Week 5 – Part I.

Lecture # 5

Signal-Space Analysis

For Digital Modulation Techniques

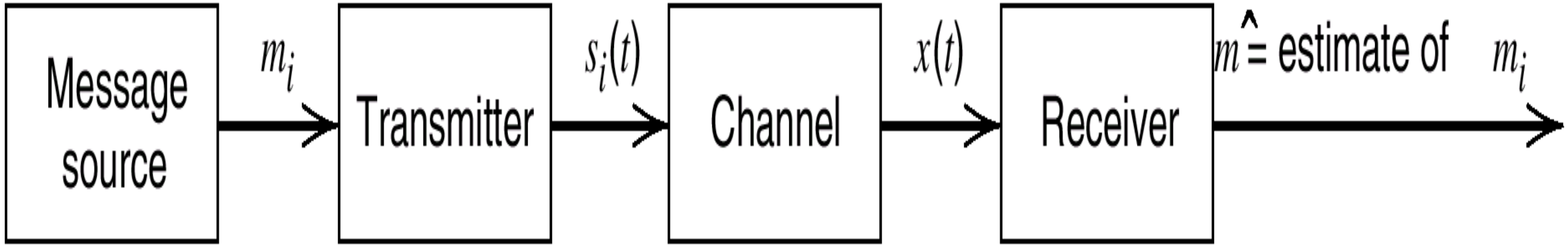


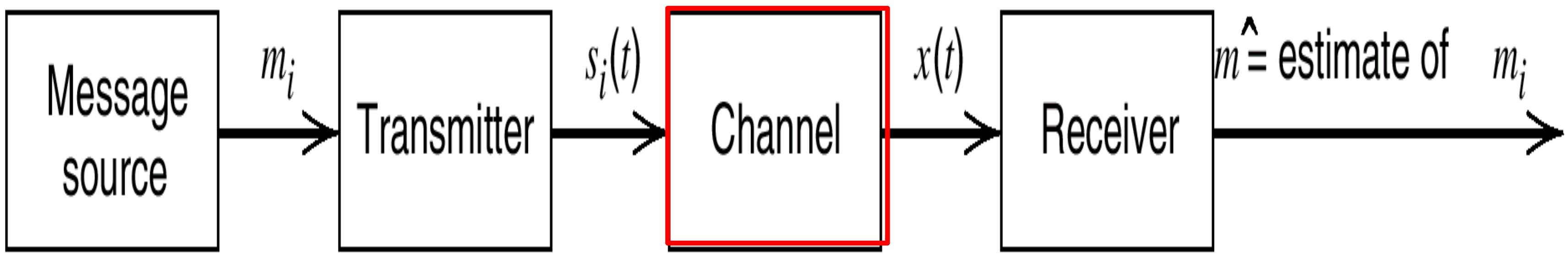
Figure 5.1 Block diagram of a generic digital communication system.

Consider the most basic form of a digital communication system depicted in Figure 5.1. A *message source* emits one *symbol* every T seconds, with the symbols belonging to an alphabet of M symbols denoted by $m_1, m_2, \ldots, m_M$. Consider, for example, the remote connection of two digital computers, with one computer acting as an information source that calculates digital outputs based on observations and inputs fed into it. The resulting computer output is expressed as a sequence of 0s and 1s, which are transmitted to a second computer over a communication channel. In this case, the alphabet consists simply of two binary symbols: 0 and 1. A second example is that of a quaternary PCM encoder with an alphabet consisting of four possible symbols: 00, 01, 10, and 11. In any event, the *a priori* probabilities $p_1, p_2, \ldots, p_M$ specify the message source output. In the absence of prior information, it is customary to assume that the M symbols of the alphabet are *equally likely*. Then we may express the probability that symbol m_i is emitted by the source as *likely*. Then we may express the probability that symbol m_i is emitted by the source as

$$p_i = P(m_i) = \frac{1}{M} \text{ for } i = 1, 2, \ldots, M \tag{5.1}$$

The transmitter takes the message source output m_i and codes it into a *distinct* signal $s_i(t)$ suitable for transmission over the channel. The signal $s_i(t)$ occupies the full duration T allotted to symbol m_i. Most important, $s_i(t)$ is a real-valued *energy signal* (i.e., a signal with finite energy), as shown by

$$E_i = \int_0^T s_i^2(t)\, dt, \qquad i = 1, 2, \ldots, M \tag{5.2}$$



The channel is assumed to have two characteristics:

1. The channel is *linear*, with a bandwidth that is wide enough to accommodate the transmission of signal $s_i(t)$ with negligible or no distortion.
2. The channel noise, $w(t)$, is the sample function of a *zero-mean white Gaussian noise process*. The reasons for this second assumption are that it makes receiver calculations tractable, and it is a reasonable description of the type of noise present in many practical communication systems.

We refer to such a channel as an *additive white Gaussian noise* (AWGN) *channel.* Accordingly, we may express the *received signal* $x(t)$ as

$$x(t) = s_i(t) + w(t), \qquad \begin{cases} 0 \le t \le T \\ i = 1, 2, \ldots, M \end{cases} \tag{5.3}$$

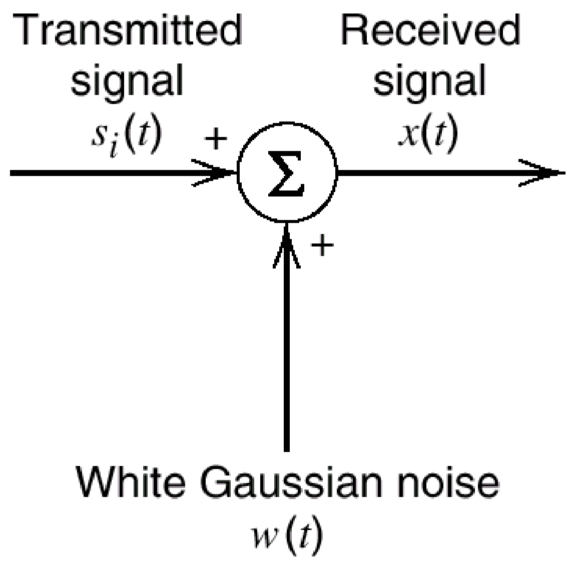
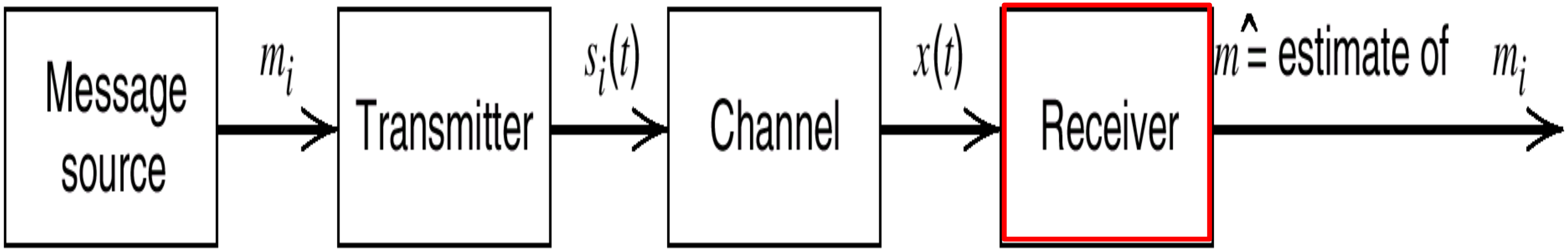


Figure 5.2 Additive white Gaussian noise (AWGN) model of a channel.



The receiver has the task of observing the received signal $x(t)$ for a duration of T seconds and making a best *estimate* of the transmitted signal $s_i(t)$ or, equivalently, the symbol m_i. However, owing to the presence of channel noise, this decision-making process is statistical in nature, with the result that the receiver will make occasional errors. The requirement is therefore to design the receiver so as to minimize the *average probability of symbol error*, defined as

$$P_e = \sum_{i=1}^{M} p_i \, P(\hat{m} \neq m_i \,|\, m_i) \tag{5.4}$$

where m_i is the transmitted symbol, $\hat{m}$ is the estimate produced by the receiver, and $P(\hat{m} \neq m_i \,|\, m_i)$ is the conditional error probability given that the ith symbol was sent. The resulting receiver is said to be *optimum in the minimum probability of error* sense.

This model provides a basis for the design of the optimum receiver, for which we will use geometric representation of the known set of transmitted signals, $\{s_i(t)\}$. This method, discussed in Section 5.2, provides a great deal of insight, with considerable simplification of detail.

Geometric Representation of Signals

Geometric Representation of Modulation Signals

The essence of *geometric representation of signals*[1] is to represent any set of M energy signals $\{s_i(t)\}$ as linear combinations of N *orthonormal basis functions*, where $N \leq M$. That is to say, given a set of real-valued energy signals $s_1(t), s_2(t), \ldots, s_M(t)$, each of duration T seconds, we write

$$s_i(t) = \sum_{j=1}^{N} s_{ij}\phi_j(t), \qquad \begin{cases} 0 \leq t \leq T \\ i = 1, 2, \ldots, M \end{cases} \tag{5.5}$$

where the coefficients of the expansion are defined by

$$s_{ij} = \int_0^T s_i(t)\phi_j(t)\, dt, \qquad \begin{cases} i = 1, 2, \ldots, M \\ j = 1, 2, \ldots, N \end{cases} \tag{5.6}$$

The real-valued basis functions $\phi_1(t), \phi_2(t), \ldots, \phi_N(t)$ are *orthonormal*, by which we mean

$$\int_0^T \phi_i(t)\phi_j(t)\, dt = \delta_{ij} = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases} \tag{5.7}$$

where δ_{ij} is the *Kronecker delta*. The first condition of Equation (5.7) states that each basis function is *normalized* to have unit energy. The second condition states that the basis functions $\phi_1(t), \phi_2(t), \ldots, \phi_N(t)$ are *orthogonal* with respect to each other over the interval $0 \leq t \leq T$.

Accordingly, we may state that each signal in the set $\{s_i(t)\}$ is completely determined by the *vector* of its coefficients

$$\mathbf{s}_i = \begin{bmatrix} s_{i1} \\ s_{i2} \\ \vdots \\ s_{iN} \end{bmatrix}, \qquad i = 1, 2, \ldots, M \tag{5.8}$$

The vector $\mathbf{s}_i$ is called a *signal vector.* Furthermore, if we conceptually extend our conventional notion of two- and three-dimensional Euclidean spaces to an *N-dimensional Euclidean space,* we may visualize the set of signal vectors $\{\mathbf{s}_i | i = 1, 2, \ldots, M\}$ as defining a corresponding set of M points in an N-dimensional Euclidean space, with N mutually perpendicular axes labeled $\phi_1, \phi_2, \ldots, \phi_N$. This N-dimensional Euclidean space is called the *signal space.*

$$\| \mathbf{s}_i \|^2 = \mathbf{s}_i^T \mathbf{s}_i = \sum_{j=1}^{N} s_{ij}^2, \qquad i = 1, 2, \ldots, M$$

$$E_i = \int_0^T s_i^2(t)\, dt$$

$$E_i = \int_0^T \left[\sum_{j=1}^{N} s_{ij}\phi_j(t) \right] \left[\sum_{k=1}^{N} s_{ik}\phi_k(t) \right] dt$$

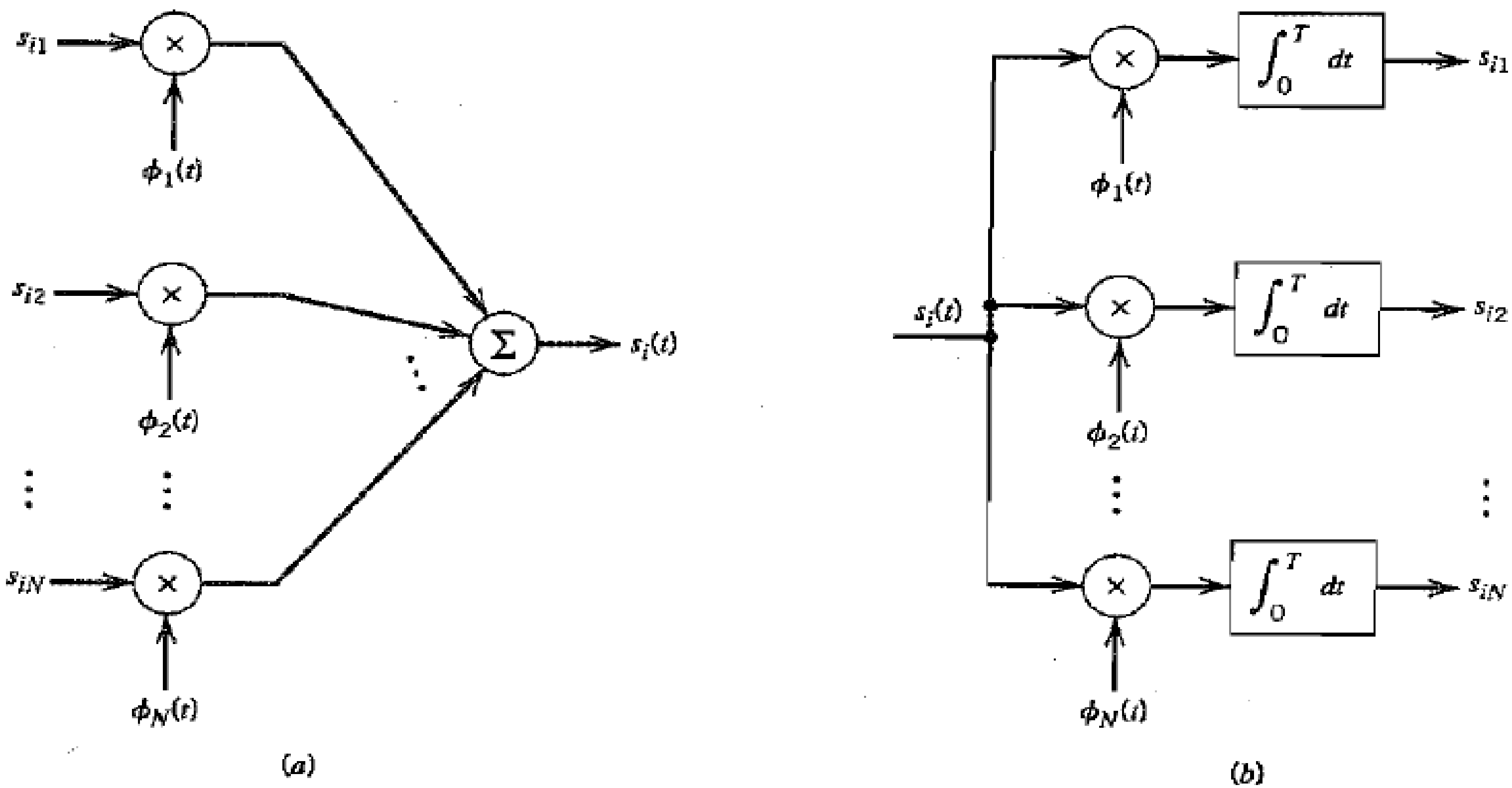


FIGURE 5.3 (a) Synthesizer for generating the signal $s_i(t)$. (b) Analyzer for generating the set of signal vectors $\{s_i\}$.

the energy of a signal $s_i(t)$ is equal to the squared length of the signal vector $s_i(t)$

$$E_i = \sum_{j=1}^{N} s_{ij}^2 = \| \mathbf{s}_i \|^2$$

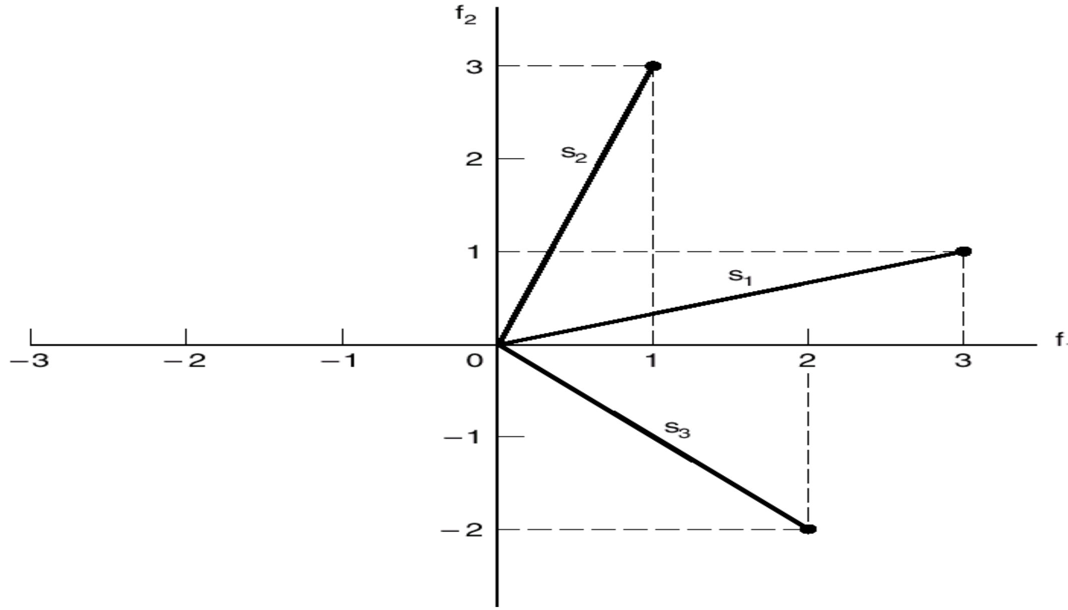


Figure 5.4
Illustrating the geometric representation of signals for the case when $N = 2$ and $M = 3$.

The idea of visualizing a set of energy signals geometrically, as just described, is of profound importance. It provides the mathematical basis for the geometric representation of energy signals, thereby paving the way for the noise analysis of digital communication systems in a conceptually satisfying manner. This form of representation is illustrated in Figure 5.4 for the case of a two-dimensional signal space with three signals, that is, $N = 2$ and $M = 3$.

In an N-dimensional Euclidean space, we may define *lengths* of vectors and *angles* between vectors. It is customary to denote the length (also called the *absolute value* or *norm*) of a signal vector $\mathbf{s}_i$ by the symbol $\| \mathbf{s}_i \|$. The squared-length of any signal vector $\mathbf{s}_i$ is defined to be the *inner product* or *dot product* of $\mathbf{s}_i$ with itself, as shown by

$$\begin{aligned} \| \mathbf{s}_i \|^2 &= \mathbf{s}_i^T \mathbf{s}_i \\ &= \sum_{j=1}^{N} s_{ij}^2, \qquad i = 1, 2, \ldots, M \end{aligned} \tag{5.9}$$

where s_{ij} is the jth element of $\mathbf{s}_i$, and the superscript T denotes matrix transposition.

There is an interesting relationship between the energy content of a signal and its representation as a vector. By definition, the energy of a signal $s_i(t)$ of duration T seconds is

$$E_i = \int_0^T s_i^2(t)\, dt \tag{5.10}$$

Therefore, substituting Equation (5.5) into (5.10), we get

$$E_i = \int_0^T \left[\sum_{j=1}^{N} s_{ij}\phi_j(t)\right]\left[\sum_{k=1}^{N} s_{ik}\phi_k(t)\right]dt$$

Interchanging the order of summation and integration, and then rearranging terms, we get

$$E_i = \sum_{j=1}^{N} \sum_{k=1}^{N} s_{ij} s_{ik} \int_0^T \phi_j(t)\phi_k(t)dt \tag{5.11}$$

But since the $\phi_j(t)$ form an orthonormal set, in accordance with the two conditions of Equation (5.7), we find that Equation (5.11) reduces simply to

$$\begin{aligned} E_i &= \sum_{j=1}^{N} s_{ij}^2 \\ &= \| \mathbf{s}_i \|^2 \end{aligned} \tag{5.12}$$

Thus Equations (5.9) and (5.12) show that the energy of a signal $s_i(t)$ is equal to the squared length of the signal vector $\mathbf{s}_i(t)$ representing it.

In the case of a pair of signals $s_i(t)$ and $s_k(t)$, represented by the signal vectors $\mathbf{s}_i$ and $\mathbf{s}_k$, respectively, we may also show that

$$\int_0^T s_i(t)s_k(t)\, dt = \mathbf{s}_i^T \mathbf{s}_k \tag{5.13}$$

Equation (5.13) states that the *inner product* of the signals $s_i(t)$ and $s_k(t)$ over the interval [0, T], using their time-domain representations, is equal to the inner product of their respective vector representations $\mathbf{s}_i$ and $\mathbf{s}_k$. Note that the inner product of $s_i(t)$ and $s_k(t)$ is *invariant* to the choice of basis functions $\{\phi_j(t)\}_{j=1}^{N}$ in that it only depends on the components of the signals $s_i(t)$ and $s_k(t)$ projected onto each of the basis functions.

5.3 Conversion of the Continuous AWGN Channel into a Vector Channel

5.5 Coherent Detection of Signals in Noise: Maximum Likelihood Decoding

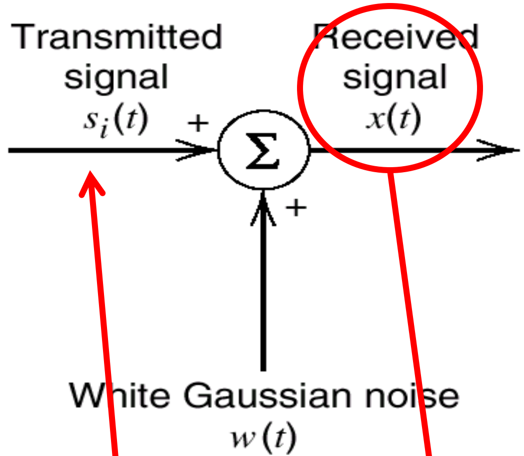


Figure 5.2 Additive white Gaussian noise (AWGN) model of a channel.

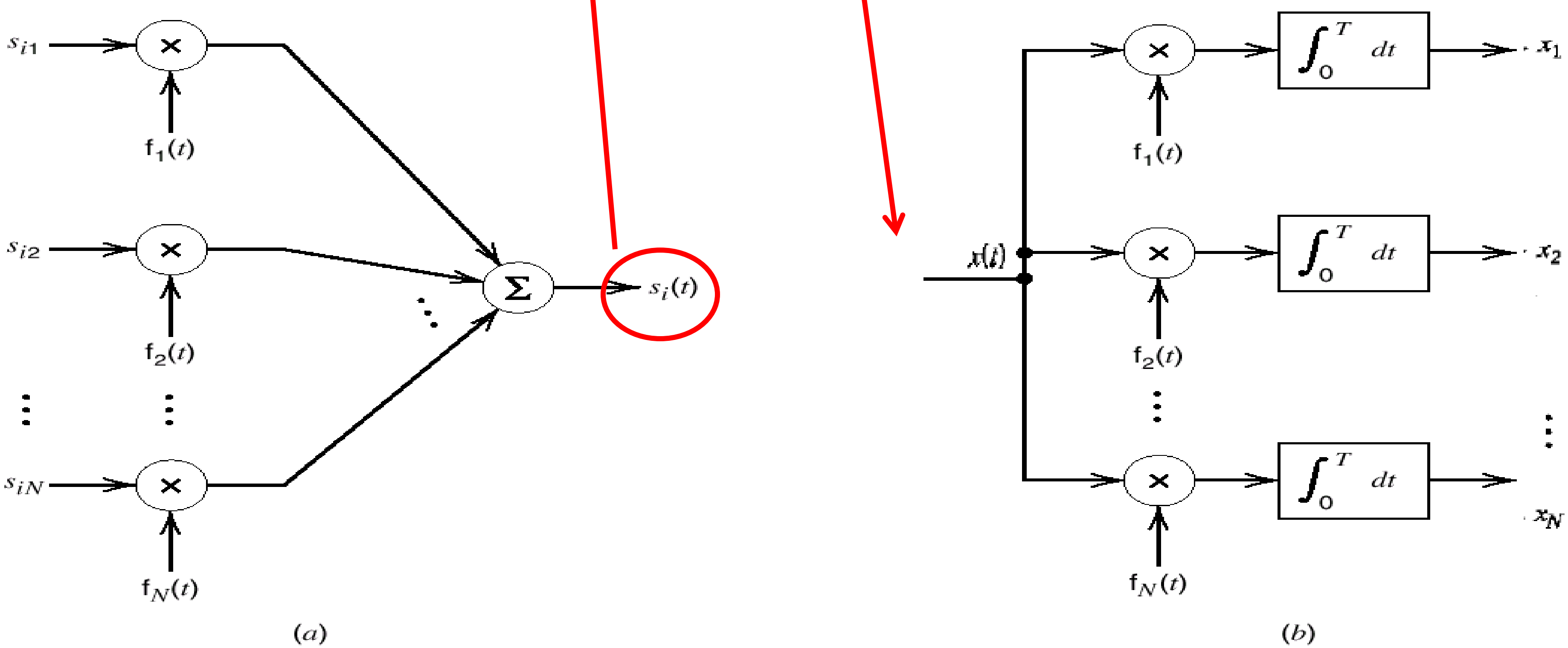


Figure 5.3 ***(a) Synthesizer for generating the signal $s_i(t)$.***
(b) Analyzer for generating the set of signal vectors $\{s_i\}$.

5.3 Conversion of the Continuous AWGN Channel into a Vector Channel

Suppose that the input to the bank of N product integrators or correlators in Figure 5.3b is not the transmitted signal $s_i(t)$ but rather the received signal $x(t)$ defined in accordance with the idealized AWGN channel of Figure 5.2. That is to say,

$$x(t) = s_i(t) + w(t), \qquad \begin{cases} 0 \le t \le T \\ i = 1, 2, \ldots, M \end{cases} \tag{5.28}$$

where $w(t)$ is a sample function of a white Gaussian noise process $W(t)$ of zero mean and power spectral density $N_0/2$. Correspondingly, we find that the output of correlator j, say, is the sample value of a random variable X_j, as shown by

$$\begin{aligned} x_j &= \int_0^T x(t)\phi_j(t)dt \\ &= s_{ij} + w_j, \qquad j = 1, 2, \ldots, N \end{aligned} \tag{5.29}$$

The first component, s_{ij}, is a deterministic quantity contributed by the transmitted signal $s_i(t)$; it is defined by

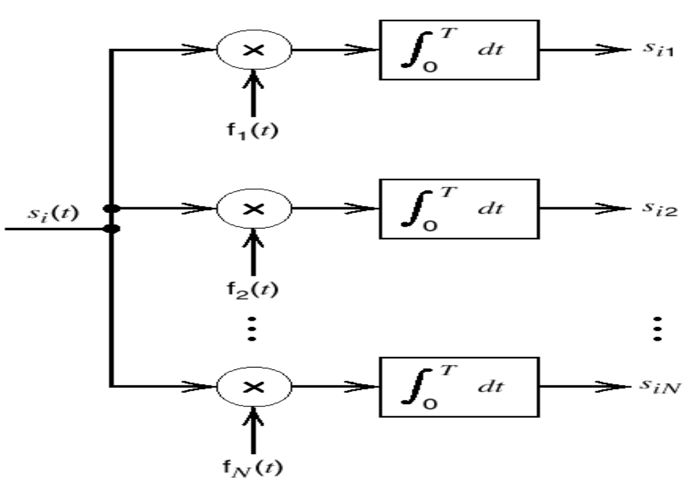
$$s_{ij} = \int_0^T s_i(t)\phi_j(t)dt \tag{5.30}$$

The second component, w_j, is the sample value of a random variable W_j that arises because of the presence of the channel noise $w(t)$; it is defined by

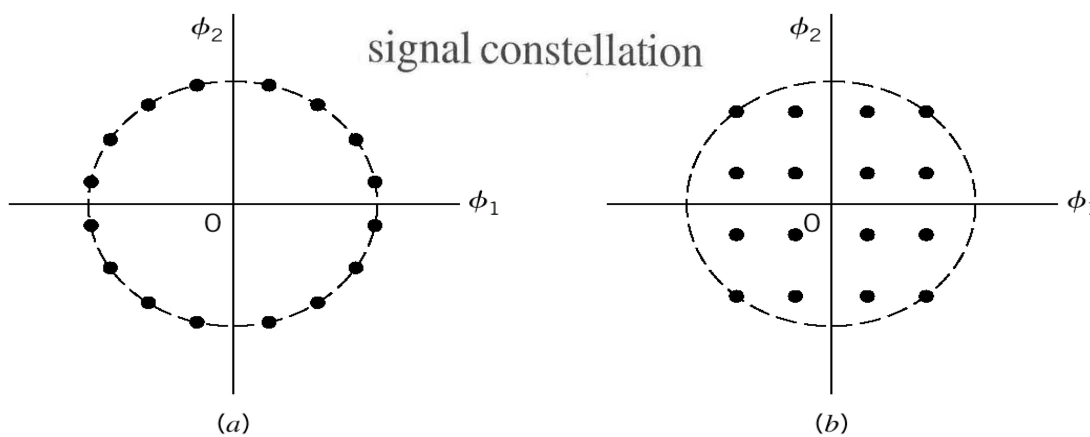
$$w_j = \int_0^T w(t)\phi_j(t)dt \tag{5.31}$$

5.5 Coherent Detection of Signals in Noise: Maximum Likelihood Decoding

Suppose that in each time slot of duration T seconds, one of the M possible signals $s_1(t)$, $s_2(t), \ldots, s_M(t)$ is transmitted with equal probability, $1/M$. For geometric signal representation, the signal $s_i(t), i = 1, 2, \ldots, M$, is applied to a bank of correlators, with a common input and supplied with an appropriate set of N orthonormal basis functions. The resulting correlator outputs define the *signal vector* $\mathbf{s}_i$. Since knowledge of the signal vector $\mathbf{s}_i$ is as good as knowing the transmitted signal $s_i(t)$ itself, and vice versa, we may represent $s_i(t)$ by a point in a Euclidean space of dimension $N \le M$. We refer to this point as the *transmitted signal point* or *message point*. The set of message points corresponding to the set of transmitted signals $\{s_i(t)\}_{i=1}^{M}$ is called a *signal constellation*.



(b)



Signal constellation for (a) M-ary PSK and (b) corresponding M-ary QAM, for M = 16.

However, the representation of the received signal $x(t)$ is complicated by the presence of additive noise $w(t)$. We note that when the received signal $x(t)$ is applied to the bank of N correlators, the correlator outputs define the observation vector $\mathbf{x}$. From Equation (5.48), the vector $\mathbf{x}$ differs from the signal vector $\mathbf{s}_i$ by the *noise vector* $\mathbf{w}$ whose orientation is completely random. The noise vector $\mathbf{w}$ is completely characterized by the noise $w(t)$;

Now, based on the observation vector $\mathbf{x}$, we may represent the received signal $x(t)$ by a point in the same Euclidean space used to represent the transmitted signal. We refer to this second point as the *received signal point*. The received signal point wanders about the message point in a completely random fashion, in the sense that it may lie anywhere inside a Gaussian-distributed "cloud" centered on the message point. This is illustrated in Figure 5.7*a* for the case of a three-dimensional signal space. For a particular realization of the noise vector $\mathbf{w}$ (i.e., a particular point inside the random cloud of Figure 5.7*a*), the relationship between the observation vector $\mathbf{x}$ and the signal vector $\mathbf{s}_i$ is as illustrated in Figure 5.7*b*.

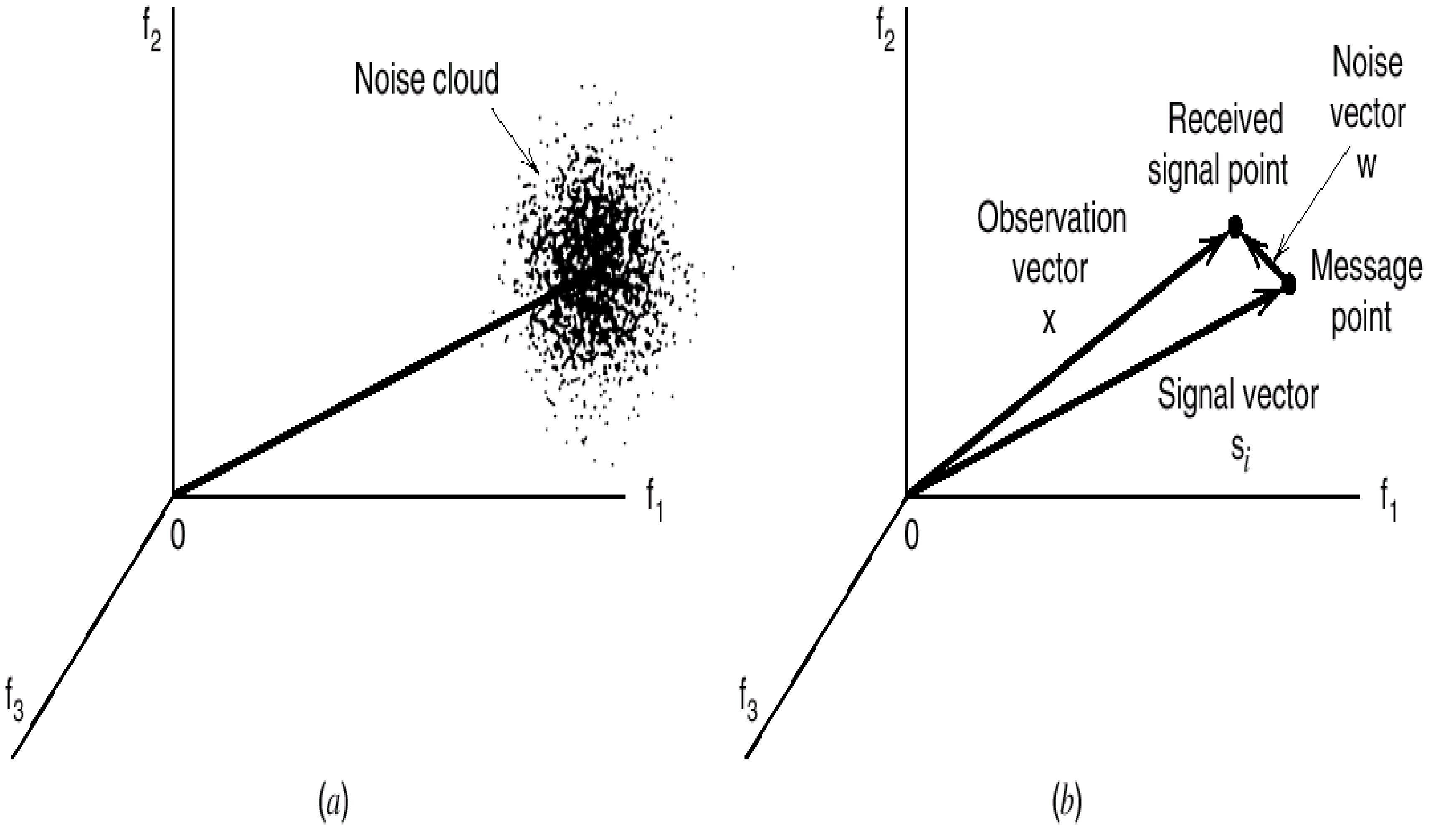


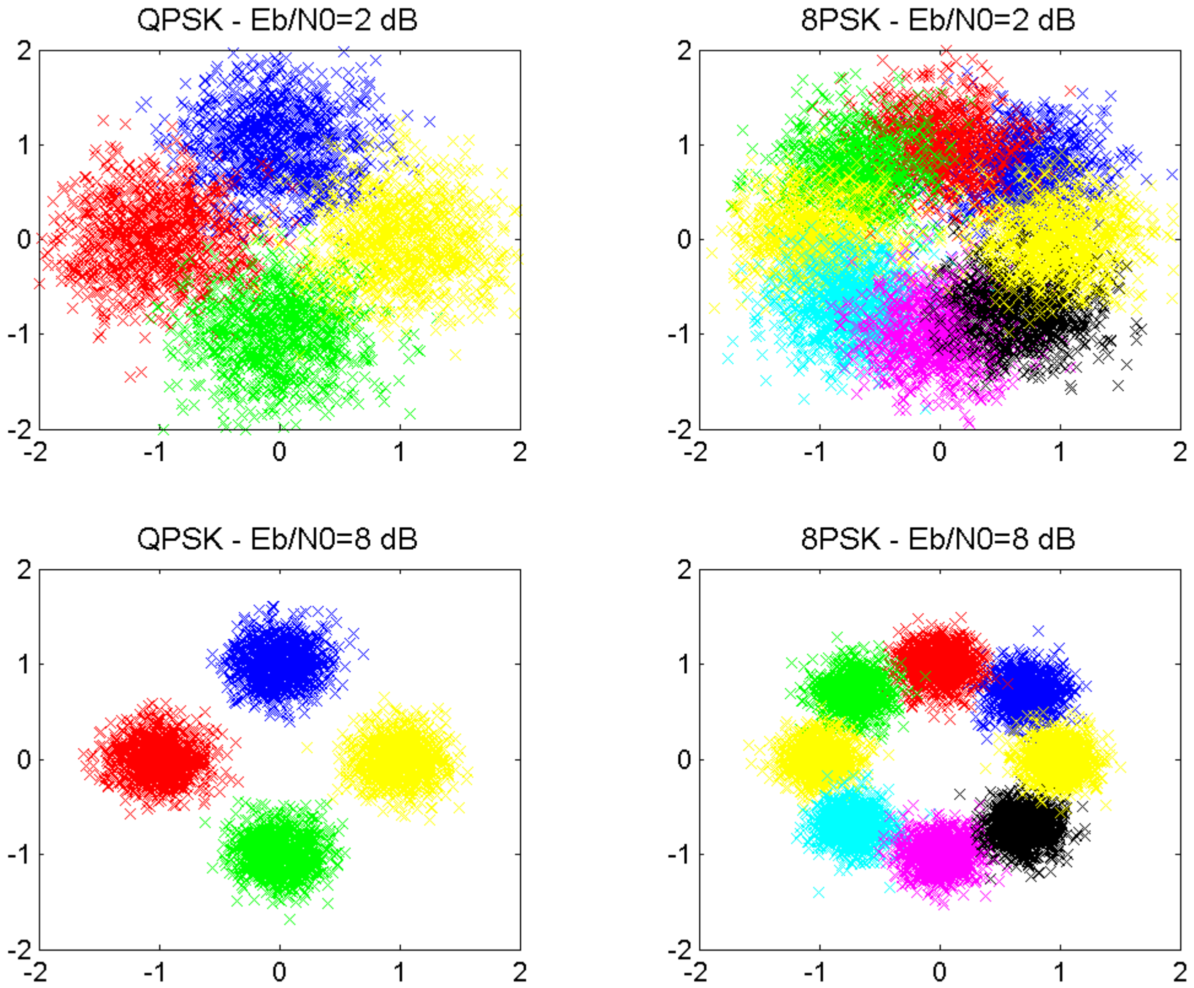
Figure 5.7 ***Illustrating the effect of noise perturbation, depicted in (a), on the location of the received signal point, depicted in (b).***

AWGN is

equivalent to an N-dimensional *vector channel* described by the observation vector

$$\mathbf{x} = \mathbf{s}_i + \mathbf{w}, \qquad i = 1, 2, \ldots, M \tag{5.48}$$

Example of samples of matched filter output for some bandpass modulation schemes



graphical interpretation of the maximum likelihood decision

rule. Let Z denote the N-dimensional space of all possible observation vectors $\mathbf{x}$. We refer to this space as the *observation space*. Because we have assumed that the decision rule must say $\hat{m} = m_i$, where $i = 1, 2, \ldots, M$, the total observation space Z is correspondingly partitioned into *M-decision regions*, denoted by $Z_1, Z_2, \ldots, Z_M$. Accordingly, we may restate the decision rule of Equation (5.55) as follows:

$$\text{Observation vector } \mathbf{x} \text{ lies in region } Z_i \text{ if the Euclidean distance } \| \mathbf{x} - \mathbf{s}_k \| \text{ is minimum for } k = i \tag{5.59}$$

Equation (5.59) states that *the maximum likelihood decision rule is simply to choose the message point closest to the received signal point*, which is intuitively satisfying.

$$\sum_{j=1}^{N} (x_j - s_{kj})^2 = \| \mathbf{x} - \mathbf{s}_k \|^2$$

$\| \mathbf{x} - \mathbf{s}_k \|$ is the Euclidean distance

$$\sum_{j=1}^{N} (x_j - s_{kj})^2 = \| \mathbf{x} - \mathbf{s}_k \|^2$$

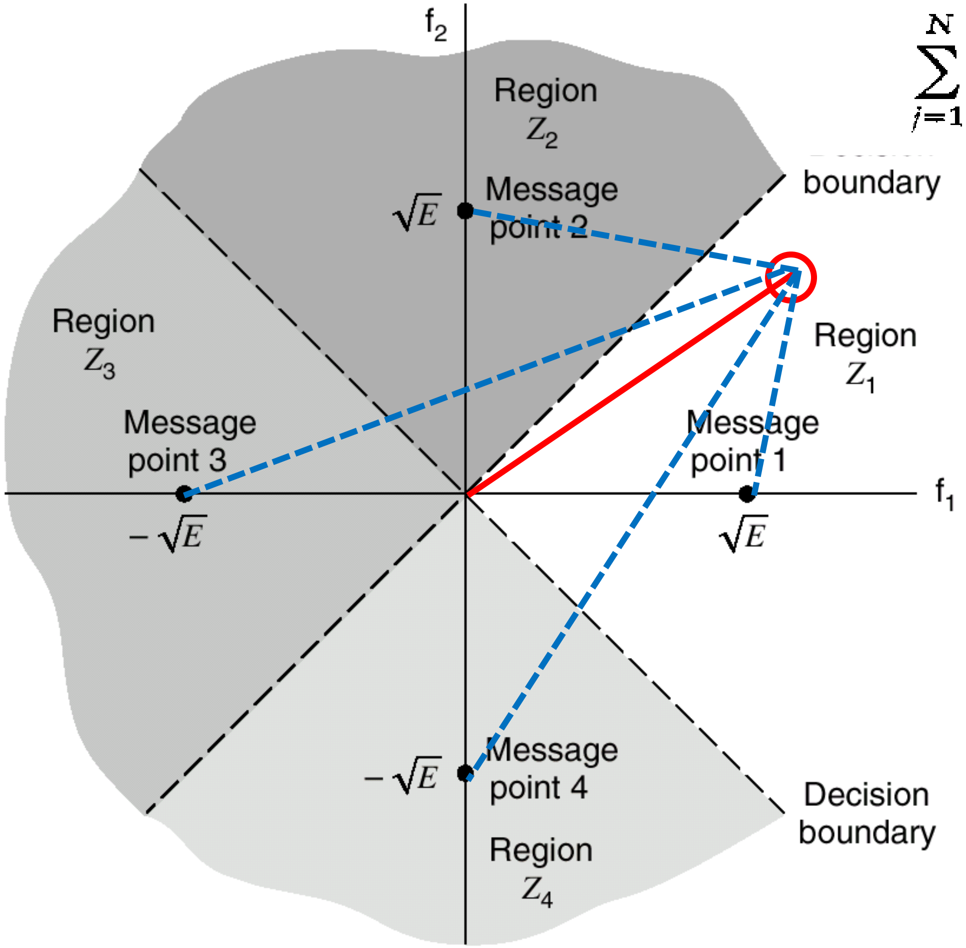


FIGURE 5.8 Illustrating the partitioning of the observation space into decision regions for the case when $N = 2$ and $M = 4$; it is assumed that the M transmitted symbols are equally likely.

$M = 4$ signals and $N = 2$ dimensions, assuming that the signals are transmitted with equal energy, E, and equal probability.

5.6 *Correlation Receiver*

$\{ s_i(t) \}$ are equally likely .

The correlation receiver consists of two parts :

* detector (correlator)
* decoder

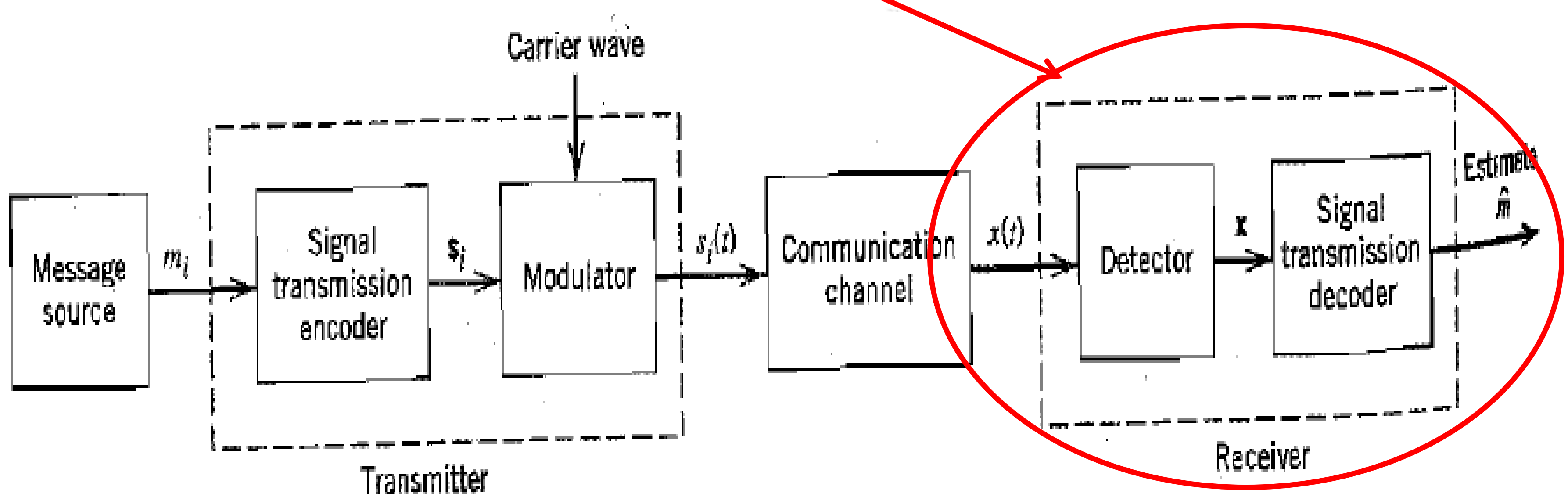
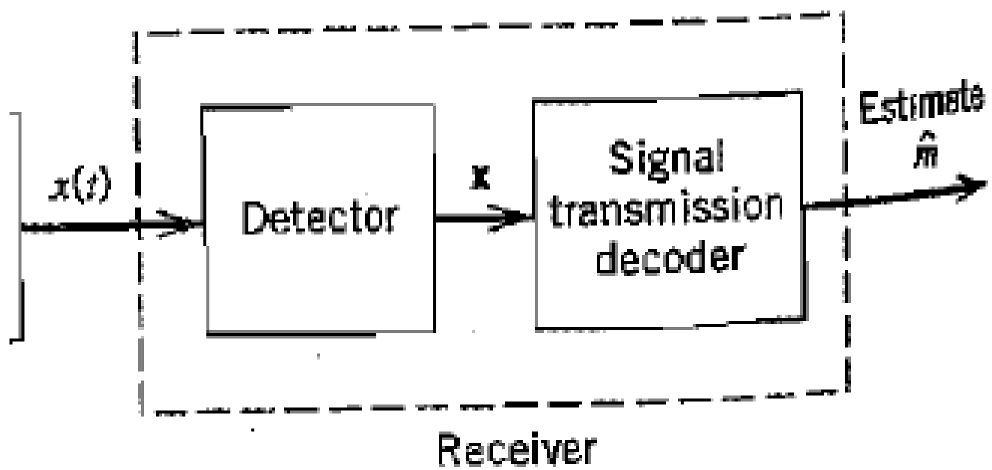


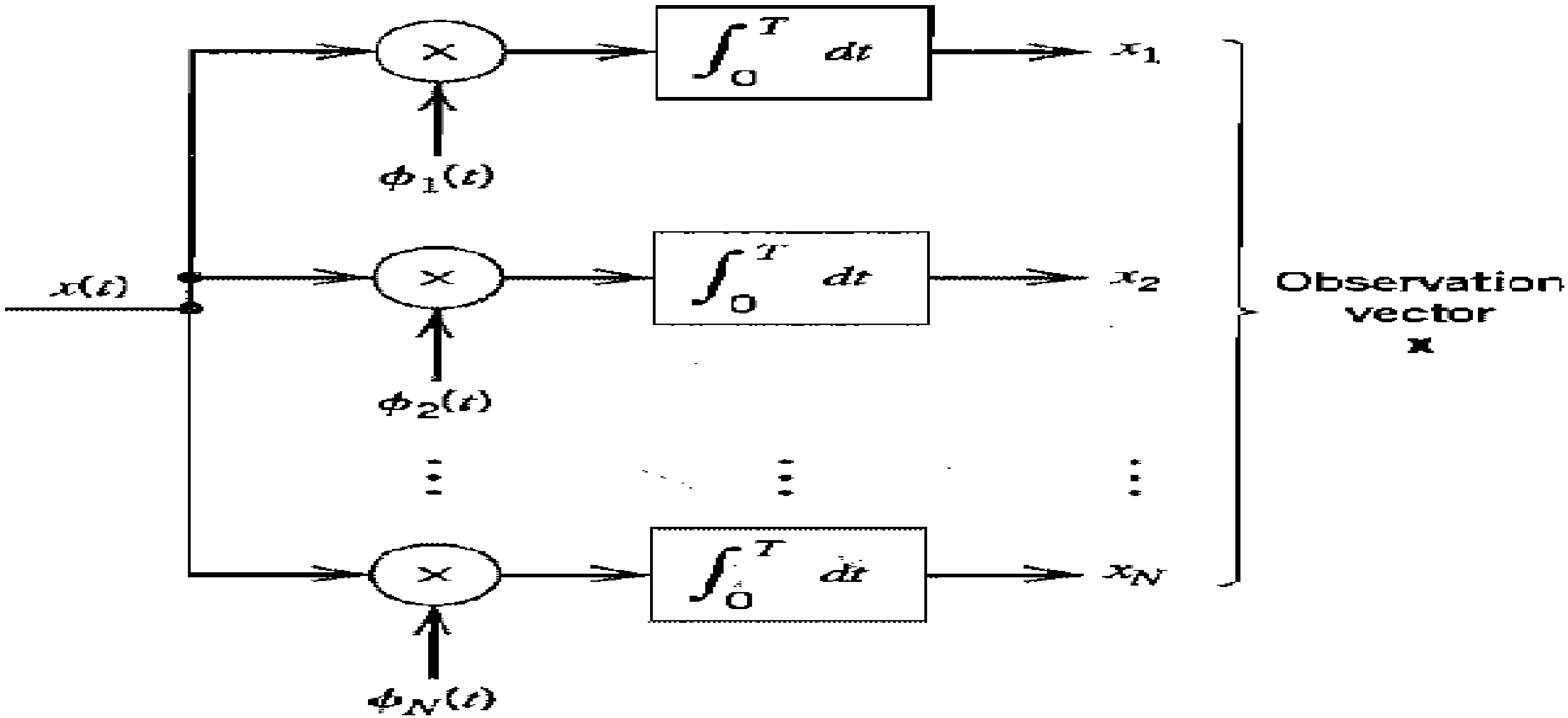
FIGURE 6.2 Functional model of passband data transmission system.

5.6 *Correlation Receiver*

From the material presented in the previous sections, we find that for an AWGN channel and for the case when the transmitted signals $s_1(t), s_2, \ldots, s_M(t)$ are equally likely, the optimum receiver consists of two subsystems, which are detailed in Figure 5.9 and described here:

1. The *detector* part of the receiver is shown in Figure 5.9*a*. It consists of a bank of M *product-integrators* or *correlators*, supplied with a corresponding set of coherent reference signals or orthonormal basis functions $\phi_1(t), \phi_2(t), \ldots, \phi_N(t)$ that are generated locally. This bank of correlators operates on the received signal $x(t)$, $0 \le t \le T$, to produce the observation vector $\mathbf{x}$.





$x(t)$
$\phi_1(t)$
$\phi_2(t)$
$\phi_N(t)$
$\int_0^T dt$
x_1
x_2
x_N
Observation vector
x