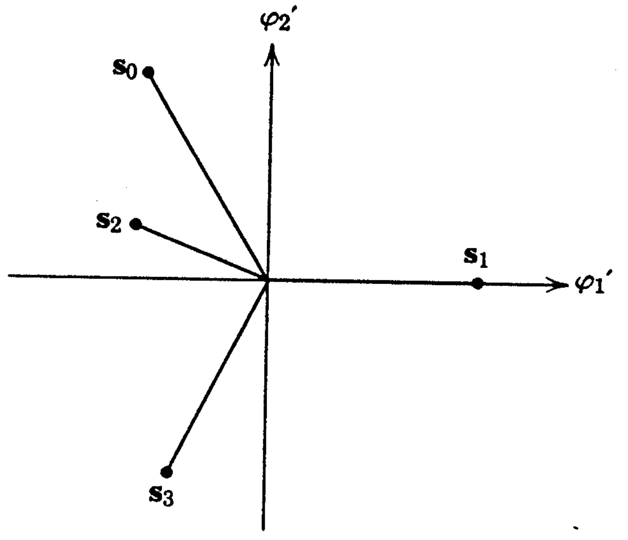


Figure 4A.4 An alternative vector diagram for the $\{s_i(t)\}$ of Fig. 4A.2.

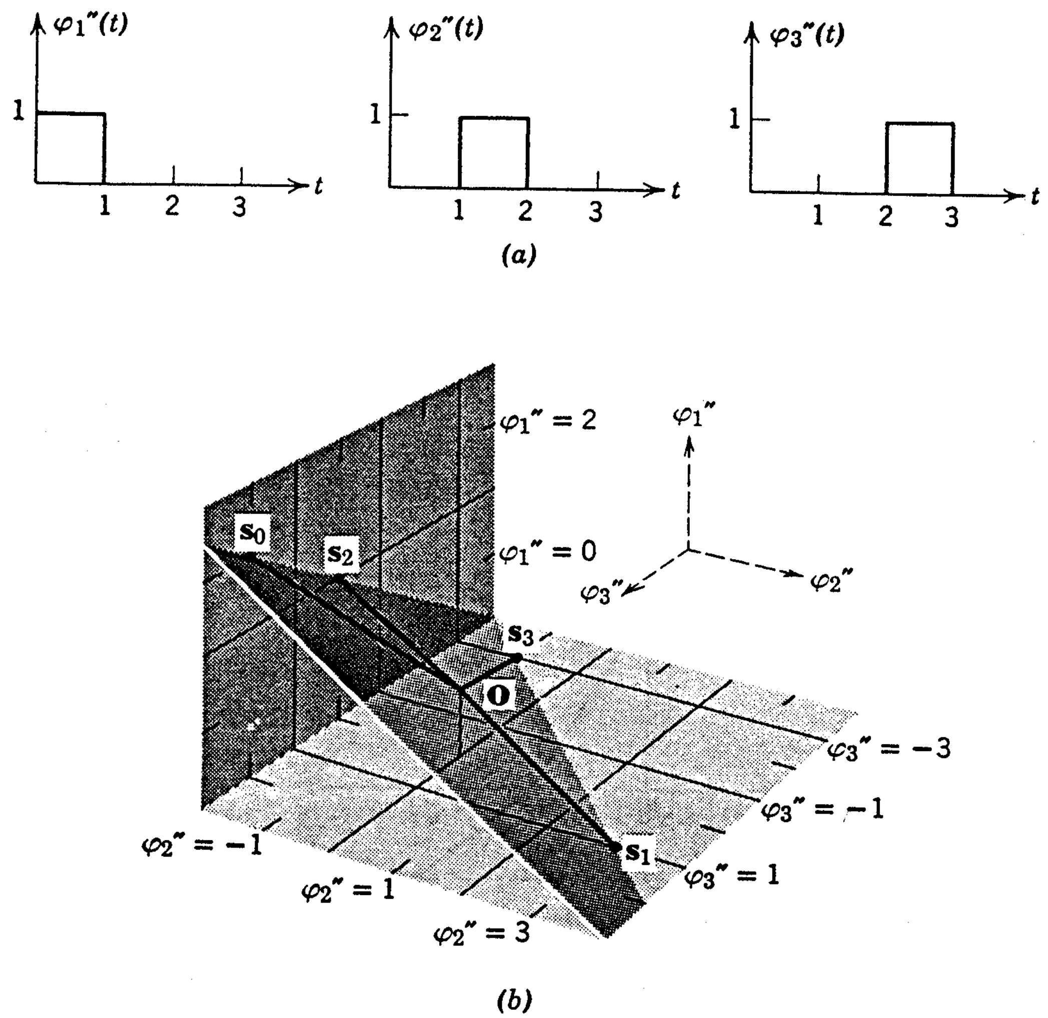


Figure 4A.5 A third vector representation of the signals of Fig. 4A.2

We have shown that it is always possible to represent a finite set of signals $\{s_i(t)\}$ by means of at least one finite weighted sum of orthonormal functions $\{\varphi_j(t)\}$ and therefore that the derivation of the optimum receivers of Figs. 4.18 and 4.19 is always valid.

Note that any given set $\{s_i(t)\}$ can be expanded in many different orthonormal sets, all of which ultimately yield the same receiver, hence the same decisions and the same probability of error. For example, if the Gram-Schmidt procedure for the signals of Fig. 4A.2 were carried out by considering signals in the order $s_1(t)$, $s_2(t)$, $s_3(t)$, $s_0(t)$, a different pair of orthonormal functions $\varphi_1'(t)$, $\varphi_2'(t)$, and a different set of coefficients $\{s_{ij}'\}$ would have been obtained. In particular, $\mathbf{s}_1$ would lie on the φ_1'-axis and $\mathbf{s}_2$ would have a positive projection on the φ_2'-axis, as shown in Fig. 4A.4. Alternatively, a set $\{\varphi_j''(t)\}$ might be obtained without use of the Gram-Schmidt procedure, although the resulting number of functions might be larger than the dimensionality, N. Such a set is shown in Fig. 4A.5*a* and the corresponding vectors in Fig. 4A.5*b*. Note that the four signal points remain coplanar and have the same relative positions. The important fact is that the signal points $\{\mathbf{s}_i\}$ always retain the same geometrical configuration, regardless of the particular set of coordinates in terms of which they are described.

PROBLEMS

4.1 The random variable n in Fig. P4.1*a* is Gaussian, with zero mean. If one of two equally likely messages is transmitted, using the signals of Fig. P4.1*b*, an optimum receiver yields $P[\mathcal{E}] = 0.01$.

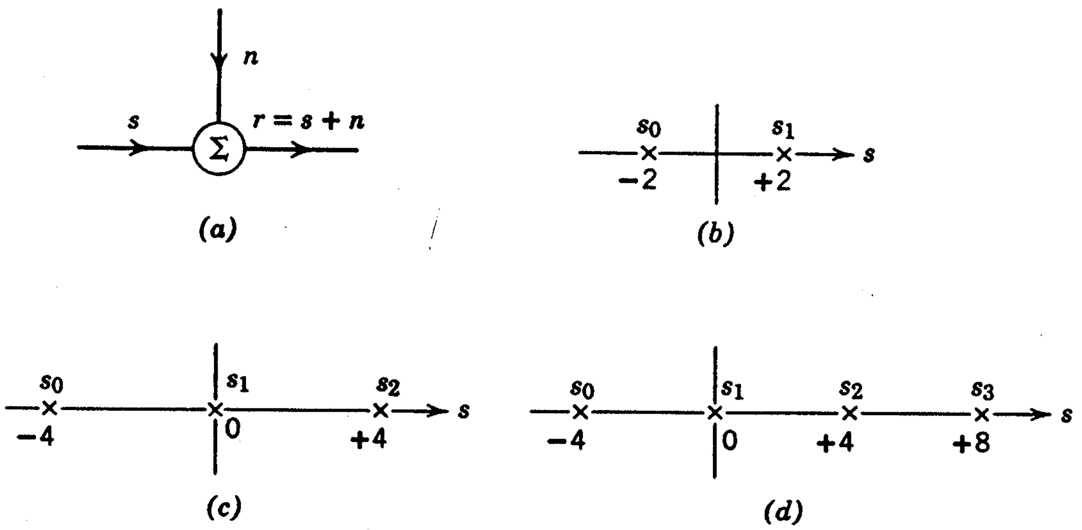


Figure P4.1

a. What is the minimum attainable probability of error, $P[\mathcal{E}]_{min}$, when the channel of Fig. P4.1*a* is used with three equally likely messages and the signals of (*c*)? With four equally likely messages and the signals of (*d*)?

b. How do the answers to part (a) change if it is known that $\bar{n} = 1$ rather than 0?

4.2 One of four equally likely messages is to be communicated over a vector channel which adds a (different) statistically independent zero-mean Gaussian random variable with variance $\mathcal{N}_0/2$ to each transmitted vector component. Assume that the transmitter uses the signal vectors shown in Fig. P4.2 and express the $P[\mathcal{E}]$ produced by an optimum receiver in terms of the function $Q(\alpha)$.

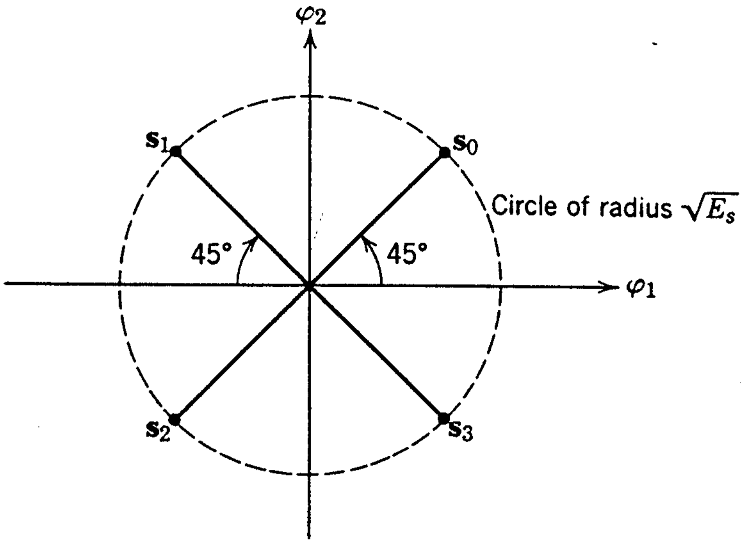


Figure P4.2

4.3 It is known that $P[\mathcal{E}]_{min} = q$ when the two signal vectors s_0 and s_1 shown in Fig. P4.3*a* are transmitted with equal probability over a channel disturbed by additive white Gaussian noise. Compute $P[\mathcal{E}]_{min}$ in terms of q, θ, and l when the nine vectors indicated by ×'s in Fig. P4.3*b* are used as signals with equal probability over the same channel.

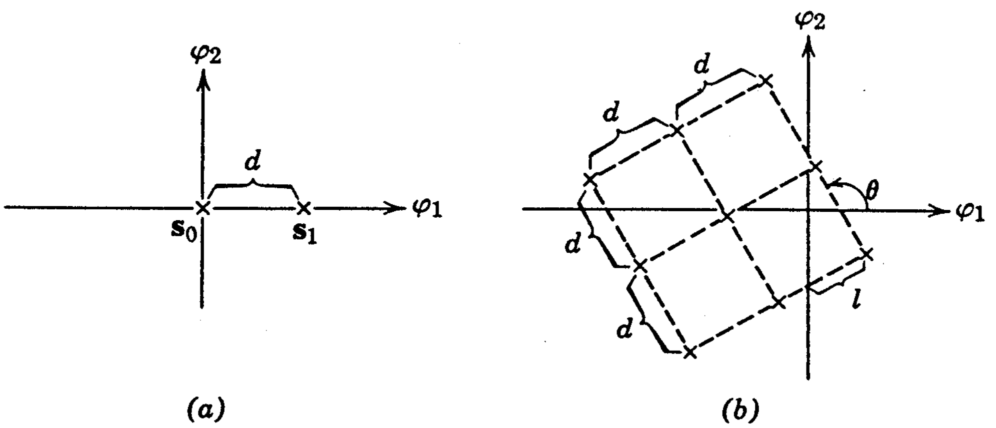


Figure P4.3

4.4 One of 16 equally likely messages is to be communicated over an additive Gaussian noise channel with $\mathcal{S}_n(f) = \mathcal{N}_0/2$. The transmitter utilizes a signal set $\{s_i(t)\}$ whose vector representation is indicated by $\times$'s in Fig. P4.4.

a. Draw the optimum decision regions.

b. Determine $P[\mathcal{E}]_{\min}$ in terms of $Q(\alpha)$.

c. Find a set of 16 two-dimensional signal vectors (not necessarily optimum) such that the transmitted energy is never greater than E_s but for which the attainable $P[\mathcal{E}]$ is less than the answer to part (b).

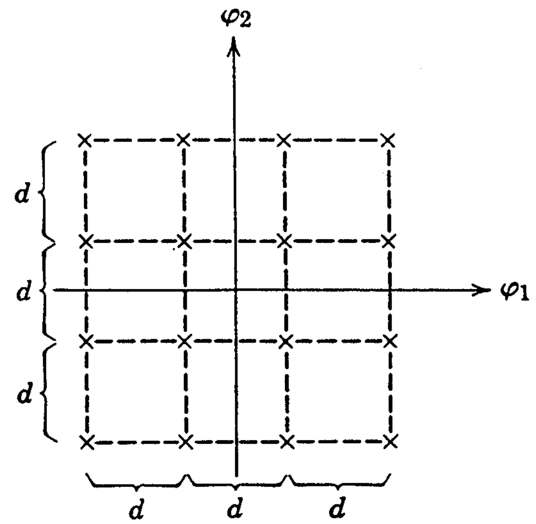


Figure P4.4

4.5 One of the two signals $s_0 = -1$, $s_1 = +1$ is transmitted over the channel shown in Fig. P4.5*a*. The two noise random variables n_1 and n_2 are statistically independent of the transmitted signal and of each other. Their density functions are

$$p_{n_1}(\alpha) = p_{n_2}(\alpha) = \tfrac{1}{2}\, e^{-|\alpha|}.$$

a. Prove that the optimum decision regions for equally likely messages are as shown in Fig. P4.5*b*. *Hint.* Use geometric reasoning and the fact that $|\rho_1 - 1| + |\rho_2 - 1| = a + b$, as shown on the next page in Fig. P4.5*d*.

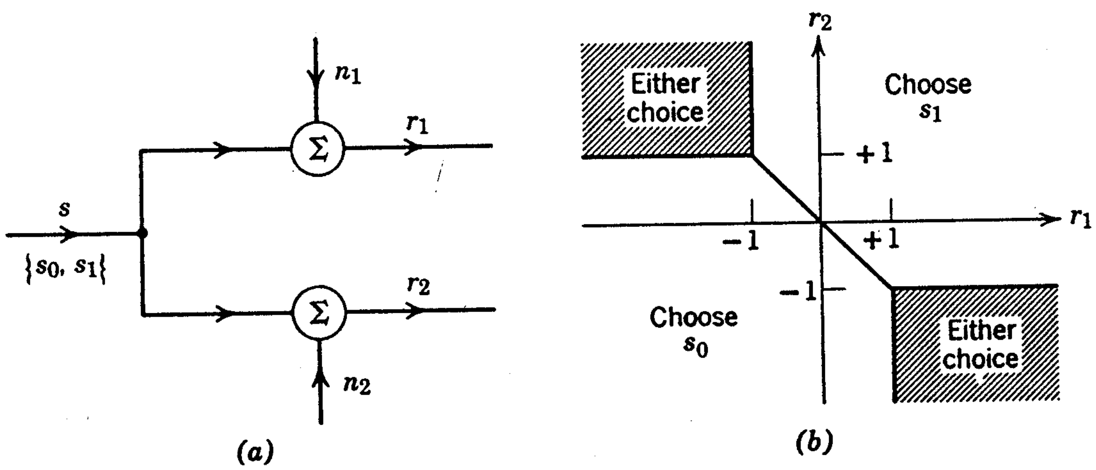
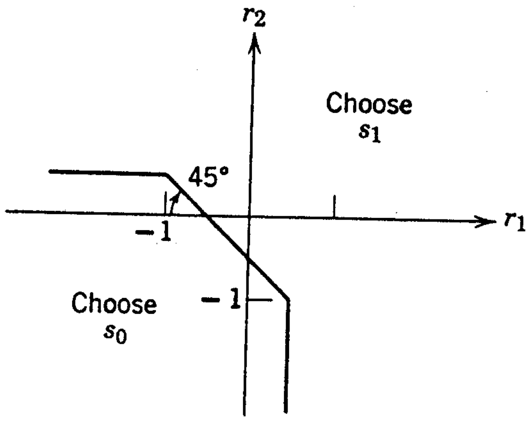
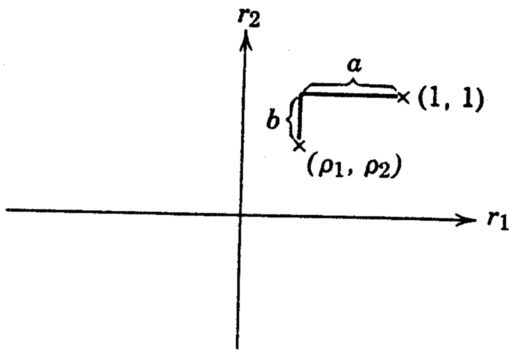


Figure P4.5



(c)



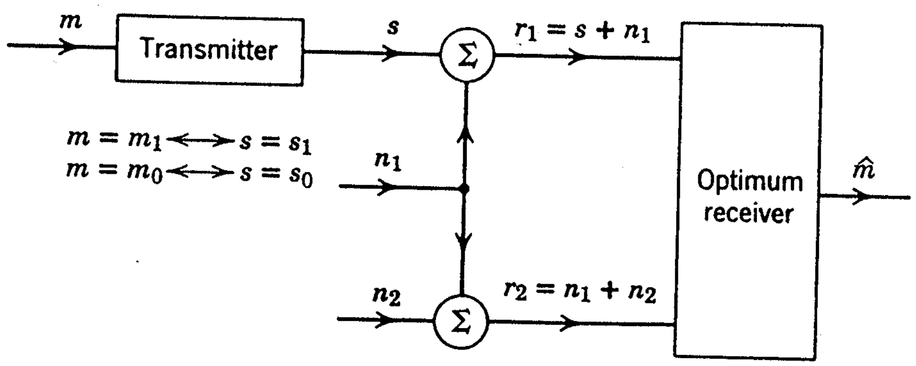
(d)

Figure P4.5 (*Continued*)

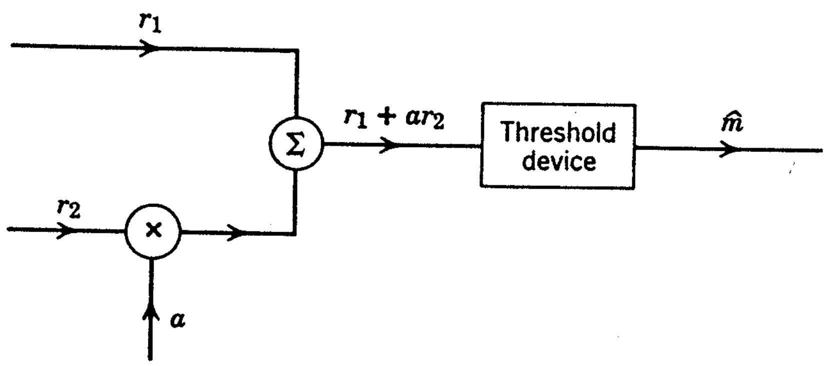
b. A receiver decides that s_1 was transmitted if and only if $(r_1 + r_2) > 0$. Is this receiver optimum for equally likely messages? What is its probability of error?

c. Prove that the optimum decision regions are modified as indicated in Fig. P4.5*c* when $P[s_1] > \frac{1}{2}$.

d. The channel may be discarded without affecting $P[\mathcal{E}]_{min}$ if $P[s_1] \geqslant q$. Evaluate q.



(a)



(b)

Figure P4.6

4.6 In the communication system diagrammed in Fig. P4.6*a*, the transmitted signal s and the noises n_1 and n_2 are all random voltages and all statistically independent. Assume that

$$P[m_0] = P[m_1] = \tfrac{1}{2},$$

$$s_1 = -s_0 = \sqrt{E_s},$$

$$p_{n_1}(\alpha) = p_{n_2}(\alpha) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\alpha^2/2\sigma^2}.$$

a. Show that the optimum receiver can be realized as diagrammed in Fig. P4.6*b*, where a is an appropriately chosen constant.

b. What is the optimum value of a?

c. What is the optimum threshold setting?

d. Express the resulting $P[\mathcal{E}]$ in terms of $Q(\alpha)$.

e. By what factor would E_s have to be increased to yield this same probability of error if the receiver were restricted to observing *only* r_1.

4.7 The voltage waveforms $x(t)$ and $y(t)$, plotted below, have the properties that when applied across a 1-ohm resistor

$$\int_0^T x^2(t)\,dt = \int_0^T y^2(t)\,dt = 16 \text{ joules.}$$

$$\int_0^T x(t)\,y(t)\,dt = 0.$$

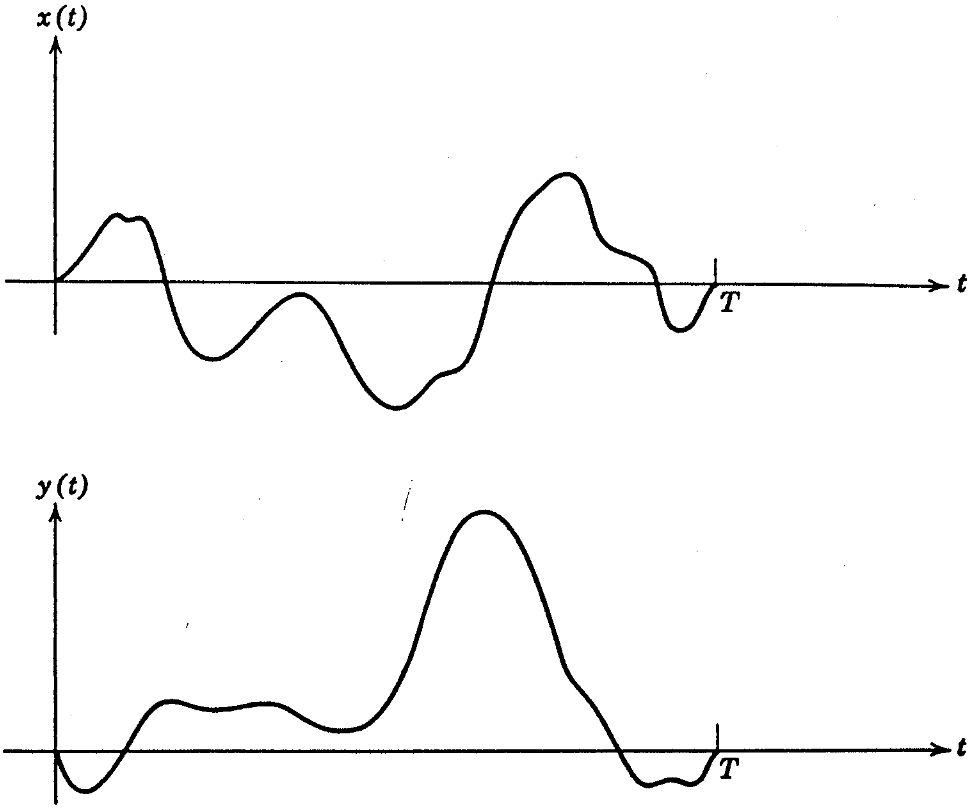


Figure P4.7

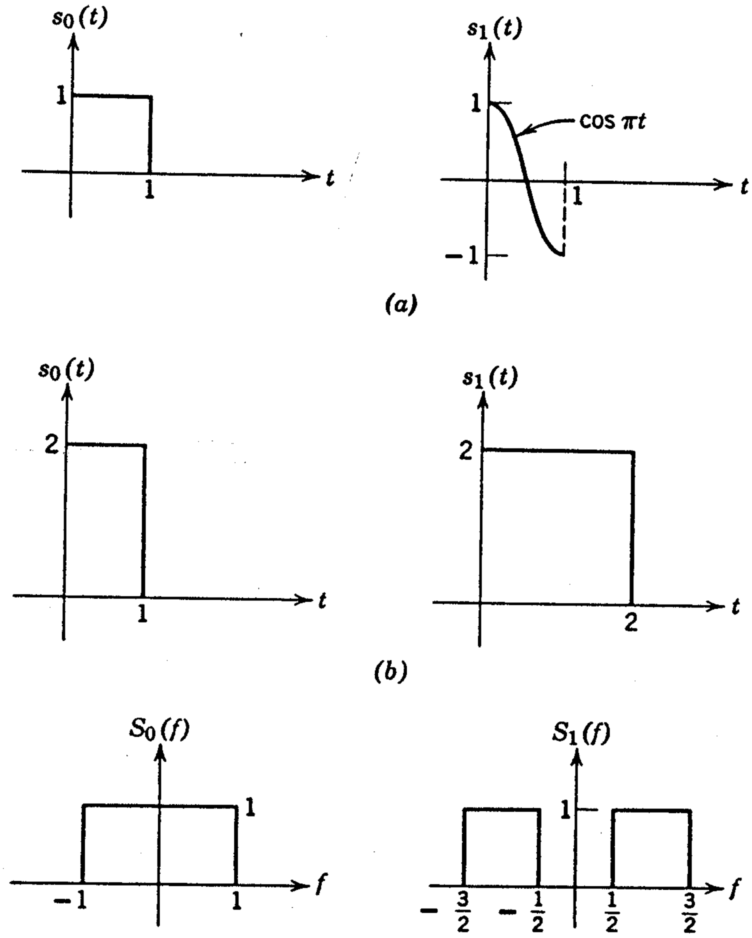
These signals can be used to communicate one of two equally likely messages over a channel perturbed by additive white Gaussian noise with power density of 4 watts/cycle/sec (on a bilateral frequency scale).

a. Calculate the minimum attainable probability of error when the two signals used are $x(t)$ and $-x(t)$.

b. Calculate the minimum attainable probability of error when the two signals used are $x(t)$ and $y(t)$.

4.8 a. Calculate $P[\mathcal{E}]_{\min}$ when the signal sets specified by Figs. P4.8*a*, *b*, and *c* are used to communicate one of two equally likely messages over a channel disturbed by additive Gaussian noise with $\mathcal{S}_n(f) = 0.15$.

b. Repeat part (a) for a priori message probabilities $(\frac{1}{4}, \frac{3}{4})$.



[$S_i(f)$, the Fourier transform of $s_i(t)$, is pure real.]

(*c*)

Figure P4.8

4.9 Express $P[\mathcal{E}]_{\min}$ in terms of $Q(\alpha)$ when the signal set shown in Fig. P4.9 is used to communicate one of eight equally likely messages over a channel disturbed by additive Gaussian noise with $\mathcal{S}_n(f) = \mathcal{N}_0/2$.

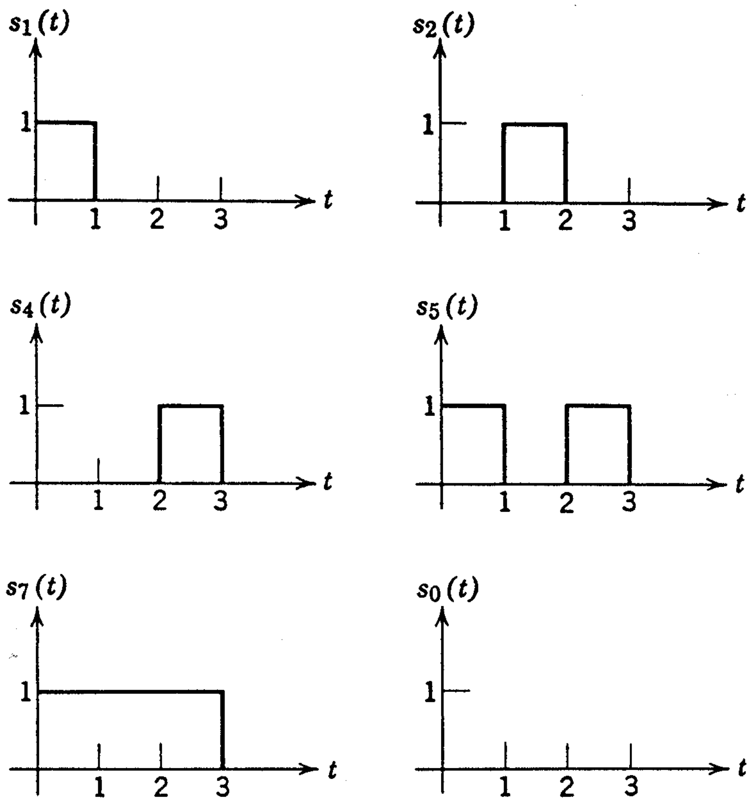


Figure P4.9

4.10 One of two equally likely messages is to be transmitted over an additive white Gaussian noise channel with $\mathcal{S}_n(f) = 0.05$ by means of binary pulse position modulation. Specifically,

$$s_0(t) = p(t),$$

$$s_1(t) = p(t - 2),$$

in which the pulse $p(t)$ is shown in Fig. P4.10.

a. What mathematical operations are performed by the optimum receiver?

b. What is the resulting probability of error?

c. Indicate two methods of implementing the receiver, each of which uses a single linear filter followed by a sampler and comparison device. Method I requires that two samples from the filter output be fed into the comparison device. Method II requires that just one sample be used. For each method

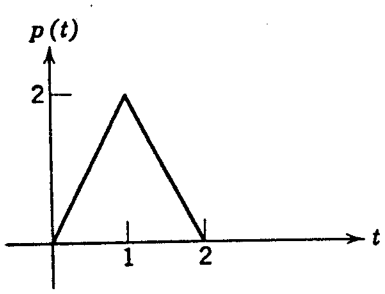


Figure P4.10

sketch the impulse response of the appropriate filter and its response to $p(t)$. Which of these methods is most easily extended to M-ary pulse position modulation, where $s_i(t) = p(t - 2i)$, $i = 0, 1, \ldots, M - 1$?

d. Suggest another pair of waveforms that require the same energy as the binary pulse-position waveforms and yield the same error probability; yield a lower error probability.

e. Calculate the minimum attainable probability of error if

$$s_0(t) = p(t) \quad \text{and} \quad s_1(t) = p(t - 1).$$

Repeat for

$$s_0(t) = p(t) \quad \text{and} \quad s_1(t) = -p(t - 1).$$

4.11 One of two equally likely messages, m_0 or m_1, is to be transmitted over an additive white Gaussian noise channel by means of the two signals

$$s_0(t) = \begin{cases} \sqrt{\dfrac{2E_s}{T}} \cos 2\pi f_1 t; & 0 \leqslant t \leqslant T \\ 0; & \text{elsewhere,} \end{cases}$$

$$s_1(t) = \begin{cases} \sqrt{\dfrac{2E_s}{T}} \cos 2\pi (f_1 + \Delta) t; & 0 \leqslant t \leqslant T \\ 0; & \text{elsewhere,} \end{cases}$$

where $T = 2$ msec, $f_1 = 1$ Mc, and $\Delta = 250$ cps. The noise has power density spectrum $\mathcal{N}_0/2$. If $E_s/\mathcal{N}_0 = 6$, calculate the probability of error to two significant digits. Repeat for $\Delta = 500$ cps.

4.12 M signals $s_0(t), s_1(t), \ldots, s_{M-1}(t)$ exist for $0 \leqslant t \leqslant T$, but each is identical to all others in the subinterval $[t_1, t_2]$, where $0 < t_1 < t_2 < T$.

a. Show that the optimum receiver may ignore this subinterval. Equivalently, show that if $\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_{M-1}$ all have the same projection in one dimension, then this dimension may be ignored. Assume an additive white Gaussian noise channel.

b. Does this result necessarily hold true if the noise is Gaussian but not white? Explain.

4.13 Consider the multipath communication model shown in Fig. P4.13a, for which $P[m_0] = \frac{1}{2}$. Assume that the three paths are characterized by the following parameters:

Constant attenuation	$\alpha_1 = 0.2$	$\alpha_2 = 0.4$	$\alpha_3 = 0.6.$
Constant delay	$\tau_1 = 1$ msec	$\tau_2 = 1.5$ msec	$\tau_3 = 2$ msec.
White noise power density	$\mathcal{S}_{n_1}(f) = 0.002$	$\mathcal{S}_{n_2}(f) = 0.006$	$\mathcal{S}_{n_3}(f) = 0.004.$

The three noise processes are Gaussian and statistically independent of each other and the signal transmitted. The transmitter is defined by the mapping

$$m = m_0 \Leftrightarrow s(t) = s_0(t) = \begin{cases} 5\cos 2\pi 10^3 t; & 0 \leqslant t \leqslant 3 \times 10^{-3} \\ 0; & \text{elsewhere.} \end{cases}$$

$$m = m_1 \Leftrightarrow s(t) = -s_0(t).$$

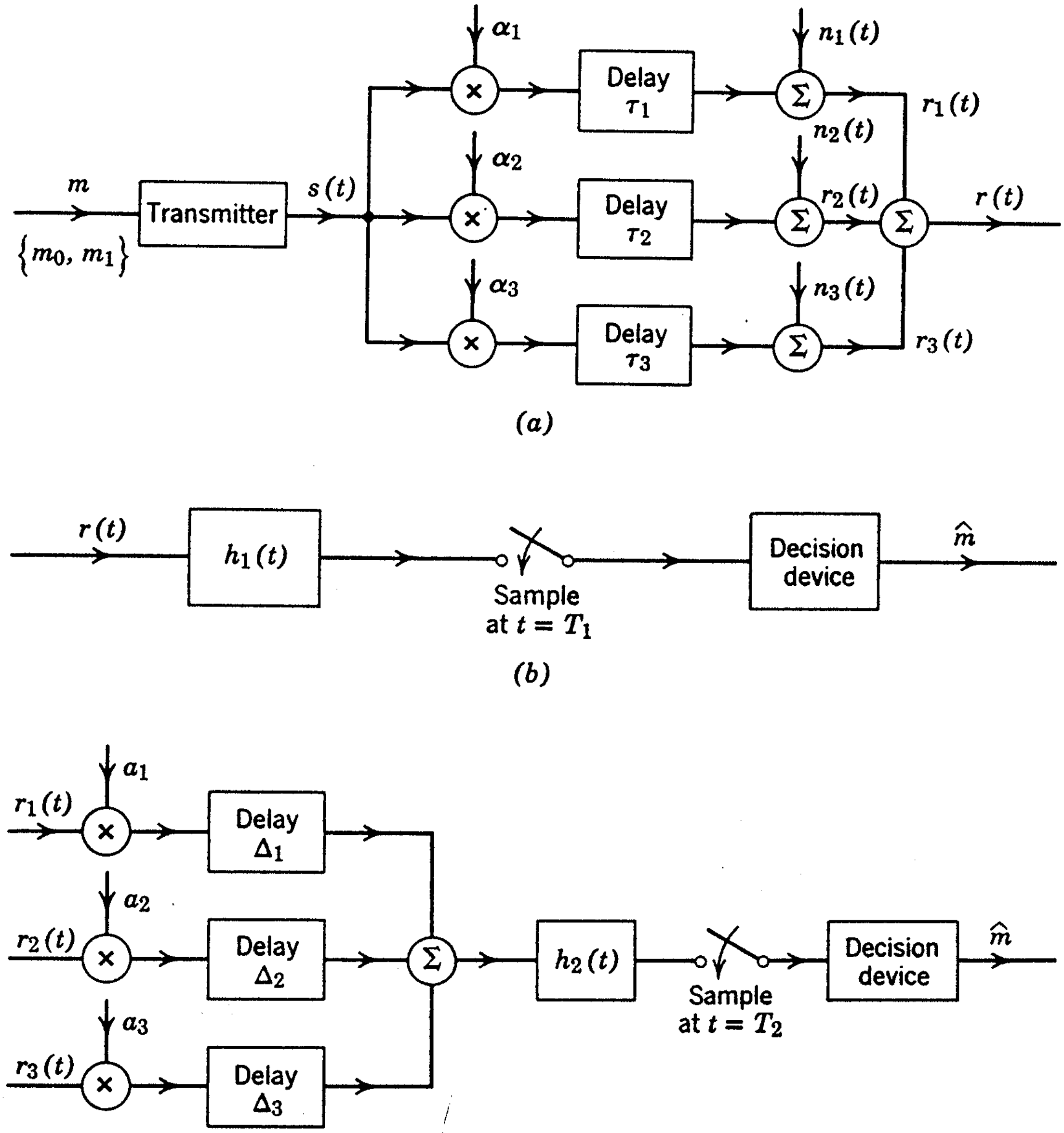


Figure P4.13

a. Show that the optimum receiver can be realized in the form illustrated in Fig. P4.13*b*. Determine $h_1(t)$, T_1, and the specification of the decision device. Suggest a reasonable implementation for $h_1(t)$. Calculate P[ℰ] to two significant digits.