Chapter 5
Optimum Receivers for the Additive White Gaussian Noise Channel
Table of Contents

5.1 Optimum Receiver for Signals Corrupted by Additive White Noise
   5.1.1 Correlation Demodulator
   5.1.2 Matched-Filter Demodulator
   5.1.3 The Optimum Detector
   5.1.4 The Maximum-Likelihood Sequence Detector
   5.1.5 A Symbol-by-Symbol MAP Detector for Signals with Memory

5.2 Performance of the Optimum Receiver for Memoryless Modulation
   5.2.1 Probability of Error for Binary Modulation
   5.2.2 Probability of Error for $M$-ary Orthogonal Signals
Table of Contents

5.2.3 Probability of Error for $M$-ary Biorthogonal Signals
5.2.4 Probability of Error for Simplex Signals
5.2.5 Probability of Error for $M$-ary Binary-Coded Signals
5.2.6 Probability of Error for $M$-ary PAM
5.2.7 Probability of Error for $M$-ary PSK
5.2.8 Differential PSK(DPSK) and Its Performance
5.2.9 Probability of Error for QAM
5.2.10 Comparison of Digital Modulation Methods

5.3 Optimum Receiver for CPM Signals
5.3.1 Optimum Demodulator and Detection of CPM
5.3.2 Performance of CPM Signals
Table of Contents

5.3.3 Symbol-by-Symbol Detection of CPM Signals
5.3.4 Suboptimum Demodulation and Detection of CPM Signals

5.4.1 Optimum Receiver for Signals with Random Phase on AWGN Channel
5.4.1 Optimum Receiver for Binary Signals
5.4.2 Optimum Receiver for $M$-ary Orthogonal Signals
5.4.3 Probability of Error for Envelope Detection of $M$-ary Orthogonal Signals
5.4.4 Probability of Error for Envelope Detection of Correlated Binary Signals
Table of Contents

5.5 Performance Analysis for Wireline and Radio Communication Systems
   5.5.1 Regenerative Repeaters
   5.5.2 Link Budget Analysis in Raid Communication Systems
5.1 Optimum Receiver for Signals Corrupted by Additive White Gaussian Noise

- We assume that the transmitter sends digital information by use of \( M \) signals waveforms \( \{s_m(t) = 1, 2, \ldots, M\} \). Each waveform is transmitted within the symbol interval of duration \( T \), i.e. \( 0 \leq t \leq T \).

- The channel is assumed to corrupt the signal by the addition of white Gaussian noise, as shown in the following figure:

\[
r(t) = s_m(t) + n(t), \quad 0 \leq t \leq T
\]

where \( n(t) \) denotes a sample function of AWGN process with power spectral density \( \Phi_{nn}(f) = \frac{1}{2} N_0 \) W/Hz.
Our object is to design a receiver that is optimum in the sense that it minimizes the probability of making an error. It is convenient to subdivide the receiver into two parts—the signal demodulator and the detector.

The function of the signal demodulator is to convert the received waveform $r(t)$ into an $N$-dimensional vector $\mathbf{r} = [r_1, r_2, \ldots, r_N]$ where $N$ is the dimension of the transmitted signal waveform.

The function of the detector is to decide which of the $M$ possible signal waveforms was transmitted based on the vector $\mathbf{r}$. 

---

**Fall, 2004.** WITS Lab, NSYSU.
Two realizations of the signal demodulator are described in the next two sections:

- One is based on the use of *signal correlators*.
- The second is based on the use of *matched filters*.

The optimum detector that follows the signal demodulator is designed to minimize the probability of error.
5.1.1 Correlation Demodulator

- We describe a *correlation demodulation* that decomposes the receiver signal and the noise into $N$-dimensional vectors. In other words, the signal and the noise are expanded into a series of linearly weighted orthonormal basis functions $\{f_n(t)\}$.

- It is assumed that the $N$ basis function $\{f_n(t)\}$ span the signal space, so every one of the possible transmitted signals of the set $\{s_m(t) = 1 \leq m \leq M\}$ can be represented as a linear combination of $\{f_n(t)\}$.

- In case of the noise, the function $\{f_n(t)\}$ do not span the noise space. However we show below that the noise terms that fall outside the signal space are irrelevant to the detection of the signal.
Suppose the receiver signal \( r(t) \) is passed through a parallel bank of \( N \) basis functions \( \{ f_n(t) \} \), as shown in the following figure:

\[
\int_0^T r(t) f_k(t) dt = \int_0^T \left[ s_m(t) + n(t) \right] f_k(t) dt
\]
\[
\Rightarrow \quad r_k = s_{mk} + n_k, \quad k = 1, 2, \ldots, N
\]

\[
s_{mk} = \int_0^T s_m(t) f_k(t) dt, \quad k = 1, 2, \ldots, N
\]

\[
n_k = \int_0^T n(t) f_k(t) dt, \quad k = 1, 2, \ldots, N
\]

The signal is now represented by the vector \( s_m \) with components \( s_{mk}, k=1,2,\ldots,N \). Their values depend on which of the \( M \) signals was transmitted.
In fact, we can express the receiver signal $r(t)$ in the interval $0 \leq t \leq T$ as:

$$r(t) = \sum_{k=1}^{N} s_{mk} f_k(t) + \sum_{k=1}^{N} n_k f_k(t) + n'(t)$$

$$= \sum_{k=1}^{N} r_k f_k(t) + n'(t)$$

The term $n'(t)$, defined as

$$n'(t) = n(t) - \sum_{k=1}^{N} n_k f_k(t)$$

is a zero-mean Gaussian noise process that represents the difference between original noise process $n(t)$ and the part corresponding to the projection of $n(t)$ onto the basis functions $\{f_k(t)\}$. 

5.1.1 Correlation Demodulator
5.1.1 Correlation Demodulator

- We shall show below that \( n'(t) \) is irrelevant to the decision as to which signal was transmitted. Consequently, the decision may be based entirely on the correlator output signal and noise components \( r_k = s_{mk} + n_k, \ k = 1, 2, \ldots, N \).

- The noise components \( \{n_k\} \) are Gaussian and mean values are:

\[
E(n_k) = \int_0^T E[n(t)] f_k(t) dt = 0 \quad \text{for all } n.
\]

and their covariances are:

\[
E(n_k n_m) = \int_0^T \int_0^T E[n(t)n(\tau)] f_k(t) f_m(\tau) dt d\tau
\]

\[
= \frac{1}{2} \int_0^T \int_0^T \delta(t-\tau) f_k(t) f_m(\tau) dt d\tau
\]

\[
= \frac{1}{2} N_0 \int_0^T f_k(t) f_m(t) dt = \frac{1}{2} N_0 \delta_{mk}
\]

**Conclusion:** The \( N \) noise Components \( \{n_k\} \) are zero-mean uncorrelated Gaussian random variables with a common variance \( \sigma_n^2 = \frac{1}{2} N_0 \).

**Power spectral density** is \( \Phi_{nn}(f) = \frac{1}{2} N_0 \text{ W/Hz} \)
5.1.1 Correlation Demodulator

- From the above development, it follows that the correlator output \( \{ r_k \} \) conditioned on the \( m \)th signal being transmitted are Gaussian random variables with mean

\[
E(r_k) = E(s_{mk} + n_k) = s_{mk}
\]

and equal variance

\[
\sigma_r^2 = \sigma_n^2 = \frac{1}{2} N_0
\]

- Since the noise components \( \{ n_k \} \) are uncorrelated Gaussian random variables, they are also statistically independent. As a consequence, the correlator outputs \( \{ r_k \} \) conditioned on the \( m \)th signal being transmitted are statistically independent Gaussian variables.
5.1.1 Correlation Demodulator

The conditional probability density functions of the random variables \( r = [r_1 \ r_2 \ \ldots \ r_N] \) are:

\[
p(r \mid s_m) = \prod_{k=1}^{N} p(r_k \mid s_{mk}), \quad m = 1, 2, \ldots, M \quad \text{----(A)}
\]

\[
p(r_k \mid s_{mk}) = \frac{1}{\sqrt{\pi N_0}} \exp \left[ - \frac{(r_k - s_{mk})^2}{N_0} \right], \quad k = 1, 2, \ldots, N \quad \text{----(B)}
\]

By substituting Equation (A) into Equation (B), we obtain the joint conditional PDFs

\[
p(r|s_m) = \frac{1}{(\pi N_0)^{N/2}} \exp \left[ - \sum_{k=1}^{N} \frac{(r_k - s_{mk})^2}{N_0} \right], \quad m = 1, 2, \ldots, M
\]
The correlator outputs \((r_1, r_2, \ldots, r_N)\) are **sufficient statistics** for reaching a decision on which of the \(M\) signals was transmitted, i.e., no additional relevant information can be extracted from the remaining noise process \(n'(t)\).

Indeed, \(n'(t)\) is uncorrelated with the \(N\) correlator outputs \(\{r_k\}\):

\[
E[n'(t)r_k] = E[n'(t)]s_{mk} + E[n'(t)n_k] = E[n'(t)n_k] = E\left\{n(t) - \sum_{j=1}^{N} n_j f_j(t)\right\} n_k
\]

\[
= \int_0^T E[n(t)n(\tau)]f_k(\tau)d\tau - \sum_{j=1}^{N} E(n_jn_k)f_j(t)
\]

\[
= \int_0^T \frac{1}{2} N_0 \delta(t-\tau)f_k(\tau)d\tau - \sum_{j=1}^{N} \frac{1}{2} N_0 \delta_{jk} f_j(t)
\]

\[
= \frac{1}{2} N_0 f_k(t) - \frac{1}{2} N_0 f_k(t) = 0
\]  

Q.E.D.
5.1.1 Correlation Demodulator

- Since $n'(t)$ and $\{r_k\}$ are Gaussian and uncorrelated, they are also statistically independent.

- Consequently, $n'(t)$ does not contain any information that is relevant to the decision as to which signal waveform was transmitted.

- All the relevant information is contained in the correlator outputs $\{r_k\}$ and, hence, $n'(t)$ can be ignored.
Example 5.1-1.

Consider an $M$-ary baseband PAM signal set in which the basic pulse shape $g(t)$ is rectangular as shown in following figure.

The additive noise is a zero-mean white Gaussian noise process.

Let us determine the basis function $f(t)$ and the output of the correlation-type demodulator.

The energy in the rectangular pulse is

$$
E_g = \int_0^T g^2(t)dt = \int_0^T a^2 dt = a^2 T
$$
Example 5.1-1.(cont.)

Since the PAM signal set has dimension $N=1$, there is only one basis function $f(t)$ given as:

$$f(t) = \frac{1}{\sqrt{a^2 T}} g(t) = \begin{cases} 1/\sqrt{T} & (0 \leq t \leq T) \\ 0 & \text{(otherwise)} \end{cases}$$

The output of the correlation-type demodulator is:

$$r = \int_0^T r(t) f(t) dt = \frac{1}{\sqrt{T}} \int_0^T r(t) dt$$

The correlator becomes a simple integrator when $f(t)$ is rectangular: if we substitute for $r(t)$, we obtain:

$$r = \frac{1}{\sqrt{T}} \left\{ \int_0^T \left[ s_m(t) + n(t) \right] \right\} dt = \frac{1}{\sqrt{T}} \left[ \int_0^T s_m(t) dt + \int_0^T n(t) dt \right]$$

$$= s_m + n$$
Example 5.1-1.(cont.)

- The noise term $E(n)=0$ and:

$$\sigma_n^2 = E[ n(t)n(t) ] = E \left[ \frac{1}{T} \int_0^T \int_0^T n(t)n(\tau)d\tau d\tau \right]$$

$$= \frac{1}{T} \int_0^T \int_0^T E[ n(t)n(\tau) ] d\tau d\tau = \frac{N_0}{2T} \int_0^T \int_0^T \delta(t-\tau)d\tau d\tau$$

$$= \frac{N_0}{2T} \int_0^T 1 \cdot d\tau = \frac{1}{2} N_0$$

- The probability density function for the sampled output is:

$$p(r \mid s_m) = \frac{1}{\sqrt{\pi N_0}} \exp \left[ - \frac{(r - s_m)^2}{N_0} \right]$$
5.1.2 Matched-Filter Demodulator

Instead of using a bank of $N$ correlators to generate the variables \( \{r_k\} \), we may use a bank of $N$ linear filters. To be specific, let us suppose that the impulse responses of the $N$ filters are:

\[
h_k(t) = f_k(T - t), \quad 0 \leq t \leq T
\]

where \( \{f_k(t)\} \) are the $N$ basis functions and $h_k(t)=0$ outside of the interval $0 \leq t \leq T$.

The outputs of these filters are:

\[
y_k(t) = r(t) * h_k(t) = \int_{0}^{t} r(\tau) h_k(t - \tau) d\tau = \int_{t}^{T} r(\tau) f_k(T - t + \tau) d\tau, \quad k = 1, 2, \ldots, N
\]
5.1.2 Matched-Filter Demodulator

If we sample the outputs of the filters at \( t = T \), we obtain

\[
y_k(T) = \int_{0}^{T} r(\tau)f_k(\tau)d\tau = r_k, \quad k = 1, 2, \ldots, N
\]

A filter whose impulse response \( h(t) = s(T-t) \), where \( s(t) \) is assumed to be confined to the time interval \( 0 \leq t \leq T \), is called **matched filter** to the signal \( s(t) \).

An example of a signal and its matched filter are shown in the following figure.

(a) Signal \( s(t) \)

(b) Impulse response of filter matched to \( s(t) \)
The response of \( h(t) = s(T-t) \) to the signal \( s(t) \) is:

\[
y(t) = s(t) * h(t) = \int_0^t s(\tau)h(t - \tau) \, d\tau = \int_0^t s(\tau)s(T - t + \tau) \, d\tau
\]

which is the time-autocorrelation function of the signal \( s(t) \).

Note that the autocorrelation function \( y(t) \) is an even function of \( t \), which attains a peak at \( t = T \).
5.1.2 Matched-Filter Demodulator

_matched filter demodulator that generates the observed variables \( \{r_k\} \)
5.1.2 Matched-Filter Demodulator

Properties of the matched filter.

- If a signal $s(t)$ is corrupted by AWGN, the filter with an impulse response matched to $s(t)$ maximizes the output signal-to-noise ratio (SNR).

Proof: let us assume the receiver signal $r(t)$ consists of the signal $s(t)$ and AWGN $n(t)$ which has zero-mean and $\Phi_{nm}(f) = \frac{1}{2} N_0$ W/Hz.

- Suppose the signal $r(t)$ is passed through a filter with impulse response $h(t)$, $0 \leq t \leq T$, and its output is sampled at time $t=T$. The output signal of the filter is:

$$y(t) = s(t) * h(t) = \int_0^t r(\tau)h(t - \tau)d\tau$$

$$= \int_0^t s(\tau)h(t - \tau)d\tau + \int_0^t n(\tau)h(t - \tau)d\tau$$
5.1.2 Matched-Filter Demodulator

Proof: (cont.)

At the sampling instant $t=T$:

$$y(T) = \int_0^T s(\tau)h(T - \tau)d\tau + \int_0^T n(\tau)h(T - \tau)d\tau$$

$$= y_s(T) + y_n(T)$$

This problem is to select the filter impulse response that maximizes the output SNR$_0$ defined as:

$$\text{SNR}_0 = \frac{y_s^2(T)}{E[y_n^2(T)]}$$

$$E[y_n^2(T)] = \int_0^T \int_0^T E[n(\tau)n(t)]h(T - \tau)h(T - t)dtd\tau$$

$$= \frac{1}{2} N_0 \int_0^T \int_0^T \delta(t - \tau)h(T - \tau)h(T - t)dtd\tau = \frac{1}{2} N_0 \int_0^T h^2(T - t)dt$$
5.1.2 Matched-Filter Demodulator

Proof: (cont.)

- By substituting for \( y_s(T) \) and \( E\left[ y_n^2(T) \right] \) into SNR₀.

\[
SNR_0 = \frac{\left[ \int_0^T s(\tau)h(T-\tau)d\tau \right]^2}{\frac{1}{2} N_0 \int_0^T h^2(T-t)dt} = \frac{\left[ \int_0^T h(T'-\tau)s(T'-T)d\tau' \right]^2}{\frac{1}{2} N_0 \int_0^T h^2(T-t)dt}
\]

- Denominator of the SNR depends on the energy in \( h(t) \).

- The maximum output SNR over \( h(t) \) is obtained by maximizing the numerator subject to the constraint that the denominator is held constant.
5.1.2 Matched-Filter Demodulator

Proof: (cont.)

- **Cauchy-Schwarz inequality**: if \( g_1(t) \) and \( g_2(t) \) are finite-energy signals, then

\[
\left[ \int_{-\infty}^{\infty} g_1(t) g_2(t) dt \right]^2 \leq \int_{-\infty}^{\infty} g_1^2(t) dt \int_{-\infty}^{\infty} g_2^2(t) dt
\]

with equality when \( g_1(t) = C g_2(t) \) for any arbitrary constant \( C \).

- If we set \( g_1(t) = h_1(t) \) and \( g_2(t) = s(T-t) \), it is clear that the SNR is maximized when \( h(t) = C s(T-t) \).
5.1.2 Matched-Filter Demodulator

Proof: (cont.)

The output (maximum) SNR obtained with the matched filter is:

\[
\text{SNR}_0 = \frac{\left[ \int_0^T s(\tau)h(T-\tau)d\tau \right]^2}{\frac{1}{2} N_0 \int_0^T h^2(T-t)dt} = \frac{2}{N_0} \left[ \int_0^T s(\tau)Cs(T-(T-\tau))d\tau \right]^2 = \frac{2}{N_0} \int_0^T s^2(t)dt = \frac{2\mathcal{E}}{N_0}
\]

Note that the output SNR from the matched filter depends on the energy of the waveform \(s(t)\) but not on the detailed characteristics of \(s(t)\).
5.1.2 Matched-Filter Demodulator

Frequency-domain interpretation of the matched filter

- Since $h(t)=s(T-t)$, the Fourier transform of this relationship is:

$$H(f) = \int_0^T s(T-t)e^{-j2\pi ft} dt \quad \text{let } \tau = T-t$$

$$= \left[ \int_0^T s(\tau)e^{j2\pi f\tau} d\tau \right] e^{-j2\pi fT} = S^*(f)e^{-j2\pi fT}$$

- The matched filter has a frequency response that is the complex conjugate of the transmitted signal spectrum multiplied by the phase factor $e^{-j2\pi fT}$ (sampling delay of $T$).

- In other worlds, $|H(f)| = |S(f)|$, so that the magnitude response of the matched filter is identical to the transmitted signal spectrum.

- On the other hand, the phase of $H(f)$ is the negative of the phase of $S(f)$. 
5.1.2 Matched-Filter Demodulator

- Frequency-domain interpretation of the matched filter
  - If the signal \( s(t) \) with spectrum \( S(f) \) is passed through the matched filter, the filter output has a spectrum
    \[
    Y(f) = |S(f)|^2 e^{-j2\pi f T}.
    \]
  - The output waveform is:
    \[
    y_s(t) = \int_{-\infty}^{\infty} Y(f) e^{j2\pi f t} df = \int_{-\infty}^{\infty} |S(f)|^2 e^{-j2\pi f T} e^{j2\pi f t} df
    \]
  - By sampling the output of the matched filter at \( t = T \), we obtain (from Parseval’s relation):
    \[
    y_s(T) = \int_{-\infty}^{\infty} |S(f)|^2 df = \int_{0}^{T} s^2(t) dt = \mathcal{E}
    \]
  - The noise at the output of the matched filter has a power spectral density (from equation 2.2-27)
    \[
    \Phi_0(f) = \frac{1}{2} |H(f)|^2 N_0
    \]
5.1.2 Matched-Filter Demodulator

- Frequency-domain interpretation of the matched filter
  - The total noise power at the output of the matched filter is
    \[ P_n = \int_{-\infty}^{\infty} \Phi_0(f)df \]
    \[ = \frac{1}{2} N_0 \int_{-\infty}^{\infty} |H(f)|^2 df = \frac{1}{2} N_0 \int_{-\infty}^{\infty} |S(f)|^2 df = \frac{1}{2} \mathcal{E} N_0 \]
  - The output SNR is simply the ratio of the signal power \( P_s \), given by \( P_s = y_s^2(T) \), to the noise power \( P_n \).
    \[ \text{SNR}_0 = \frac{P_s}{P_n} = \frac{\mathcal{E}^2}{1/2 \mathcal{E} N_0} = \frac{2 \mathcal{E}}{N_0} \]
Example 5.1-2:

$M=4$ biorthogonal signals are constructed from the two orthogonal signals shown in the following figure for transmitting information over an AWGN channel. The noise is assumed to have a zero-mean and power spectral density $\frac{1}{2} N_0$. 

![Diagram of biorthogonal signals and matched filters]
Example 5.1-2: (cont.)

- The $M=4$ biorthogonal signals have dimensions $N=2$ (shown in figure a):

$$f_1(t) = \begin{cases} \sqrt{2/T} & (0 \leq t \leq \frac{1}{2} T) \\ 0 & \text{(otherwise)} \end{cases}$$

$$f_2(t) = \begin{cases} \sqrt{2/T} & \left(\frac{1}{2} T \leq t \leq T\right) \\ 0 & \text{(otherwise)} \end{cases}$$

- The impulse responses of the two matched filters are (figure b):

$$h_1(t) = f_1(T - t) = \begin{cases} \sqrt{2/T} & \left(\frac{1}{2} T \leq t \leq T\right) \\ 0 & \text{(otherwise)} \end{cases}$$

$$h_2(t) = f_2(T - t) = \begin{cases} \sqrt{2/T} & \left(0 \leq t \leq \frac{1}{2} T\right) \\ 0 & \text{(otherwise)} \end{cases}$$
Example 5.1-2: (cont.)

- If $s_1(t)$ is transmitted, the responses of the two matched filters are as shown in figure c, where the signal amplitude is $A$.

- Since $y_1(t)$ and $y_2(t)$ are sampled at $t=T$, we observe that $y_{1S}(T)=\sqrt{\frac{1}{2}}A^2T$, $y_{2S}(T)=0$.

- From equation 5.1-27, we have $\frac{1}{2}A^2T=\varepsilon$ and the received vector is:

$$r = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \left[ \sqrt{\varepsilon} + n_1 \\ n_2 \right]$$

where $n_1(t)=y_{1n}(T)$ and $n_2=y_{2n}(T)$ are the noise components at the outputs of the matched filters, given by

$$y_{kn}(T) = \int_0^T n(t) f_k(t) dt, \quad k = 1,2$$
5.1.2 Matched-Filter Demodulator

**Example 5.1-2 (cont.)**

- Clearly, $E(n_k) = E[y_{kn}(T)]$ and their variance is

  $$
  \sigma_n^2 = E[y_{kn}^2(T)] = \int_0^T \int_0^T E[n(t)n(\tau)]f_k(t)f_k(\tau)dtd\tau
  $$

  $$
  = \frac{1}{2} N_0 \int_0^T \int_0^T \delta(t-\tau)f_k(\tau)f_k(t)dtd\tau
  $$

  $$
  = \frac{1}{2} N_0 \int_0^T f_k^2(t)dt = \frac{1}{2} N_0
  $$

- Observe that the $\text{SNR}_0$ for the first matched filter is

  $$
  \text{SNR}_0 = \frac{\left(\sqrt{\mathcal{E}}\right)^2}{\frac{1}{2} N_0} = \frac{2\mathcal{E}}{N_0}
  $$
Example 5.1-2: (cont.)

- This result agrees with our previous result.
- We can also note that the four possible outputs of the two matched filters, corresponding to the four possible transmitted signals are:

\[ (r_1, r_2) = \left( \sqrt{\varepsilon} + n_1, n_2 \right), \left( n_1, \sqrt{\varepsilon} + n_2 \right), \left( -\sqrt{\varepsilon} + n_1, n_2 \right) \text{ and } \left( n_1, -\sqrt{\varepsilon} + n_2 \right) \]
5.1.3 The Optimum Detector

- Our goal is to design a signal detector that makes a decision on the transmitted signal in each signal interval based on the observation of the vector \( r \) in each interval such that the probability of a correct decision is maximized.
- We assume that there is no memory in signals transmitted in successive signal intervals.
- We consider a decision rule based on the computation of the *posterior probabilities* defined as
  \[ P(s_m|r) ≜ P(\text{signal } s_m \text{ was transmitted}|r), \quad m=1,2,\ldots,M. \]
- The decision criterion is based on selecting the signal corresponding to the maximum of the set of posterior probabilities \( \{ P(s_m|r) \} \). This decision criterion is called the *maximum a posterior probability (MAP) criterion*. 
Using Bayes’ rule, the posterior probabilities may be expressed as

\[ P(s_m | r) = \frac{p(r | s_m)P(s_m)}{p(r)} \quad \ldots (A) \]

where \( P(s_m) \) is the a priori probability of the \( m \)th signal being transmitted.

The denominator of \((A)\), which is independent of which signal is transmitted, may be expressed as

\[ p(r) = \sum_{m=1}^{M} p(r | s_m)P(s_m) \]

Some simplification occurs in the MAP criterion when the \( M \) signal are equally probable a priori, i.e., \( P(s_m) = 1/M \).

The decision rule based on finding the signal that maximizes \( P(s_m | r) \) is equivalent to finding the signal that maximizes \( P(r | s_m) \).
The conditional PDF $P(r|s_m)$ or any monotonic function of it is usually called the *likelihood function*.

The decision criterion based on the maximum of $P(r|s_m)$ over the $M$ signals is called *maximum-likelihood (ML) criterion*.

We observe that a detector based on the MAP criterion and one that is based on the ML criterion make the same decisions as long as a priori probabilities $P(s_m)$ are all equal.

In the case of an AWGN channel, the likelihood function $p(r|s_m)$ is given by:

$$p(r|s_m) = \frac{1}{(\pi N_0)^{N/2}} \exp\left[ -\sum_{k=1}^{N} \frac{(r_k - s_{mk})^2}{N_0} \right], \quad m = 1, 2, \ldots, M$$

$$\ln p(r \mid s_m) = -\frac{1}{2} N \ln(\pi N_0) - \frac{1}{N_0} \sum_{k=1}^{N} (r_k - s_{mk})^2$$
5.1.3 The Optimum Detector

- The maximum of \( \ln p(r|s_m) \) over \( s_m \) is equivalent to finding the signal \( s_m \) that minimizes the Euclidean distance:

\[
D(r, s_m) = \sum_{k=1}^{N} (r_k - s_{mk})^2
\]

- We called \( D(r,s_m) \), \( m=1,2,\ldots,M \), the distance metrics.

- Hence, for the AWGN channel, the decision rule based on the ML criterion reduces to finding the signal \( s_m \) that is closest in distance to the receiver signal vector \( r \). We shall refer to this decision rule as minimum distance detection.
Expanding the distance metrics:

\[
D(r, s_m) = \sum_{n=1}^{N} r_n^2 - 2 \sum_{n=1}^{N} r_n s_{mn} + \sum_{n=1}^{N} s_{mn}^2
\]

\[
= \|r\|^2 - 2r \cdot s_m + \|s_m\|^2, \quad m = 1, 2, \ldots, M
\]

The term \(\|r\|^2\) is common to all distance metrics, and, hence, it may be ignored in the computations of the metrics.

The result is a set of modified distance metrics.

\[
D'(r, s_m) = -2r \cdot s_m + \|s_m\|^2
\]

Note that selecting the signal \(s_m\) that minimizes \(D'(r, s_m)\) is equivalent to selecting the signal that maximizes the metrics

\[
C(r, s_m) = -D'(r, s_m),
\]

\[
C(r, s_m) = 2r \cdot s_m - \|s_m\|^2
\]
5.1.3 The Optimum Detector

- The term \( r \cdot s_m \) represents the projection of the signal vector onto each of the \( M \) possible transmitted signal vectors.

- The value of each of these projection is a measure of the correlation between the receiver vector and the \( m \)th signal. For this reason, we call \( C(r, s_m), m=1,2,\ldots,M \), the correlation metrics for deciding which of the \( M \) signals was transmitted.

- Finally, the terms \( \| s_m \|^2 = \varepsilon_m, m=1,2,\ldots,M \), may be viewed as bias terms that serve as compensation for signal sets that have unequal energies.

- If all signals have the same energy, \( \| s_m \|^2 \) may also be ignored.

- Correlation metrics can be expressed as:

\[
C(r, s_m) = 2 \int_0^T r(t)s_m(t)dt - \mathcal{E}_m, \quad m = 0,1,\ldots, M
\]
5.1.3 The Optimum Detector

These metrics can be generated by a demodulator that cross-correlates the received signal \( r(t) \) with each of the \( M \) possible transmitted signals and adjusts each correlator output for the bias in the case of unequal signal energies.
5.1.3 The Optimum Detector

We have demonstrated that the optimum ML detector computes a set of $M$ distances $D(r,s_m)$ or $D'(r,s_m)$ and selects the signal corresponding to the smallest (distance) metric.

Equivalently, the optimum ML detector computes a set of $M$ correlation metrics $C(r, s_m)$ and selects the signal corresponding to the largest correlation metric.

The above development for the optimum detector treated the important case in which all signals are equal probable. In this case, the MAP criterion is equivalent to the ML criterion.

When the signals are not equally probable, the optimum MAP detector bases its decision on the probabilities given by:

$$P(s_m | r) = \frac{p(r | s_m)P(s_m)}{p(r)}$$

or

$$PM(r, s_m) = p(r | s_m)P(s_m)$$
Example 5.1-3:

Consider the case of binary PAM signals in which the two possible signal points are \( s_1 = -s_2 = \sqrt{\varepsilon_b} \), where \( \varepsilon_b \) is the energy per bit. The priori probabilities are \( P(s_1) = p \) and \( P(s_2) = 1 - p \). Let us determine the metrics for the optimum MAP detector when the transmitted signal is corrupted with AWGN.

The receiver signal vector for binary PAM is:

\[
r = \pm \sqrt{\varepsilon_b} + y_n(T)
\]

where \( y_n(T) \) is a zero mean Gaussian random variable with variance \( \sigma_n^2 = \frac{1}{2} N_0 \).
5.1.3 The Optimum Detector

Example 5.1-3: (cont.)

The conditional PDF $P(r|s_m)$ for two signals are

$$p(r | s_1) = \frac{1}{\sqrt{2\pi\sigma_n}} \exp \left[ -\frac{(r - \sqrt{\mathcal{E}_b})^2}{2\sigma_n^2} \right]$$

$$p(r | s_2) = \frac{1}{\sqrt{2\pi\sigma_n}} \exp \left[ -\frac{(r + \sqrt{\mathcal{E}_n})^2}{2\sigma_n^2} \right]$$

Then the metrics $PM(r,s_1)$ and $PM(r,s_2)$ are

$$PM(r,s_1) = p \cdot p(r | s_1) = \frac{p}{\sqrt{2\pi\sigma_n}} \exp \left[ -\frac{(r - \sqrt{\mathcal{E}_b})^2}{2\sigma_n^2} \right]$$

$$PM(r,s_2) = (1-p) \cdot p(r | s_2) = \frac{1-p}{\sqrt{2\pi\sigma_n}} \exp \left[ -\frac{(r + \sqrt{\mathcal{E}_b})^2}{2\sigma_n^2} \right]$$
5.1.3 The Optimum Detector

Example 5.1-3: (cont.)

If \( PM(r, s_1) > PM(r, s_2) \), we select \( s_1 \) as the transmitted signal; otherwise, we select \( s_2 \). This decision rule may be expressed as:

\[
\frac{PM(r, s_1)}{PM(r, s_2)} \begin{cases} 
& s_1 \\
& s_2 
\end{cases} 1
\]

\[
PM(r, s_1) = p \exp \left[ \frac{(r + \sqrt{\mathcal{E}_b})^2 - (r - \sqrt{\mathcal{E}_b})^2}{2\sigma_n^2} \right]
\]

\[
\geq \ln \frac{1-p}{p}
\]

\[
\sqrt{\mathcal{E}_b} r \begin{cases} 
& s_1 \\
& s_2 
\end{cases} \frac{1}{2\sigma_n^2} \ln \frac{1-p}{p} = \frac{1}{4} N_0 \ln \frac{1-p}{p}
\]
Example 5.1-3: (cont.)

- The threshold is \( \frac{1}{4} N_0 \ln \frac{1-p}{p} \), denoted by \( \tau_h \), divides the real line into two regions, say \( R_1 \) and \( R_2 \), where \( R_1 \) consists of the set of points that are greater than \( \tau_h \) and \( R_2 \) consists of the set of points that are less than \( \tau_h \).

- If \( r \sqrt{\varepsilon_b} > \tau_h \), the decision is made that \( s_1 \) was transmitted.

- If \( r \sqrt{\varepsilon_b} < \tau_h \), the decision is made that \( s_2 \) was transmitted.
Example 5.1-3: (cont.)

- The threshold $\tau_h$ depends on $N_0$ and $p$. If $p=1/2$, $\tau_h=0$.
- If $p>1/2$, the signal point $s_1$ is more probable and, hence, $\tau_h<0$. In this case, the region $R_1$ is larger than $R_2$, so that $s_1$ is more likely to be selected than $s_2$.
- The average probability of error is minimized
- It is interesting to note that in the case of unequal priori probabilities, it is necessary to know not only the values of the priori probabilities but also the value of the power spectral density $N_0$, or equivalently, the noise-to-signal ratio, in order to compute the threshold.
- When $p=1/2$, the threshold is zero, and knowledge of $N_0$ is not required by the detector.
5.1.3 The Optimum Detector

Proof of “the decision rule based on the maximum-likelihood criterion minimizes the probability of error when the $M$ signals are equally probable a priori”.

- Let us denote by $R_m$ the region in the $N$-dimensional space for which we decide that signal $s_m(t)$ was transmitted when the vector $r = [r_1 r_2 \ldots r_N]$ is received.

- The probability of a correct decision given that $s_m(t)$ was transmitted is:

$$P(c \mid s_m) \cdot p(s_m) = \int_{R_m} p(r \mid s_m) \cdot p(s_m) dr$$

- The average probability of a correct decision is:

$$P(c) = \sum_{m=1}^{M} \frac{1}{M} P(c \mid s_m) = \sum_{m=1}^{M} \frac{1}{M} \int_{R_m} p(r \mid s_m) dr$$

- Note that $P(c)$ is maximized by selecting the signal $s_m$ if $p(r \mid s_m)$ is larger than $p(r \mid s_k)$ for all $m \neq k$. Q.E.D.
5.1.3 The Optimum Detector

Similarly for the MAP criterion, when the $M$ signals are not equally probable, the average probability of a correct decision is

$$P(c) = \sum_{m=1}^{M} \int_{R_m} P(s_m | r) p(r) dr$$

same for all $s_m$.

In order for $P(c)$ to be as large as possible, the points that are to be included in each particular region $R_m$ are those for which $P(s_m | r)$ exceeds all the other posterior probabilities. Q.E.D.

We conclude that MAP criterion maximize the probability of correct detection.
5.1.4 The Maximum-Likelihood Sequence Detector

- When the signal has no memory, the symbol-by-symbol detector described in the preceding section is optimum in the sense of minimizing the probability of a symbol error.

- When the transmitted signal has memory, i.e., the signals transmitted in successive symbol intervals are interdependent, the optimum detector is a detector that bases its decisions on observation of a sequence of received signals over successive signal intervals.

- In this section, we describe a maximum-likelihood sequence detection algorithm that searches for minimum Euclidean distance path through the trellis that characterizes the memory in the transmitted signal.
To develop the maximum-likelihood sequence detection algorithm, let us consider, as an example, the NRZI signal described in Section 4.3-2. Its memory is characterized by the trellis shown in the following figure:

The signal transmitted in each signal interval is binary PAM.

Hence, there are two possible transmitted signals corresponding to the signal points \( s_1 = -s_2 = \sqrt{\varepsilon_b} \), where \( \varepsilon_b \) is the energy per bit.
5.1.4 The Maximum-Likelihood Sequence Detector

The output of the matched-filter or correlation demodulator for binary PAM in the $k$th signal interval may be expressed as

$$r_k = \pm \sqrt{E_b} + n_k$$

where $n_k$ is a zero-mean Gaussian random variable with variance $\sigma_n^2 = \frac{N_0}{2}$.

The conditional PDFs for the two possible transmitted signals are

$$p(r_k \mid s_1) = \frac{1}{\sqrt{2\pi\sigma_n}} \exp\left[ -\frac{(r_k - \sqrt{E_b})^2}{2\sigma_n^2} \right]$$

$$p(r_k \mid s_2) = \frac{1}{\sqrt{2\pi\sigma_n}} \exp\left[ -\frac{(r_k + \sqrt{E_b})^2}{2\sigma_n^2} \right]$$
Since the channel noise is assumed to be white and Gaussian and \( f(t = iT), \, f(t = jT) \) for \( i \neq j \) are orthogonal, it follows that 

\[ E(n_k n_j) = 0, \, k \neq j. \]

Hence, the noise sequence \( n_1, n_2, \ldots, n_k \) is also white.

Consequently, for any given transmitted sequence \( s^{(m)} \), the joint PDF of \( r_1, r_2, \ldots, r_k \) may be expressed as a product of \( K \) marginal PDFs,

\[
p(r_1, r_2, \ldots, r_k \mid s^{(m)}) = \prod_{k=1}^{K} p(r_k \mid s_k^{(m)})
\]

\[
= \prod_{k=1}^{K} \frac{1}{\sqrt{2\pi} \sigma_n} \exp \left[ -\frac{(r_k - s_k^{(m)})^2}{2\sigma_n^2} \right]
\]

\[
= \left( \frac{1}{\sqrt{2\pi} \sigma_n} \right)^K \exp \left[ -\sum_{k=1}^{K} \frac{(r_k - s_k^{(m)})^2}{2\sigma_n^2} \right]
\]

\[(A)\]
5.1.4 The Maximum-Likelihood Sequence Detector

Then, given the received sequence \( r_1, r_2, \ldots, r_k \) at the output of the matched filter or correlation demodulator, the detector determines the sequence \( s^{(m)} = \{ s_1^{(m)}, s_2^{(m)}, \ldots, s_K^{(m)} \} \) that maximizes the conditional PDF \( p(r_1, r_2, \ldots, r_k | s^{(m)}) \). Such a detector is called the maximum-likelihood (ML) sequence-detector.

By taking the logarithm of Equation (4) and neglecting the terms that are independent of \( (r_1, r_2, \ldots, r_k) \), we find that an equivalent ML sequence detector selects the sequence \( s^{(m)} \) that minimizes the Euclidean distance metric

\[
D(r, s^{(m)}) = \sum_{k=1}^{K} (r_k - s_k^{(m)})^2
\]
In searching through the trellis for the sequence that minimizes the Euclidean distance, it may appear that we must compute the distance for every possible sequence.

For the NRZI example, which employs binary modulation, the total number of sequence is $2^K$, where $K$ is the number of outputs obtained from the demodulator.

However, this is not the case. We may reduce the number of the sequences in the trellis search by using the Viterbi algorithm to eliminate sequences as new data is received from the demodulator.

The Viterbi algorithm is a sequence trellis search algorithm for performing ML sequence detection. We describe it below in the context of the NRZI signal. We assume that the search process begins initially at state $S_0$. 

5.1.4 The Maximum-Likelihood Sequence Detector
Viterbi Decoding Algorithm

Basic concept

- Generate the code trellis at the decoder
- The decoder penetrates through the code trellis *level by level* in search for the transmitted code sequence
- At each level of the trellis, the decoder computes and compares the metrics of all the partial paths entering a node
- The decoder *stores* the partial path with the larger metric and *eliminates* all the other partial paths. The stored partial path is called the *survivor*. 
The corresponding trellis is shown in the following figure:

At time \( t = T \), we receive \( r_1 = s_1^{(m)} + n_1 \) from the demodulator, and at \( t = 2T \), we receive \( r_2 = s_2^{(m)} + n_2 \). Since the signal memory is one bit, which we denote by \( L = 1 \), we observe that the trellis reaches its regular (steady state) form after two transitions.
Thus, upon receipt of \( r_2 \) at \( t=2T \), we observe that there are two signal paths entering each of the nodes and two signal paths leaving each node. The two paths entering node \( S_0 \) at \( t=2T \) correspond to the information bits (0,0) and (1,1) or, equivalently, to the signal points \( (-\sqrt{\epsilon_b},-\sqrt{\epsilon_b}) \) and \( (\sqrt{\epsilon_b},-\sqrt{\epsilon_b}) \), respectively.

The two paths entering node \( S_1 \) at \( t=2T \) correspond to the information bits (0,1) and (1,0) or, equivalently, to the signal points \( (-\sqrt{\epsilon_b},\sqrt{\epsilon_b}) \) and \( (\sqrt{\epsilon_b},\sqrt{\epsilon_b}) \), respectively.

For the two paths entering node \( S_0 \), we compute the two Euclidean distance metrics.

\[
D_0(0,0) = (r_1 + \sqrt{\epsilon_b})^2 + (r_2 + \sqrt{\epsilon_b})^2
\]

\[
D_0(1,1) = (r_1 - \sqrt{\epsilon_b})^2 + (r_2 + \sqrt{\epsilon_b})^2
\]

by using the outputs \( r_1 \) and \( r_2 \) from demodulator.
5.1.4 The Maximum-Likelihood Sequence Detector

The Viterbi algorithm compares these two metrics and discards the path having the larger metrics. The other path with the lower metric is saved and is called survivor at $t=2T$.

Similarly, for two paths entering node $S_1$ at $t=2T$, we compute the two Euclidean distance metrics

$$D_1(0,1) = (r_1 + \sqrt{\mathcal{E}_b})^2 + (r_2 - \sqrt{\mathcal{E}_b})^2$$

$$D_1(1,0) = (r_1 - \sqrt{\mathcal{E}_b})^2 + (r_2 - \sqrt{\mathcal{E}_b})^2$$

by using the output $r_1$ and $r_2$ from the demodulator. The two metrics are compared and the signal path with the larger metric is eliminated.

Thus, at $t=2T$, we are left with two survivor paths, one at node $S_0$ and the other at node $S_1$, and their corresponding metrics.
5.1.4 The Maximum-Likelihood Sequence Detector

Upon receipt of $r_3$ at $t=3T$, we compute the metrics of the two paths entering state $S_0$. Suppose the survivors at $t=2T$ are the paths $(0,0)$ at $S_0$ and $(0,1)$ at $S_1$. Then, the two metrics for the paths entering $S_0$ at $t=3T$ are

$$D_0(0,0,0) = D_0(0,0) + (r_3 + \sqrt{\mathcal{E}_b})^2$$

$$D_0(0,1,1) = D_1(0,1) + (r_3 + \sqrt{\mathcal{E}_b})^2$$

These two metrics are compared and the path with the larger (greater-distance) metric is eliminated.

Similarly, the metrics for the two paths entering $S_1$ at $t=3T$ are

$$D_1(0,0,1) = D_0(0,0) + (r_3 - \sqrt{\mathcal{E}_b})^2$$

$$D_1(0,1,0) = D_1(0,1) + (r_3 - \sqrt{\mathcal{E}_b})^2$$

These two metrics are compared and the path with the larger metrics is eliminated.
It is relatively easy to generalize the trellis search performed by the Viterbi algorithm for $M$-ary modulation.

Delay modulation with $M=4$ signals: such signal is characterized by the four-state trellis shown in the following figure.
Delay modulation with $M=4$ signals: (cont.)

- We observe that each state has two signal paths entering and two signals paths leaving each node.
- The memory of the signal is $L=1$.
- The Viterbi algorithm will have four survivors at each stage and their corresponding metrics.
- Two metrics corresponding to the two entering paths are computed at each node, and one of the two signal paths entering the node is eliminated at each state of the trellis.
- The Viterbi algorithm minimizes the number of trellis paths searched in performing ML sequence detection.

From the description of the Viterbi algorithm given above, it is unclear as to how decisions are made on the individual detected information symbols given the surviving sequences.
5.1.4 The Maximum-Likelihood Sequence Detector

- If we have advanced to some stage, say $K$, where $K \gg L$ in the trellis, and we compare the surviving sequences, we shall find that with probability approaching one all surviving sequences will be identical in bit (or symbol) positions $K-5L$ and less.

- In a practical implementation of the Viterbi algorithm, decisions on each information bit (or symbol) are forced after a delay of $5L$ bits, and, hence, the surviving sequences are truncated to the $5L$ most recent bits (or symbol). Thus, a variable delay in bit or symbol detection is avoided.

- The loss in performance resulting from the sub-optimum detection procedure is negligible if the delay is at least $5L$. 
In this section, we describe a maximum a posteriori probability algorithm that makes decisions on a symbol-by-symbol basis, but each symbol decision is based on an observation of a sequence of received signal vectors. Hence, this detection is optimum in the sense that is minimizes the probability of a symbol error.

The detection algorithm that is presented below is due to Abend and Fritchman (1970), who developed it as a detection algorithm for channels with intersymbol interference, i.e., channels with memory.
We illustrate the algorithm in the context of detection a PAM signal with $M$ possible levels. Suppose that it is desired to detect the information symbol transmitted in the $k$th signal interval, and let $r_1, r_2, \ldots, r_{k+D}$ be the observed received sequence, where $D$ is the delay parameter which is chosen to exceed the signal memory, i.e., $D >> L$, where $L$ is the inherent memory in the signal. On the basis of the received sequence, we compute the posterior probabilities

\[
P(s^{(k)} = A_m \mid r_{k+D}, r_{k+D-1}, \ldots, r_1)
\]

for the $M$ possible symbol values and choose the symbol with the largest probability.
Since
\[
P\left(s^{(k)} = A_m \mid r_{k+D}, \ldots, r_1 \right) = \frac{p(r_{k+D}, \ldots, r_1 \mid s^{(k)} = A_m) \cdot P(s^{(k)} = A_m)}{p(r_{k+D}, r_{k+D-1}, \ldots, r_1)} \tag{A}
\]

and since the denominator is common for all $M$ probabilities, the MAP criterion is equivalent to choosing the value of $s^{(k)}$ that maximizes the numerator of (A). Thus, the criterion for deciding on the transmitted symbol $s^{(k)}$ is

\[
\hat{s}^{(k)} = \arg\max_{s^{(k)}} \left\{ p(r_{k+D}, \ldots, r_1 \mid s^{(k)} = A_m) \cdot P(s^{(k)} = A_m) \right\} \tag{B}
\]

When the symbol are equally probable, the probability $P(s^{(k)} = A_m)$ may be dropped from the computation.
The algorithm for computing the probabilities in Equation (B) recursively begins with the first symbol \( s^{(1)} \). We have

\[
\tilde{s}^{(1)} = \arg\left\{ \max_{s^{(1)}} p(r_{1+D}, \ldots, r_1 \mid s^{(1)} = A_m) P(s^{(1)} = A_m) \right\}
\]

\[
= \arg\left\{ \max_{s^{(1)}} \sum_{s^{1+D}} \cdots \sum_{s^{(2)}} p(r_{1+D}, \ldots, r_1 \mid s^{(1+D)}, \ldots, s^{(1)}) P(s^{(1+D)}, \ldots, s^{(1)}) \right\}
\]

\[
= \arg\left\{ \max_{s^{(1)}} \sum_{s^{(1+D)}} \cdots \sum_{s^{(2)}} p_1(s^{(1+D)}, \ldots, s^{(2)}, s^{(1)}) \right\}
\]

where \( \tilde{s}^{(1)} \) denotes the decision on \( s^{(1)} \).
For mathematical convenience, we have defined

\[ P_1(s^{(1+D)}, \ldots, s^{(2)}, s^{(1)}) = p(r_{1+D}, \ldots, r_1 \mid s^{(1+D)}, \ldots, s^{(1)})P(s^{(1+D)}, \ldots, s^{(1)}) \]

The joint probability \( P_1(s^{(1+D)}, \ldots, s^{(2)}, s^{(1)}) \) may be omitted if the symbols are equally probable and statistically independent. As a consequence of the statistical independent of the additive noise sequence, we have

\[
p(r_{1+D}, \ldots, r_1 \mid s^{(1+D)}, \ldots, s^{(1)})
\]

\[
= p(r_{1+D} \mid s^{(1+D)}, \ldots, s^{(1+D-L)}) p(r_D \mid s^{(D)}, \ldots, s^{(D-L)})
\]

\[
\ldots p(r_2 \mid s^{(2)}, s^{(1)}) p(r_1 \mid s^{(1)})
\]

where we assume that \( s^{(k)} = 0 \) for \( k \leq 0 \)
For detection of the symbol \( s^{(2)} \), we have

\[
\tilde{s}^{(2)} = \arg\left\{ \max_{s^{(2)}} p\left( r_{2+D}, \ldots, r_1 \mid s^{(2)} = A_m \right) P\left( s^{(2)} = A_m \right) \right\} \quad --(C)
\]

\[
= \arg\left\{ \max_{s^{(2)}} \sum_{s^{(2+D)}} \cdots \sum_{s^{(3)}} p\left( r_{2+D}, \ldots, r_1 \mid s^{(2+D)}, \ldots, s^{(2)} \right) P\left( s^{(2+D)}, \ldots, s^{(2)} \right) \right\}
\]

The joint conditional probability in the multiple summation can be expressed as

\[
p\left( r_{2+D}, \ldots, r_1 \mid s^{(2+D)}, \ldots, s^{(2)} \right) = p\left( r_{2+D} \mid s^{(2+D)}, \ldots, s^{(2+D-L)} \right) p\left( r_{1+D}, \ldots, r_1 \mid s^{(1+D)}, \ldots, s^{(2)} \right) \quad --(D)
\]
Furthermore, the joint probability

\[
p\left(r_{1+D}, \ldots, r_1 \mid s^{(1+D)}, \ldots, s^{(2)}\right)P\left(s^{(1+D)}, \ldots, s^{(2)}\right)
\]

can be obtained from the probabilities computed previously in the detection of \(s^{(1)}\). That is,

\[
p\left(r_{1+D}, \ldots, r_1 \mid s^{(1+D)}, \ldots, s^{(2)}\right) = \sum_{s^{(1)}} p\left(r_{1+D}, \ldots, r_1 \mid s^{(1+D)}, \ldots, s^{(1)}\right)P\left(s^{(1+D)}, \ldots, s^{(1)}\right) \quad \text{--(E)}
\]

\[
= \sum_{s^{(1)}} p_1\left(s^{(1+D)}, \ldots, s^{(2)}, s^{(1)}\right)
\]
By combining Equation (D) and (E) and then substituting into Equation (C), we obtain

\[ \tilde{s}^{(2)} = \text{arg} \left\{ \max_{s^{(2)}} \sum_{s^{(2+D)}} \ldots \sum_{s^{(3)}} p_2 \left( s^{(2+D)}, \ldots, s^{(3)}, s^{(2)} \right) \right\} \]

where, by definition,

\[
p_2 \left( s^{(2+D)}, \ldots, s^{(3)}, s^{(2)} \right) =
\]

\[ p \left( r_{2+D} \mid s^{(2+D)}, \ldots, s^{(2+D-L)} \right) P \left( s^{(2+D)} \right) \sum_{s^{(1)}} p_1 \left( s^{(1+D)}, \ldots, s^{(2)}, s^{(1)} \right) \]
In general, the recursive algorithm for detection the symbol $s^{(k)}$ is as follows: upon reception of $r_{K+D}, \ldots, r_2, r_1$, we compute

$$\tilde{s}^{(k)} = \arg \max_{s^{(k)}} p(r_{k+D}, \ldots, r_1 | s^{(k)}) P(s^{(k)})$$

$$= \arg \max_{s^{(k)}} \sum_{s^{(k+D)}} \cdots \sum_{s^{(k+1)}} p_k(s^{(k+D)}, \ldots, s^{(k+1)}, s^{(k)})$$

where, by definition,

$$p_k(s^{(k+D)}, \ldots, s^{(k+1)}, s^{(k)})$$

$$= p(r_{k+D} | s^{(k+D)}, \ldots, s^{(k+D-L)}) P(s^{(k+D)}) \sum_{s^{(k-1)}} p_{k-1}(s^{(k-1+D)}, \ldots, s^{(k-1)})$$

Thus, the recursive nature of the algorithm is established by relations $(F)$ and $(G)$. 

---

5.1.5 A Symbol-by-Symbol MAP Detector for Signals with Memory
The major problem with the algorithm is its computational complexity. In particular, the averaging performed over the symbol $s^{(k+D)}, \ldots, s^{(k+1)}, s^{(k)}$ in Equation (F) involves a large amount of computation per received signal, especially if the number $M$ of amplitude levels $\{A_m\}$ is large. On the other hand, if $M$ is small and the memory $L$ is relatively short, this algorithm is easily implemented.
5.2.1 Probability of Error for Binary Modulation

- Let us consider binary PAM signals where the two signal waveforms are \( s_1(t) = g(t) \) and \( s_1(t) = -g(t) \), and \( g(t) \) is an arbitrary pulse that is nonzero in the interval \( 0 \leq t \leq T_b \) and zero elsewhere.
- Since \( s_1(t) = -s_2(t) \), these signals are said to be antipodal.
- The energy in the pulse \( g(t) \) is \( \varepsilon_b \).
- As indicated in section 4.3.1, PAM signals are one-dimensional, and, their geometric representation is simply the one-dimensional vector \( s_1 = \sqrt{\varepsilon_b} \), \( s_2 = -\sqrt{\varepsilon_b} \).

Figure (A)
5.2.1 Probability of Error for Binary Modulation

Let us assume that the two signals are equally likely and that signal $s_1(t)$ was transmitted. Then, the received signal from the (matched filter or correlation) demodulator is

$$ r = s_1 + n = \sqrt{\varepsilon_b} + n $$

where $n$ represents the additive Gaussian noise component, which has zero mean and variance $\sigma_n^2 = \frac{1}{2} N_0$.

In this case, the decision rule based on the correlation metric given by Equation 5.1-44 compares $r$ with the threshold zero. If $r > 0$, the decision is made in favor of $s_1(t)$, and if $r < 0$, the decision is made that $s_2(t)$ was transmitted.
5.2.1 Probability of Error for Binary Modulation

The two conditional PDFs of $r$ are:

$$p(r \mid s_1) = \frac{1}{\sqrt{\pi N_0}} e^{-\frac{(r - \sqrt{\varepsilon_b})^2}{N_0}}$$

$$p(r \mid s_2) = \frac{1}{\sqrt{\pi N_0}} e^{-\frac{(r + \sqrt{\varepsilon_b})^2}{N_0}}$$
5.2.1 Probability of Error for Binary Modulation

Given that $s_1(t)$ was transmitted, the probability of error is simply the probability that $r < 0$.

$$P(e | s_1) = \int_{-\infty}^{0} p(r | s_1) dr = \frac{1}{\sqrt{\pi N_0}} \int_{-\infty}^{0} \exp \left[ - \frac{(r - \sqrt{E_b})^2}{N_0} \right] dr$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\sqrt{2E_b/N_0}} e^{-x^2/2} dx \quad x = \frac{r - \sqrt{E_b}}{\sqrt{N_0}/2}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{\sqrt{2E_b/N_0}}^{\infty} e^{-x^2/2} dx \quad (x = -x)$$

$$= Q \left( \frac{\sqrt{2E_b}}{N_0} \right)$$

$$Q(t) = \frac{1}{\sqrt{2\pi}} \int_{t}^{\infty} e^{-x^2/2} dx \quad t \geq 0$$
5.2.1 Probability of Error for Binary Modulation

If we assume that $s_2(t)$ was transmitted, $r = -\sqrt{\epsilon_b} + n$ and the probability that $r > 0$ is also $P(e|s_2) = Q\left(\sqrt{2\epsilon_b / N_0}\right)$. Since the signal $s_1(t)$ and $s_2(t)$ are equally likely to be transmitted, the average probability of error is

$$P_b = \frac{1}{2} P(e|s_1) + \frac{1}{2} P(e|s_2) = Q\left(\sqrt{\frac{2\epsilon_b}{N_0}}\right) \quad -(A)$$

Two important characteristics of this performance measure:

- First, we note that the probability of error depends only on the ratio $\epsilon_b / N_0$.
- Secondly, we note that $2\epsilon_b / N_0$ is also the output SNR$_0$ from the matched-filter (and correlation) demodulator.
- The ratio $\epsilon_b / N_0$ is usually called the signal-to-noise ratio per bit.
5.2.1 Probability of Error for Binary Modulation

- We also observe that the probability of error may be expressed in terms of the distance between the two signals $s_1$ and $s_2$.

- From figure (A), we observe that the two signals are separated by the distance $d_{12} = 2 \sqrt{\epsilon_b}$. By substituting $\epsilon_b = \frac{1}{4} d_{12}^2$ into Equation (A), we obtain

$$P_b = Q\left( \sqrt{\frac{d_{12}^2}{2N_0}} \right)$$

- This expression illustrates the dependence of the error probability on the distance between the two signals points.
5.2.1 Probability of Error for Binary Modulation

- Error probability for binary orthogonal signals
- The signal vectors $s_1$ and $s_2$ are two-dimensional.

\[
s_1 = \begin{bmatrix} \sqrt{\varepsilon_b} \\ 0 \end{bmatrix}
\]

\[
s_2 = \begin{bmatrix} 0 \\ \sqrt{\varepsilon_b} \end{bmatrix}
\]

where $\varepsilon_b$ denote the energy for each of the waveforms. Note that the distance between these signal points is

\[
d_{12} = \sqrt{2\varepsilon_b}
\]
5.2.1 Probability of Error for Binary Modulation

- **Error probability for binary orthogonal signals**
  - To evaluate the probability of error, let us assume that $s_1$ was transmitted. Then, the received vector at the output of the demodulator is $r = [\sqrt{\varepsilon_b} + n_1 \quad n_2]$.
  - We can now substitute for $r$ into the correlation metrics given by $C(r, s_m) = 2r \cdot s_m - \|s_m\|^2$ to obtain $C(r, s_1)$ and $C(r, s_2)$.
  - The probability of error is the probability that $C(r, s_2) > C(r, s_1)$.

\[
\Rightarrow \left(2n_2\sqrt{\varepsilon_b} - \varepsilon_b\right) > \left(2\varepsilon_b + 2n_1\sqrt{\varepsilon_b} - \varepsilon_b\right)
\]

\[
P(e | s_1) = P[C(r, s_2) > C(r, s_2)] = P[n_2 - n_1 > \sqrt{\varepsilon_b}]
\]
5.2.1 Probability of Error for Binary Modulation

Error probability for binary orthogonal signals

Since $n_1$ and $n_2$ are zero-mean statistically independent Gaussian random variables each with variance $\frac{1}{2} N_0$, the random variable $x = n_2 - n_1$ is zero-mean Gaussian with variance $N_0$. Hence,

$$P\left(n_2 - n_1 > \sqrt{\varepsilon_b}\right) = \frac{1}{\sqrt{2\pi N_0}} \int_{\sqrt{\varepsilon_b}}^{\infty} e^{-x^2/2N_0} \, dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{\sqrt{\varepsilon_b/N_0}}^{\infty} e^{-x^2/2} \, dx = Q\left(\sqrt{\frac{\varepsilon_b}{N_0}}\right)$$

The same error probability is obtained when we assume that $s_2$ is transmitted:

$$P_b = Q\left(\sqrt{\frac{\varepsilon_b}{N_0}}\right) = Q\left(\sqrt{\gamma_b}\right) \text{ where } \gamma_b \text{ is the SNR per bit.}$$
If we compare the probability of error for binary antipodal signals with that for binary orthogonal signals, we find that orthogonal signals required a factor of 2 increase in energy to achieve the same error probability as antipodal signals.

Since $10 \log_{10} 2 = 3$ dB, we say that orthogonal signals are 3 dB poorer than antipodal signals. The difference of 3 dB is simply due to the distance between the two signal points, which is $d_{12}^2 = 2 \varepsilon_b$ for orthogonal signals, whereas $d_{12}^2 = 4 \varepsilon_b$ for antipodal signals.

The error probability versus $10 \log_{10} \varepsilon_b/N_0$ for these two types of signals is shown in the following figure ($B$). As observed from this figure, at any given error probability, the $\varepsilon_b/N_0$ required for orthogonal signals is 3 dB more than that for antipodal signals.
5.2.1 Probability of Error for Binary Modulation

- Probability of error for binary signals

![Graph showing the probability of error for binary signals with respect to SNR per bit, $\gamma_b$ (dB)].

Figure (B)
5.2.2 Probability of Error for M-ary Orthogonal Signals

For equal-energy orthogonal signals, the optimum detector selects the signal resulting in the largest cross correlation between the received vector \( \mathbf{r} \) and each of the \( M \) possible transmitted signals vectors \( \{s_m\} \), i.e.,

\[
C(r, s_m) = r \cdot s_m = \sum_{k=1}^{M} r_k s_{mk}, \quad m = 1, 2, \ldots, M
\]

To evaluate the probability of error, let us suppose that the signal \( s_1 \) is transmitted. Then the received signal vector is

\[
\mathbf{r} = \left[ \sqrt{\varepsilon_s} + n_1, n_2, n_3, \ldots, n_M \right]
\]

where \( \varepsilon_s \) denotes the symbol energy and \( n_1, n_2, \ldots, n_M \) are zero-mean, mutually statistically independent Gaussian random variable with equal variance \( \sigma_n^2 = \frac{1}{2} N_0 \).
5.2.2 Probability of Error for M-ary Orthogonal Signals

In this case, the outputs from the bank of $M$ correlations are

$$C(r, s_1) = \sqrt{\varepsilon_s} \left( \sqrt{\varepsilon_s} + n_1 \right)$$

$$C(r, s_2) = \sqrt{\varepsilon_s} n_2$$

$$\vdots$$

$$C(r, s_M) = \sqrt{\varepsilon_s} n_M$$

Note that the scale factor $\varepsilon_s$ may be eliminated from the correlator outputs dividing each output by $\sqrt{\varepsilon_s}$. With this normalization, the PDF of the first correlator output is

$$p_n(x_1) = \frac{1}{\sqrt{\pi N_0}} \exp \left[ -\frac{(x_1 - \sqrt{\varepsilon_s})^2}{N_0} \right]$$
5.2.2 Probability of Error for M-ary Orthogonal Signals

And the PDFs of the other $M-1$ correlator outputs are

$$p_{r_m}(x_m) = \frac{1}{\sqrt{\pi N_0}} e^{-x_m^2/N_0}, \quad m = 2, 3, \ldots, M$$

It is mathematically convenient to first derive the probability that the detector makes a correct decision. This is the probability that $r_1$ is large than each of the other $M-1$ correlator outputs $n_2, n_3, \ldots, n_M$. This probability may be expressed as

$$P_c = \int_{-\infty}^{\infty} P(n_2 < r_1, n_3 < r_1, \ldots, n_M < r_1 | r_1) p(r_1) dr_1$$

where $P(n_2 < r_1, n_3 < r_1, \ldots, n_M < r_1 | r_1)$ denotes the joint probability that $n_2, n_3, \ldots, n_M$ are all less than $r_1$, conditioned on any given $r_1$. Then this joint probability is averaged over all $r_1$. 
5.2.2 Probability of Error for M-ary Orthogonal Signals

Since the \( \{r_m\} \) are statistically independent, the joint probability factors into a product of \( M-1 \) marginal probabilities of the form:

\[
P(n_m < r_1 | r_1) = \int_{-\infty}^{r_1} p_{r_m}(x_m) dx_m \quad m = 2,3,\ldots,M \quad -(B)
\]

\[
= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{r_1\sqrt{2/N_0}} e^{-x^2/2} dx \quad x = \sqrt{2/N_0} x_m
\]

This probabilities are identical for \( m=2,3,\ldots,M \), and, the joint probability under consideration is simply the result in Equation (B) raised to the \( (M-1) \)th power. Thus, the probability of a correct decision is

\[
P_c = \int_{-\infty}^{\infty} \left( \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{r_1\sqrt{2/N_0}} e^{-x^2/2} dx \right)^{M-1} p(r_1) dr_1
\]
The probability of a \((k\text{-bit})\) symbol error is

\[
P_M = 1 - P_c
\]

\[
P_M = 1 - \left\{ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left( \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-x^2/2} \, dx \right)^{M-1} \exp \left[ -\frac{1}{2} \left( y - \sqrt{\frac{2\varepsilon_s}{N_0}} \right)^2 \right] \, dy \right\}
\]

--- (C)

The same expression for the probability of error is obtained when any one of the other \(M-1\) signals is transmitted. Since all the \(M\) signals are equally likely, the expression for \(P_M\) given above is the average probability of a symbol error.

This expression can be evaluated numerically.
In comparing the performance of various digital modulation methods, it is desirable to have the probability of error expressed in terms of the SNR per bit, $\varepsilon_b/N_0$, instead of the SNR per symbol, $\varepsilon_s/N_0$.

With $M=2^k$, each symbol conveys $k$ bits of information, and hence $\varepsilon_s = k \varepsilon_b$. Thus, Equation (C) may be expressed in terms of $\varepsilon_b/N_0$ by substituting for $\varepsilon_s$.

It is also desirable to convert the probability of a symbol error into an equivalent probability of a binary digit error. For equiprobable orthogonal signals, all symbol errors are equiprobable and occur with probability

$$\frac{P_M}{M - 1} = \frac{P_M}{2^k - 1}$$
Furthermore, there are \( \binom{k}{n} \) ways in which \( n \) bits out of \( k \) may be in error. Hence, the average number of bit errors per \( k \)-bit symbol is

\[
\sum_{n=1}^{k} \binom{k}{n} \frac{P_M}{2^k - 1} = k \frac{2^{k-1}}{2^k - 1} P_M \tag{D}
\]

and the average bit error probability is just the result in Equation (D) divided by \( k \), the number of bits per symbol. Thus,

\[
P_b = \frac{2^{k-1}}{2^k - 1} P_M \approx \frac{P_M}{2} \quad k \gg 1
\]

The graphs of the probability of a binary digit error as a function of the SNR per bit, \( \varepsilon_b/N_0 \), are shown in Figure (C) for \( M=2,4,8,16,32, \) and 64. This figure illustrates that, by increasing the number \( M \) of waveforms, one can reduce the SNR per bit required to achieve a given probability of a bit error.
5.2.2 Probability of Error for M-ary Orthogonal Signals

For example, to achieve a $P_b=10^{-5}$, the required SNR per bit is a little more than 12dB for $M=2$, but if $M$ is increased to 64 signal waveforms, the required SNR per bit is approximately 6dB. Thus, a savings of over 6dB is realized in transmitter power required to achieve a $P_b=10^{-5}$ by increasing $M$ from $M=2$ to $M=64$. 

Figure (C)
5.2.2 Probability of Error for M-ary Orthogonal Signals

- What is the minimum required $\varepsilon_b/N_0$ to achieve an arbitrarily small probability of error as $M \to \infty$?

- A union bound on the probability of error.

Let us investigate the effect of increasing $M$ on the probability of error for orthogonal signals. To simplify the mathematical development, we first derive an upper bound on the probability of a symbol error that is much simpler than the exact form given in the following Equation (5.2-21)

\[
P_M = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left[ 1 - \left( \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-x^2/2} \, dx \right)^{M-1} \right] \exp \left[ -\frac{1}{2} \left( y - \sqrt{\frac{2\varepsilon_s}{N_0}} \right)^2 \right] \, dy
\]
5.2.2 Probability of Error for \( M \)-ary Orthogonal Signals

A union bound on the probability of error (cont.)

Recall that the probability of error for binary orthogonal signals is given by 5.2-11:

\[
P_b = Q \left( \sqrt{\frac{\varepsilon_b}{N_0}} \right) = Q \left( \sqrt{\gamma_b} \right)
\]

Now, if we view the detector for \( M \) orthogonal signals as one that makes \( M - 1 \) binary decisions between the correlator outputs \( C(r,s_1) \) that contains the signal and the other \( M - 1 \) correlator outputs \( C(r,s_m) \), \( m=2,3,\ldots,M \), the probability of error is upper-bounded by union bound of the \( M - 1 \) events. That is, if \( E_i \) represents the event that \( C(r,s_i) > C(r,s_1) \) for \( i \neq 1 \), then we have

\[
P_M = P \left( \bigcup_{i=2}^{M} E_i \right) \leq \sum_{i=2}^{M} P(E_i)
\]

Hence,

\[
P_M \leq (M - 1) P_b = (M - 1) Q \left( \sqrt{\frac{\varepsilon_s}{N_0}} \right) < MQ \left( \sqrt{\frac{\varepsilon_s}{N_0}} \right)
\]
5.2.2 Probability of Error for M-ary Orthogonal Signals

A union bound on the probability of error (cont.)

This bound can be simplified further by upper-bounding

$$Q\left(\sqrt{\frac{\varepsilon_s}{N_0}}\right)$$

We have

$$Q\left(\sqrt{\frac{\varepsilon_s}{N_0}}\right) < e^{-\varepsilon_s/2N_0} \quad \text{--(E)}$$

thus,

$$P_M < Me^{-\varepsilon_s/2N_0} = 2^k e^{-k\varepsilon_b/2N_0} \quad \text{--(F)}$$

$$P_M < e^{-k(\varepsilon_b/N_0 - 2\ln2)/2}$$

As $k \to \infty$ or equivalently, as $M \to \infty$, the probability of error approaches zero exponentially, provided that $\varepsilon_b/N_0$ is greater than $2\ln2$,

$$\frac{\varepsilon_b}{N_0} > 2\ln2 = 1.39 \quad (1.42\text{dB})$$
A union bound on the probability of error (cont.)

The simple upper bound on the probability of error given by Equation (F) implies that, as long as \( \text{SNR} > 1.42 \, \text{dB} \), we can achieve an arbitrarily low \( P_M \).

However, this union bound is not a very tight upper bound as a sufficiently low SNR due to the fact that upper bound for the \( Q \) function in Equation (E) is loose.

In fact, by more elaborate bounding techniques, it is shown in Chapter 7 that the upper bound in Equation (F) is sufficiently tight for \( \frac{\varepsilon_b}{N_0} > 4 \ln 2 \).

For \( \frac{\varepsilon_b}{N_0} < 4 \ln 2 \), a tighter upper bound on \( P_M \) is

\[
P_M < 2e^{-k\left(\sqrt{\frac{\varepsilon_0}{N_0}} - \sqrt{\ln 2}\right)^2}
\]
5.2.2 Probability of Error for M-ary Orthogonal Signals

- **A union bound on the probability of error** (cont.)
  - Consequently, $P_M \rightarrow 0$ as $k \rightarrow \infty$, provided that
  \[
  \frac{\varepsilon_b}{N_0} > \ln 2 = 0.693 \quad (-1.6 \text{dB})
  \]

- Hence, $-1.6$ dB is the minimum required SNR per bit to achieve an arbitrarily small probability of error in the limit as $k \rightarrow \infty (M \rightarrow \infty)$. This minimum SNR per bit is called the *Shannon limit* for an additive Gaussian noise channel.
5.2.3 Probability of Error for M-ary Biorthogonal Signals

As indicated in Section 4.3, a set of $M=2^k$ biorthogonal signals are constructed from $\frac{1}{2}M$ orthogonal signals by including the negatives of the orthogonal signals. Thus, we achieve a reduction in the complexity of the demodulator for the biorthogonal signals relative to that for orthogonal signals, since the former is implemented with $\frac{1}{2}M$ cross correlation or matched filters, whereas the latter required $M$ matched filters or cross correlators.

Let us assume that the signal $s_1(t)$ corresponding to the vector $s_1 = [\sqrt{\varepsilon_s} \ 0 \ 0 \ldots 0]$ was transmitted. The received signal vector is

$$r = [\sqrt{\varepsilon_s} + n_1 \ n_2 \ldots n_{M/2}]$$

where the $\{n_m\}$ are zero-mean, mutually statistically independent and identically distributed Gaussian random variables with variance $\sigma_n^2 = \frac{1}{2} N_0$. 

Fall, 2004.
The optimum detector decides in favor of the signal corresponding to the largest in magnitude of the cross correlators

\[ C(r, s_m) = r \cdot s_m = \sum_{k=1}^{M/2} r_k s_{mk}, \quad m = 1, 2, \ldots, \frac{1}{2} M \]

while the sign of this largest term is used to decide whether \( s_m(t) \) or \( -s_m(t) \) was transmitted.

According to this decision rule, the probability of a correct decision is equal to the probability that \( r_1 = \sqrt{\varepsilon_s} + n_1 > 0 \) and \( r_1 \) exceeds \( |r_m| = |n_m| \) for \( m = 2, 3, \ldots, \frac{1}{2} M \). But

\[
P(|n_m| < r_1 | r_1 > 0) = \frac{1}{\sqrt{\pi N_0}} \int_{-r_1}^{r_1} e^{-x^2/N_0} dx = \frac{1}{\sqrt{2\pi}} \int_{-r_1/\sqrt{N_0/2}}^{r_1/\sqrt{N_0/2}} e^{-y^2/2} dy
\]
Then, the probability of a correct decision is

\[ P_c = \int_0^\infty \left( \frac{1}{\sqrt{2\pi}} \int_{-n/\sqrt{N_0/2}}^{n/\sqrt{N_0/2}} e^{-x^2/2} \right)^{M/2-1} p(r_1) \, dr_1 \]

from which, upon substitution for \( p(r_1) \), we obtain

\[ P_c = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{2\varepsilon_s/N_0}}^{\sqrt{2\varepsilon_s/N_0}} \left( \frac{1}{\sqrt{2\pi}} \int_{-(v+\sqrt{2\varepsilon_s/N_0})}^{v+\sqrt{2\varepsilon_s/N_0}} e^{-x^2/2} \, dx \right)^{M/2-1} e^{-v^2/2} \, dv \quad \text{--(G)} \]

\[ p_{r_1}(x_1) = \frac{1}{\sqrt{\pi N_0}} \exp \left[ -\frac{(x_1 - \sqrt{\varepsilon_s})^2}{N_0} \right] \quad v = \left( r_1 - \sqrt{\varepsilon_s} \right) / \sqrt{\frac{N_0}{2}} \]

where we have used the PDF of \( r_1 \) given in Equation 5.2-15.

The probability of a symbol error \( P_M = 1 - P_c \)
5.2.3 Probability of Error for M-ary Biorthogonal Signals

- $P_c$, and $P_M$ may be evaluated numerically for different values of $M$. The graph shown in the following figure (D) illustrates $P_M$ as a function of $\varepsilon_b/N_0$, where $\varepsilon_s = k \varepsilon_b$, for $M=2,4,8,16$ and 32.
5.2.3 Probability of Error for M-ary Biorthogonal Signals

- In this case, the probability of error for $M=4$ is greater than that for $M=2$. This is due to the fact that we have plotted the symbol error probability $P_M$ in Figure(D).

- If we plotted the equivalent bit error probability, we should find that the graphs for $M=2$ and $M=4$ coincide.

- As in the case of orthogonal signals, as $M \to \infty$ ($k \to \infty$), the minimum required $\varepsilon_b/N_0$ to achieve an arbitrarily small probability of error is $-1.6$ dB, the Shannon limit.
Next we consider the probability of error for $M$ simplex signals. Recall from Section 4.3 that simplex signals are a set of $M$ equally correlated with mutual cross-correlation coefficient $\rho_{mn} = -1/(M-1)$.

These signals have the same minimum separation of $\sqrt{2\varepsilon_s}$ between adjacent signal points in $M$-dimensional space as orthogonal signals. They achieve this mutual separation with a transmitted energy of $\varepsilon_s (M-1)/M$, which is less than that required for orthogonal signals by a factor of $(M-1)/M$.

Consequently, the probability of error for simplex signals is identical to the probability of error for orthogonal signals, but this performance is achieved with saving of

$$10 \log (1 - \rho) = 10 \log \frac{M}{M - 1} \text{ dB}$$

For $M=2$, the saving is 3dB. As $M$ is increased, the saving in SNR approaches 0 dB.
5.2.5 Probability of Error for $M$-ary Binary-Coded Signals

In Section 4.3, we have shown that binary-coded signal waveform are represented by signal vectors (4.3-38):

$$s_m = [s_{m1} s_{m2} \cdots s_{mN}], \quad m = 1,2,\ldots,M$$

where $s_{mj} = \pm \sqrt{\mathcal{E}/N}$ for all $m$ and $j$. $N$ is the block length of the code and is also the dimension of the $M$ signal waveform.

If $d_{\text{min}}^{(e)}$ is the minimum Euclidean distance, then the probability of a symbol error is upper-bounded as

$$P_M < (M-1)P_b = (M-1)Q\left(\sqrt{\frac{(d_{\text{min}}^{(e)})^2}{2N_0}}\right)$$

$$< 2^k \exp\left[-\frac{(d_{\text{min}}^{(e)})^2}{4N_0}\right]$$
5.2.6 Probability of Error for \( M \)-ary PAM

Recall (4.3-6) that \( M \)-ary PAM signals are represented geometrically as \( M \) one-dimensional signal points with value:

\[
s_m = \sqrt{\frac{1}{2}} E_g A_m \quad m = 1, 2, \ldots, M
\]

where \( A_m = (2m - 1 - M)d \), \( m = 1, 2, \ldots, M \)

The Euclidean distance between adjacent signal points is \( d \sqrt{2E_g} \).

Assuming equally probable signals, the average energy is:

\[
E_{av} = \frac{1}{M} \sum_{m=1}^{M} E_m = \frac{1}{M} \sum_{m=1}^{M} s_m^2 = \frac{d^2 E_g}{2M} \sum_{m=1}^{M} (2m - 1 - M)^2
\]

\[
= \frac{d^2 E_g}{2M} \left[ \frac{1}{3} M (M^2 - 1) \right] = \frac{1}{6} (M^2 - 1) d^2 E_g
\]

\[
\sum_{m=1}^{M} m = \frac{M(M+1)}{2}
\]

\[
\sum_{m=1}^{M} m^2 = \frac{M(M+1)(2M+1)}{6}
\]
5.2.6 Probability of Error for $M$-ary PAM

Equivalently, we may characterize these signals in terms of their average power, where is:

$$P_{av} = \frac{\mathcal{E}_{av}}{T} = \frac{1}{6}(M^2 - 1)\frac{d^2 \mathcal{E}_g}{T}$$  \hspace{1cm} (5.2-40)

The average probability of error for $M$-ary PAM:

- The detector compares the demodulator output $r$ with a set of $M-1$ thresholds, which are placed at the midpoints of successive amplitude level and decision is made in favor of the amplitude level that is close to $r$.

- We note that if the $m$th amplitude level is transmitted, the demodulator output is

$$r = s_m + n = \sqrt{\frac{1}{2} \mathcal{E}_g A_m} + n$$

where the noise variable $n$ has zero-mean and variance $\sigma_n^2 = \frac{1}{2} N_0$
5.2.6 Probability of Error for $M$-ary PAM

- Assuming all amplitude levels are equally likely a priori, the average probability of a symbol error is the probability that the noise variable $n$ exceeds in magnitude one-half of the distance between levels.

- However, when either one of the two outside levels $\pm(M-1)$ is transmitted, an error can occur in one direction only.

- As a result, we have the error probability:

$$P_M = \frac{M-1}{M} P\left( |r - s_m| > d \sqrt{\frac{1}{2} E_g} \right) = \frac{M-1}{M} \frac{2}{\sqrt{2\pi N_0/2}} \int_{-\infty}^{\infty} e^{-x^2/N_0} dx$$

$$= \frac{M-1}{M} \frac{2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx$$

$$= 2(M-1) Q\left( \frac{\sqrt{d^2 E_g/N_0}}{M} \right)$$

$$Q(t) = \frac{1}{\sqrt{2\pi}} \int_t^{\infty} e^{-x^2/2} dx \quad t \geq 0$$

$$= \frac{1}{M} (M-2) + \frac{1}{M} \cdot \frac{1}{2} \cdot 2$$
From (5.2-40), we note that \( d^2 E_g = \frac{6}{M^2 - 1} P_{av} T \)

By substituting for \( d^2 E_g \), we obtain the average probability of a symbol error for PAM in terms of the average power:

\[
P_M = \frac{2(M - 1)}{M} Q \left( \sqrt{\frac{6P_{av} T}{(M^2 - 1) N_0}} \right) = \frac{2(M - 1)}{M} Q \left( \sqrt{\frac{6E_{av}}{(M^2 - 1) N_0}} \right)
\]

It is customary for us to use the SNR per bit as the basic parameter, and since \( T = kT_b \) and \( k = \log_2 M \):

\[
P_M = \frac{2(M - 1)}{M} Q \left( \sqrt{\frac{6 \log_2 M E_{bav}}{(M^2 - 1) N_0}} \right)
\]

where \( E_{bav} = P_{av} T_b \) is the average bit energy and \( E_{bav}/N_0 \) is the average SNR per bit.
The case $M=2$ corresponds to the error probability for binary antipodal signals.

- The SNR per bit increase by over 4 dB for every factor-of-2 increase in $M$.

- For large $M$, the additional SNR per bit required to increase $M$ by a factor of 2 approaches 6 dB.
Recall from section 4.3 that digital phase-modulated signal waveforms may be expressed as (4.3-11):

\[ s_m(t) = g(t) \cos \left[ 2\pi f_c t + \frac{2\pi}{M} (m-1) \right], \quad 1 \leq m \leq M, \quad 0 \leq t \leq T \]

and have the vector representation:

\[ s_m(t) = \begin{bmatrix} \sqrt{\mathcal{E}_s} \cos \frac{2\pi}{M} (m-1) \\ \sqrt{\mathcal{E}_s} \sin \frac{2\pi}{M} (m-1) \end{bmatrix}, \quad \mathcal{E}_s = \frac{1}{2} \mathcal{E}_g \]

Since the signal waveforms have equal energy, the optimum detector for the AWGN channel given by Equation 5.1-44 computes the correlation metrics

\[ C(r, s_m) = r \cdot s_m, \quad m = 1, 2, \ldots, M \]
5.2.7 Probability of Error for $M$-ary PSK

- In other word, the received signal vector $\mathbf{r} = [r_1, r_2]$ is projected onto each of the $M$ possible signal vectors and a decision is made in favor of the signal with the largest projection.

- This correlation detector is equivalent to a phase detector that computes the phase of the received signal from $\mathbf{r}$. We selects the signal vector $\mathbf{s}_m$ whose phase is closer to $\mathbf{r}$.

- The phase of $\mathbf{r}$ is $\Theta_r = \tan^{-1} \frac{r_2}{r_1}$

- We will determine the PDF of $\Theta_r$, and compute the probability of error from it.

- Consider the case in which the transmitted signal phase is $\Theta_r = 0$, corresponding to the signal $s_1(t)$. 

5.2.7 Probability of Error for $M$-ary PSK

The transmitted signal vector is $s_1 = \begin{bmatrix} \sqrt{E_s} & 0 \end{bmatrix}$, and the received signal vector has components:

$$r_1 = \sqrt{E_s} + n_1 \quad r_2 = n_2$$

Because $n_1$ and $n_2$ are jointly Gaussian random variables, it follows that $r_1$ and $r_2$ are jointly Gaussian random variables with $E(r_1) = \sqrt{E_s}$, $E(r_2) = 0$, and $\sigma^2_{r_1} = \sigma^2_{r_2} = \frac{1}{2} N_0 = \sigma_r^2$.

$$p_r(r_1, r_2) = \frac{1}{2\pi\sigma_r^2} \exp \left[ - \frac{(r_1 - \sqrt{E_s})^2 + r_2^2}{2\sigma_r^2} \right]$$

The PDF of the phase $\Theta_r$ is obtained by a change in variables from $(r_1, r_2)$ to:

$$V = \sqrt{r_1^2 + r_2^2} \quad \Theta_r = \tan^{-1} \frac{r_2}{r_1}$$
5.2.7 Probability of Error for $M$-ary PSK

The joint PDF of $V$ and $\Theta_r$:

$$p_{V,\Theta_r}(V,\Theta_r) = \frac{V}{2\pi\sigma_r^2} \exp\left(-\frac{V^2 + \mathcal{E}_s - 2\sqrt{\mathcal{E}_s} V \cos \Theta_r}{2\sigma_r^2}\right)$$

Integration of $p_{V,\Theta_r}(V,\Theta_r)$ over the range of $V$ yields $p_{\Theta_r}(\Theta_r)$

$$p_{\Theta_r}(\Theta_r) = \int_0^\infty p_{V,\Theta_r}(V,\Theta_r) dV$$

$$= \frac{1}{2\pi\sigma_r^2} e^{-\mathcal{E}_s/N_0 \sin^2 \Theta_r} \int_0^\infty \frac{V}{N_0/2} e^{-\left(\frac{V}{N_0/2}\sqrt{2\mathcal{E}_s/N_0 \cos \Theta_r}\right)^2} dV$$

$$= \frac{1}{2\pi\sigma_r^2} e^{-\gamma_s \sin^2 \Theta_r} \int_0^\infty V' e^{-\left(V'\sqrt{2\gamma_s} \cos \Theta_r\right)^2} dV'$$

where we define the symbol SNR as $\gamma_s = \mathcal{E}_s/N_0$. 

Fall, 2004. 

WITS Lab, NSYSU.
5.2.7 Probability of Error for $M$-ary PSK

Probability density function $p_{\Theta_r}(\Theta_r)$ for $\gamma_s = 1, 2, 4,$ and $10$. $f_{\Theta_r}(\Theta_r)$ becomes narrower and more peaked about $\Theta_r = 0$ as the SNR $\gamma_s$ increases.
When $s_1(t)$ is transmitted, a decision error is made if the noise causes the phase to fall outside the range $-\pi/M \leq \Theta_r \leq \pi/M$. Hence, the probability of a symbol error is

$$P_M = 1 - \int_{-\pi/M}^{\pi/M} p_{\Theta_r}(\Theta_r) d\Theta_r$$

In general, the integral of $p_{\Theta_r}(\Theta_r)$ doesn’t reduced to a simple form and must be evaluated numerically, except for $M=2$ and $M=4$.

For binary phase modulation, the two signals $s_1(t)$ and $s_2(t)$ are antipodal. Hence, the error probability is

$$P_2 = Q\left(\sqrt{\frac{2\mathcal{E}_b}{N_0}}\right)$$
5.2.7 Probability of Error for $M$-ary PSK

- When $M = 4$, we have in effect two binary phase-modulation signals in phase quadrature. Since there is no crosstalk or interference between the signals on the two quadrature carriers, the bit error probability is identical to that of $M = 2$. (5.2-57)

- Then the symbol error probability for $M=4$ is determined by noting that

\[
P_c = (1 - P_2)^2 = \left[1 - Q\left(\sqrt{\frac{2\mathcal{E}_b}{N_0}}\right)\right]^2
\]

where $P_c$ is the probability of a correct decision for the 2-bit symbol.

- There, the symbol error probability for $M = 4$ is

\[
P_4 = 1 - P_c = 2Q\left(\sqrt{\frac{2\mathcal{E}_b}{N_0}}\right)\left[1 - \frac{1}{2} Q\left(\sqrt{\frac{2\mathcal{E}_b}{N_0}}\right)\right]
\]
For $M > 4$, the symbol error probability $P_M$ is obtained by numerically integrating Equation $P_M = 1 - \int_{-\pi/M}^{\pi/M} p_{\Theta \mid r} \Theta \d\Theta$.

For large values of $M$, doubling the number of phases requires an additional 6 dB/bit to achieve the same performance.
An approximation to the error probability for large $M$ and for large SNR may be obtained by first approximating $p_{\Theta_r}(\Theta)$.

For $E_s/N_0 >> 1$ and $|\Theta_r| \leq 0.5\pi$, $p_{\Theta_r}(\Theta)$ is well approximated as:

$$P_{\Theta_r}(\Theta_r) \approx \sqrt{\frac{\gamma_s}{\pi}} \cos\Theta_r e^{-\gamma_s \sin^2 \Theta_r}$$

Performing the change in variable from $\Theta_r$ to $u = \sqrt{\nu_s} \sin \Theta_r$,

$$P_M \approx 1 - \int_{-\pi/M}^{\pi/M} \sqrt{\frac{\gamma_s}{\pi}} \cos\Theta_r e^{-\gamma_s \sin^2 \Theta_r} d\Theta_r$$

$$\approx \frac{2}{\sqrt{\pi}} \int_{-\gamma_s \sin(\pi/M)}^{\gamma_s \sin(\pi/M)} e^{-u^2} du$$

$$= 2Q\left(\sqrt{2\gamma_s \sin \frac{\pi}{M}}\right) = 2Q\left(\sqrt{2k\gamma_b \sin \frac{\pi}{M}}\right)$$

$k = \log_2 M$

$\nu_s = k \nu_b$
When a Gray Code is used in the mapping of $k$-bits symbols into the corresponding signal phases, two $k$-bit symbols corresponding to adjacent signal phases differ in only a signal bit.

The most probable error result in the erroneous selection of an adjacent phase to the true one.

Most $k$-bit symbol error contain only a single-bit error. The equivalent bit error probability for $M$-ary PSK is well approximated as:

$$P_b \approx \frac{1}{k} P_M$$

In practice, the carrier phase is extracted from the received signal by performing some nonlinear operation that introduces a phase ambiguity.
For binary PSK, the signal is often squared in order to remove the modulation, and the double-frequency component that is generated is filtered and divided by 2 in frequency in order to extract an estimate of the carrier frequency and phase $\phi$. This operation result in a phase ambiguity of 180° in the carrier phase.

For four-phase PSK, the received signal is raised to the fourth power to remove the digital modulation, and the resulting fourth harmonic of the carrier frequency is filtered and divided by 4 in order to extract the carrier component.

These operations yield a carrier frequency component containing $\phi$, but there are phase ambiguities of $\pm90^\circ$ and $180^\circ$ in the phase estimate.

Consequently, we do not have an absolute estimate of the carrier phase for demodulation.
The phase ambiguity problem can be overcome by encoding the information in *phase differences* between successive signal transmissions as opposed to absolute phase encoding.

For example, in binary PSK, the information bit 1 may be transmitted by shifting the phase of carrier by 180° relative to the previous carrier phase. Bit 0 is transmitted by a zero phase shift relative to the phase in the previous signaling interval.

In four-phase PSK, the relative phase shifts between successive intervals are 0, 90°, 180°, and -90°, corresponding to the information bits 00, 01, 11, and 10, respectively.

The PSK signals resulting from the encoding process are said to be *differentially encoded*. 

---

*Fall, 2004. 123 WITS Lab, NSYSU.*
The detector is relatively simple phase comparator that compares the phase of the demodulated signal over two consecutive interval to extract the information.

Coherent demodulation of differentially encoded PSK results in a higher probability of error than that derived for absolute phase encoding.

With differentially encoded PSK, an error in the demodulated phase of the signal in any given interval will usually result in decoding errors over two consecutive signaling intervals.

The probability of error in differentially encoded $M$-ary PSK is approximately twice the probability of error for $M$-ary PSK with absolute phase encoding.
5.2.8 Differential PSK (DPSK) and Its Performance

- The received signal of a differentially encoded phase-modulated signal in any given signaling interval is compared to the phase of the received signal from the preceding signaling interval.
- We demodulate the differentially encoded signal by multiplying \( r(t) \) by \( \cos 2\pi f_c t \) and \( \sin 2\pi f_c t \) integrating the two products over the interval \( T \).
- At the \( k \)th signaling interval, the demodulator output:

\[
\begin{bmatrix}
\sqrt{E_s} \cos(\theta_k - \phi) + n_{k_1} \\
\sqrt{E_s} \sin(\theta_k - \phi) + n_{k_2}
\end{bmatrix}
\]

or equivalently,

\[
r_k = \sqrt{E_s} e^{j(\theta_k - \phi)} + n_k
\]

where \( \theta_k \) is the phase angle of the transmitted signal at the \( k \)th signaling interval, \( \phi \) is the carrier phase, and \( n_k = n_{k_1} + jn_{k_2} \) is the noise vector.
Similarly, the received signal vector at the output of the demodulator in the preceding signaling interval is:

\[ r_{k-1} = \sqrt{E_s} e^{j(\theta_{k-1}-\phi)} + n_{k-1} \]

The decision variable for the phase detector is the phase difference between these two complex numbers. Equivalently, we can project \( r_k \) onto \( r_{k-1} \) and use the phase of the resulting complex number:

\[ r_k r_{k-1}^* = E_s e^{j(\theta_k-\theta_{k-1})} + \sqrt{E_s} e^{j(\theta_k-\phi)} n_{k-1}^* + \sqrt{E_s} e^{-j(\theta_{k-1}-\phi)} n_k + n_k n_{k-1}^* \]

which, in the absence of noise, yields the phase difference \( \theta_k - \theta_{k-1} \).

Differentially encoded PSK signaling that is demodulated and detected as described above is called **differential PSK** (DPSK).
If the pulse $g(t)$ is rectangular, the matched filter may be replaced by integrate-and-dump filter.
The error probability performance of a DPSK demodulator and detector

- The derivation of the exact value of the probability of error for $M$-ary DPSK is extremely difficult, except for $M = 2$.
- Without loss of generality, suppose the phase difference $\theta_k - \theta_{k-1} = 0$. Furthermore, the exponential factor $e^{-j(\theta_{k-1} - \phi)}$ and $e^{j(\theta_k - \phi)}$ can be absorbed into Gaussian noise components $n_{k-1}$ and $n_k$, without changing their statistical properties.

$$r_k r_{k-1}^* = E_s + \sqrt{E_s} \left( n_k + n_{k-1}^* \right) + n_k n_{k-1}^*$$

- The complication in determining the PDF of the phase is the term $n_k n_{k-1}^*$. 
5.2.8 Differential PSK (DPSK) and Its Performance

- However, at SNRs of practical interest, the term $n_k n^*_k$ is small relative to the dominant noise term $\sqrt{E_s} (n_k + n^*_k)$.

- We neglect the term $n_k n^*_k$ and normalize $r_k r^*_k$ by dividing through by $\sqrt{E_s}$, the new set of decision metrics becomes:

\[
\begin{align*}
  x &= \sqrt{E_s} + \text{Re}(n_k + n^*_k) \\
  y &= \text{Im}(n_k + n^*_k)
\end{align*}
\]

- The variables $x$ and $y$ are uncorrelated Gaussian random variable with identical variances $\sigma^2 = N_0$. The phase is

\[
\Theta_r = \tan^{-1} \frac{y}{x}
\]

- The noise variance is now twice as large as in the case of PSK. Thus we can conclude that the performance of DPSK is 3 dB poorer than that for PSK.
This result is relatively good for $M \geq 4$, but it is pessimistic for $M = 2$ that the loss in binary DPSK relative to binary PSK is less than 3 dB at large SNR.

In binary DPSK, the two possible transmitted phase differences are 0 and $\pi$ rad. Consequently, only the real part of $r_{kr_{k-1}}^*$ is needed for recovering the information.

$$\text{Re}(r_{kr_{k-1}}^*) = \frac{1}{2} (r_{kr_{k-1}}^* + r_{kr_{k-1}}^*)$$

Because the phase difference between two successive signaling intervals is zero, an error is made if $\text{Re}(r_{kr_{k-1}}^*) < 0$.

The probability that $r_{kr_{k-1}}^* + r_{kr_{k-1}}^* < 0$ is a special case of a derivation, given in Appendix B.
Appendix B concerned with the probability that a general quadratic form in complex-valued Gaussian random variable is less than zero. According to Equation B-21, we find it depend entirely on the first and second moments of the complex-valued Gaussian random variables $r_k$ and $r_{k-1}$.

We obtain the probability of error for binary DPSK in the form

$$P_b = \frac{1}{2} e^{-\varepsilon_b/N_0}$$

where $\varepsilon_b/N_0$ is the SNR per bit.
The probability of a binary digit error for four-phase DPSK with Gray coding can be express in terms of well-known functions, but it’s derivation is quite involved.

According to Appendix C, it is expressed in the form:

$$P_b = Q_1(a, b) - \frac{1}{2} I_0(ab) \exp \left[ -\frac{1}{2} \left( a^2 + b^2 \right) \right]$$

where $Q_1(a, b)$ is the Marcum $Q$ function (2.1-122), $I_0(x)$ is the modified Bessel function of order zero(2.1-123), and the parameters $a$ and $b$ are defined as

$$a = \sqrt{2\gamma_b \left( 1 - \sqrt{\frac{1}{2}} \right)}, \text{ and } b = \sqrt{2\gamma_b \left( 1 + \sqrt{\frac{1}{2}} \right)}$$
5.2.8 Differential PSK (DPSK) and Its Performance

Because binary DPSK is only slightly inferior to binary PSK at large SNR, and DPSK does not require an elaborate method for estimate the carrier phase, it is often used in digital communication system.
5.2.9 Probability of Error for QAM

Recall from Section 4.3 that QAM signal waveforms may be expressed as (4.3-19)

\[ s_m(t) = A_{mc} g(t) \cos 2\pi f_c t - A_{ms}(t) \sin 2\pi f_c t \]

where \( A_{mc} \) and \( A_{ms} \) are the information-bearing signal amplitudes of the quadrature carriers and \( g(t) \) is the signal pulse.

The vector representation of these waveform is

\[ s_m = \begin{bmatrix} A_{mc} \sqrt{\frac{1}{2} E_g} \\ A_{ms} \sqrt{\frac{1}{2} E_g} \end{bmatrix} \]

To determine the probability of error for QAM, we must specify the signal point constellation.
QAM signal sets that have $M = 4$ points.

Figure (a) is a four-phase modulated signal and Figure (b) is with two amplitude levels, labeled $A_1$ and $A_2$, and four phases.

Because the probability of error is dominated by the minimum distance between pairs of signal points, let us impose the condition that $d_{\text{min}}^{(e)} = 2A$ and we evaluate the average transmitter power, based on the premise that all signal points are equally probable.
5.2.9 Probability of Error for QAM

- For the four-phase signal, we have
  \[ P_{av} = \frac{1}{4}(4)2A^2 = 2A^2 \]

- For the two-amplitude, four-phase QAM, we place the points on circles of radii \( A \) and \( \sqrt{3}A \). Thus, \( d_{min}^{(e)} = 2A \), and
  \[ P_{av} = \frac{1}{2}[2(3A^2) + 2A^2] = 2A^2 \]
  which is the same average power as the \( M = 4 \)-phase signal constellation.

- Hence, for all practical purposes, the error rate performance of the two signal sets is the same.

- There is no advantage of the two-amplitude QAM signal set over \( M = 4 \)-phase modulation.
QAM signal sets that have \( M = 8 \) points.

We consider the four signal constellations:

Assuming that the signal points are equally probable, the average transmitted signal power is:

\[
P_{av} = \frac{1}{M} \sum_{m=1}^{M} \left( A_{mc}^2 + A_{ms}^2 \right)
\]

\[
= \frac{A^2}{M} \sum_{m=1}^{M} \left( a_{mc}^2 + a_{ms}^2 \right)
\]
The two sets (a) and (c) contain signal points that fall on a rectangular grid and have $P_{av} = 6A^2$.

The signal set (b) requires an average transmitted power $P_{av} = 6.83A^2$, and (d) requires $P_{av} = 4.73A^2$.

The fourth signal set (d) requires approximately 1 dB less power than the first two and 1.6 dB less power than the third to achieve the same probability of error.

The fourth signal constellation is known to be the best eight-point QAM constellation because it requires the least power for a given minimum distance between signal points.
5.2.9 Probability of Error for QAM

- QAM signal sets for $M \geq 16$
  - For 16-QAM, the signal points at a given amplitude level are phase-rotated by relative to the signal points at adjacent amplitude levels.
  - However, the circular 16-QAM constellation is not the best 16-point QAM signal constellation for the AWGN channel.
- Rectangular $M$-ary QAM signal are most frequently used in practice. The reasons are:
  - Rectangular QAM signal constellations have the distinct advantage of being easily generated as two PAM signals impressed on phase-quadrature carriers.
5.2.9 Probability of Error for QAM

- The average transmitted power required to achieve a given minimum distance is only slightly greater than that of the best $M$-ary QAM signal constellation.

- For rectangular signal constellations in which $M = 2^k$, where $k$ is even, the QAM signal constellation is equivalent to two PAM signals on quadrature carriers, each having $\sqrt{M} = 2^{k/2}$ signal points.

- The probability of error for QAM is easily determined from the probability of error for PAM.

- Specifically, the probability of a correct decision for the $M$-ary QAM system is

$$P_c = \left(1 - P_p^{\sqrt{M}}\right)^2$$

where $P_p^{\sqrt{M}}$ is the probability of error of an $\sqrt{M}$-ary PAM.
By appropriately modifying the probability of error for $M$-ary PAM, we obtain

\[
P_{\sqrt{M}} = 2 \left( 1 - \frac{1}{\sqrt{M}} \right) Q\left( \sqrt{\frac{3}{M-1}} \frac{\varepsilon_{av}}{N_0} \right)
\]

where $\varepsilon_{av}/N_0$ is the average SNR per symbol.

Therefore, the probability of a symbol error for $M$-ary QAM is

\[
P_M = 1 - \left( 1 - P_{\sqrt{M}} \right)^2
\]

Note that this result is exact for $M = 2^k$ when $k$ is even.

When $k$ is odd, there is no equivalent $\sqrt{M}$-ary PAM system. There is still ok, because it is rather easy to determine the error rate for a rectangular signal set.
We employ the optimum detector that bases its decisions on the optimum distance metrics (5.1-43), it is relatively straightforward to show that the symbol error probability is tightly upper-bounded as

$$P_M \leq 1 - 2Q\left(\sqrt{\frac{3\varepsilon_{av}}{(M-1)N_0}}\right)^2$$

$$\leq 4Q\left(\sqrt{\frac{3k\varepsilon_{bav}}{(M-1)N_0}}\right)$$

for any $k \geq 1$, where $\varepsilon_{bav}/N_0$ is the average SNR per bit.
5.2.9 Probability of Error for QAM

- For nonrectangular QAM signal constellation, we may upper-bound the error probability by use of a union bound:

\[ P_M < (M - 1)Q\left(\sqrt{d_{\text{min}}^{(e)}}^2 / 2N_0\right) \]

where \( d_{\text{min}}^{(e)} \) is the minimum Euclidean distance between signal points.

- This bound may be loose when \( M \) is large.

- We approximate \( P_M \) by replacing \( M-1 \) by \( M_n \), where \( M_n \) is the largest number of neighboring points that are at distance \( d_{\text{min}}^{(e)} \) from any constellation point.

- It is interesting to compare the performance of QAM with that of PSK for any given signal size \( M \).
For $M$-ary PSK, the probability of a symbol error is approximate as

$$P_M \approx 2Q\left(\sqrt{2\gamma_s \sin \frac{\pi}{M}}\right)$$

where $\gamma_s$ is the SNR per symbol.

For $M$-ary QAM, the error probability is:

$$P_{\sqrt{M}} = 2\left(1 - \frac{1}{\sqrt{M}}\right)Q\left(\sqrt{\frac{3}{M-1} \frac{\mathcal{E}_{av}}{N_0}}\right)$$

We simply compare the arguments of Q function for the two signal formats. The ratio of these two arguments is

$$R_M = \frac{3/(M-1)}{2\sin^2\left(\frac{\pi}{M}\right)}$$
5.2.9 Probability of Error for QAM

- We can find out that when $M = 4$, we have $R_M = 1$.
- 4-PSK and 4-QAM yield comparable performance for the same SNR per symbol.
- For $M > 4$, $M$-ary QAM yields better performance than $M$-ary PSK.

<table>
<thead>
<tr>
<th>$M$</th>
<th>$10 \log_{10} R_M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>1.65</td>
</tr>
<tr>
<td>16</td>
<td>4.20</td>
</tr>
<tr>
<td>32</td>
<td>7.02</td>
</tr>
<tr>
<td>64</td>
<td>9.95</td>
</tr>
</tbody>
</table>
5.2.10 Comparison of Digital Modulation Methods

One can compare the digital modulation methods on the basis of the SNR required to achieve a specified probability of error.

However, such a comparison would not be very meaningful, unless it were made on the basis of some constraint, such as a fixed data rate of transmission or, on the basis of a fixed bandwidth.

For multiphase signals, the channel bandwidth required is simply the bandwidth of the equivalent low-pass signal pulse $g(t)$ with duration $T$ and bandwidth $W$, which is approximately equal to the reciprocal of $T$.

Since $T = k/R = (\log_2 M)/R$, it follows that $W = \frac{R}{\log_2 M}$.
5.2.10 Comparison of Digital Modulation Methods

- As $M$ is increased, the channel bandwidth required, when the bit rate $R$ is fixed, decreases. The *bandwidth efficiency* is measured by the bit rate to bandwidth ratio, which is:

$$\frac{R}{W} = \log_2 M$$

- The bandwidth-efficient for transmitting PAM is single-sideband. The channel bandwidth required to transmit the signal is approximately equal to $1/2T$ and,

$$\frac{R}{W} = 2 \log_2 M$$

  this is a factor of 2 better than PSK.

- For QAM, we have two orthogonal carriers, with each carrier having a PAM signal.
Thus, we double the rate relative to PAM. However, the QAM signal must be transmitted via double-sideband. Consequently, QAM and PAM have the same bandwidth efficiency when the bandwidth is referenced to the band-pass signal.

As for orthogonal signals, if the $M = 2^k$ orthogonal signals are constructed by means of orthogonal carriers with minimum frequency separation of $1/2T$, the bandwidth required for transmission of $k = \log_2 M$ information bits is

$$W = \frac{M}{2T} = \frac{M}{2(k/R)} = \frac{M}{2\log_2 M^R}$$

In the case, the bandwidth increases as $M$ increases.

In the case of biorthogonal signals, the required bandwidth is one-half of that for orthogonal signals.
5.2.10 Comparison of Digital Modulation Methods

- A compact and meaningful comparison of modulation methods is one based on the normalized data rate $R/W$ (bits per second per hertz of bandwidth) versus the SNR per bit ($\epsilon_b/N_0$) required to achieve a given error probability.

- In the case of PAM, QAM, and PSK, increasing $M$ results in a higher bit-to-bandwidth ratio $R/W$. 

5.2.10 Comparison of Digital Modulation Methods

- However, the cost of achieving the higher data rate is an increase in the SNR per bit.

- Consequently, these modulation methods are appropriate for communication channels that are bandwidth limited, where we desire a $R/W > 1$ and where there is sufficiently high SNR to support increases in $M$.

- Telephone channels and digital microwave ratio channels are examples of such band-limited channels.

- In contrast, $M$-ary orthogonal signals yield a $R/W \leq 1$. As $M$ increases, $R/W$ decreases due to an increase in the required channel bandwidth.

- The SNR per bit required to achieve a given error probability decreases as $M$ increases.
Consequently, $M$-ary orthogonal signals are appropriate for power-limited channels that have sufficiently large bandwidth to accommodate a large number of signals.

As $M \to \infty$, the error probability can be made as small as desired, provided that $\text{SNR} > 0.693$ (-1.6dB). This is the minimum SNR per bit required to achieve reliable transmission in the limit as the channel bandwidth $W \to \infty$ and the corresponding $R/W \to 0$.

The figure above also shown the normalized capacity of the band-limited AWGN channel, which is due to Shannon (1948).

The ratio $C/W$, where $C (= R)$ is the capacity in bits/s, represents the highest achievable bit rate-to-bandwidth ratio on this channel.

Hence, it serves the upper bound on the bandwidth efficiency of any type of modulation.
5.3 Optimum Receiver for CPM Signals

- CPM is a modulation method with memory. The memory results from the continuity of the transmitted carrier phase from one signal interval to the next.

- The transmitted CPM signal may be expressed as

\[ s(t) = \sqrt{\frac{2E}{T}} \cos[2\pi f_c t + \phi(t ; I)] \]

where \( \phi(t ; I) \) is the carrier phase.

- The filtered received signal for an additive Gaussian noise channel is

\[ r(t) = s(t) + n(t) \]

where \( n(t) = n_c(t)\cos 2\pi f_c t - n_s(t)\sin 2\pi f_c t \)
5.3.1 Optimum Demodulation and Detection of CPM

- The optimum receiver for this signal consists of a correlator followed by a maximum-likelihood sequence detector that searches the paths through the state trellis for the minimum Euclidean distance path. And the Viterbi algorithm is an efficient method for performing this search.

- The carrier phase for a CPM signal with a fixed modulation index $h$ may be expressed as

$$\phi(t; I) = 2\pi h \sum_{k=-\infty}^{n} I_k q(t - kT)$$

$$= \pi h \sum_{k=-\infty}^{n-L} I_k + 2\pi h \sum_{k=n-L+1}^{n} I_k q(t - kT)$$

$$= \theta_n + \theta(t; I) \quad nT \leq t \leq (n+1)T$$

where $q(t)=0$ for $t < 0$, $q(t)=1/2$ for $t \geq LT$, and $q(t) = \int_{0}^{t} g(\tau)d\tau$.
5.3.1 Optimum Demodulation and Detection of CPM

- The signal pulse $g(t) = 0$ for $t < 0$ and $t \geq LT$.
- For $L = 1$, we have a full response CPM, and for $L > 1$, where $L$ is a positive integer, we have a partial response CPM signal.
- $h$ is rational, i.e., $h = m/p$ where $m$ and $p$ are relatively prime positive integers, the CPM scheme can be represented by a trellis.

There are $p$ phase states when $m$ is even (4.3-60):

$$\Theta_s = \left\{ 0, \frac{\pi m}{p}, \frac{2\pi m}{p}, \ldots, \frac{(p-1)\pi m}{p} \right\}$$

There are $2p$ phase states when $m$ is odd (4.3-61):

$$\Theta_s = \left\{ 0, \frac{\pi m}{p}, \ldots, \frac{(2p-1)\pi m}{p} \right\}$$
These phase diagrams are called *phase tree*. CPFSK with binary symbols $I_n = \pm 1$, the set of phase trajectories beginning at time $t=0$. Phase trajectories for quaternary CPFSK.
5.3.1 Optimum Demodulation and Detection of CPM

We know if \( L > 1 \), we have an additional number of states due to the partial response character of the signal \( g(t) \). These additional states can be identified by expressing \( \theta(t;\mathbf{I}) \):

\[
\theta(t;\mathbf{I}) = 2\pi h \sum_{k=n-L+1}^{n-1} I_k q(t-kT) + 2\pi h I_n q(t-nT)
\]

- The first term on the right-hand side depends on the information symbols \((I_{n-1}, I_{n-2}, \ldots, I_{n-L+1})\), which is called the **correlative state vector**, and represents the phase term corresponding to signal pulse that have not reached their final value.
- The second term represents the **phase contribution** due to the most recent symbol \( I_n \).
5.3.1 Optimum Demodulation and Detection of CPM

Hence, the state of the CPM signal (or the modulator) at time $t = nT$ may be expressed as the combined phase state and correlative state, denoted as $S_n = \{\theta_n, I_{n-1}, I_{n-2}, \ldots, I_{n-L+1}\}$ for a partial response signal pulse of length $LT$, where $L > 1$. In this case, the number of states is

$$N_s = \begin{cases} pM^{L-1} & \text{(even } m\text{)} \\ 2pM^{L-1} & \text{(odd } m\text{)} \end{cases}$$

where $h = m/p$.

Suppose the state of the modulator at $t = nT$ is $S_n$. The effect of the new symbol in the time interval $nT \leq t \leq (n+1)T$ is to change the state from $S_n$ to $S_{n+1}$. At $t = (n+1)T$, the state

$$S_{n+1} = (\theta_{n+1}, I_n, I_{n-1}, \ldots, I_{n-L+2})$$

where $\theta_{n+1} = \theta_n + \pi h I_{n-L+1}$.
Example 5.3-1.

A binary CPM scheme with a modulation index \( h = \frac{3}{4} \) and a partial response pulse with \( L = 2 \). Determine the states \( S_n \) of the CPM scheme and sketch the phase tree and state trellis.

First, we note that there are \( 2^p = 8 \) phase states, namely,

\[
\Theta_s = \left\{ 0, \pm \frac{1}{4} \pi, \pm \frac{1}{2} \pi, \pm \frac{3}{4} \pi, \pi \right\}
\]

For each of these phase states, there are two states that result from the memory of the CPM scheme.

Hence, the total number of states is \( N_s = 16 \), namely,

\[
(0, 1), (0, -1), (\pi, 1), (\pi, -1), \left( \frac{\pi}{4}, 1 \right), \left( \frac{\pi}{4}, -1 \right), \left( \frac{\pi}{2}, 1 \right), \left( \frac{\pi}{2}, -1 \right), \left( \frac{3\pi}{4}, 1 \right), \left( \frac{3\pi}{4}, -1 \right), \left( -\frac{\pi}{4}, 1 \right), \left( -\frac{\pi}{4}, -1 \right), \left( -\frac{\pi}{2}, 1 \right), \left( -\frac{\pi}{2}, -1 \right), \left( -\frac{3\pi}{4}, 1 \right), \left( -\frac{3\pi}{4}, -1 \right)
\]
Example 5.3-1.

If the system is in phase state $\theta_n = -\pi/4$ and $I_{n-1} = -1$, then

$$\theta_{n+1} = \theta_n + \pi h I_{n-1}$$

$$= -\frac{1}{4}\pi - \frac{3}{4}\pi = -\pi$$

State trellis for partial response ($L = 2$) CPM with $h = 3/4$
5.3.1 Optimum Demodulation and Detection of CPM

Example 5.3-1.

A path through the state trellis corresponding to the sequence (1, -1, -1, -1, 1, 1)

Phase tree for \( L = 2 \) partial response CPM with \( h = \frac{3}{4} \)

\( g(t) \) is a rectangular pulse of duration \( 2T \) with initial state (0,1)
5.3.1 Optimum Demodulation and Detection of CPM

**Metric computations.**

- It is easy to show that the logarithm of the probability of the observed signal \( r(t) \) conditioned on a particular sequence of transmitted symbols \( I \) is proportional to the cross-correlation metric

\[
CM_n(I) = \int_{-\infty}^{(n+1)T} r(t) \cos[\omega_c t + \phi(t; I)] dt
\]

\[
= CM_{n-1}(I) + \int_{nT}^{(n+1)T} r(t) \cos[\omega_c t + \theta(t; I) + \theta_n] dt
\]

- The term \( CM_{n-1}(I) \) represents the metrics for the surviving sequences up to time \( nT \), and the term

\[
v_n(I; \theta_n) = \int_{nT}^{(n+1)T} r(t) \cos[\omega_c t + \theta(t; I) + \theta_n] dt
\]

represents the additional increments to the metrics contributed by the signal in the time interval \( nT \leq t \leq (n+1)T \).
5.3.1 Optimum Demodulation and Detection of CPM

There are $M^L$ possible sequences $I = (I_n, I_{n-1}, \ldots, I_{n-L+1})$ of symbols and $p$ (or $2P$) possible phase states $\{\theta_n\}$.

Therefore, there are $pM^L$ (or $2pM^L$) different values of $v_n(I, \theta_n)$ computed in each signal interval, and each value is used to increment the metrics corresponding to the $pM^{L-1}$ surviving sequences from the previous signaling interval.
The number of surviving sequences at each state of the Viterbi decoding process is $pM^{L-1}$ (or $2pM^{L-1}$).

For each surviving sequence, we have $M$ new increments of $v_n(I; \theta_n)$ that are added to the existing metrics to yield $pM^L$ (or $2pM^L$) sequences with $pM^L$ (or $2pM^{L-1}$) metrics.

However, this number is then reduced back to $pM^{L-1}$ (or $2pM^{L-1}$) survivors with corresponding metrics by selecting the most probable sequence of the $M$ sequences merging at each node of the trellis and discarding the other $M-1$ sequences.
5.3.2 Performance of CPM Signals

In evaluating the performance of CPM signals achieved with maximum-likelihood sequence detection, we must determine the minimum Euclidean distance of path through the trellis that separate at the node at $t = 0$ and remerge at a later time at the same node.

Suppose two signals $s_i(t)$ and $s_j(t)$ corresponding to two phase trajectories $\phi(t; I_i)$ and $\phi(t; I_j)$.

The sequence $I_i$ and $I_j$ must be different in their first symbol.
The Euclidean distance between the two signals over an interval of length $NT$ ($1/T$ is the symbol rate) is defined as

$$d_{ij}^2 = \int_0^{NT} [s_i(t) - s_j(t)]^2 dt = \int_0^{NT} s_i^2(t) dt + \int_0^{NT} s_j^2(t) dt - 2\int_0^{NT} s_i(t)s_j(t) dt$$

$$= 2N\mathcal{E} - 2\frac{2\mathcal{E}}{T} \int_0^{NT} \cos[\omega c t + \phi(t; I_i)]\cos[\omega c t + \phi(t; I_j)] dt$$

$$= 2N\mathcal{E} - \frac{2\mathcal{E}}{T} \int_0^{NT} \cos[\phi(t; I_i) - \phi(t; I_j)] dt$$

$$= \frac{2\mathcal{E}}{T} \int_0^{NT} \{1 - \cos[\phi(t; I_i) - \phi(t; I_j)]\} dt$$

Hence the Euclidean distance is related to the phase difference between the paths in the state trellis.
Since \( \varepsilon = \varepsilon_b \log_2 M \), \( d_{ij}^2 \) may be expressed as

\[
d_{ij}^2 = 2\varepsilon_b \delta_{ij}^2
\]

where \( d_{ij}^2 \) is defined as

\[
\delta_{ij}^2 = \frac{\log_2 M}{T} \int_0^{NT} \left\{ 1 - \cos[\phi(t; I_i) - \phi(t; I_j)] \right\} dt
\]

Furthermore, since \( \phi(t; I_i) - \phi(t; I_j) = \phi(t; I_i - I_j) \), so that, with \( \xi = I_i - I_j \), \( \delta_{ij}^2 \) may be written as

\[
\delta_{ij}^2 = \frac{\log_2 M}{T} \int_0^{NT} \{ 1 - \cos \phi(t; \xi) \} dt
\]

where any element of \( \xi \) can take the values 0, \( \pm 2 \), \( \pm 4 \), \ldots, \( \pm 2(M-1) \), except that \( \xi_0 \neq 0 \).
5.3.2 Performance of CPM Signals

The error rate performances for CPM is dominated by the term corresponding to the minimum Euclidean distance:

\[ P_M = K_{\delta_{\text{min}}} Q \left( \sqrt{\frac{E_b}{N_0}} \delta_{\text{min}}^2 \right) \]

where \( K_{\delta_{\text{min}}} \) is the number of paths having the minimum distance

\[ \delta_{\text{min}}^2 = \lim_{N \to \infty} \min_{i,j} \delta_{ij}^2 \]

\[ = \lim_{N \to \infty} \min_{i,j} \left\{ \log_2 \frac{M}{T} \int_0^{NT} [1 - \cos \phi(t; I_i - I_j)] dt \right\} \]

We note that for conventional binary PSK with no memory, \( N = 1 \) and \( \delta_{\text{min}}^2 = \delta_{12}^2 = 2 \).
5.3.2 Performance of CPM Signals

Since $\delta_{\text{min}}^2$ characterizes the performance of PCM, we can investigate the effect on $\delta_{\text{min}}^2$ resulting from varying the alphabet size $M$, the modulation index $h$, and the length of the transmitted pulse in partial response CPM.

First, we consider full response ($L = 1$) CPM. For $M = 2$, we note that the sequence

$$I_i = +1, -1, I_2, I_3$$
$$I_j = -1, +1, I_2, I_3$$

which differ for $k = 0, 1$ and agree for $k \geq 2$, result in two phase trajectories that merge after the second symbol. The corresponds to the difference sequence

$$\xi = \{2, -2, 0, 0, \ldots\}$$
The Euclidean distance for this sequence is easily calculated from

\[ \delta_{ij}^2 = \log_2 M \int_0^{NT} \{1 - \cos(\phi(t; \xi))\} dt \]

and provides an upper bound on \( \delta_{\text{min}}^2 \). This upper bound for CPSK with \( M = 2 \) is

\[ d_B^2(h) = 2 \left( 1 - \frac{\sin 2\pi h}{2\pi h} \right), \quad M = 2 \]

For example, where \( h = 1/2 \), which corresponds to MSK, we have

\[ d_B^2 \left( \frac{1}{2} \right) = 2, \quad \text{so that} \quad \delta_{\text{min}}^2 \left( \frac{1}{2} \right) \leq 2. \]

For \( M > 2 \) and full response CPM, an upper bound on \( \delta_{\text{min}}^2 \) can be obtained by considering the phase difference sequence \( \xi = \{ \alpha, -\alpha, 0, 0, \ldots \} \) where \( \alpha = \pm 2, \pm 4, \ldots, \pm 2(M-1) \).
5.3.2 Performance of CPM Signals

- This sequence \( \xi = \{ \alpha, -\alpha, 0, 0, \ldots \} \) yields the upper bound for \( M \)-ary CPFSK as

\[
d_B^2(h) = \min_{1 \leq k \leq M-1} \left( 2 \log_2 M \left( 1 - \frac{\sin 2k\pi h}{2k\pi h} \right) \right)
\]

- Large gains in performance can be achieved by increasing the alphabet size \( M \).

- The upper bound may not be achievable for all values of \( h \) because of \( \delta_{\text{min}}^2(h) \leq d_B^2(h) \).
In this section, we consider the design of the optimum receiver for carrier modulated signals when the carrier phase is unknown and no attempt is made to estimate its value.

Uncertainty in the carrier phase of the receiver signal may be due to one or more of the following reasons:

- The oscillators that are used at the transmitter and the receiver to generate the carrier signals are generally not phase synchronous.
- The time delay in the propagation of the signal from the transmitter to the receiver is not generally known precisely.

Assuming a transmitted signal of the form

\[ s(t) = \text{Re}[g(t)e^{j2\pi f_c t}] \]

that propagates through a channel with delay \( t_0 \) will be received as:

\[ s(t - t_0) = \text{Re}[g(t - t_0)e^{j2\pi f_c (t - t_0)}] = \text{Re}[g(t - t_0)e^{-j2\pi f_c t_0}e^{j2\pi f_c}] \]
5.4 Optimum Receiver For Signals with Random Phase In AWGN Channel

- The carrier phase shift due to the propagation delay $t_0$ is
  \[ \phi = -2\pi \frac{f_c}{c} t_0 \]

- Note that large changes in the carrier phase can occur due to relatively small changes in the propagation delay.

- For example, if the carrier frequency $f_c = 1$ MHz, an uncertainty or a change in the propagation delay of 0.5 $\mu$s will cause a phase uncertainty of $\pi$ rad.

- In some channels the time delay in the propagation of the signal from the transmitter to the receiver may change rapidly and in an apparently random fashion.

- In the absence of the knowledge of the carrier phase, we may treat this signal parameter as a random variable and determine the form of the optimum receiver for recovering the transmitted information from the received signal.
5.4.1 Optimum Receiver for Binary Signals

We consider a binary communication system that uses the two carrier modulated signal $s_1(t)$ and $s_2(t)$ to transmit the information, where:

$$s_m(t) = \text{Re}\left[s_{lm}(t)e^{j2\pi f_c t}\right], \quad m = 1, 2, \quad 0 \leq t \leq T$$

and $s_{lm}(t)$, $m=1,2$ are the equivalent low-pass signals.

The two signals are assumed to have equal energy

$$\varepsilon = \int_0^T s_m^2(t)dt = \frac{1}{2} \int_0^T |s_{ml}(t)|^2 dt$$

The two signals are characterized by the complex-valued correlation coefficient

$$\rho_{12} \equiv \rho = \frac{1}{2\varepsilon} \int_0^T s_{l1}^*(t)s_{l2}(t)dt$$
5.4.1 Optimum Receiver for Binary Signals

The received signal is assumed to be a phase-shifted version of the transmitted signal and corrupted by the additive noise

\[ n(t) = \text{Re}\left\{n_c(t) + jn_s(t)e^{j2\pi f_c t}\right\} = \text{Re}\left[z(t)e^{j2\pi f_c t}\right] \]

The received signal may be expressed as

\[ r(t) = \text{Re}\left\{s_{lm}(t)e^{j\phi} + z(t)e^{j2\pi f_c t}\right\} \]

where \( r_l(t) = s_{lm}(t)e^{j\phi} + z(t), \quad 0 \leq t \leq T \)

\( r_l(t) \) is the equivalent low-pass received signal.

This received signal is now passed through a demodulator whose sampled output at \( t = T \) is passed to the detector.
5.4.1 Optimum Receiver for Binary Signals

The optimum demodulator

- In section 5.1.1, we demonstrated that if the received signals was correlated with a set of orthogonal functions \( \{f_n(t)\} \) that spanned the signal space, the outputs from the bank of correlators provide a set of sufficient statistics for the detector to make a decision that minimizes the probability error.
- We also demonstrated that a bank of matched filters could be substituted for the bank of correlations.
- A similar orthogonal decomposition can be employed for a received signal with an unknown carrier phase.
- It is mathematically convenient to deal with the equivalent low-pass signal and to specify the signal correlators or matched filters in terms of the equivalent low-pass signal waveforms.
5.4.1 Optimum Receiver for Binary Signals

The optimum demodulator

- The impulse response \( h_l(t) \) of a filter that is matched to the complex-valued equivalent low-pass signal \( s_l(t) \), \( 0 \leq t \leq T \) is given as (from Problem 5.6) \( h_l(t) = s_l^*(T-t) \) and the output of such filter at \( t=T \) is simply (4.1-24) \( \int_0^T |s_l(t)|^2 \, dt = 2\varepsilon \) where \( \varepsilon \) is the signal energy.

- A similar result is obtained if the signal \( s_l(t) \) is correlated with \( s_l^*(t) \) and the correlator is sampled signal \( s_l(t) \) at \( t=T \).

- The optimum demodulator for the equivalent low-pass received signal \( s_l(t) \) may be realized by two matcher filters in parallel, one matched to \( s_{l1}(t) \) and the other to \( s_{l2}(t) \).
5.4.1 Optimum Receiver for Binary Signals

- The optimum demodulator
  - Optimum receiver for binary signals

- The output of the matched filters or correlators at the sampling instant are the two complex numbers

\[ r_m = r_{mc} + j r_{ms}, \quad m = 1, 2 \]
The optimum demodulator

Suppose that the transmitted signal is $s_1(t)$. It can be shown that (Problem 5.41)

\[
\begin{align*}
  r_1 &= 2\varepsilon \cos \phi + n_{1c} + j(2\varepsilon \sin \phi + n_{1s}), \\
  r_2 &= 2\varepsilon |\rho| \cos(\phi - \alpha_0) + n_{2c} + j[2\varepsilon |\rho| \sin(\phi - \alpha_0) + n_{2s}]
\end{align*}
\]

where $\rho$ is the complex-valued correlation coefficient of two signals $s_{l1}(t)$ and $s_{l2}(t)$, which may be expressed as $\rho = |\rho| \exp(j\alpha_0)$.

The random noise variables $n_{1c}$, $n_{1s}$, $n_{2c}$, and $n_{2s}$ are jointly Gaussian, with zero-mean and equal variance.
5.4.1 Optimum Receiver for Binary Signals

The optimum detector

The optimum detector observes the random variables \( [r_{1c} r_{1s} r_{2c} r_{2s}] = r \), where \( r_1 = r_{1c} + j r_{1s} \) and \( r_2 = r_{2c} + j r_{2s} \), and bases its decision on the posterior probabilities \( P(s_m | r) \), \( m=1,2 \).

These probabilities may be expressed as

\[
P(s_m | r) = \frac{p(r | s_m)P(s_m)}{p(r)}, \quad m = 1,2
\]

The optimum decision rule may be expressed as

\[
P(s_1 | r) \geq P(s_2 | r) \quad \text{or} \quad \frac{p(r | s_1)}{p(r | s_2)} \geq \frac{P(s_2)}{P(s_1)}
\]

Likelihood ratio.
5.4.1 Optimum Receiver for Binary Signals

The optimum detector

- The ratio of PDFs on the left-hand side is the likelihood ratio, which we denote as
  \[ \Lambda(r) = \frac{p(r | s_1)}{p(r | s_2)} \]

- The right-hand side is the ratio of the two prior probabilities, which takes the value of unity when the two signals are equally probable.

- The probability density functions \( p(r | s_1) \) and \( p(r | s_2) \) can be obtained by averaging the PDFs \( p(r | s_m, \phi) \) over the PDF of the random carrier phase, i.e.,
  \[ p(r | s_m) = \int_0^{2\pi} p(r | s_m, \phi)p(\phi)d\phi \]
5.4.1 Optimum Receiver for Binary Signals

The optimum detector

For the special case in which the two signals are orthogonal, i.e., \( \rho = 0 \), the outputs of the demodulator are (from 5.4-10):

\[
\begin{align*}
    r_1 &= r_{1c} + jr_{1s} \\
    &= 2\varepsilon \cos \phi + n_{1c} + j(2\varepsilon \sin \phi + n_{1s}) \\
    r_2 &= r_{2c} + jr_{2s} \\
    &= n_{2c} + jn_{2s}
\end{align*}
\]

where \((n_{1c}, n_{1s}, n_{2c}, n_{2s})\) are mutually uncorrelated and, statistically independent, zero-mean Gaussian random variable.
5.4.1 Optimum Receiver for Binary Signals

The optimum detector

- The joint PDF of \( r = [r_{1c}, r_{1s}, r_{2c}, r_{2s}] \) may be expressed as a product of the marginal PDFs:

\[
p(r_{1c}, r_{1s} \mid s_1, \phi) = \frac{1}{2\pi\sigma^2} \exp \left[ -\frac{(r_{1c} - 2\varepsilon\cos\phi)^2 + (r_{1s} - 2\varepsilon\sin\phi)^2}{2\sigma^2} \right]
\]

\[
p(r_{2c}, r_{2s} \mid s_1, \phi) = \frac{1}{2\pi\sigma^2} \exp \left( -\frac{r_{2c}^2 + r_{2s}^2}{2\sigma^2} \right)
\]

where \( \sigma^2 = 2 \varepsilon N_0 \).

- The uniform PDF for the carrier phase \( \phi \) \( (p(\phi) = 1/2 \pi) \) represents the most ignorance that can be exhibited by the detector.

- This is called the least favorable PDF for \( \phi \).
5.4.1 Optimum Receiver for Binary Signals

The optimum detector

With \( p(\phi) = \frac{1}{2} \pi, 0 \leq \phi \leq 2 \pi \), substituted into the integral in \( p(r|s_m) \), we obtain:

\[
\frac{1}{2\pi} \int_0^{2\pi} p(r_{1c}, r_{1s} | s_1, \phi) d\phi
\]

\[
= \frac{1}{2\pi\sigma^2} \exp \left( -\frac{r_{1c}^2 + r_{1s}^2 + 4\varepsilon^2}{2\sigma^2} \right) \frac{1}{2\pi} \int_0^{2\pi} \exp \left[ \frac{2\varepsilon (r_{1c} \cos \phi + r_{1s} \sin \phi)}{\sigma^2} \right] d\phi
\]

\[
\frac{1}{2\pi} \int_0^{2\pi} \exp \left[ \frac{2\varepsilon (r_{1c} \cos \phi + r_{1s} \sin \phi)}{\sigma^2} \right] d\phi = I_0 \left( \frac{2\varepsilon \sqrt{r_{1c}^2 + r_{1s}^2}}{\sigma^2} \right)
\]

where \( I_0(x) \) is the modified Bessel function of zeroth order.
5.4.1 Optimum Receiver for Binary Signals

The optimum detector

- By performing a similar integration under the assumption that the signal $s_2(t)$ was transmitted, we obtain the result

$$p(r_{2c}, r_{2s} \mid s_2) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{r_{2c}^2 + r_{2s}^2 + 4\varepsilon^2}{2\sigma^2}\right) I_0\left(\frac{2\varepsilon \sqrt{r_{2c}^2 + r_{2s}^2}}{\sigma^2}\right)$$

- When we substitute these results into the likelihood ratio given by equation $\Lambda(r)$, we obtain the result

$$\Lambda(r) = \frac{I_0\left(2\varepsilon \sqrt{r_{1c}^2 + r_{1s}^2} / \sigma^2\right)_{s_1} P(s_2)}{I_0\left(2\varepsilon \sqrt{r_{2c}^2 + r_{2s}^2} / \sigma^2\right)_{s_2} P(s_1)}$$

- The optimum detector computes the two envelopes $\sqrt{r_{1c}^2 + r_{1s}^2}$ and $\sqrt{r_{2c}^2 + r_{2s}^2}$ and the corresponding values of the Bessel function $I_0\left(2\varepsilon \sqrt{r_{1c}^2 + r_{1s}^2} / \sigma^2\right)$ and $I_0\left(2\varepsilon \sqrt{r_{2c}^2 + r_{2s}^2} / \sigma^2\right)$ to form the likelihood ratio.
5.4.1 Optimum Receiver for Binary Signals

- The optimum detector
  - we observe that this computation requires knowledge of the noise variance $\sigma^2$.
  - The likelihood ratio is then compared with the threshold $P(s_2)/P(s_1)$ to determine which signal was transmitted.
- A significant simplification in the implementation of the optimum detector occurs when the two signals are equally probable. In such a case the threshold becomes unity, and, due to the monotonicity of Bessel function shown in figure, the optimum detection rule simplifies to

$$\sqrt{r_{1c}^2 + r_{1s}^2} \geq \sqrt{r_{2c}^2 + r_{2s}^2}$$
The optimum detector

- The optimum detector bases its decision on the two envelopes and $\sqrt{r_{1c}^2 + r_{1s}^2}$, and $\sqrt{r_{2c}^2 + r_{2s}^2}$, it is called an envelope detector.

- We observe that the computation of the envelopes of the received signal samples at the output of the demodulator renders the carrier phase irrelevant in the decision as to which signal was transmitted.

- Equivalently, the decision may be based on the computation of the squared envelope $r_{1c}^2 + r_{1s}^2$ and $r_{2c}^2 + r_{2s}^2$, in which case the detector is call a square-law detector.
5.4.1 Optimum Receiver for Binary Signals

Detection of binary FSK signal

- Recall that in binary FSK we employ two different frequencies, say $f_1$ and $f_2 = f_1 + \Delta f$, to transmit a binary information sequence.

- The choice of minimum frequency separation $\Delta f = f_2 - f_1$ is considered below:

$$s_1(t) = \frac{2\varepsilon_b}{T_b} \cos 2\pi f_1 t, \quad s_2(t) = \frac{2\varepsilon_b}{T_b} \cos 2\pi f_2 t, \quad 0 \leq t \leq T_b$$

- The equivalent low-pass counterparts are:

$$s_{l1}(t) = \frac{2\varepsilon_b}{T_b}, \quad s_{l2}(t) = \frac{2\varepsilon_b}{T_b} e^{j2\pi\Delta ft}, \quad 0 \leq t \leq T_b$$

- The received signal may be expressed as:

$$r(t) = \sqrt{\frac{2\varepsilon_b}{T_b}} \cos (2\pi f_m t + \phi_m) + n(t), \quad m = 0, 1$$
5.4.1 Optimum Receiver for Binary Signals

- Detection of binary FSK signal
- Demodulation and square-law detection:

\[ f_{kc}(t) = \sqrt{\frac{2}{T_b}} \cos \left[ (2\pi f_1 + 2\pi k\Delta f)t \right], \quad k = 0, 1 \]
\[ f_{ks}(t) = \sqrt{\frac{2}{T_b}} \sin \left[ (2\pi f_1 + 2\pi k\Delta f)t \right], \quad k = 0, 1 \]
Detection of binary FSK signal

If the $m$th signal is transmitted, the four samples at the detector may be expressed as:

$$r_{kc} = \int_0^{T_b} r(t) \cdot f_{kc}(t) \, dt = \int_0^{T_b} \left\{ \frac{2\epsilon_b}{T_b} \cos \left( 2\pi \left( f_1 + m\Delta f \right) t + \phi_m \right) + n(t) \right\} \cdot \left\{ \frac{2}{T_b} \cos \left( 2\pi f_1 + 2\pi k\Delta f \right) t \right\} \, dt$$

$$= \frac{2\sqrt{\epsilon_b}}{T_b} \int_0^{T_b} \left\{ \cos \left( 2\pi \left( f_1 + m\Delta f \right) t \right) \cos \phi_m - \sin \left( 2\pi \left( f_1 + m\Delta f \right) t \right) \sin \phi_m \right\} \left\{ \cos \left( 2\pi \left( f_1 + k\Delta f \right) t \right) \right\} \, dt + n_{kc}$$

$$= \frac{\sqrt{\epsilon_b}}{T_b} \int_0^{T_b} \left\{ \cos \left[ 2\pi \left( m - k \right) \Delta f \right] + \cos \left[ 2\pi \left( 2f_1 + (m + k)\Delta f \right) t \right] \right\} \cos \phi_m -$$

$$\left\{ \sin \left[ 2\pi \left( m - k \right) \Delta f \right] - \sin \left[ 2\pi \left( 2f_1 + (m + k)\Delta f \right) t \right] \right\} \sin \phi_m \, dt + n_{kc}$$

$$= \frac{\sqrt{\epsilon_b}}{T_b} \left\{ \frac{\sin \left[ 2\pi \left( m - k \right) \Delta f \right]}{2\pi \left( m - k \right) \Delta f} \cos \phi_m + \frac{\cos \left[ 2\pi \left( m - k \right) \Delta f \right]}{2\pi \left( m - k \right) \Delta f} \sin \phi_m \right\} _0^{T_b} + n_{kc}$$

$$= \sqrt{\epsilon_b} \left\{ \frac{\sin \left[ 2\pi \left( m - k \right) \Delta f T_b \right]}{2\pi \left( m - k \right) \Delta f T_b} \cos \phi_m + \frac{\cos \left[ 2\pi \left( m - k \right) \Delta f T_b \right] - 1}{2\pi \left( m - k \right) \Delta f T_b} \sin \phi_m \right\} + n_{kc} \quad k, m = 0, 1$$
5.4.1 Optimum Receiver for Binary Signals

- Detection of binary FSK signal

If the $m$th signal is transmitted, the four samples at the detector may be expressed as (cont.):

$$r_{ks} = \int_{0}^{T_b} r(t) \cdot f_{ks}(t) \, dt = \int_{0}^{T_b} \left\{ \sqrt{\frac{2\varepsilon_b}{T_b}} \cos \left[ 2\pi \left( f_i + m\Delta f \right) t + \phi_m \right] + n(t) \right\} \cdot \left\{ \sqrt{\frac{2}{T_b}} \sin \left[ (2\pi f_i + 2\pi k\Delta f) t \right] \right\} \, dt$$

$$= \frac{2\sqrt{\varepsilon_b}}{T_b} \int_{0}^{T_b} \left\{ \cos \left[ 2\pi \left( f_i + m\Delta f \right) t \right] \cos \phi_m - \sin \left[ 2\pi \left( f_i + m\Delta f \right) t \right] \sin \phi_m \right\} \left\{ \sin \left[ 2\pi \left( f_i + k\Delta f \right) t \right] \right\} \, dt + n_{ks}$$

$$= \frac{\sqrt{\varepsilon_b}}{T_b} \int_{0}^{T_b} \left\{ -\sin \left[ 2\pi \left( m - k \right) \Delta f t \right] + \sin \left[ 2\pi \left( 2f_i + (m + k) \Delta f \right) t \right] \right\} \cos \phi_m -$$

$$\left\{ \cos \left[ 2\pi \left( m - k \right) \Delta f t \right] - \cos \left[ 2\pi \left( 2f_i + (m + k) \Delta f \right) t \right] \right\} \sin \phi_m \, dt + n_{ks}$$

$$= \frac{\sqrt{\varepsilon_b}}{T_b} \left\{ \cos \left[ 2\pi \left( m - k \right) \Delta f \right] \cos \phi_m - \sin \left[ 2\pi \left( m - k \right) \Delta f \right] \sin \phi_m \right\} \int_{0}^{T_b} + n_{ks}$$

$$= \sqrt{\varepsilon_b} \left\{ \frac{\cos \left[ 2\pi \left( m - k \right) \Delta f T_b \right] - 1}{2\pi \left( m - k \right) \Delta f T_b} \cos \phi_m - \frac{\sin \left[ 2\pi \left( m - k \right) \Delta f T_b \right]}{2\pi \left( m - k \right) \Delta f T_b} \sin \phi_m \right\} + n_{ks} \quad k, m = 0, 1$$
5.4.1 Optimum Receiver for Binary Signals

- Detection of binary FSK signal
  - We observe that when $k = m$, the sampled values to the detector are
    \[ r_{mc} = \sqrt{\varepsilon_b} \cos \phi_m + n_{mc} \quad k=m \]
    \[ r_{ms} = -\sqrt{\varepsilon_b} \sin \phi_m + n_{ms} \]
  - We observe that when $k \neq m$, the signal components in the samples $r_{kc}$ and $r_{ks}$ will vanish, independently of the values of the phase shifts $\phi_k$, provided that the frequency separation between successive frequency is $\Delta f = 1/T$.
  - In such case, the other two correlator outputs consist of noise only, i.e.,
    \[ r_{kc} = n_{kc}, \quad r_{ks} = n_{ks}, \quad k \neq m. \]
5.4.2 Optimum Receiver for $M$-ary Orthogonal Signals

- If the equal energy and equally probable signal waveforms are represented as
  $$s_m(t) = \text{Re} \left[ s_{lm}(t) e^{j2\pi ft} \right], \quad m = 1, 2, \ldots, M, \quad 0 \leq t \leq T$$
  where $s_{lm}(t)$ are the equivalent low-pass signals.

- The optimum correlation-type or matched-filter-type demodulator produces the $M$ complex-valued random variables
  $$r_m = r_{mc} + j r_{ms} = \int_0^T r_l(t) h_{lm}^*(T-t) dt$$
  $$= \int_0^T r_l(t) s_{lm}^*(T-(T-t)) dt = \int_0^T r_l(t) s_{lm}^*(t) dt, \quad m = 1, 2, \ldots, M$$
  where $r_l(t)$ is the equivalent low-pass signals.

- The optimum detector, based on a random, uniformly distributed carrier phase, computes the $M$ envelopes
  $$|r_m|^2 = \sqrt{r_{mc}^2 + r_{ms}^2}, \quad m = 1, 2, \ldots, M$$
  or, the squared envelopes $|r_m|^2$, and selects the signal with the largest envelope.
Optimum receiver for \( M \)-ary orthogonal FSK signals.

- There are \( 2M \) correlators: two for each possible transmitted frequency.
- The minimum frequency separation between adjacent frequencies to maintain orthogonality is \( \Delta f = \frac{1}{T} \).

Demodulation of \( M \)-ary signals for noncoherent detection.
5.4.3 Probability of Error for Envelope Detection of $M$-ary Orthogonal Signals

❖ We assume that the $M$ signals are equally probable a priori and that the signal $s_1(t)$ is transmitted in the signal interval $0 \leq t \leq T$.

❖ The $M$ decision metrics at the detector are the $M$ envelopes

\[ |r_m| = \sqrt{r_{mc}^2 + r_{ms}^2}, \quad m = 1, 2, \ldots, M \]

where

\[ r_{1c} = \sqrt{\varepsilon_s} \cos \phi_1 + n_{1c} \]

\[ r_{1s} = \sqrt{\varepsilon_s} \sin \phi_1 + n_{1s} \]

and

\[ r_{mc} = n_{mc}, \quad r_{ms} = n_{ms}, \quad m = 2, 3, \ldots, M \]

❖ The additive noise components $\{n_{mc}\}$ and $\{n_{ms}\}$ are mutually statistically independent zero-mean Gaussian variables with equal variance $\sigma^2=N_0/2$. 
5.4.3 Probability of Error for Envelope Detection of $M$-ary Orthogonal Signals

- The PDFs of the random variables at the input to the detector are

$$p_{r_1}(r_{1c}, r_{1s}) = \frac{1}{2\pi \sigma^2} \exp \left( - \frac{r_{1c}^2 + r_{1s}^2 + \varepsilon_s}{2\sigma^2} \right) I_0 \left( \sqrt{\varepsilon_s \left( r_{1c}^2 + r_{1s}^2 \right)} \right)$$

$$p_{r_m}(r_{mc}, r_{ms}) = \frac{1}{2\pi \sigma^2} \exp \left( - \frac{r_{mc}^2 + r_{ms}^2}{2\sigma^2} \right), \quad m = 2, 3, \ldots, M$$

- We define the normalized variables

$$R_m = \frac{\sqrt{r_{mc}^2 + r_{ms}^2}}{\sigma}$$

$$\Theta_m = \tan^{-1} \frac{r_{ms}}{r_{mc}}$$
5.4.3 Probability of Error for Envelope Detection of $M$-ary Orthogonal Signals

Clearly, $r_{mc} = \sigma R_m \cos \Theta_m$ and $r_{ms} = \sigma R_m \sin \Theta_m$. The Jacobian of this transformation is

$$
|J| = \begin{vmatrix}
\sigma \cos \Theta_m & \sigma \sin \Theta_m \\
-\sigma R_m \sin \Theta_m & \sigma R_m \cos \Theta_m
\end{vmatrix} = \sigma^2 R_m
$$

Consequently,

$$
p( R_1, \Theta_1 ) = \frac{R_1}{2\pi} \exp \left[ -\frac{1}{2} \left( R_1^2 + 2\frac{\varepsilon_s}{N_0} \right) \right] I_0 \left( \sqrt{\frac{2\varepsilon_s}{N_0}} R_1 \right)
$$

$$
p( R_m, \Theta_m ) = \frac{R_m}{2\pi} \exp \left( -\frac{1}{2} R_m^2 \right), \quad m = 2,3,\ldots,M
$$

Finally, by averaging $p( R_m, \Theta_m )$ over $\Theta_m$, the factor of $2 \pi$ is eliminated.

Thus, we find that $R_1$ has a Rice probability distribution and $R_m$, $m=2,3,\ldots,M$, are each Rayleigh-distributed.
5.4.3 Probability of Error for Envelope Detection of $M$-ary Orthogonal Signals

The probability of a correct decision is simply the probability that $R_1 > R_2$, and $R_1 > R_3$, …, and $R_1 > R_m$. Hence,

$$P_c = P\left( R_2 < R_1, R_3 < R_1, \ldots, R_M < R_1 \right)$$

$$= \int_0^\infty P\left( R_2 < R_1, R_3 < R_1, \ldots, R_M < R_1 \mid R_1 = x \right) p_{R_1}(x) \, dx$$

Because the random variables $R_m$, $m=2,3,\ldots,M$, are statistically independent and identically distributed, the joint probability conditioned on $R_1$ factors into a product of $M-1$ identical terms.

where

$$P_c = \int_0^\infty \left[ P\left( R_2 < R_1 \mid R_1 = x \right) \right]^{M-1} p_{R_1}(x) \, dx \quad (5.4-42)$$

$$P\left( R_2 < R_1 \mid R_1 = x \right) = \int_0^x p_{R_2}(r_2) \, dr_2 = 1 - e^{-x^2/2} \quad (\text{From Eq 2.1-129})$$
The \((M-1)\)th power may be expressed as
\[
\left(1-e^{-x^2/2}\right)^{M-1} = \sum_{n=0}^{M-1} (-1)^n \binom{M-1}{n} e^{-nx^2/2}
\]

Substitution of this result into Equation 5.4-42 and integration over \(x\) yield the probability of a correct decision as
\[
P_c = \sum_{n=0}^{M-1} (-1)^n \binom{M-1}{n} \frac{1}{n+1} \exp\left[-\frac{n\varepsilon_s}{(n+1)N_0}\right]
\]
where \(\varepsilon_s/N_0\) is the SNR per symbol.

The probability of a symbol error, which is \(P_m = 1 - P_c\), becomes
\[
P_m = \sum_{n=1}^{M-1} (-1)^{n+1} \binom{M-1}{n} \frac{1}{n+1} \exp\left[-\frac{nk\varepsilon_b}{(n+1)N_0}\right]
\]
where \(\varepsilon_b/N_0\) is the SNR per bit.
5.4.3 Probability of Error for Envelope Detection of $M$-ary Orthogonal Signals

- For binary orthogonal signals ($M=2$), Equation reduces to the simple form
  \[ P_2 = \frac{1}{2} e^{-\varepsilon_b/2N_0} \]

- For $M>2$, we may compute the probability of a bit error by making use of the relationship
  \[ P_b = \frac{2^{k-1}}{2^k - 1} P_M \]
  which was established in Section 5.2.

- Figure shows the bit-error probability as a function of SNR per bit $\gamma_b$ for $M=2, 4, 8, 16, \text{ and } 32$. 

\[ \text{Fall, 2004.} \quad 199 \quad \text{WITS Lab, NSYSU.} \]
5.4.3 Probability of Error for Envelope Detection of \( M \)-ary Orthogonal Signals

- Just as in the case of coherent detection of \( M \)-ary orthogonal signals, we observe that for any given bit-error probability, the SNR per bit decreases as \( M \) increases.
- It will be shown in Chapter 7 that, in the limit as \( M \to \infty \), the probability of bit error \( P_b \) can be made arbitrarily small provided that the SNR per bit is greater than the Shannon limit of -1.6dB.
- The cost for increasing \( M \) is the bandwidth required to transmit the signals.
- For \( M \)-ary FSK, the frequency separation between adjacent frequencies is \( \Delta f = 1/T \) for signal orthogonality.
- The bandwidth required for the \( M \) signals is \( W = M \Delta f = M/T \).
5.4.3 Probability of Error for Envelope Detection of $M$-ary Orthogonal Signals

- The bit rate is $R = k/T$, where $k = \log_2 M$.
- Therefore, the bit rate-to-bandwidth ratio is

$$\frac{R}{W} = \frac{\log_2 M}{M}$$
5.4.4 Probability of Error for Envelope Detection of Correlated Binary Signals

- In this section, we consider the performance of the envelope detector for binary, equal-energy correlated signals.
- When the two signals are correlated, the input to the detector is the complex-valued random variables given by Equation 5.4-10. We assume that the detector bases its decision on the envelopes $|r_1|$ and $|r_2|$, which are correlated (statistically dependent).
- The marginal PDFs of $R_1=|r_1|$ and $R_2=|r_2|$ are Ricean distributed and may be expressed as

$$p(R_m) = \begin{cases} \frac{R_m}{2\varepsilon N_0} \exp\left(-\frac{R_m^2 + \beta_m^2}{4\varepsilon N_0}\right) I_0\left(\frac{\beta_m R_m}{2\varepsilon N_0}\right) & (R_m > 0) \\ 0 & (R_m < 0) \end{cases}$$

- $m=1,2$, where $\beta_1=2\varepsilon$ and $\beta_2=2\varepsilon|\rho|$, based on the assumption that signal $s_1(t)$ was transmitted.
Since \( R_1 \) and \( R_2 \) are statistically dependent as a consequence of the nonorthogonality of the signals, the probability of error may be obtained by evaluating the double integral

\[
P_b = P\left( R_2 > R_1 \right) = \int_{0}^{\infty} \int_{x_1}^{\infty} p(x_1, x_2) \, dx_1 \, dx_2
\]

where \( p(x_1, x_2) \) is the joint PDF of the envelopes \( R_1 \) and \( R_2 \).

This approach was first used by Helstrom (1995), who determined the joint PDF of \( R_1 \) and \( R_2 \) and evaluated the double integral.

An alternative approach is based on the observation that the probability of error may also be expressed as

\[
P_b = P\left( R_2 > R_1 \right) = P\left( R_2^2 > R_1^2 \right) = P\left( R_2^2 - R_1^2 > 0 \right)
\]
5.4.4 Probability of Error for Envelope Detection of Correlated Binary Signals

- But $R_2^2 - R_1^2$ is a special case of a general quadratic form in complex-valued Gaussian random variable, treated later in Appendix B.

- For the special case under consideration, the derivation yields the error probability in the form

$$P_b = Q_1(a, b) - \frac{1}{2} e^{-(a^2 + b^2)/2} I_0(ab)$$

where

$$a = \sqrt{\frac{\varepsilon_b}{2N_0}} \left(1 - \sqrt{1 - |\rho|^2}\right) \quad b = \sqrt{\frac{\varepsilon_b}{2N_0}} \left(1 + \sqrt{1 - |\rho|^2}\right)$$

- $Q_1(a, b)$ is the Marcum $Q$ function defined in 2.1-123 and $I_0(x)$ is the modified Bessel function of order zero.
The error probability $P_b$ is illustrated in Figure for several values of $|\rho|$. $P_b$ is minimized when $\rho = 0$; that is, when the signals are orthogonal.
5.4.4 Probability of Error for Envelope Detection of Correlated Binary Signals

For this case, \( a = 0, b = \sqrt{\varepsilon_b / N_0} \), and Equation reduces to

\[
P_b = Q_1\left(0, \sqrt{\frac{\varepsilon_b}{N_0}}\right) - \frac{1}{2} e^{-\varepsilon_b / 2N_0}
\]

From the definition of \( Q_1(a, b) \) in Equation 2.1-123, it follows that

\[
Q_1\left(0, \sqrt{\frac{\varepsilon_b}{N_0}}\right) = e^{-\varepsilon_b / 2N_0}
\]

Substitution of these relations yields the desired result given previously in Equation 5.4-47. On the other hand, when \( |\rho| = 1 \), the error probability becomes \( P_b = 1/2 \), as expected.
5.5 Performance Analysis for Wireline and Radio Communication Systems

- In the communication of digital signals through an AWGN channel, we have observed that the performance of the communication system, measured in terms of the probability of error, depends solely on the received SNR, $\varepsilon_b/\sqrt{N_0}$, where $\varepsilon_b$ is the transmitted energy per bit and $N_0/2$ is the power spectral density of the additive noise.

- The additive noise ultimately limits the performance of the communication system.

- Another factor that affects the performance of communication system is channel attenuation. All physical channels, including wire lines and radio channels, are lossy.
The signal is attenuated as it travels through the channel. The simple mathematical model for the attenuation shown in the following figure may be used for the channel.

If the transmitted signal is $s(t)$, the received signal, with $0 < \alpha \leq 1$ is

$$r(t) = \alpha s(t) + n(t)$$
5.5 Performance Analysis for Wireline and Radio Communication Systems

- If the energy in the transmitted signal is $\varepsilon_b$, the energy in the received signal is $\alpha^2 \varepsilon_b$. Consequently, the received signal has an SNR $\alpha^2 \varepsilon_b / N_0$.

- The effect of signal attenuation is to reduce the energy in the received signal and thus to render the communication system more vulnerable to additive noise.

- In the analog communication systems, amplifiers called *repeaters* are used to periodically boost the signal strength in transmission through the channel. However, each amplifier also boosts the noise in the system.

- In contrast, digital communication systems allow us to detect and regenerate a clean (noise-free) signal in a transmission channel.

- Such devices, called *regenerative repeaters*, are frequently used in wireline and fiber-optic communication channels.
5.5.1 Regenerative Repeaters

The front end of each regenerative repeater consists of a demodulator/detector that demodulates and detects the transmitted digital information sequence sent by the preceding repeater.

Once detected, the sequence is passed to the transmitter side of the repeater, which maps the sequence into signal waveforms that are transmitted to the next repeater.

This type of repeater is called a regenerative repeater.

Since a noise-free signal is regenerated at the each repeater, the additive noise does not accumulate. However, when errors occur in the detector of a repeater, the errors are propagated forward to the following repeaters in the channel.
5.5.1 Regenerative Repeaters

To evaluate the effect of errors on the performance of overall system, suppose that the modulation is binary PAM, so that the probability of a bit error for one hop is

\[ P_b = Q\left(\sqrt{\frac{2\varepsilon_b}{N_0}}\right) \]

Since errors occur with low probability, the overall probability of error over \( K \) regenerative repeaters may be approximated as

\[ P_b = 1 - P_c = 1 - \left[ 1 - Q\left(\sqrt{\frac{2\varepsilon_b}{N_0}}\right) \right]^K \]

\[ = 1 - \left[ 1 - KQ\left(\sqrt{\frac{2\varepsilon_b}{N_0}}\right) + o^2 \left\{ Q\left(\sqrt{\frac{2\varepsilon_b}{N_0}}\right) \right\} \right] \approx KQ\left(\sqrt{\frac{2\varepsilon_b}{N_0}}\right) \]
5.5.1 Regenerative Repeaters

- In contrast, the use of $K$ analog repeaters in the channel reduces the received SNR by $K$, and hence, the bit error probability is

$$P_b \approx Q\left(\sqrt{\frac{2\varepsilon_b}{KN_0}}\right)$$

- For the same probability of error performance, the use of regenerative repeaters results in a significant saving in transmitter power compared with analog repeaters.

- In digital communication systems, regenerative repeaters are preferable. However, in wireline telephone channels that are used to transmit both analog and digital signals, analog repeaters are generally employed.
5.5.1 Regenerative Repeaters

Example 5.5-1

A binary digital communication system transmits data over a wireline channel of length 1000 km. Repeaters are used every 10 km to offset the effect of channel attenuation. Let us determine the $\varepsilon_b/N_0$ that is required to achieve a probability of bit error of $10^{-5}$ if (a) analog repeaters are employed, and (b) regenerative repeaters are employed.

The number of repeaters used in the system is $K=100$. If regenerative repeaters are used, $\varepsilon_b/N_0$ obtained is:

$$10^{-5} = 100 Q\left(\sqrt{\frac{2\varepsilon_b}{N_0}}\right)$$
$$10^{-7} = Q\left(\sqrt{\frac{2\varepsilon_b}{N_0}}\right)$$

which is approximately 11.3 dB.
Example 5.5-1 (cont.)

If analog repeaters are used, the $\frac{\varepsilon_b}{N_0}$ obtained is

$$10^{-5} = Q\left(\sqrt{\frac{2\varepsilon_b}{100N_0}}\right)$$

Which yields $\frac{\varepsilon_b}{N_0} \approx 29.6$ dB.

The difference in the required SNR is about 18.3 dB, or approximately 70 times the transmitter power of the digital communication system.
5.5.2 Link Budget Analysis in Radio Communication System

In the design of radio communication systems that transmit over line-of-sight microwave channels and satellite channels, the system designer must specify the size of the transmit and receive antennas, the transmitted power, and the SNR required to achieve a given level of performance at some desired data rate.

Consider a transmit antenna that radiates isotropically in the free space at a power level of $P_T$ watts:
5.5.2 Link Budget Analysis in Radio Communication System

- The power density at a distance $d$ from the antenna is $\frac{P_T}{4 \pi d^2}$ W/m².

- If the transmitting antenna has some directivity in a particular direction, the power density in the direction is increased by factor called the antenna gain and denoted by $G_T$. In such case, the power density at the distance $d$ is $\frac{P_T G_T}{4 \pi d^2}$ W/m².

- The product $P_T G_T$ is usually called the effective radiated power (ERP or EIRP), which is basically the radiated power relative to an isotropic antenna, for which $G_T=1$.

- A receiving antenna pointed in the direction of the radiated power gathers a portion of the power that is proportional to its cross-sectional area.
The received power extracted by the antenna may be expressed as

\[ P_R = \frac{P_T G_T A_R}{4\pi d^2} \]

where \( A_R \) is the effective area of the antenna.

From electromagnetic field theory, we obtain the basic relationship between the gain \( G_R \) of an antenna and its effective area as

\[ A_R = \frac{G_R \lambda^2}{4\pi} \text{ m}^2 \]

where \( \lambda = c/f \) is the wavelength of the transmitted signal, \( c \) is the speed of light (3×10^8 m/s), and \( f \) is the frequency of the transmitted signal.
5.5.2 Link Budget Analysis in Radio Communication System

If we substitute $A_R$ into $P_R$, we obtain an expression for the received power in the form

$$P_R = \frac{P_T G_T G_R}{(4\pi d/\lambda)^2}$$

$$L_s = \left(\frac{\lambda}{4\pi d}\right)^2$$

The factor $L_s$ is called the free-space path loss.

If other losses, such as atmospheric losses, are encountered in the transmission of the signal, they may be accounted for by introducing an additional loss factor, say $L_a$.

Therefore, the received power may be written in general as

$$P_R = P_T G_T G_R L_s L_a$$
5.5.2 Link Budget Analysis in Radio Communication System

The important characteristics of an antenna are its gain and its effective area. These generally depend on the wavelength of the radiated power and the physical dimensions of the antenna.

For example, a *parabolic (dish) antenna* of diameter $D$ has an effective area

$$A_R = \frac{G_R \lambda^2}{4\pi} = \frac{1}{4} \pi D^2 \eta$$

where $\pi D^2/4$ is the physical area and $\eta$ is the illumination efficiency factor, which falls in the range $0.5 \leq \eta \leq 0.6$.

The antenna gain for a parabolic antenna of diameter $D$ is

$$G_R = \eta \left(\frac{\pi D}{\lambda}\right)^2$$
5.5.2 Link Budget Analysis in Radio Communication System

As a second example, a *horn antenna* of physical area $A$ has an efficiency factor of 0.8, an effective area of $A_R = 0.8A$, and an antenna gain of

$$G_R = \frac{10A}{\lambda^2}$$

Another parameter that is related to the gain (directivity) of an antenna is its beamwidth, which we denote as $\Theta_B$ and which is illustrated graphically in the following figure.
5.5.2 Link Budget Analysis in Radio Communication System

- The -3dB beamwidth of a parabolic antenna is approximately
  \[ \Theta_B = 70\left(\frac{\lambda}{D}\right) \]
  so that \( G_T \) is inversely proportional to \( \Theta_B^2 \).
- That is, a decrease of the beamwidth by a factor of 2, which is
  obtained by doubling the diameter \( D \), increases the antenna gain
  by a factor of 4 (6dB).
- Based on the general relationship for the received signal power
  given by \( P_R = P_T G_T G_R L_s L_a \), the system designer can compute \( P_R \)
  from a specification of the antenna gains and the distance
  between the transmitter and the receiver.
- Such computations are usually done on a power basis, so that
  \[ (P_R)_{dB} = (P_T)_{dB} + (G_T)_{dB} + (G_R)_{dB} + (L_s)_{dB} + (L_a)_{dB} \]
Example 5.5-2

Suppose that we have a satellite in geosynchronous orbit that radiates 100W of power, i.e., 20dB above 1W (20dBW).

Also, suppose that the each station employs a 3-m parabolic antenna and that the downlink is operating at a frequency of 4GHz.

The efficiency factor is \( \eta = 0.5 \).

By substituting these numbers into \( G_R = \eta \left( \frac{\pi D}{\lambda} \right)^2 \), we obtain the value of the antenna gain as 39 dB.

The free-space path loss is \( L_s = 195.6 \) dB.
Example 5.5-2 (cont.)

- No other losses are assumed.
- Therefore, the received signal power is

\[
(P_R)_{dB} = 20 + 17 + 39 - 195.6
\]

\[
= -119.6 \text{ dBW}
\]

Or, equivalently,

\[
P_R = 1.1 \times 10^{-12} \text{ W}
\]
5.5.2 Link Budget Analysis in Radio Communication System

To complete the link budget computation, we must consider the effect of the additive noise at the receiver front end. Thermal noise that arises at the receiver front end has a relatively flat power density spectrum up to about $10^{12}$ Hz, and is given as

$$N_0 = k_B T_0 \quad \text{W/Hz}$$

where $k_B$ is Boltzmann’s constant ($1.38 \times 10^{-23}$ W-s/K) and $T_0$ is the noise temperature in Kelvin.

The total noise power in the signal bandwidth $W$ is $N_0 W$.

The performance of the digital communication system is specified by the $\varepsilon_b/N_0$ required to keep the error rate performance below some given value.
5.5.2 Link Budget Analysis in Radio Communication System

Since \[ \frac{\varepsilon_b}{N_0} = \frac{T_b P_R}{N_0} = \frac{1}{R} \frac{P_R}{N_0} \]

it follows that \[ \frac{P_R}{N_0} = R \left( \frac{\varepsilon_b}{N_0} \right)_{\text{req}} \]

where \( (\varepsilon_b/N_0)_{\text{req}} \) is the required SNR per bit.

If we have \( P_R/N_0 \) and the required SNR per bit, we can determine the maximum data rate that is possible.
5.5.2 Link Budget Analysis in Radio Communication System

Example 5.5-3

For the link consider in Example 5.5-2, the received signal power is

\[ P_R = 1.1 \times 10^{-12} \text{ W} \quad (-119.6 \text{ dBW}) \]

Now, suppose the receiver front end has a noise temperature of 300 K, which is typical for a receiver in the 4-GHz range. Then, \( N_0 = 4.1 \times 10^{-21} \text{ W/Hz}, \) or, equivalently, \(-203.9 \text{ dBW/Hz}\).

Therefore,

\[ \frac{P_R}{N_0} = -119.6 + 203.9 = 84.3 \text{ dB Hz} \]

\[ N_0 = 4.1 \times 10^{-21} \text{ W/Hz} \]
Exapmle 5.5-3 (cont.)

If the required SNR per bit is 10 dB, then, form Equation 5.5-16, we have the available rate as

\[ R_{dB} = 84.3 - 10 \]
\[ = 74.3 \text{ dB} \] (with respect to 1 bit/s)

This corresponds to the rate of 26.9 megabits/s, which is equivalent to about 420 PCM channels, each operating at 64000 bits/s.
5.5.2 Link Budget Analysis in Radio Communication System

It is a good idea to introduce some safety margin, which is we shall the link margin $M_{dB}$, in the above computations for the capacity of the communication links. Typically, this may be selected as $M_{dB}=6$ dB.

Then, the link budget computation for the link capacity may be expressed in sample form:

$$
R_{dB} = \left( \frac{P_R}{N_0} \right)_{dB \text{ Hz}} - \left( \frac{\varepsilon_b}{N_0} \right)_{req} - M_{dB}
$$

$$
= (P_T)_{dBW} + (G_T)_{dB} + (G_R)_{dB}
$$

$$
+ (L_a)_{dB} + (L_s)_{dB} - \left( \frac{\varepsilon_b}{N_0} \right)_{req} - M_{dB}
$$