

Section 3: Signaling over AWGN Channels

Contents

Section 3: Signaling over AWGN Channels	1
3.1 Basic Concept of Digital Modulation	4
3.2 Geometric Representation of Signals	6
3.2.1 Properties of the Signal Vectors.....	13
3.2.2 Gram-Schmidt (GS) Orthogonalization Procedure.....	14
3.3 Relevant and Irrelevant Noise	24
3.4 Optimum (Coherent) Receiver in AWGN	30
3.4.1 Illustration using a One-Dimension Example.....	31
3.4.2 Generalization to M-ary Modulations.....	36
3.4.3 From Geometric Receiver to Correlation Receiver and Matched Filter Receiver.....	39
3.5 Decision Error Probability	45
3.5.1 Symbol Error Probability (SEP)	45

3.5.2 Bit Error Probability (BEP).....	61
3.6 M-ary Phase-Shift Keying (PSK) Modulation with Coherent Detection ...	63
3.7 Quadrature Amplitude Modulation (QAM)	71
3.8 Frequency Shift Keying (FSK) with Coherent Detection	78
3.9 Noncoherent Detection of FSK and PSK.....	89
3.9.1 Binary FSK with non-coherent detection	91
3.9.2 MPSK with Differential Detection	97

- The objectives of this section are:
 1. Introduce the concept of digital modulation and learn how to represent modulation waveforms and received waveform as vectors in a signal space.
 2. From the signal-space concept developed above, demonstrate that the optimal demodulator in an additive white Gaussian noise (AWGN) channel is one that finds the signal vector closet to the received vector.
 3. Present a number of commonly used digital modulation schemes (PSK, FSK, QAM) and derive their bit-error probabilities (BEPs) in AWGN channels with coherent detection.
 4. Introduce the concept of non-coherent detection; present the BEPs of non-coherent FSK and differential-coherent PSK.
- The so-called **one-shot transmission model** is adopted throughout this section. The next section will address successive pulse transmissions.

3.1 Basic Concept of Digital Modulation

- A data bit can assume two possible logic levels, a “0” or a “1”.
- Digital modulation is the process of mapping a data bit, or a group of data bits, into a waveform for transmission.
- A M -ary modulation scheme maps each group of $m = \log_2 M$ bits into a waveform from the set $\{s_1(t), s_2(t), \dots, s_M(t)\}$ for transmission.
 - Binary modulation implies $m = 1$ and $M = 2$, with two modulation waveforms $s_1(t)$ and $s_2(t)$ representing logic levels “0” and “1” respectively.
 - The duration of the waveforms is denoted as T_b for binary modulations (T for $M > 2$ modulations).

○ The energies of the waveforms are $E_i = \int_0^{T_b} s_i^2(t) dt$, $i = 1, 2$

○ The correlation of the two waveforms is $C_{12} = \int_0^{T_b} s_1(t)s_2(t) dt$.

Intuitively, we want the two waveforms to be very dissimilar, i.e. a low, or even negative correlation.

○ The two waveforms are sent with probabilities of π_1 and π_2 . Most of the times, we assume equi-probable waveforms.

- **Transmission model:** Throughout this section, we adopt a one-shot (or isolated pulse) transmission model with an AWGN channel:

$$r(t) = s(t) + n(t)$$

where $s(t) \in \{s_1(t), s_2(t), \dots, s_M(t)\}$ and $n(t)$ is AWGN with a PSD of $N_0 / 2$.

- After observing $r(t)$, the receiver makes a decision on which of the M waveforms (equivalently the data bits) was transmitted. Naturally, we want to design a receiver that maximizes the chance of making a correct decision.

But how to convert a continuous-time waveform into a decision variable?

3.2 Geometric Representation of Signals

Reference: Section 7.2 of textbook

- It can be shown that each of M modulation waveforms $s_1(t), s_2(t), \dots, s_M(t)$ can be expressed as a linear combination of N orthonormal basis functions $\phi_1(t), \phi_2(t), \dots, \phi_N(t)$ as

$$s_i(t) = \sum_{j=1}^N s_{ij} \phi_j(t) \quad (\text{signal synthesis})$$

where all the basis functions have a pulse duration of T and all have unit energy, i.e.

$$\int_0^T \phi_j^2(t) dt = 1.$$

Furthermore, the correlation of any two different basis functions is zero, i.e.

$$\int_0^T \phi_i(t) \phi_j(t) dt = 0; \quad i \neq j$$

- From the orthonormal properties shown above, we can deduce that

$$s_{ij} = \int_0^T s_i(t) \phi_j(t) dt, \quad (\text{signal analysis})$$

for all $i = 1, 2, \dots, M$ and all $j = 1, 2, \dots, N$.

- The set of correlation coefficients, $s_{i1}, s_{i2}, \dots, s_{iN}$ is viewed as the vector representation of $s_i(t)$ and is given the notation

$$\mathbf{s}_i = \begin{bmatrix} s_{i1} \\ s_{i2} \\ \vdots \\ s_{iN} \end{bmatrix}$$

- The M signal vectors $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_M$ are points in a N -dimensional signal space. Together they form a signal constellation.
- The signal synthesis and analysis are summarized nicely in the following block diagram:

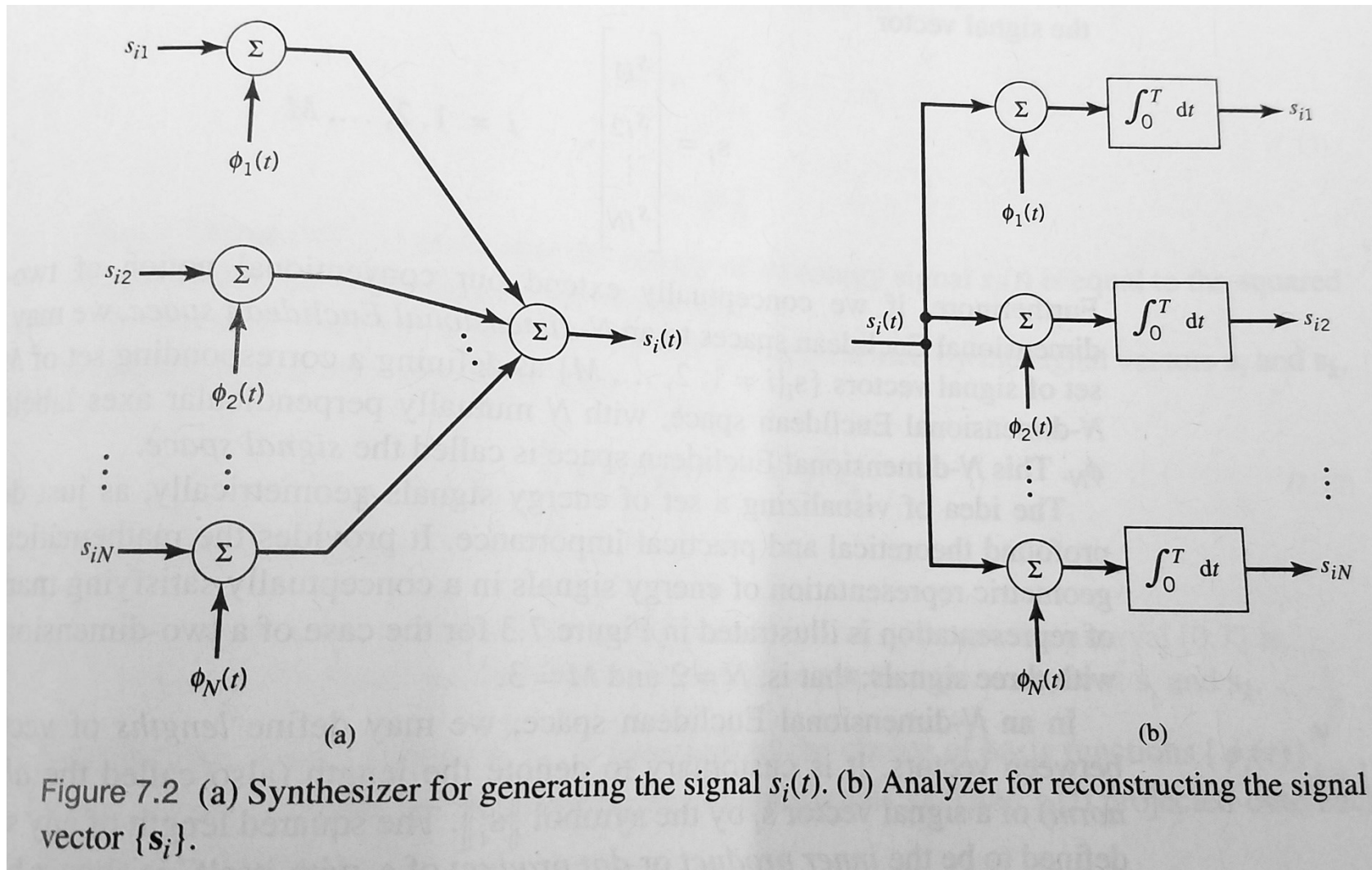


Figure 7.2 (a) Synthesizer for generating the signal $s_i(t)$. (b) Analyzer for reconstructing the signal vector $\{\mathbf{s}_i\}$.

- The signal constellation of a hypothetical $N = 2$, $M = 3$ modulation scheme is shown below

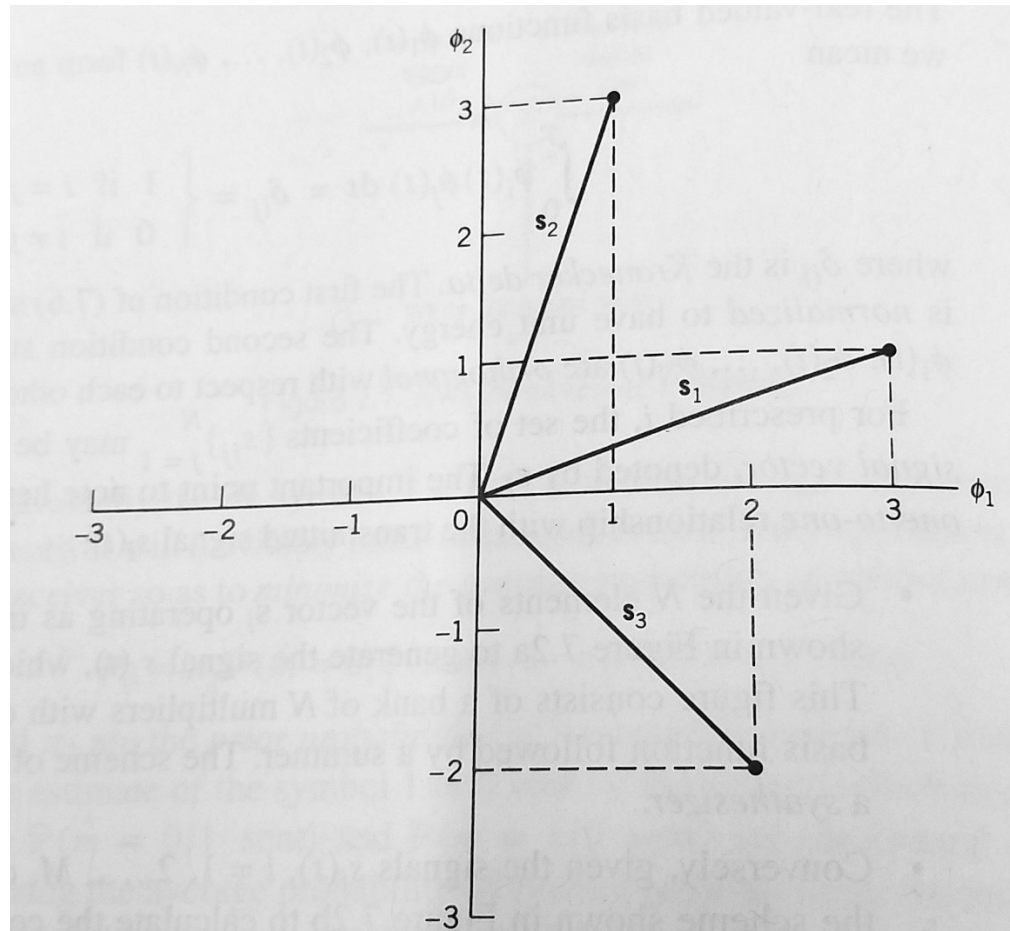


Figure 7.3 Illustrating the geometric representation of signals for the case when $N = 2$ and $M = 3$.

- **Example MPSK:** The modulation waveforms of a M -ary Phase-Shift-Keying (MPSK) scheme are given by

$$s_m(t) = \begin{cases} \sqrt{\frac{2E}{T}} \cos\left(2\pi f_c t + \frac{2\pi}{M} \cdot (m-1)\right), & 0 \leq t < T; \\ 0, & \text{otherwise} \end{cases}, \quad m = 1, 2, \dots, M.$$

Using

$$\phi_1(t) = \begin{cases} \sqrt{2/T} \cos(2\pi f_c t), & 0 \leq t < T; \\ 0, & \text{otherwise} \end{cases}$$

and

$$\phi_2(t) = \begin{cases} -\sqrt{2/T} \sin(2\pi f_c t), & 0 \leq t < T; \\ 0, & \text{otherwise} \end{cases}$$

as basis functions, determine the corresponding signal vectors. Sketch the signal constellation for the case $M = 8$. Note that f_c is the carrier frequency and is much much greater than 1, making $\phi_1(t)$ and $\phi_2(t)$ orthonormal.

- **Example 16QAM:** This scheme encodes 4 bits of information at a time. The 16 modulation waveforms are given by

$$s(t) = a \cdot \sqrt{\frac{2}{T}} \cos(2\pi f_c t) - b \cdot \sqrt{\frac{2}{T}} \sin(2\pi f_c t) ,$$

where $a \in \{-3, -1, +1, +3\}$ and $b \in \{-3, -1, +1, +3\}$ are independent random data symbols in the cosine channel and sine channel respectively. Since there are 16 different combinations of the a and b symbols, so there are 16 different modulation waveforms, conveying 4 bits of information at a time. It is understood that the modulation waveforms are all time-limited to the interval $[0, T]$.

Using the same two orthonormal basis functions as in the last example, sketch the signal constellation of the 16QAM scheme.

3.2.1 Properties of the Signal Vectors

- **Energy:** $E_i = \int_0^T s_i^2(t) dt = \|\mathbf{s}_i\|^2$

Proof:

- **Correlation and inner product:** $\int_0^T s_i(t)s_k(t) dt = \mathbf{s}_i^T \mathbf{s}_k$

Proof:

- The last property is significant in the sense that we can measure similarity of waveforms (implicit in the demodulation process) using vector processing of representative samples (as opposed to processing waveforms).

3.2.2 Gram-Schmidt (GS) Orthogonalization Procedure

- We now show how to derive a set of orthonormal basis functions from the modulation waveforms. It will be shown that:
 1. The number of orthonormal basis functions, N , can not be greater than the number of modulation waveforms, M .
 2. The set of orthonormal basis functions are not unique. Different sets of basis functions can be viewed as different rotations of the signal constellations.

- The GS procedure:

1. Index the modulation waveforms from 1 to M in any order you want.

2. The first orthonormal function is

$$\phi_1(t) = \frac{s_1(t)}{\sqrt{E_1}}$$

where E_1 is the energy of $s_1(t)$. Since $s_1(t) = \sqrt{E_1}\phi_1(t)$, so $s_{11} = \sqrt{E_1}$ and the first signal vector is

$$\mathbf{s}_1 = \left[\sqrt{E_1}, 0, \dots, 0 \right]^T.$$

At this point, we don't know what the dimension of the signal vectors is.

3. Next, let's project the second modulation waveform $s_2(t)$ onto $\phi_1(t)$ to obtain

$$s_{21} = \int_0^T s_2(t) \phi_1(t) dt.$$

This makes $s_{21}\phi_1(t)$ one component of $s_2(t)$. The remaining components is

$$g_2(t) = s_2(t) - s_{21}\phi_1(t).$$

It can be shown that $g_2(t)$ is orthogonal to $\phi_1(t)$ (see below as well). So we can define our second orthonormal basis function to be

$$\phi_2(t) = \frac{g_2(t)}{\sqrt{\int_0^T g_2^2(t) dt}} = \frac{g_2(t)}{\sqrt{E_2 - s_{21}^2}},$$

where $\int_0^T g_2^2(t) dt$ is the energy of $g_2(t)$. Note that

$$\begin{aligned}
\int_0^T \phi_2(t) \phi_1(t) dt &= \int_0^T \frac{g_2(t)}{\sqrt{\int_0^T g_2^2(t) dt}} \phi_1(t) dt \\
&= \int_0^T \frac{s_2(t) - s_{21} \phi_1(t)}{\sqrt{E_2 - s_{21}^2}} \phi_1(t) dt \\
&= \frac{1}{\sqrt{E_2 - s_{21}^2}} \int_0^T \{s_2(t) - s_{21} \phi_1(t)\} \phi_1(t) dt \\
&= \frac{1}{\sqrt{E_2 - s_{21}^2}} \left(\int_0^T s_2(t) \phi_1(t) dt - \int_0^T s_{21} \phi_1(t) \phi_1(t) dt \right) \\
&= 0
\end{aligned}$$

Since $g_2(t) = \sqrt{E_2 - s_{21}^2} \phi_2(t)$ and $g_2(t) = s_2(t) - s_{21} \phi_1(t)$, so we can rewrite the second modulation waveform as

$$s_2(t) = s_{21} \phi_1(t) + \underbrace{g_2(t)}_{\text{orthogonal to } \phi_1(t)}$$

With a corresponding signal vector of

$$\mathbf{s}_2 = \left[s_{21}, \sqrt{E_2 - s_{21}^2}, 0, \dots, 0 \right]^T.$$

4. Repeat the above procedure. At the beginning of the i – th iteration, we have already computed the orthonormal basis functions $\phi_1(t)$ to $\phi_{i-1}(t)$. We then project $s_i(t)$ onto these basis functions to obtain

$$s_{ij} = \int_0^T s_i(t) \phi_j(t) dt, \quad j = 1, 2, \dots, i-1.$$

The signals $s_{i1}\phi_1(t)$, $s_{i2}\phi_2(t)$, ..., $s_{i,i-1}\phi_{i-1}(t)$ are the first $i-1$ components of $s_i(t)$. The remaining component is

$$g_i(t) = s_i(t) - \sum_{j=1}^{i-1} s_{ij} \phi_j(t).$$

and the i -th basis function is

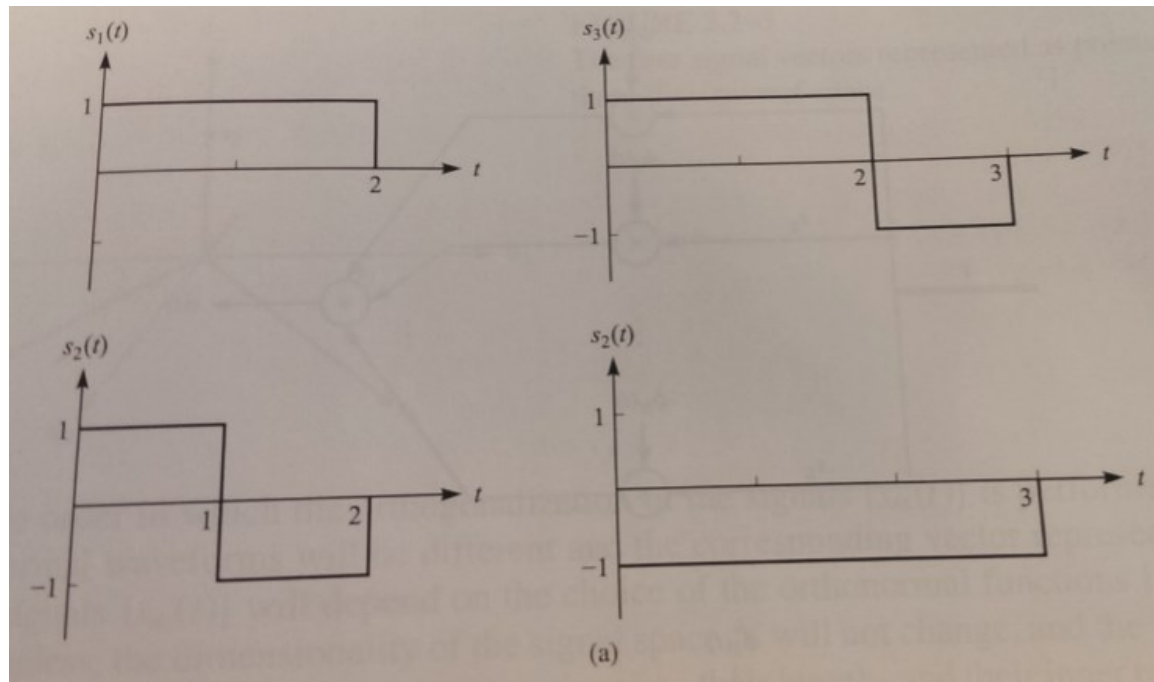
$$\phi_i(t) = \frac{g_i(t)}{\sqrt{\int_0^T g_i^2(t) dt}} = \frac{s_i(t) - \sum_{j=1}^{i-1} s_{ij} \phi_j(t)}{\sqrt{E_i - \sum_{j=1}^{i-1} s_{ij}^2}}$$

The i -th signal vector is

$$\mathbf{s}_i = \left[s_{i1}, s_{i2}, \dots, s_{i,i-1}, \sqrt{E_i - \sum_{j=1}^{i-1} s_{ij}^2}, 0, \dots, 0 \right]^T.$$

- **Question:** does each iteration of the GS procedure necessarily add one new basis function?

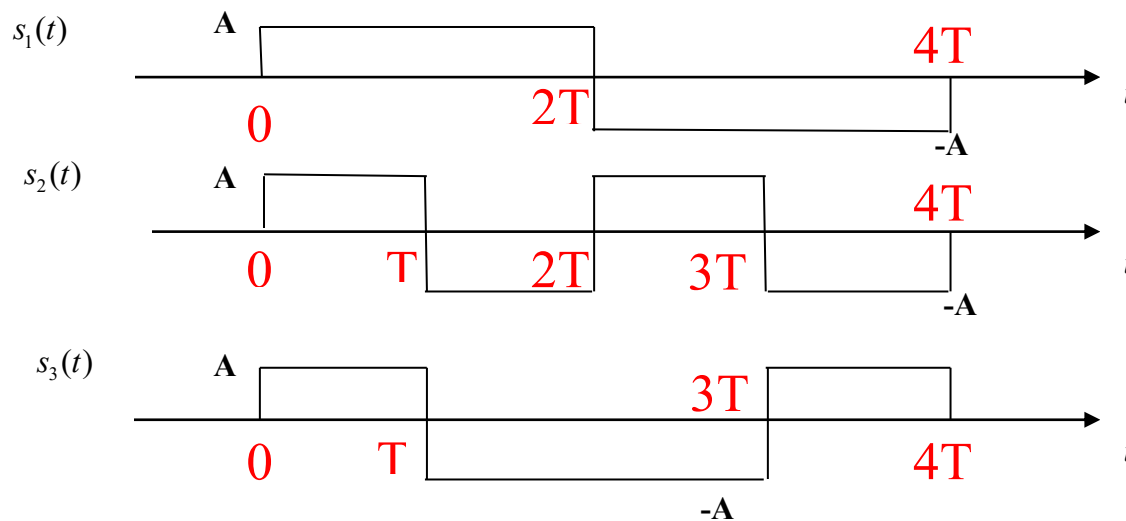
- **Example:** Carry out the GS procedure in the order of $s_1(t), s_2(t), s_3(t), s_4(t)$.



This page is deliberately left blank

- **Example:** A ternary ($M = 3$) modulation scheme employs the following waveforms for transmission:

1. What are the energies of the three waveforms?
2. Use the Gram-Schmidt procedure to determine the orthonormal basis functions of this modulation scheme.
3. Sketch the signal constellation.



This page is deliberately left blank

3.3 Relevant and Irrelevant Noise

Reference: Section 7.3 Textbook

- Fig. 7.2b shows how to compute the signal vector associated with a modulation waveform. The structure there represents a correlator bank; something we briefly discussed at end of Section 2.
- Now suppose the input to the correlator bank is used as part of a receiver. Its input is the received signal

$$x(t) = s_i(t) + w(t)$$

where $s_i(t)$ is the transmitted modulation waveform (corresponding to the message/symbol m_i) and $w(t)$ is AWGN with a PSD of $N_0 / 2$.

- The AWGN has infinite bandwidth, equivalently, infinite number of dimensions. The modulation waveforms, however, have only N dimensions. For this reason, we rewrite the AWGN as:

$$w(t) = \underbrace{\sum_{i=1}^N w_i \phi_i(t)}_{w_e(t), \text{ RELEVANT (IN-BAND NOISE)}} + \underbrace{w(t) - \sum_{i=1}^N w_i \phi_i(t)}_{w'(t), \text{ IRRELEVANT (OUT-OF-BAND NOISE)}},$$

where

$$w_j = \int_0^T w(t) \phi_j(t) dt$$

is the projection of the AWGN onto the j -th basis function.

The first component, $w_e(t)$, lies within the signal space of the modulation waveforms, while the second component, $w'(t)$, lies in an orthogonal space.

- **Question:** what are the statistical properties of the w_j s (see example in Section 2)?

Answer: the w_j s are independent and identically distributed (iid) Gaussian RVs with zero mean and variance $\sigma_w^2 = N_0 / 2$.

Proof:

- **The Theorem of Irrelevance (in simple terms):** Noise outside the signal space can be ignored in formulating the optimal receiver. Intuitively, it does not make sense to design a receiver that admits more noise than necessarily.
- From the Theorem of Irrelevance, $x(t)$ is equivalent to

$$\begin{aligned}
 x_e(t) &= s_i(t) + w_e(t) \\
 &= \sum_{j=1}^N s_{ij} \phi_j(t) + \sum_{j=1}^N w_j \phi_j(t) \\
 &= \sum_{j=1}^N (s_{ij} + w_j) \phi_j(t) \\
 &= \sum_{j=1}^N X_j \phi_j(t)
 \end{aligned}$$

where

$$X_j = s_{ij} + w_j$$

is a Gaussian random variable with mean s_{ij} and variance $\sigma_w^2 = N_0 / 2$. The pdf of X_j is thus

$$f_{X_j}(x_j | m_i) = \frac{1}{\sqrt{2\pi(N_0 / 2)}} \exp \left\{ -\frac{(x_j - s_{ij})^2}{2(N_0 / 2)} \right\}$$

where the notation $f_{X_j}(x_j | m_i)$ is used to emphasize this pdf is conditioned on that the i -th message/symbol m_i was sent.

- Now, whether we are passing $r(t)$ or its equivalence $r_e(t)$ to the correlator bank in Fig. 7.2b, we obtain the $X_1, X_2, \dots, X_N$ because
- Following the notation in the textbook, we collect the correlator outputs into the vector

$$\mathbf{X} = [X_1 \quad X_2 \quad \dots$$

We call $\mathbf{X}$ a Gaussian random vector. The pdf of $\mathbf{X}$ is the joint pdf of $X_1, X_2, \dots, X_N$. Since the X_j s are independent, so their joint pdf is

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}|m_i) &= \prod_{j=1}^N f_{X_j}(x_j|m_i) \\ &= \prod_{j=1}^N \frac{1}{\sqrt{N_0\pi}} \exp\left\{-\frac{(x_j - s_{ij})^2}{N_0}\right\} \\ &= (N_0\pi)^{-N/2} \prod_{j=1}^N \exp\left\{-\frac{(x_j - s_{ij})^2}{N_0}\right\} \\ &= (N_0\pi)^{-N/2} \exp\left\{-\frac{1}{N_0} \sum_{j=1}^N (x_j - s_{ij})^2\right\} \end{aligned}$$

We will next derive the optimal demodulator based on this statistical description of the correlators' outputs.

3.4 Optimum (Coherent) Receiver in AWGN

Reference: Section 7.4 of Text

- We now learn that given the message sent was m_i , the outputs of the correlators $\mathbf{X} = (x_1, x_2, \dots, x_N)^T$ has a pdf of

$$f_{\mathbf{X}}(\mathbf{x}|m_i) = (N_0\pi)^{-N/2} \exp\left\{-\frac{1}{N_0} \sum_{j=1}^N (x_j - s_{ij})^2\right\}$$

- However, when we observe $\mathbf{X} = (x_1, x_2, \dots, x_N)^T$ at the outputs of the correlators, it can come from the transmission of any of the M messages. So how should we decide which was the one actually sent?
Use a one-dimension modulation to explain first.

3.4.1 Illustration using a One-Dimension Example

- Suppose we have a one-dimensional ($N = 1$) binary modulation scheme with signal levels

$$s_1 = +5 \text{ and } s_2 = -5.$$

When we observe $X_1 = 6$, and

1. if we assume it is due to the transmission of s_1 , then implicitly we are assuming that noise is $W_1 = 1$,
2. if we assume it is due to the transmission of s_2 , then implicitly we are assuming that noise is $W_1 = 11$.

Which of these two implicit assumptions are more likely? Answer can be found by looking at the pdf of noise!

Based on the above observation, we decide that s_1 was sent when the correlator output is 6.

Draw pdf here

- The general decision rule seems to be that, always pick the signal level whose implied noise level is the lowest.
- The above decision rule will work well most of the time, except that we ignore one possibility – the two signal levels, in general, are not sent with equal probabilities. So how should the decision strategy be modified when s_1 and s_2 are sent with (priori) probabilities π_1 and π_2 ?

Answer: pick the signal level that has a larger **posterior probability** after observing the correlator output, i.e.

$$\max_{i=1,2} \Pr[s_i | X_1 = x] .$$

This will maximize your chance of making a correct decision.

- Treating pdf as ‘probability’ and applying Bayes’ Rule, we have

$$\Pr[s_i | X_1 = x] f_{X_1}(x) = \Pr[s_i] f_{X_1}(x | s_i)$$

where

$$f_{X_1}(x | s_i) = \frac{1}{\sqrt{2\pi(N_0/2)}} \exp\left\{-\frac{(x-s_i)^2}{2(N_0/2)}\right\}$$

is the pdf of noise conditioned on s_i . Consequently, the posteriori probabilities are:

$$\Pr[s_i | X_1 = x] = \frac{\Pr[s_i] f_{X_1}(x | s_i)}{f_{X_1}(x)} = \frac{\pi_i f_{X_1}(x | s_i)}{f_{X_1}(x)}; \quad i = 1, 2.$$

- Since $f_{X_1}(x)$ at the denominator is common to both $\Pr[s_1 | X_1 = x]$ and $\Pr[s_2 | X_1 = x]$, it can be ignored when we compare the two. As a result, the larger of the two posteriori probabilities is the larger of

$$\pi_1 f_{X_1}(x | s_1) = \pi_1 \frac{1}{\sqrt{\pi N_0}} \exp \left\{ -\frac{(x - s_1)^2}{N_0} \right\}$$

and

$$\pi_2 f_{X_1}(x|s_2) = \pi_2 \frac{1}{\sqrt{\pi N_0}} \exp \left\{ -\frac{(x - s_2)^2}{N_0} \right\}.$$

- **Double Check:** for equally-probable symbols, i.e. when $\pi_1 = \pi_2 = 1/2$, the terms in front of the two exponential functions above are identical and so they can be ignored. As a result, the decision rule is equivalent to selecting the symbol/message with the larger exponential function.

Furthermore, since the arguments inside the two exponential functions are both negative, selecting the larger exponential function is equivalent to selecting the smaller of

$$(x - s_1)^2 \quad \text{and} \quad (x - s_2)^2$$

i.e the message whose signal level is closer to the correlator output is chosen as the transmitted symbol when the symbols are equiprobable.

3.4.2 Generalization to M -ary Modulations

- In general, the decision rule that maximizes the probability of making a correct decision is to select the message that maximizes

$$\pi_i f_{\mathbf{X}}(\mathbf{x}|m_i) = \pi_i \cdot (N_0 \pi)^{-N/2} \exp \left\{ -\frac{1}{N_0} \sum_{j=1}^N (x_j - s_{ij})^2 \right\}.$$

In the event of equiprobable messages, this is equivalent to select the message whose signal vector $\mathbf{s}_i = (s_{i1}, s_{i2}, \dots, s_{iN})^T$ is closest to the received vector $\mathbf{x} = (x_1, x_2, \dots, x_N)^T$, i.e. minimizing

$$d_i^2 = \sum_{j=1}^N (x_j - s_{ij})^2 = \|\mathbf{x} - \mathbf{s}_i\|^2.$$

This is a nice geometric description of the optimal demodulator.

- Even when the messages are not equiprobable, we can still use the above geometric demodulator. It just won't be optimal.

This is also known as Maximum-likelihood (ML) Demodulation.

- For equiprobable messages, the decision boundaries are perpendicular bisectors between pairs of signal points; see example on the left below.
- For non-equiprobable messages, the decisions boundary between a pair of signal points is the perpendicular bisectors shifted towards the signal point with a higher probability of being transmitted; see diagram on the right.

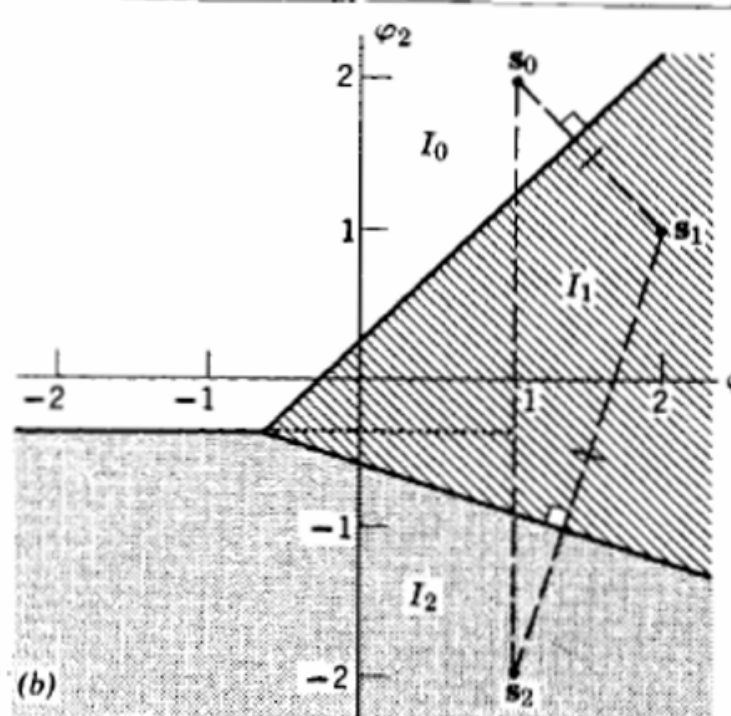
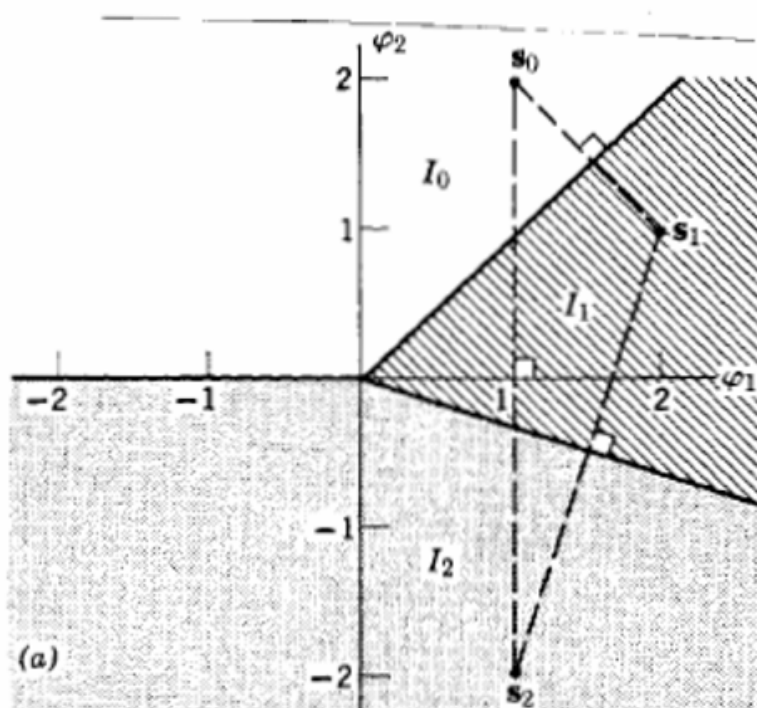


Figure 4.5 Optimum decision regions for additive Gaussian noise: (a) the boundaries of the $\{I_i\}$ are the perpendicular bisectors of the sides of the signal triangle whenever $P[m_0] = P[m_1] = P[m_2]$; (b) the boundaries of the $\{I_i\}$ are displaced when $P[m_1] > P[m_0] > P[m_2]$.

- **Example:** A 4-ary modulation scheme has signal vectors $\mathbf{s}_1 = [1, 0]$, $\mathbf{s}_2 = [0, 1]$, $\mathbf{s}_3 = [-1, 0]$, and $\mathbf{s}_4 = [0, -1]$. Sketch the signal constellations. Provide the decision boundaries for equiprobable messages.

Draw signal constellation and decision boundaries here

3.4.3 From Geometric Receiver to Correlation Receiver and Matched Filter Receiver

- We focus on the ML (i.e. geometric) detector and the Euclidean distance decision metric from hereon:

$$d_i^2 = \sum_{j=1}^N (x_j - s_{ij})^2 = \|\mathbf{x} - \mathbf{s}_i\|^2, \quad i = 1, 2, \dots, M$$

The square distance can be rewritten as

$$\begin{aligned} d_i^2 &= \sum_{j=1}^N (x_j - s_{ij})^2 \\ &= \sum_{j=1}^N (x_j^2 + s_{ij}^2 - 2x_j s_{ij}) \\ &= \sum_{j=1}^N x_j^2 + \sum_{j=1}^N s_{ij}^2 - 2 \sum_{j=1}^N x_j s_{ij} \\ &= \|\mathbf{x}\|^2 + \|\mathbf{s}_i\|^2 - 2\mathbf{x}^T \mathbf{s}_i \end{aligned}$$

- But $\|\mathbf{x}\|^2$ is common to all the square distances, so it can be ignored.
Furthermore,

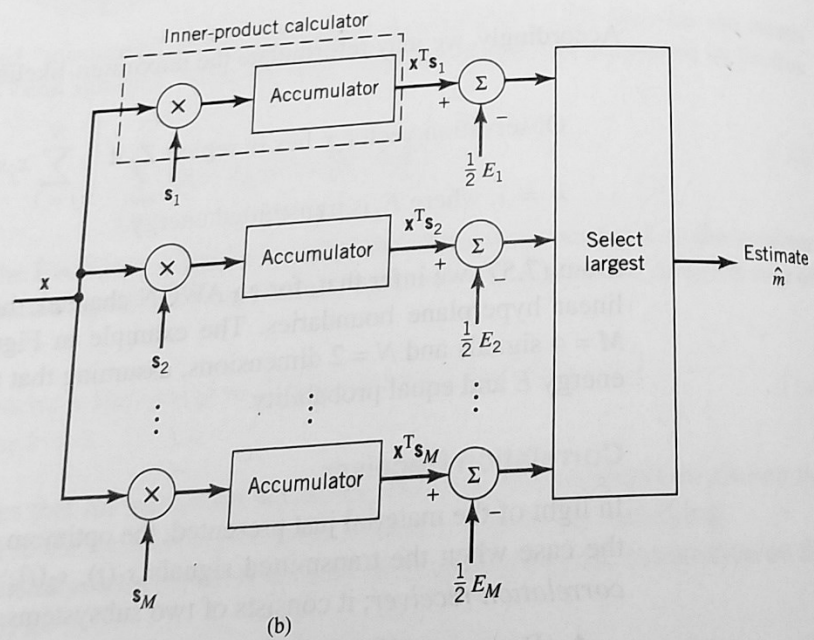
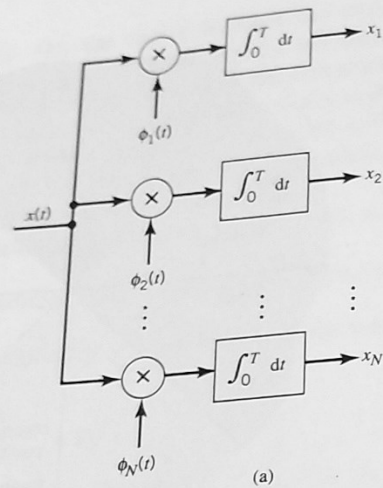
$$\begin{aligned}
E_i &= \int_0^T s_i^2(t) dt \\
&= \int_0^T \left(\sum_{j=1}^N s_{ij} \phi_j(t) \right)^2 dt \\
&= \int_0^T \left(\sum_{j=1}^N s_{ij} \phi_j(t) \right) \left(\sum_{k=1}^N s_{ik} \phi_k(t) \right) dt \\
&= \sum_{j=1}^N \sum_{k=1}^N \int_0^T s_{ij} s_{ik} \phi_j(t) \phi_k(t) dt \\
&= \sum_{j=1}^N s_{ij}^2 = \|\mathbf{s}_i\|^2
\end{aligned}$$

So minimizing d_i^2 is equivalent to maximizing

$$\mathbf{x}^T \mathbf{s}_i - \frac{1}{2} E_i$$

This leads to the receiver in Fig. 7.8

Figure 7.8
(a) Detector or demodulator. (b) Signal
transmission decoder.



- Derive $\mathbf{x}^T \mathbf{s}_i \equiv \int_0^T x(t)s_i(t)dt$ here and sketch the alternative correlation receiver structure

- Show here that $\int_0^T x(t)s_i(t)dt = x(t) * h_i(t) \big|_{t=T}$, where $h_i(t) = s_i(T-t)$ is a filter matched to the i -th modulation waveform, and $*$ denotes convolution. Sketch matched filter receiver here

3.5 Decision Error Probability

Reference: Section 7.5 of Text.

3.5.1 Symbol Error Probability (SEP)

- From the geometric view of the demodulation process, we learn how to carve up the signal space into decision regions:

Notation: let Z_i be the decision region of the message m_i , $i = 1, 2, \dots, M$, and from hereon, we assume that all m_i s are sent with equal probability of $\pi_i = 1 / M$.

- When m_i was sent, but the correlators' outputs $\mathbf{x} = (x_1, x_2, \dots, x_N)^T$ lies outside Z_i , the demodulator will make a wrong decision. So in general, the error probability is

$$\begin{aligned}
 P_e &= \Pr[\hat{m} \neq m] = \sum_{i=1}^M \pi_i \Pr[\mathbf{x} \notin Z_i | m = m_i] \\
 &= \frac{1}{M} \sum_{i=1}^M \Pr[\mathbf{x} \notin Z_i | m = m_i] = 1 - \frac{1}{M} \sum_{i=1}^M \Pr[\mathbf{x} \in Z_i | m = m_i]
 \end{aligned}$$

Conceptually, the error probability expression is easy to explain. But its exact value could be difficult to evaluate because of the shapes of the decision regions. This is especially true for $M > 2$.

So use an approximation instead.

- When $\mathbf{x} = (x_1, x_2, \dots, x_N)^T$ does not lie in the decision region Z_i when m_i was sent, it must be closer to at least one other signal point than to $\mathbf{s}_i$. Define the event A_{ij} as the event that $\mathbf{x}$ is closer to $\mathbf{s}_j$ than to $\mathbf{s}_i$ when the message is m_i , i.e.

$A_{ij} : \|\mathbf{x} - \mathbf{s}_i\|^2 > \|\mathbf{x} - \mathbf{s}_j\|^2$ when $\mathbf{x} = \mathbf{s}_i + \mathbf{W}$; $\mathbf{s}_j \neq \mathbf{s}_i$, $\mathbf{w}$ being Gaussian noise

Then the probability of making a wrong decision given m_i was sent is

$$P_e(m_i) = \Pr[A_{i1} \cup \dots A_{i,i-1} \cup A_{i,i+1} \cup \dots A_{i,M}]$$

where $\cup$ is the “or” operator. Since $\Pr[A \cup B] \leq \Pr[A] + \Pr[B]$, repeated use of this property on $P_e(m_i)$ yields

$$P_e(m_i) \leq \sum_{\substack{j=1 \\ j \neq i}}^M \Pr[A_{ij}] = \sum_{\substack{j=1 \\ j \neq i}}^M \Pr[\|\mathbf{x} - \mathbf{s}_i\|^2 > \|\mathbf{x} - \mathbf{s}_j\|^2 \mid \mathbf{x} = \mathbf{s}_i + \mathbf{W}]$$

Finally, the decision or symbol error probability (SEP) can be approximated by summing over all the conditional error probabilities above

$$P_e = \frac{1}{M} \sum_{i=1}^M P_e(m_i) \leq \frac{1}{M} \sum_{j=1}^M \sum_{\substack{j=1 \\ j \neq i}}^M \Pr \left[\|\mathbf{x} - \mathbf{s}_i\|^2 > \|\mathbf{x} - \mathbf{s}_j\|^2 \mid \mathbf{x} = \mathbf{s}_i + \mathbf{W} \right]$$

This is called the **Union Bound on the symbol error probability**.

- So how to determine $\Pr \left[\|\mathbf{x} - \mathbf{s}_i\|^2 > \|\mathbf{x} - \mathbf{s}_j\|^2 \mid \mathbf{x} = \mathbf{s}_i + \mathbf{W} \right]$?

Solution: Since

$$\mathbf{x} = \mathbf{s}_i + \mathbf{W},$$

$$\|\mathbf{x} - \mathbf{s}_i\|^2 = \|\mathbf{W}\|^2$$

and

$$\begin{aligned}
\|\mathbf{x} - \mathbf{s}_j\|^2 &= \|(\mathbf{s}_i - \mathbf{s}_j) + \mathbf{W}\|^2 \\
&= \|\mathbf{d}_{ij} + \mathbf{W}\|^2 \\
&= \|\mathbf{d}_{ij}\|^2 + \|\mathbf{W}\|^2 + 2\mathbf{W}^T \mathbf{d}_{ij}
\end{aligned}$$

where $\mathbf{d}_{ij} = \mathbf{s}_i - \mathbf{s}_j$ is the displacement vector. As a result, $\|\mathbf{x} - \mathbf{s}_i\|^2 > \|\mathbf{x} - \mathbf{s}_j\|^2$ is equivalent to

$$\|\mathbf{W}\|^2 \geq \|\mathbf{d}_{ij}\|^2 + \|\mathbf{W}\|^2 + 2\mathbf{W}^T \mathbf{d}_{ij}, \text{ or}$$

$$0 \geq \|\mathbf{d}_{ij}\|^2 + 2\mathbf{W}^T \mathbf{d}_{ij}, \text{ or}$$

$$\mathbf{W}^T \mathbf{d}_{ij} \leq -\frac{1}{2}\|\mathbf{d}_{ij}\|^2, \text{ or}$$

$$W = \mathbf{W}^T \mathbf{d}_{ij} = (W_1, W_2, \dots, W_N) \begin{pmatrix} s_{i1} - s_{j1} \\ s_{i2} - s_{j2} \\ \vdots \\ s_{iN} - s_{jN} \end{pmatrix} \quad \sum_{n=1}^N W_n (s_{in} - s_{jn}) \leq -\frac{1}{2} \|\mathbf{d}_{ij}\|^2$$

Now, since the W_n s are iid Gaussian random variables with zero mean and variance $N_0 / 2$, so

$$W = \sum_{n=1}^N W_n (s_{in} - s_{jn})$$

is zero-mean Gaussian with a variance of

$$\begin{aligned}
\sigma_W^2 &= E[W^2] = E\left[\left(\sum_{n=1}^N W_n (s_{in} - s_{jn})\right)\left(\sum_{k=1}^N W_k (s_{ik} - s_{jk})\right)\right] \\
&= E\left[\sum_{n=1}^N \sum_{k=1}^N W_n W_k (s_{in} - s_{jn})(s_{ik} - s_{jk})\right] \\
&= \sum_{n=1}^N E[W_n^2] (s_{in} - s_{jn})^2 \\
&= \frac{N_0}{2} \sum_{n=1}^N (s_{in} - s_{jn})^2 \\
&= \frac{N_0}{2} \|\mathbf{d}_{ij}\|^2
\end{aligned}$$

As a result,

$$\begin{aligned}
\Pr\left[W \leq -\frac{1}{2}\|\mathbf{d}_{ij}\|^2\right] &= \int_{-\infty}^{-\frac{1}{2}\|\mathbf{d}_{ij}\|^2} \frac{1}{\sqrt{2\pi\sigma_W^2}} \exp\left\{-\frac{W^2}{2\sigma_W^2}\right\} dW \\
&= \int_{-\infty}^{-\frac{1}{2}\|\mathbf{d}_{ij}\|^2} \frac{1}{\sqrt{\pi N_0 \|\mathbf{d}_{ij}\|^2}} \exp\left\{-\frac{W^2}{N_0 \|\mathbf{d}_{ij}\|^2}\right\} dW \\
&= Q\left(\sqrt{\frac{\|\mathbf{d}_{ij}\|^2}{2N_0}}\right)
\end{aligned}$$

where $Q(x)$ is the Q-function defined in Section 2. Finally, the union bound on symbol error probability is

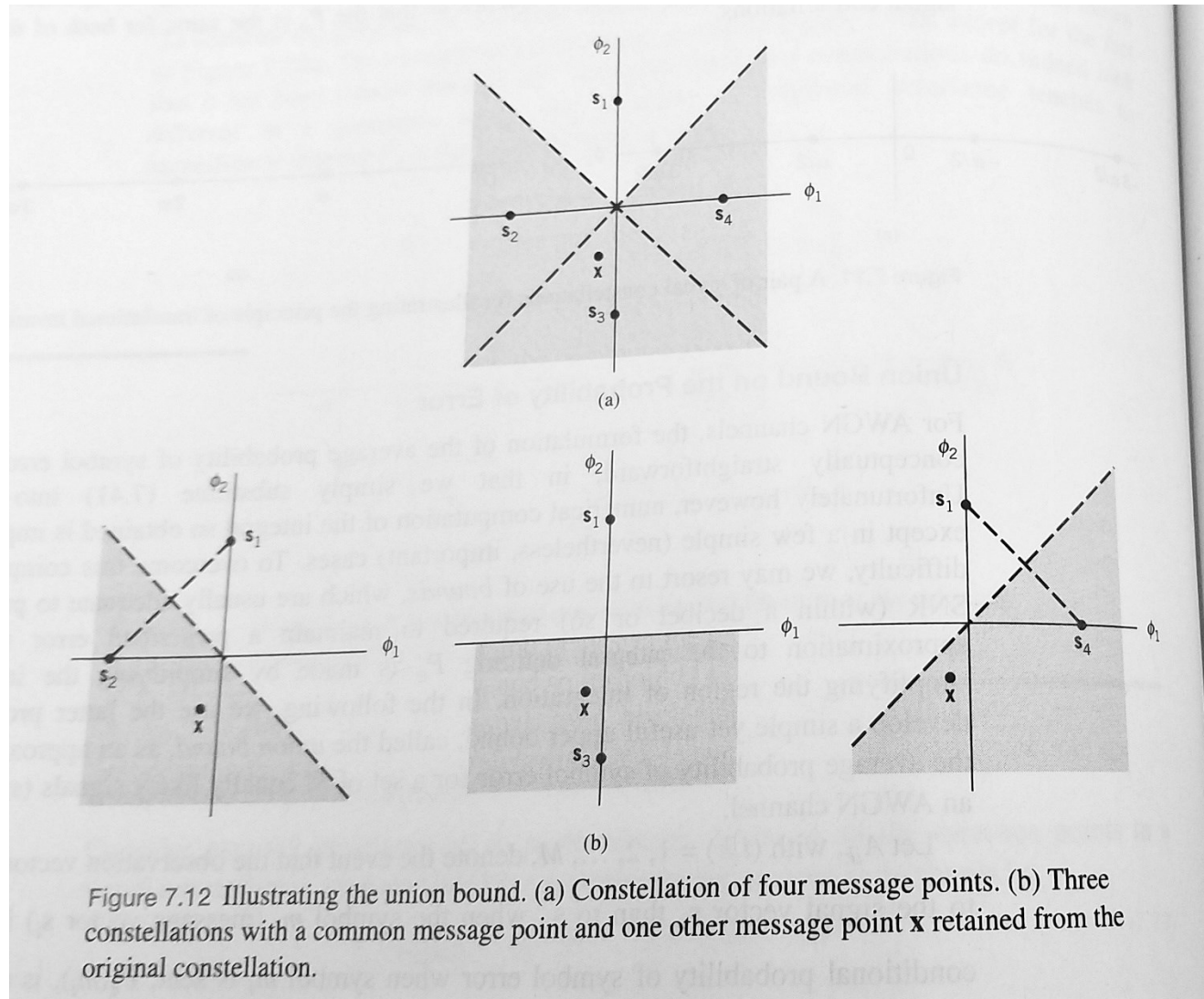
$$P_e \leq \frac{1}{M} \sum_{i=1}^M \sum_{\substack{j=1 \\ j \neq i}}^M \Pr\left[\|\mathbf{x} - \mathbf{s}_i\|^2 > \|\mathbf{x} - \mathbf{s}_j\|^2 \mid \mathbf{x} = \mathbf{s}_i + \mathbf{w}\right] = \frac{1}{M} \sum_{i=1}^M \sum_{\substack{j=1 \\ j \neq i}}^M Q\left(\sqrt{\frac{\|\mathbf{d}_{ij}\|^2}{2N_0}}\right)$$

This is a very nice result in that the error probability of the geometric demodulator depends on the separation between signal points – sort of make sense.

- Because of the properties of the Q-function, the larger the distances, the smaller the symbol error probability.
- **Example:** Consider the 4-ary constellation in Fig. 7.12, where all the signal points lie on a circle with radius $\sqrt{E}$, where

$$E = \|\mathbf{s}_i\|^2 = \int_0^T s_i^2(t) dt; \quad i = 1, 2, 3, 4$$

is the energy of the modulation waveforms. Determine the Union Bound on the SEP.



Solution:

- From the symmetry of the constellation, we can deduce that the SEP is independent of the message/signal point being transmitted, i.e.

$$P_e(m_1) = P_e(m_2) = P_e(m_3) = P_e(m_4) = P_e.$$

So without loss of generality, let's assume m_1 was sent and just focus on $P_e(m_1)$ and express the union bound as

$$P_e \leq \sum_{j=2}^M Q \left(\sqrt{\frac{\|\mathbf{d}_{1j}\|^2}{2N_0}} \right)$$

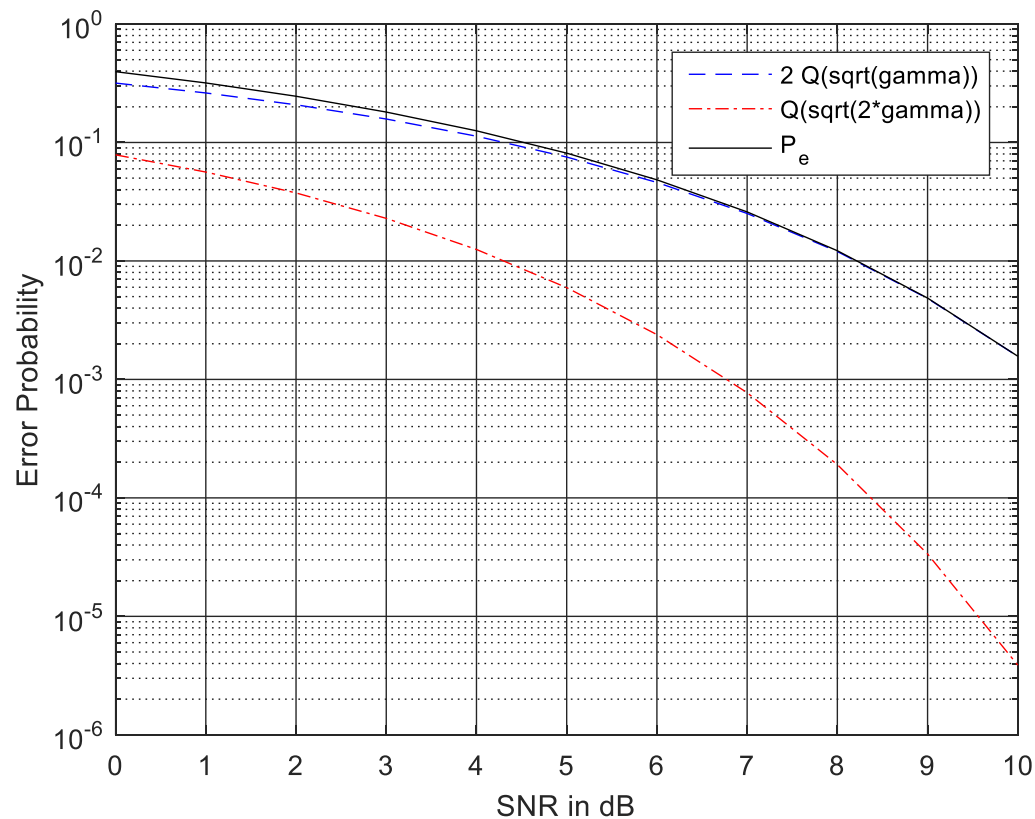
- From the signal constellation, we found $\|\mathbf{d}_{12}\|^2 = \|\mathbf{d}_{14}\|^2 = 2E$ and $\|\mathbf{d}_{13}\|^2 = 4E$. Therefore, the union bound on the symbol-error probability is

$$P_e \leq 2Q \left(\sqrt{\frac{E}{N_0}} \right) + Q \left(\sqrt{2 \frac{E}{N_0}} \right)$$

- Define the symbol signal-to-noise ratio (SNR) as

$$\gamma = E / N_0$$

We show below a plot of $2Q(\sqrt{\gamma})$, $Q(\sqrt{2\gamma})$, and $P_e = 2Q(\sqrt{\gamma}) + Q(\sqrt{2\gamma})$ (all in log scale) versus γ in dB scale.



- What the figure indicates is that the error probability is dominated by the minimum distant term, i.e. by the term $2Q\left(\sqrt{\gamma}\right)$.

This property had been found to be true for more complex modulation scheme. As such we have the following simplified, but important approximation of the error probability of a modulation scheme:

$$P_e \approx C \cdot Q\left(\sqrt{\frac{d_{\min}^2}{2N_0}}\right)$$

where

$$d_{\min}^2 = \min_{\forall i,j} \|\mathbf{d}_{ij}\|^2$$

Is the minimum square distance between pairs of signal points, and C is a constant that depends on the size M of the constellation and the average number of nearest neighbours a signal point has.

- **Example:** The signal points of a one-dimensional modulation scheme are:

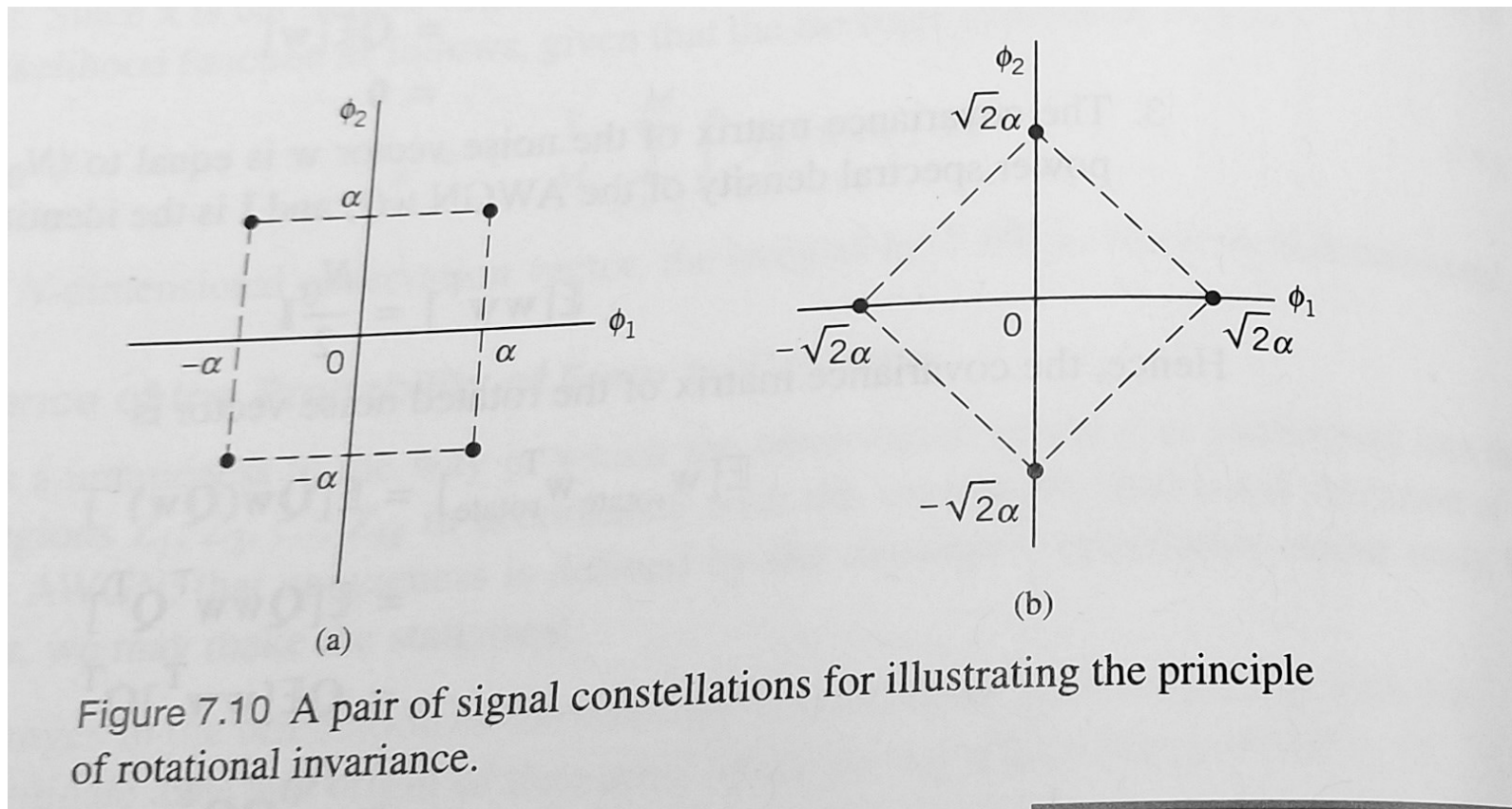
$$s_1 = -3A, s_2 = -A, s_3 = +A, s_4 = +3A.$$

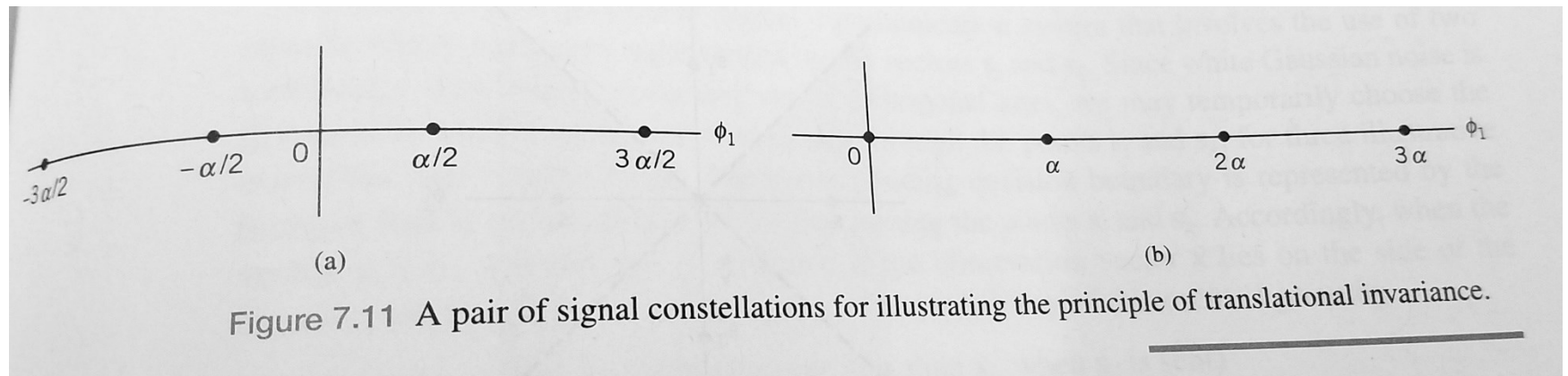
- Sketch the signal constellation.
- Determine the energies of the corresponding modulation waveforms.
- Determine the average energy of the modulation scheme.
- What is the minimum square distance, $d_{\min}^2$, of the signal constellation?
Express it in terms of the average energy.
- Use the minimum square distance to arrive at an estimate of the symbol error probability.

- **Rotation and Translation:** recall the union bound on SEP and the approximation based on minimum distance:

$$P_e \leq \frac{1}{M} \sum_{i=1}^M \sum_{\substack{j=1 \\ j \neq i}}^M Q \left(\sqrt{\frac{\|\mathbf{d}_{ij}\|^2}{2N_0}} \right) \quad \text{or} \quad P_e \approx C \cdot Q \left(\sqrt{\frac{d_{\min}^2}{2N_0}} \right).$$

Both are based on separations between signal points. This means if the entire constellation is rotated by the same angle or translated by the same amount, the separations between pairs of points would not change, so the error performance would not change.





- **Comment about translation:** while the error probability after translation in Fig. 7.11 remains the same when expressed in terms of α , it will be different when expressed in terms of the average energy.
- **Example:** Express the symbol error probability estimate for the constellations in Fig. 7.11 in terms of the average symbol energy.

3.5.2 Bit Error Probability (BEP)

- The symbol error probability (SEP) tells you on average how often the receiver makes a wrong decision about the transmitted symbol.
- When the size of the constellation, M , is greater than 2, then each symbol carries more than 1 bit of information.

Often we are more interested in the bit error probability (BEP), or the chance that the demodulated bit is not the same as the transmitted bit.

- The BEP depends on how we map information bits into signal points in the constellation.
- In our SEP analysis, we learnt that the error probability is denominated by the minimum distance error event. So intuitively, neighbouring points in the signal

constellation should only differ in 1 bit (of course, sometimes it is not possible to do so).

- **Example:** For the signal constellations in Fig. 7.10a, each symbol carries 2 bits of information. Devise a rule for mapping different combinations of binary bit pair onto the 4 points in the signal constellation. Repeat the exercise for the constellation in Fig. 7.11a.

Note: *these mapping rules are termed* **Gray Mapping or Gray Coding**.

- Now, if we assume that each nearest-neighbour error event introduces only one bit of error, then the BEP is approximately

$$P_b \approx \frac{1}{\log_2 M} C \cdot Q\left(\sqrt{\frac{d_{\min}^2}{2N_0}}\right) \approx \frac{1}{\log_2 M} P_e$$

where the 1 in $1/\log_2 M$ is the number of bit errors per wrong decision, and $\log_2 M$ is the number of information bits per symbol.

3.6 M-ary Phase-Shift Keying (PSK) Modulation with Coherent Detection

Reference: Section 7.6 of Textbook.

- M -PSK is a commonly used digital modulation technique (satellite, cellular).
The signal waveforms are

$$s_i(t) = \sqrt{\frac{2E}{T}} \cos \left[2\pi f_c t + \frac{2\pi}{M}(i-1) \right], \quad i = 1, 2, \dots, M$$

where E is the signal energy per symbol/waveform and f_c is the carrier frequency. Each waveform is assumed to have a duration of T , and the carrier

frequency is n_c / T , where n_c is a very large integer (reflecting that typically the carrier is much larger than the transmission rate in a real system)

- Using the trigonometric identity $\cos(A + B) = \cos A \cos B - \sin A \sin B$, the M-PSK waveforms can be rewritten as

$$\begin{aligned}
 s_i(t) &= \sqrt{\frac{2E}{T}} \cos \left[2\pi f_c t + \frac{2\pi}{M}(i-1) \right] \\
 &= \underbrace{\sqrt{E} \cos \left(\frac{2\pi}{M}(i-1) \right)}_{s_{i1}} \cdot \underbrace{\sqrt{2} \cos(2\pi f t)}_{\phi_1(t)} - \underbrace{\sqrt{E} \sin \left(\frac{2\pi}{M}(i-1) \right)}_{s_{i2}} \cdot \underbrace{\sqrt{2} \sin(2\pi f t)}_{-\phi_2(t)} \\
 &= s_{i1} \phi_1(t) + s_{i2} \phi_2(t)
 \end{aligned}$$

where s_{i1} , $\phi_1(t)$, s_{i2} , and $\phi_2(t)$ are as shown in the second line of the equation.

- Now, focus on the correlation of $\phi_1(t)$ and $\phi_2(t)$

$$\begin{aligned}
 \int_0^T \phi_1(t)\phi_2(t)dt &= -\int_0^T \sqrt{\frac{2}{T}} \cos(2\pi f_c t) \cdot \sqrt{\frac{2}{T}} \sin(2\pi f_c t) dt \\
 &= -\frac{1}{T} \int_0^T 2 \cos(2\pi f_c t) \sin(2\pi f_c t) dt \\
 &= -\frac{1}{T} \int_0^T \sin(4\pi f_c t) dt \\
 &= -\frac{1}{4\pi f_c T} [-\cos(4\pi f_c t)]_0^T \\
 &= \frac{1}{4\pi f_c T} [\cos(4\pi f_c T) - 1] \\
 &= \frac{1}{4\pi n_c} [\cos(4\pi n_c) - 1] \quad \text{since we assumed } f_c = n_c / T \\
 &= 0
 \end{aligned}$$

In other word, they are uncorrelated.

- Using same approach, we can show that

$$\int_0^T \phi_1^2(t) dt = \int_0^T \phi_2^2(t) dt = 1.$$

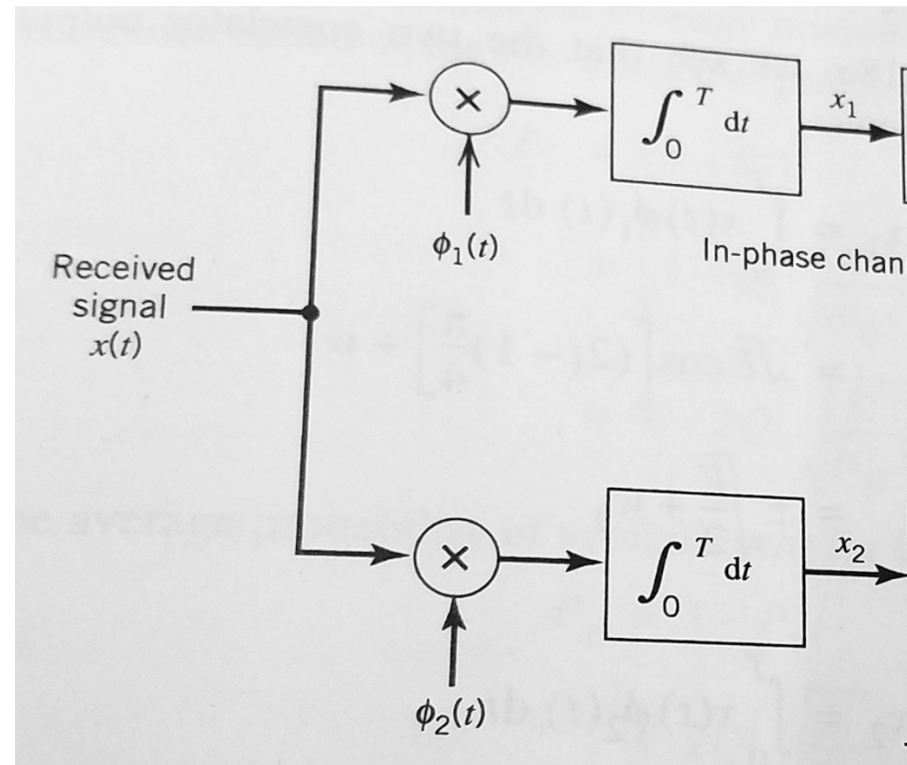
- Summarizing the results in the last three bullet items, we can conclude that

1. $\phi_1(t)$ and $\phi_2(t)$ are orthonormal, and

2. The signal vectors of M-PSK are:

$$\mathbf{s}_i = [s_{i1}, s_{i2}]^T = \left[\sqrt{E} \cos\left(\frac{2\pi}{M}(i-1)\right), \sqrt{E} \sin\left(\frac{2\pi}{M}(i-1)\right) \right]^T, \quad i = 1, 2, \dots, M$$

- The structure below generates the received vector $\mathbf{x} = [x_1, x_2]^T$ for making decision.



The signal constellation and decision boundaries for 8-PSK are shown in Fig 7.21b, where the minimum distance error events are highlighted.

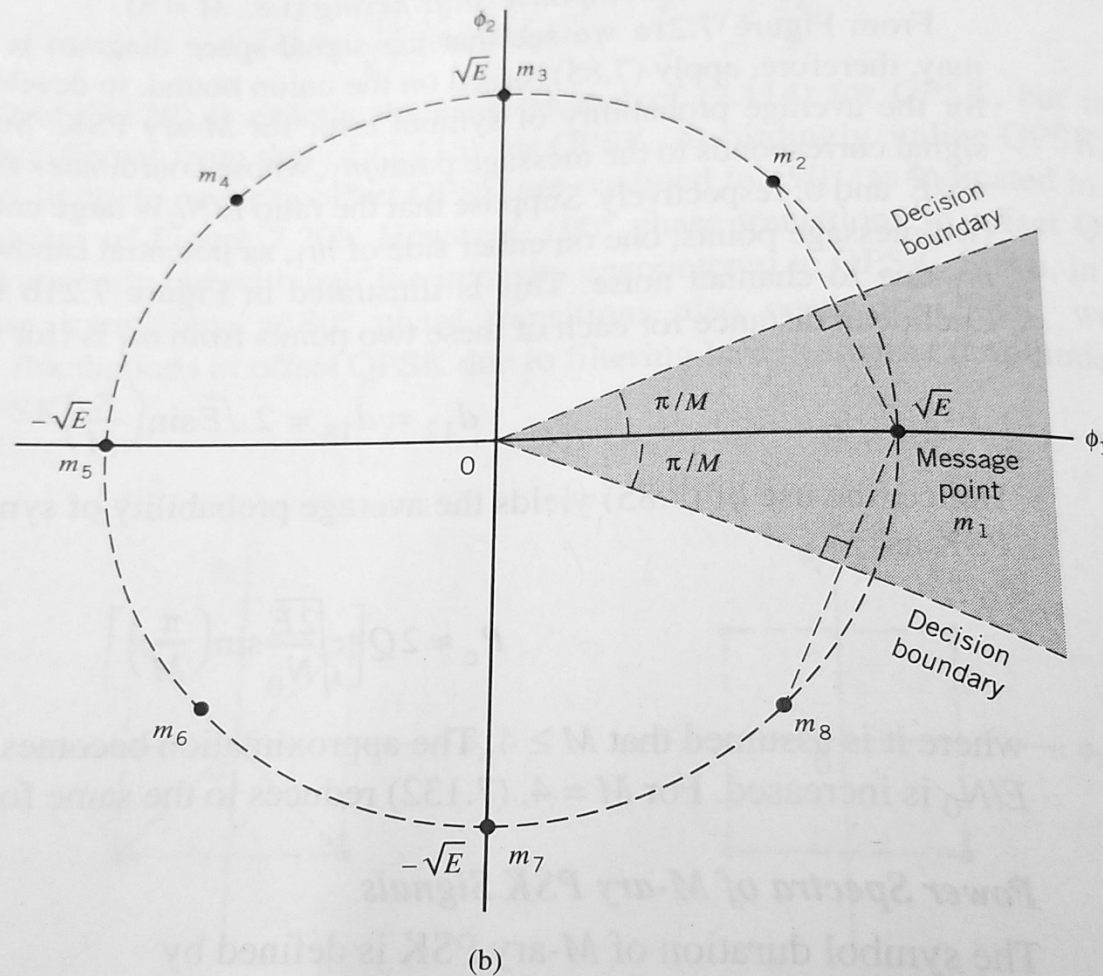


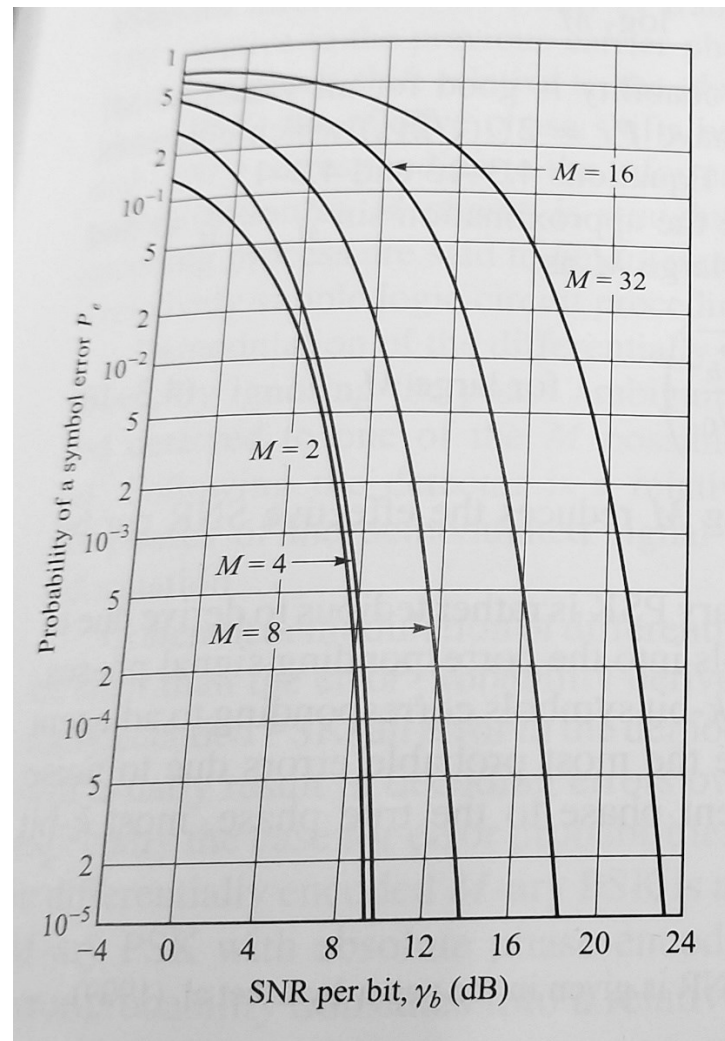
Figure 7.21 (a) Signal-space diagram for octaphase-shift keying (i.e., $M = 8$). The decision boundaries are shown as dashed lines. (b) Signal-space diagram illustrating the application of the union bound for octaphase-shift keying.

- **Example:** For the 8PSK constellation on the previous page,
 1. Derive the Gray coding rule for mapping different binary 3-tuples to points in the signal constellation.
 2. Determine the minimum square distance, $d_{\min}^2$, and express it in terms of E_b the average energy per information bit.
 3. Determine the SEP and BEP.

- **Example:** Sketch the constellations of 2PSK and 4PSK and determine their SEPs and BEPs.

- **SEP comparison of different MPSK schemes.** The bit SNR is defined as

$$\gamma_b = E_b / N_0 = \frac{1}{\log_2 M} (E / N_0)$$



Using a SEP of 10^{-5} as benchmark, 2 and 4 PSK require a SNR of approximately 9.6 and 10 dB respectively, whereas 8, 16, and 32 PSK require 14, 18, and 23 dB.

Huge penalty in SNR when M increases. However, (raw) transmission efficiency increases at a rate of $\log_2 M$. This power-rate trade-off is typical of digital communication.

3.7 Quadrature Amplitude Modulation (QAM)

Reference: Section 7.7 of Text

- MPSK has a circular constellation, with M points uniformly distributed on a circle of radius $\sqrt{E}$ in a 2-dimensional signal space.

- QAM, on the other hand, has a rectangular signal constellation. In this section, we use 16QAM as an illustrative example of the signal structure.
- The 16 waveforms of 16QAM are:

$$\begin{aligned}
 s_k(t) &= \sqrt{\frac{2E_0}{T}} a_k \cos[2\pi f_c t] + \sqrt{\frac{2E_0}{T}} b_k \sin[2\pi f_c t] \\
 &= a_k \underbrace{\sqrt{E_0} \cdot \sqrt{2} \cos[2\pi f t]}_{\phi_1(t)} - b_k \underbrace{\sqrt{E_0} \cdot \sqrt{2} \sin[2\pi f t]}_{-\phi_2(t)} \\
 &= a_k \phi_1(t) + b_k \phi_2(t), \quad k = 1, 2, \dots, 16
 \end{aligned}$$

where each waveform is time-limited to the interval $[0, T]$, and

$$\phi_1(t) = \sqrt{\frac{2}{T}} \cos[2\pi f_c t]$$

$$\phi_2(t) = -\sqrt{\frac{2}{T}} \sin[2\pi f_c t]$$

are the same orthonormal functions as in MPSK (except for opposite polarities in $\phi_2(t)$ between the two schemes).

- The corresponding signal vectors are

$$\mathbf{s}_k = \left[a_k \sqrt{E_0}, b_k \sqrt{E_0} \right]^T, \quad k = 1, 2, \dots, 16$$

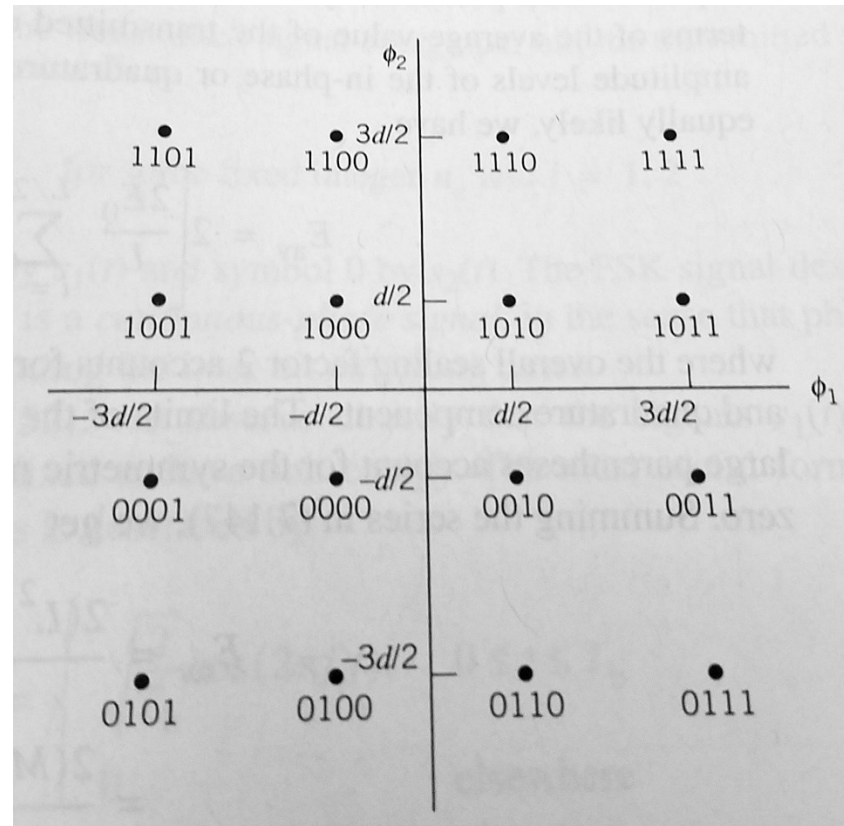
where the set of (a_k, b_k) is

$$\{(a_k, b_k)\} = \begin{bmatrix} (-3, +3) & (-1, +3) & (+1, +3) & (+3, +3) \\ (-3, +1) & (-1, +1) & (+1, +1) & (+3, +1) \\ (-3, -1) & (-1, -1) & (+1, -1) & (+3, -1) \\ (-3, -3) & (-1, -3) & (+1, -3) & (+3, -3) \end{bmatrix}$$

- It can be concluded that the signal points form a rectangular constellation with horizontal and vertical spacing between immediate neighbours being

$$d = 2\sqrt{E_0}$$

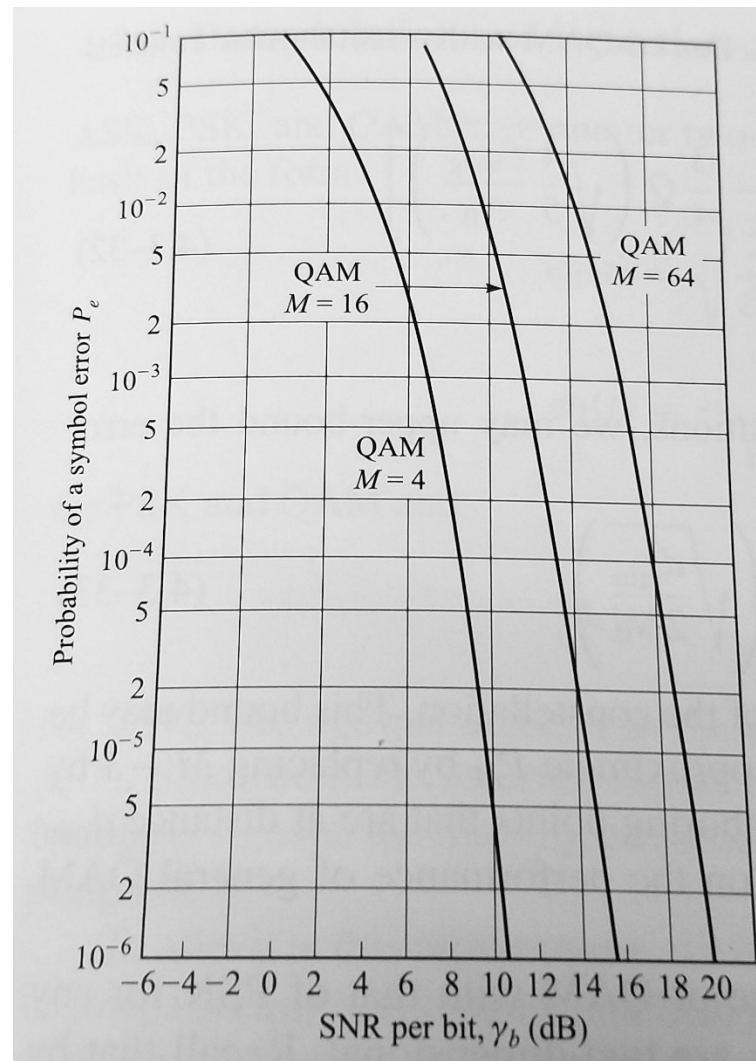
Obviously, d is also the minimum distance between signal points.



This page deliberately left blank

- **SEP comparison of different QAM schemes.** The bit SNR is defined as

$$\gamma_b = E_b / N_0 = \frac{1}{\log_2 M} (E / N_0).$$



3.8 Frequency Shift Keying (FSK) with Coherent Detection

Reference: Section 7.8 of Text

- Both MPSK and MQAM are examples of linear modulation. In this section, we look at MFSK, which is a class of **non-linear modulations**.
- MFSK are also examples of **orthogonal modulation**, i.e. the signal vectors in a MFSK constellation are mutually orthogonal.

Implication of orthogonal signal vectors: can detect which waveform was transmitted by detecting which signal dimension was excited – energy detection is a form of non-coherent detection, i.e. the receiver does not need to know the phase of the incoming FSK waveform.

- Start with $M=2$, or binary FSK. The two modulation waveforms are

$$s_1(t) = \sqrt{\frac{2E_b}{T_b}} \cos\left(2\pi \cdot \frac{n_c + 1}{T_b} \cdot t\right) = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi \cdot f_1 \cdot t),$$

$$s_2(t) = \sqrt{\frac{2E_b}{T_b}} \cos\left(2\pi \cdot \frac{n_c + 2}{T_b} \cdot t\right) = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi \cdot f_2 \cdot t)$$

where both waveforms are time-limited to the interval $[0, T_b]$, E_b is energy of each waveform (also the energy per bit), $f_1 = (n_c + 1) / T_b$ and $f_2 = (n_c + 2) / T_b$ are the frequencies of the two waveforms.

In other word, in FSK, information is encoded into the frequency of the transmitted signal.

- Consider the correlation of the two modulation waveforms:

$$\begin{aligned}
\int_0^{T_b} s_1(t)s_2(t)dt &= \int_0^{T_b} \sqrt{\frac{2E_b}{T_b}} \cos\left(2\pi \cdot \frac{n_c+1}{T_b} \cdot t\right) \sqrt{\frac{2E_b}{T_b}} \cos\left(2\pi \cdot \frac{n_c+2}{T_b} \cdot t\right) dt \\
&= \frac{E_b}{T_b} \int_0^{T_b} 2 \cos\left(2\pi \cdot \frac{n_c+1}{T_b} \cdot t\right) \cos\left(2\pi \cdot \frac{n_c+2}{T_b} \cdot t\right) dt, \\
&= \frac{E_b}{T_b} \int_0^{T_b} \cos\left(2\pi \cdot \frac{1}{T_b} \cdot t\right) + \cos\left(2\pi \cdot \frac{2n_c+3}{T_b} \cdot t\right) dt, \\
&= 0
\end{aligned}$$

that is, the two waveforms are orthogonal.

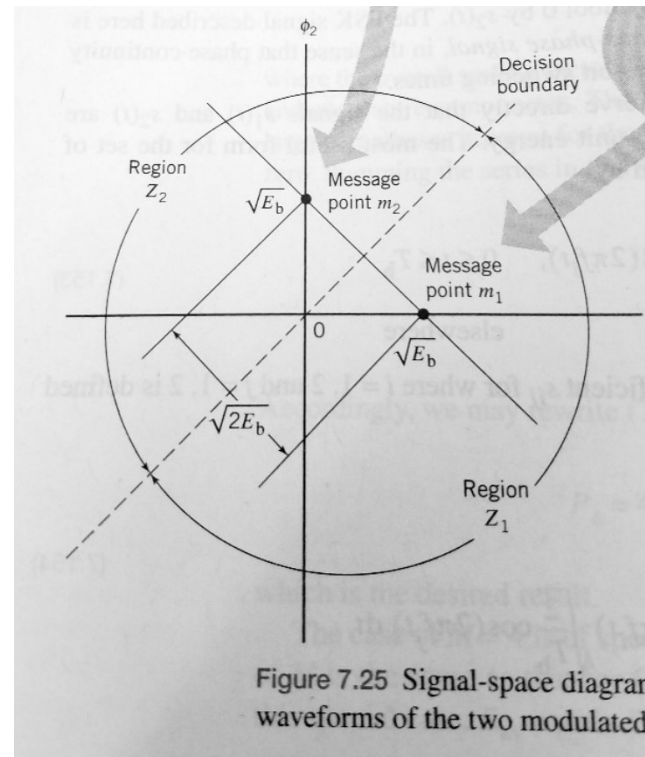
- Since the two waveforms are orthogonal, and since both have the same energy of E_b , we can deduce that

1. The basis functions are:

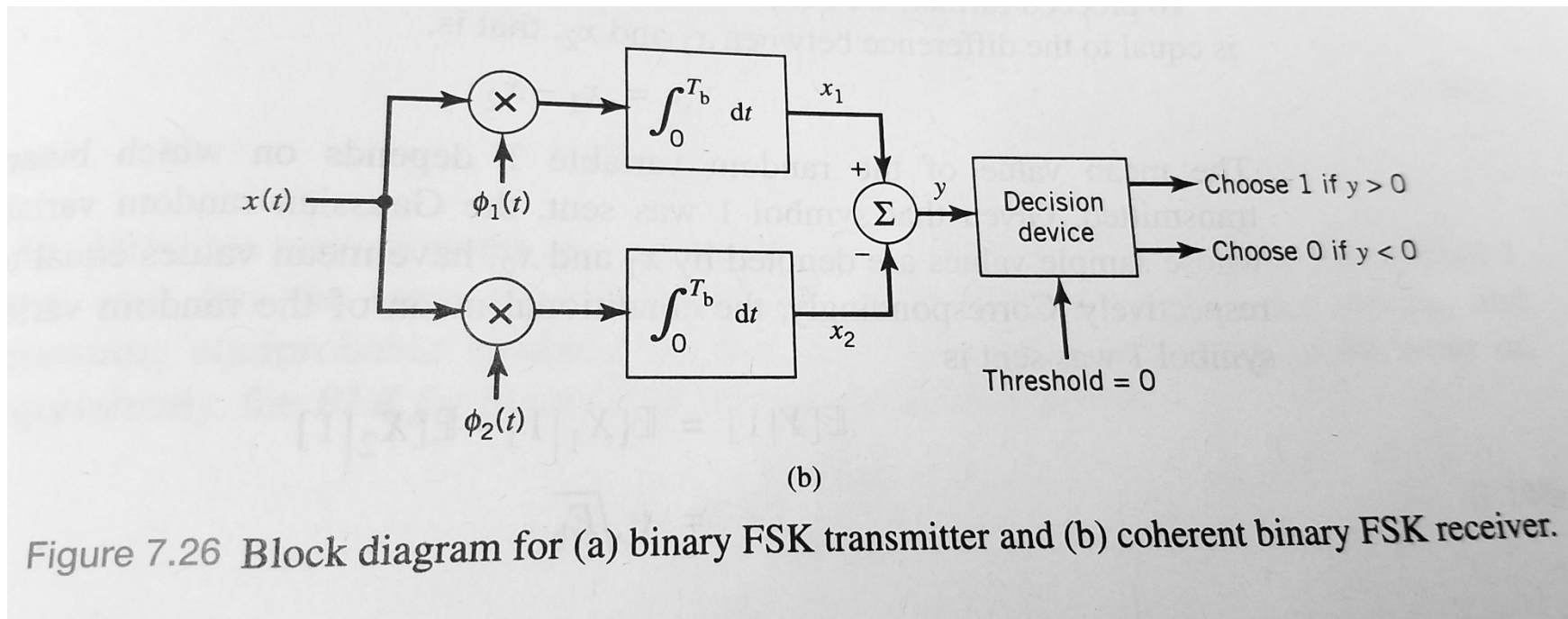
$$\phi_1(t) = \frac{s_1(t)}{\sqrt{E_b}} = \sqrt{\frac{2}{T_b}} \cos\left(2\pi \cdot \frac{n_c + 1}{T_b} \cdot t\right), \quad \phi_2(t) = \frac{s_2(t)}{\sqrt{E_b}} = \sqrt{\frac{2}{T_b}} \cos\left(2\pi \cdot \frac{n_c + 2}{T_b} \cdot t\right)$$

2. the corresponding signal vectors are:

$$\mathbf{s}_1 = \begin{bmatrix} \sqrt{E_b} \\ 0 \end{bmatrix}, \quad \mathbf{s}_2 = \begin{bmatrix} 0 \\ \sqrt{E_b} \end{bmatrix}$$



- The received vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ is obtained as shown in Fig. 7.26b. The decision rule is



$$\left\| \mathbf{x} - \mathbf{s}_1 \right\|^2 \begin{matrix} < \\ > \end{matrix} \left\| \mathbf{x} - \mathbf{s}_2 \right\|^2, \text{ or}$$

$$\left(x_1 - \sqrt{E_b} \right)^2 + x_2^2 \begin{matrix} < \\ > \end{matrix} x_1^2 + \left(x_2 - \sqrt{E_b} \right)^2, \text{ or}$$

$$x_1^2 + E_b - 2x_1\sqrt{E_b} + x_2^2 \begin{matrix} < \\ > \end{matrix} x_2^2 + x_1^2 + E_b - 2x_2\sqrt{E_b}, \text{ or}$$

$$2(x_2 - x_1)\sqrt{E_b} \begin{matrix} < \\ > \end{matrix} 0, \text{ or}$$

$$x_2 \begin{matrix} < \\ > \end{matrix} x_1$$

- BEP analysis is straight forward. From the signal constellation in Fig. 7.25, the square distance between the two signal points is

$$d_{\min}^2 = \|\mathbf{s}_1 - \mathbf{s}_2\|^2 = 2E_b$$

As a result, the BEP of 2FSK is

$$P_b(2\text{FSK}) = Q\left(\sqrt{\frac{d_{\min}^2}{2N_0}}\right) = Q\left(\sqrt{\frac{E_b}{N_0}}\right)$$

In comparison, the BEP of 2PSK is

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

This means 2FSK requires twice as much transmit power to attain the same BEP as 2PSK.

- In general, the MFSK waveforms are:

$$s_i(t) = \sqrt{\frac{2E}{T}} \cos(2\pi f_i t) = \sqrt{E} \phi_i(t), \quad i = 1, 2, \dots, M$$

where

$$\phi_i(t) = \sqrt{\frac{2}{T}} \cos(2\pi f_i t) \quad \text{is the } i\text{-th basis function,}$$

$$f_i = \frac{n_c + i}{T} \quad \text{is the frequency of the } i\text{-th tone,}$$

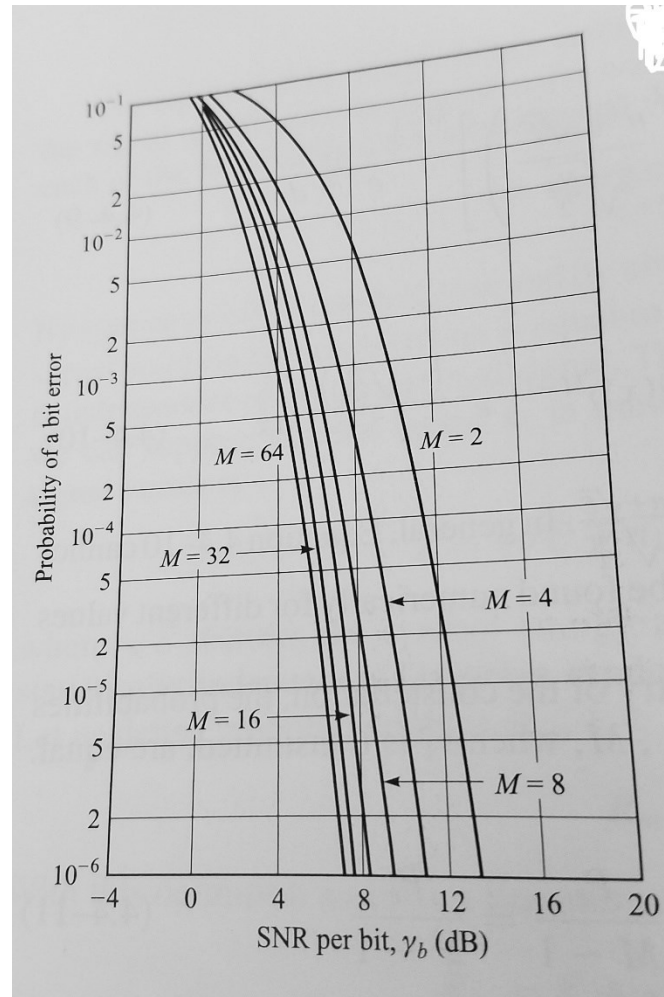
E is the energy of all the waveforms, T is the waveform duration, and n_c is an integer. The corresponding signal vectors are

$$\mathbf{s}_i = [0, \dots, 0, \underset{\substack{i\text{-th} \\ \text{position}}}{\sqrt{E}}, 0, \dots, 0]^T$$

- **Example:** How many nearest neighbours a MFSK signal point has? What is the average number of bit errors when the receiver makes a wrong decision? Use this information to determine the SEP and BEP of MFSK.

- **BEP comparison of different MFSK schemes.** The bit SNR is defined as

$$\gamma_b = E_b / N_0 = \frac{1}{\log_2 M} (E / N_0).$$



- **It is observed that as M increases, the BEP actually decreases.** This is the complete opposite of MPSK and MQAM.

The reason is that the number of signal dimensions in MFSK actually expands as M increases, and so does the bandwidth required for transmission.

In the cases of MPSK and MQAM, the number of signal dimension remains at 2, even when M increases. So the signal space is more crowded, and hence easier to make wrong decisions.

In conclusion, in MFSK, we trade off bandwidth for a higher data reliability.

3.9 Noncoherent Detection of FSK and PSK

Reference: Sections 7.12 and 7.13 of Text

- The binary FSK receiver in the last section correlates the received signal with

$$\phi_1(t) = \frac{s_1(t)}{\sqrt{E_b}} = \sqrt{\frac{2}{T_b}} \cos\left(2\pi \cdot \frac{n_c + 1}{T_b} \cdot t\right), \quad \phi_2(t) = \frac{s_2(t)}{\sqrt{E_b}} = \sqrt{\frac{2}{T_b}} \cos\left(2\pi \cdot \frac{n_c + 2}{T_b} \cdot t\right),$$

whereas the MPSK receiver in Section 7.6 correlates the received signal with

$$\phi_1(t) = \sqrt{\frac{2}{T}} \cos(2\pi f_c t) \quad \text{and} \quad \phi_2(t) = -\sqrt{\frac{2}{T}} \sin(2\pi f_c t)$$

The sinusoidal basis functions generated at the receiver are termed “**local carriers**”.

- An implicit assumption in the formulation of the binary FSK and MPSK receivers in previous sections is that the local carriers at the receivers are synchronized in both frequencies and phases to those used in the transmitter. A receiver that performs phase and frequency synchronization is called a **coherent receiver/detector**.
- Coherent detection can be expensive, e.g. in a mobile environment when the mobility of the user terminal introduces a Doppler effect which makes the phase of the received signal changes from one symbol to the next.
- In this section, we look at **non-coherent detection** of binary FSK and MPSK. These non-coherent detectors do not attempt to synchronize to the phase of the received signal (we still make the assumption that frequency is perfectly synchronized, a realistic assumption given the stability of oscillators).

3.9.1 Binary FSK with non-coherent detection

- With a phase difference of θ between the transmitter's and local carriers, we modeled the two FSK tones generated at the transmitter as

$$s_1(t) = \sqrt{\frac{2E_b}{T_b}} \cos\left(2\pi \cdot \frac{n_c + 1}{T_b} \cdot t - \theta\right) = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi \cdot f_1 \cdot t - \theta),$$

$$s_2(t) = \sqrt{\frac{2E_b}{T_b}} \cos\left(2\pi \cdot \frac{n_c + 2}{T_b} \cdot t - \theta\right) = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi \cdot f_2 \cdot t - \theta)$$

- The non-coherent binary FSK receiver correlates the received signal with

$$\phi_{11}(t) = \sqrt{\frac{2}{T_b}} \cos(2\pi \cdot f_1 \cdot t), \quad \phi_{12}(t) = \sqrt{\frac{2}{T_b}} \sin(2\pi \cdot f_1 \cdot t),$$

$$\phi_{21}(t) = \sqrt{\frac{2}{T_b}} \cos(2\pi \cdot f_2 \cdot t), \quad \phi_{22}(t) = \sqrt{\frac{2}{T_b}} \sin(2\pi \cdot f_2 \cdot t)$$

It is clear that $\phi_{11}(t)$ and $\phi_{21}(t)$ are those used in the coherent detector, and we introduce two additional signals, $\phi_{12}(t)$ and $\phi_{22}(t)$.

- **Exercise:** Show that when we correlate $s_1(t)$ with $\phi_{11}(t)$, $\phi_{21}(t)$, $\phi_{12}(t)$, and $\phi_{22}(t)$, the results are

$$s_{1,11} = \int_0^{T_b} s_1(t)\phi_{11}(t)dt = \sqrt{E_b} \cos \theta, \quad s_{1,12} = \int_0^{T_b} s_1(t)\phi_{12}(t)dt = \sqrt{E_b} \sin \theta$$

$$s_{1,21} = \int_0^{T_b} s_1(t)\phi_{21}(t)dt = 0, \quad s_{1,22} = \int_0^{T_b} s_1(t)\phi_{22}(t)dt = 0$$

Similarly, show that the results of correlating $s_2(t)$ with $\phi_{11}(t)$, $\phi_{21}(t)$, $\phi_{12}(t)$ are

$$\begin{aligned}
s_{2,11} &= \int_0^{T_b} s_2(t) \phi_{11}(t) dt = 0, & s_{2,12} &= \int_0^{T_b} s_2(t) \phi_{12}(t) dt = 0 \\
s_{2,21} &= \int_0^{T_b} s_2(t) \phi_{21}(t) dt = \sqrt{E_b} \cos \theta, & s_{2,22} &= \int_0^{T_b} s_2(t) \phi_{22}(t) dt = \sqrt{E_b} \sin \theta
\end{aligned}$$

- The results from the above exercise means that when $s_1(t)$ was transmitted, and the channel introduces AWGN $w(t)$ with a power spectral density of $N_0 / 2$, the results of correlating the received signal $r(t) = s_1(t) + w(t)$ with $\phi_{11}(t)$, $\phi_{21}(t)$, $\phi_{12}(t)$, $\phi_{22}(t)$ are

$$\begin{aligned}
r_{11} &= \int_0^{T_b} r(t) \phi_{11}(t) dt = \sqrt{E_b} \cos \theta + w_{11}, & r_{12} &= \int_0^{T_b} r(t) \phi_{12}(t) dt = \sqrt{E_b} \sin \theta + w_{12} \\
r_{21} &= \int_0^{T_b} r(t) \phi_{21}(t) dt = w_{21}, & r_{22} &= \int_0^{T_b} r(t) \phi_{22}(t) dt = w_{22}
\end{aligned}$$

where $w_{11}, w_{12}, w_{21}, w_{22}$ are iid Gaussian noise with zero mean and variance $N_0 / 2$.

We notice that

$$\begin{aligned}
 x_1 = r_{11}^2 + r_{12}^2 &= \left(\sqrt{E_b} \cos \theta + w_{11} \right)^2 + \left(\sqrt{E_b} \sin \theta + w_{12} \right)^2 \\
 &= \underbrace{E_b}_{\text{signal}} + \underbrace{2\sqrt{E_b}(\cos \theta w_{11} + \sin \theta w_{12})}_{\text{linear noise term}} + \underbrace{w_{11}^2 + w_{12}^2}_{\text{square noise term}}
 \end{aligned}$$

and

$$x_2 = r_{21}^2 + r_{22}^2 = w_{21}^2 + w_{22}^2.$$

Important points:

1. The signal component, E_b , is independent of the unknown phase difference θ simply disappears!

2. The mean and variance of the linear noise term, $w_{11} \cos \theta + w_{12} \sin \theta$, is also independent of the unknown phase difference. **(Show this!)**

- Similarly, when $s_2(t)$ was transmitted, the results of the correlation become

$$\begin{aligned} r_{11} &= \int_0^{T_b} r(t) \phi_{11}(t) dt = w_{11}, & r_{12} &= \int_0^{T_b} r(t) \phi_{12}(t) dt = w_{12} \\ r_{21} &= \int_0^{T_b} r(t) \phi_{21}(t) dt = \sqrt{E_b} \cos \theta + w_{21}, & r_{22} &= \int_0^{T_b} r(t) \phi_{22}(t) dt = \sqrt{E_b} \sin \theta + w_{22} \end{aligned}$$

and

$$\begin{aligned} x_1 &= r_{11}^2 + r_{12}^2 = w_{11}^2 + w_{12}^2, \\ x_2 &= r_{21}^2 + r_{22}^2 = \left(\sqrt{E_b} \cos \theta + w_{21} \right)^2 + \left(\sqrt{E_b} \sin \theta + w_{22} \right)^2 \\ &= E_b + 2\sqrt{E_b} (w_{21} \cos \theta + w_{22} \sin \theta) + w_{21}^2 + w_{22}^2 \end{aligned}$$

- Base on the difference in signal structures of x_1 and x_2 when $s_1(t)$ and $s_2(t)$ are transmitted, we arrive at the decision rule:

$$\begin{array}{c} s_1(t) \\ > \\ x_1 \qquad x_2 \\ < \\ s_2(t) \end{array}$$

(Non-coherent binary FSK)

- It turns out that the above non-coherent detection strategy is optimal, i.e. it minimizes the bit-error probability in the absence of knowledge about θ . The BEP is

$$P_b = \frac{1}{2} \exp\left(-\frac{1}{2} \frac{E_b}{N_0}\right)$$

(Non-coherent binary FSK)

3.9.2 MPSK with Differential Detection

- With coherent detection, the MPSK modulation waveforms, basis functions, and signal vectors are respectively

$$s_i(t) = \sqrt{\frac{2E}{T}} \cos \left[2\pi f_c t + \frac{2\pi}{M} (i-1) \right], \quad i = 1, 2, \dots, M,$$

$$\phi_1(t) = \sqrt{\frac{2}{T}} \cos[2\pi f_c t], \quad \phi_2(t) = -\sqrt{\frac{2}{T}} \sin[2\pi f_c t],$$

$$\mathbf{s}_i = \sqrt{E} \begin{bmatrix} \cos \left(\frac{2\pi}{M} (i-1) \right) \\ \sin \left(\frac{2\pi}{M} (i-1) \right) \end{bmatrix}, \quad i = 1, 2, \dots, M.$$

The receiver correlates the received signal with $\phi_1(t)$ and $\phi_2(t)$ to generate the received vector required for minimum distance detection. The received vector is the transmitted vector, plus a noise vector.

- As in FSK, when non-coherent detection is employed, the modulation waveforms are modelled as

$$s_{i,new}(t) = \sqrt{\frac{2E}{T}} \cos \left[2\pi f_c t + \frac{2\pi}{M}(i-1) - \theta \right], \quad i = 1, 2, \dots, M,$$

where θ is the phase difference between the transmitter's and receiver's carriers.

The receiver **still** correlates the received signal with $\phi_1(t)$ and $\phi_2(t)$.

- Correlating the new $s_{i,new}(t) = \sqrt{\frac{2E}{T}} \cos\left[2\pi f_c t + \frac{2\pi}{M}(i-1) - \theta\right]$ with $\phi_1(t)$ and $\phi_2(t)$ yields

$$\begin{aligned}
 a_{i1} &= \int_0^T \underbrace{\sqrt{\frac{2E}{T}} \cos\left[2\pi f_c t + \frac{2\pi(i-1)}{M} - \theta\right]}_{s_{i,new}(t)} \cdot \underbrace{\sqrt{2} \cos(2\pi f t)}_{\phi_1(t)} dt \\
 &= \sqrt{E} \cos\left(\frac{2\pi(i-1)}{M} - \theta\right)
 \end{aligned}$$

and

$$\begin{aligned}
 a_{i2} &= - \int_0^T \underbrace{\sqrt{\frac{2E}{T}} \cos\left[2\pi f_c t + \frac{2\pi(i-1)}{M} - \theta\right]}_{s_{i,new}(t)} \cdot \underbrace{\sqrt{2} \sin(2\pi f t)}_{-\phi_2(t)} dt \\
 &= \sqrt{E} \sin\left(\frac{2\pi(i-1)}{M} - \theta\right)
 \end{aligned}$$

- Now, let's stack a_{i1} and a_{i2} to obtain the new signal vector

$$\mathbf{a}_i = \begin{bmatrix} a_{i1} \\ a_{i2} \end{bmatrix} = \sqrt{E} \begin{bmatrix} \cos\left(\frac{2\pi(i-1)}{M} - \theta\right) \\ \sin\left(\frac{2\pi(i-1)}{M} - \theta\right) \end{bmatrix}$$

Add rotated constellation during lecture here

- Let $s_{i,new}(t)$ and $s_{j,new}(t)$ be the modulation waveforms sent in the $(k-1)$ -th and the k -th symbol interval respectively.

The phase difference in these two waveforms is

$$\Delta_k = \left(\frac{2\pi j}{M} - \theta \right) - \left(\frac{2\pi i}{M} - \theta \right) = \frac{2\pi}{M}(j - i)$$

Remember this as it is fundamental to understand how the detector works.

- In the **absence of noise**, the correlator outputs in these two intervals are:

$$\mathbf{x}_{k-1} = \begin{bmatrix} x_{k-1,1} \\ x_{k-1,2} \end{bmatrix} = \mathbf{a}_i = \sqrt{E} \begin{bmatrix} \cos\left(\frac{2\pi(i-1)}{M} - \theta\right) \\ \sin\left(\frac{2\pi(i-1)}{M} - \theta\right) \end{bmatrix}$$

and

$$\mathbf{x}_k = \begin{bmatrix} x_{k1} \\ x_{k2} \end{bmatrix} = \mathbf{a}_j = \sqrt{E} \begin{bmatrix} \cos\left(\frac{2\pi(j-1)}{M} - \theta\right) \\ \sin\left(\frac{2\pi(j-1)}{M} - \theta\right) \end{bmatrix}$$

- Now examine what happens when we take the inner product of the two vectors

$$\begin{aligned} a_k &= \mathbf{x}_k^T \mathbf{x}_{k-1} = \mathbf{a}_j^T \mathbf{a}_i = x_{k,1}x_{k-1,1} + x_{k,2}x_{k-1,2} \\ &= E \cdot \left\{ \cos\left(\frac{2\pi(j-1)}{M} - \theta\right) \cos\left(\frac{2\pi(i-1)}{M} - \theta\right) + \sin\left(\frac{2\pi(j-1)}{M} - \theta\right) \sin\left(\frac{2\pi(i-1)}{M} - \theta\right) \right\} \\ &= E \cdot \cos\left(\frac{2\pi(j-i)}{M}\right) \end{aligned}$$

- Similarly, perform a cross product of $\mathbf{x}_k$ and $\mathbf{x}_{k-1}$ (interested in amplitude of cross product only)

$$\begin{aligned}
b_k &= x_{k,2}x_{k-1,1} - x_{k,1}x_{k-1,2} \\
&= E \cdot \left\{ \sin\left(\frac{2\pi(j-1)}{M} - \theta\right) \cos\left(\frac{2\pi(i-1)}{M} - \theta\right) - \cos\left(\frac{2\pi(j-1)}{M} - \theta\right) \sin\left(\frac{2\pi(i-1)}{M} - \theta\right) \right\} \\
&= E \cdot \sin\left(\frac{2\pi(j-i)}{M}\right)
\end{aligned}$$

- It is observed that neither a_k nor b_k is a function of the unknown phase difference θ . Furthermore,

$$\tan^{-1}\left(\frac{b_k}{a_k}\right) = \frac{2\pi}{M}(j-i) = \Delta_k$$

In other words, the receiver is able to recover the phase difference of successive transmitted waveforms by computing the inner and cross products of successive received signal vectors. **This procedure is known as differential detection!**

- Differential detection on its own does not enable the detection of the data bits. It has to be used in conjunction with **differential encoding** on the transmitter side.
- Instead of mapping information bit(s) onto the ‘absolute’ phase of the transmitted waveform, we use it to adjust the phase of the next transmitted waveform.
- **Illustrative Example:** Differential 8PSK

The 8 modulation waveforms generated at the transmitter are

$$s_i(t) = \sqrt{\frac{2E}{T}} \cos(2\pi f_c t + \psi_i - \theta), \quad i = 1, 2, \dots, 8$$

where

i	1	2	3	4	5	6	7	8
ψ_i	0	$\pi / 4$	$2\pi / 4$	$3\pi / 4$	$4\pi / 4$	$5\pi / 4$	$6\pi / 4$	$7\pi / 4$

Each waveform is divided into segments of duration T seconds. Each segment is called a symbol interval. The k -th interval is $[kT, kT + T]$. The modulator may switch from one waveform in one interval to another waveform in the following interval, depending on the 3 input bits during the interval.

Let (b_1, b_2, b_3) be the information bits input to the differential encoder during a symbol interval and let δ denotes the corresponding phase change in the transmitted DPSK signal from the previous symbol interval (each symbol interval is of duration T seconds). The differential encoding rule is:

(b_1, b_2, b_3)	(0,0,0)	(0,0,1)	(0,1,1)	(0,1,0)	(1,1,0)	(1,1,1)	(1,0,1)	(1,0,0)
δ	0	$\pi / 4$	$2\pi / 4$	$3\pi / 4$	$4\pi / 4$	$5\pi / 4$	$6\pi / 4$	$7\pi / 4$

Note: Gray coding is applied to the differentially encoded phase.

Without loss of generality, let us assume that $s_1(t)$ is transmitted in the 0-th interval. This signal has a modulation phase $\psi = \psi_1 = 0$. Subsequent transmitted signals in the example are:

Interval index k	0	1	2	3	4	5	6
Input bits (b_1, b_2, b_3)		(0,1,1)	(1,0,0)	(0,0,0)	(0,1,0)	(1,1,1)	(0,1,1)
Phase change δ		$2\pi / 4$	$7\pi / 4$	0	$3\pi / 4$	$5\pi / 4$	$2\pi / 4$
New phase ψ_i		$2\pi / 4$	$\pi / 4$	$\pi / 4$	$4\pi / 4$	$\pi / 4$	$3\pi / 4$
Transmitted signal	$s_1(t)$	$s_3(t)$	$s_2(t)$	$s_2(t)$	$s_5(t)$	$s_2(t)$	$s_4(t)$

- **Exercise:** Repeat the above for differential 2PSK and 4PSK.
- The BEP of differential 2PSK is found to be

$$P_b = \frac{1}{2} \exp\left(-\frac{E_b}{N_0}\right)$$

(differential 2PSK)

This is 3 dB more power efficient than non-coherent binary FSK.

- The table below summarizes the BEPs of 2PSK and 2FSK with coherent and non-coherent detection:

Modulation Type	Coherent Detection	Non-coherent Detection
2PSK	$Q\left(\sqrt{2E_b / N_0}\right)$	$\frac{1}{2} \exp(-E_b / N_0)$
2FSK	$Q\left(\sqrt{E_b / N_0}\right)$	$\frac{1}{2} \exp(-\frac{1}{2} E_b / N_0)$

- Finally, the BEP comparison:

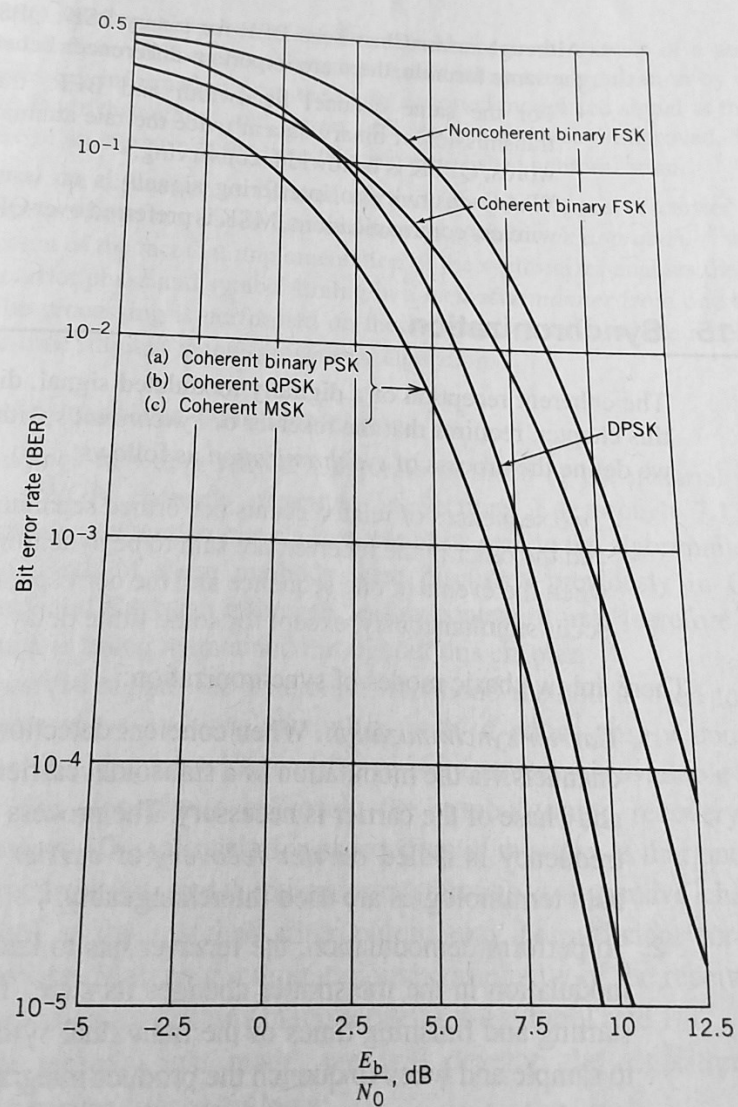


Figure 7.47 Comparison of the noise performance of different PSK and FSK schemes.