1. Dimensionality of Signals

HTE - 12.09.2012

Given a signal set of *M* signals (waveforms) of $s_1(t)\cdots s_M(t)$, we wish to find a set of *N* orthonormalized basis functions $\psi_1(t)\cdots\psi_N(t)$, so that we can write $s_1(t)\cdots s_M(t)$ in terms of $\psi_1(t)\cdots\psi_N(t)$. Note that after this operation of orthogonalization, we can represent $s_1(t)\cdots s_M(t)$ in an *N* dimensional space where each (orthogonal) axis corresponds to one of $\psi_1(t)\cdots\psi_N(t)$. We chose *m* and *n* indexes such that $1\le m\le M$, $1\le n\le N$ and it is reasonable to assume that $N\le M$.

A formal method of finding $\psi_1(t)\cdots\psi_N(t)$ given $s_1(t)\cdots s_M(t)$ is to use Gram-Schmidt Orthogonalization Procedure. This is described below in steps

Gram-Schmidt Orthogonalization Procedure

1. We begin with $s_1(t)$ and set $\psi_1(t)$ as follows

$$\psi_1(t)=\frac{s_1(t)}{\sqrt{\varepsilon_1}} \tag{1.1}$$

where ε_1 is the energy in $s_1(t)$ and found from

$$\varepsilon_1=\int_{-\infty}^{\infty} s_1^2(t)\,dt \tag{1.2}$$

This way $\psi_1(t)$ is simply $s_1(t)$ with normalized energy, i.e. unity energy. Note that we demand orthonormalized basis functions $\psi_1(t)\cdots\psi_N(t)$ to have unit energy to avoid scaling problems (similar to Frequency Transforms)

2. To find $\psi_2(t)$ we proceed as follows. We take $s_2(t)$ and find its projection onto $\psi_1(t)$ (axis) from

$$c_{21}=\int_{-\infty}^{\infty} s_2(t)\;\psi_1(t)\,dt \tag{1.3}$$

Then we subtract $c_{21}\psi_1(t)$ from $s_2(t)$ to get

$$d_2(t)=s_2(t)-c_{21}\psi_1(t) \tag{1.4}$$

Now $d_2(t)$ is orthogonal to $\psi_1(t)$ and $\psi_2(t)$ can be found by normalizing its energy, hence

$$\psi_2(t)=\frac{d_2(t)}{\sqrt{\varepsilon_2}} \quad , \quad \varepsilon_2=\int_{-\infty}^{\infty} d_2^2(t)\, dt \tag{1.5}$$

Note that unlike ε_1, ε_2 does not correspond to the energy in $s_2(t)$, but rather it refers to the energy in $d_2(t)$.

3. In general, the *n* th orthonormalized basis function is obtained from

$$\psi_n(t)=\frac{d_n(t)}{\sqrt{\varepsilon_n}} \quad , \quad d_n(t)=s_n(t)-\sum_{i=1}^{n-1} c_{ni}\psi_i(t) \quad ,$$
$$\varepsilon_n=\int_{-\infty}^{\infty} d_n^2(t)\, dt \quad , \quad c_{ni}=\int_{-\infty}^{\infty} s_n(t)\ \psi_i(t) dt \quad i=1\cdots n-1 \tag{1.6}$$

4. So, the process in 3. is continued until we reach $n=M$, i.e. when all *M* waveforms are exhausted.

Example 1.1 : Four signals, waveforms, i.e. , $M=4$ named as $s_1(t)\cdots s_4(t)$ are given in Fig. 1.1, we are asked to find a set of orthonormalized basis functions $\psi_1(t)\cdots\psi_N(t)$, where $N\le 4$. Before we tackle the solution, it is important to remind that our signals are defined piecewise over the given time intervals. This means that mathematical calculations must be made taking into these intervals, otherwise we will get inaccurate results. To be precise, we give the expressions for $s_1(t)\cdots s_4(t)$, based on the plots in Fig. 1.1a and split into the time intervals they occupy.

$$s_1(t)=\begin{cases}1 & 0\le t\le 2\\ 0 & \text{otherwise}\end{cases} \qquad s_2(t)=\begin{cases}1 & 0\le t\le 1\\ -1 & 1<t\le 2\\ 0 & \text{otherwise}\end{cases}$$
$$s_3(t)=\begin{cases}-1 & 0\le t\le 1\\ 1 & 1<t\le 3\\ 0 & \text{otherwise}\end{cases} \qquad s_4(t)=\begin{cases}1 & 0\le t\le 3\\ 0 & \text{otherwise}\end{cases} \tag{1.7}$$

Solution : As described above, first we find $\psi_1(t)$

$$\psi_1(t)=\frac{s_1(t)}{\sqrt{\varepsilon_1}} \quad , \text{ where } \quad \varepsilon_1=\int_{-\infty}^{\infty} s_1^2(t)\, dt=2 \ , \text{ hence } \psi_1(t)=\frac{s_1(t)}{\sqrt{2}} \tag{1.8}$$

Now we proceed to find $c_{21}=0$ from

$$c_{21}=\int_{-\infty}^{\infty} s_2(t)\ \psi_1(t) dt=0 \tag{1.9}$$

This is because the overlapping parts of $s_2(t)$ and $\psi_1(t)$ are orthogonal. Then $d_2(t)=s_2(t)$ and

$$\psi_2(t) = \frac{d_2(t)}{\sqrt{\varepsilon_2}} = \frac{s_2(t)}{\sqrt{\varepsilon_2}} = \frac{s_2(t)}{\sqrt{2}} \quad , \quad \text{since} \quad \varepsilon_2 = \int_{-\infty}^{\infty} s_2^2(t)\, dt = 2 \tag{1.10}$$

To evaluate $\psi_3(t)$, we need to compute c_{31} and c_{32} using (1.6), this way

$$c_{31} = \int_{-\infty}^{\infty} s_3(t)\ \psi_1(t) dt = 0 \quad , \quad c_{32} = \int_{-\infty}^{\infty} s_3(t)\ \psi_2(t) dt = -\sqrt{2} \tag{1.11}$$

Then

$$d_3(t) = s_3(t) - c_{32}\psi_2(t) = \ s_3(t) + \sqrt{2}\psi_2(t) \tag{1.12}$$

As seen from (1.12), $d_3(t)$ is a waveform extending from $t = 2$ to $t = 3$ with unit energy, thus $\psi_3(t) = d_3(t)$.

Finally (again using (1.6)), we find that $c_{41} = \sqrt{2}$, $c_{42} = 0$, $c_{43} = 1$, then

$$d_4(t) = s_4(t) - c_{43}\psi_3(t) - c_{42}\psi_2(t) - c_{41}\psi_1(t) = \ s_4(t) - \psi_3(t) - \sqrt{2}\psi_1(t) = 0 \tag{1.13}$$

As can be verified by plotting $d_4(t)$. (1.13) means that we have reached the end of the orthogonalization process and the waveforms $s_1(t) \cdots s_4(t)$ can adequately be represented by orthonormalized basis functions of $\psi_1(t) \cdots \psi_3(t)$. So in this example $M = 4,\ N = 3$. For general M and N, the waveforms $s_1(t) \cdots s_4(t)$ can be plotted in an N dimensional signal space, where any of $s_m(t)$ in $s_1(t) \cdots s_M(t)$ can be written in terms of the orthonormalized basis functions of $\psi_1(t) \cdots \psi_N(t)$ as follows

$$s_m(t) = \sum_{n=1}^{N} s_{mn}\psi_n(t) \quad , \quad m = 1 \cdots M \quad , \quad s_{mn} = \int_{-\infty}^{\infty} s_m(t)\ \psi_n(t) dt \tag{1.14}$$

(1.14) means $s_m(t)$ can be constructed from the components s_{mn} along different $\psi_n(t)$. Or alternatively, we can say that s_{mn} coefficients are the projections of $s_m(t)$ along the axis of $\psi_n(t)$. In this manner we define the vectorial representation of $s_m(t)$ as

$$\mathbf{s}_m = [s_{m1},\ s_{m2},\ \cdots\ s_{mN}] \quad , \quad \int_{-\infty}^{\infty} s_m(t)\, s_n(t) dt = \mathbf{s}_m \bullet \mathbf{s}_n \quad , \quad d_{mn} = |\mathbf{s}_m - \mathbf{s}_n| \tag{1.15}$$

The integral in the middle means that the product of two signal waveforms corresponds to the vectorial inner product. For instance if $s_m(t)$ and $s_n(t)$ are orthogonal, then the inner product will be zero. d_{mn} refers to the distance between the ends of vectors $\mathbf{s}_m$ and $\mathbf{s}_n$, which will be important from probability of error performance.

Since $s_m(t)$ and $\mathbf{s}_m$ refer to the same signal, the energy ε_m in $s_m(t)$ can be obtained either from $s_m(t)$ or from $\mathbf{s}_m$, thus

$$\varepsilon_m = \int_{-\infty}^{\infty} s_m^2(t)\, dt = \sum_{n=1}^{N} s_{mn}^2 = \|\mathbf{s}_m\|^2 \tag{1.16}$$

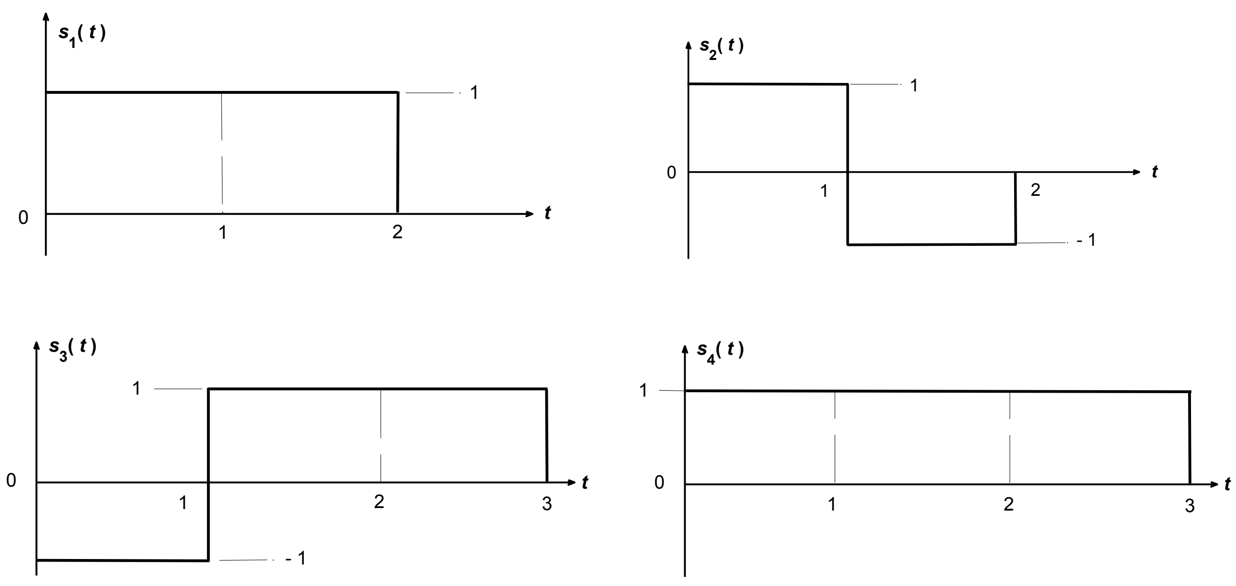
Note that it is important to comprehend these concepts, since they will be used in the detection process at receiver.

For the example above, signal vectors $\mathbf{s}_1 \cdots \mathbf{s}_4$ are shown in Fig. 1.2 in the three dimensional space of $\psi_1(t) \cdots \psi_3(t)$. By using (1.14) and (1.15) it is possible to calculate $\mathbf{s}_1 \cdots \mathbf{s}_4$ and their respective distances as

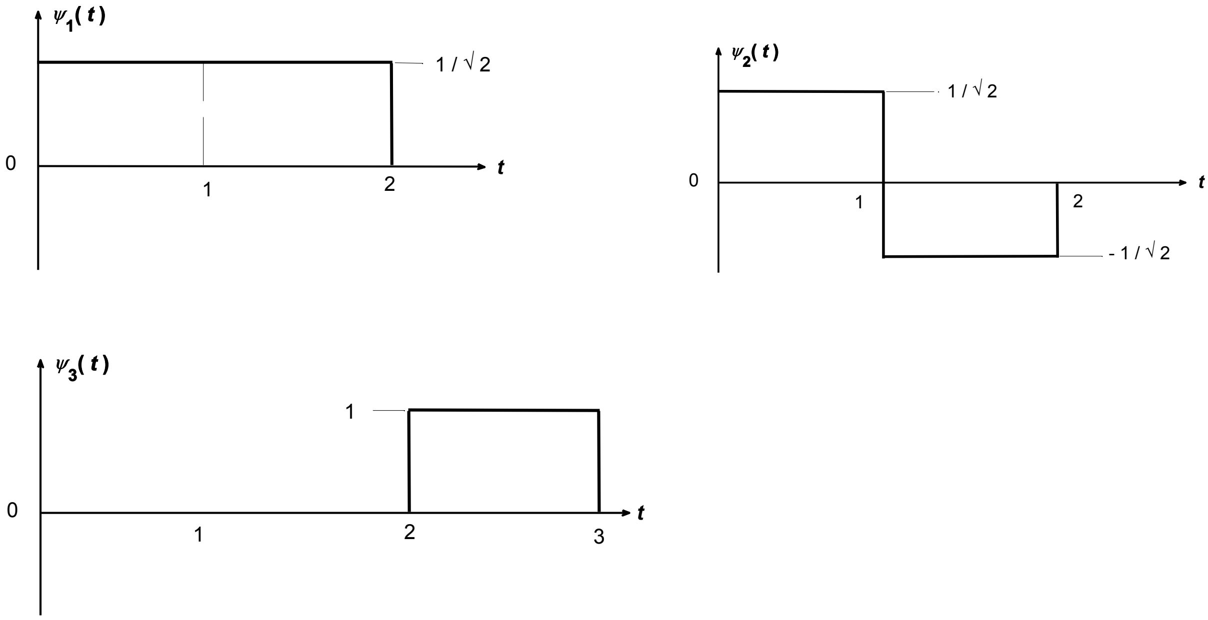
$$\mathbf{s}_1 = \left[\sqrt{2},\ 0,\ 0\right] \ ,\ \mathbf{s}_2 = \left[0,\ \sqrt{2},\ 0\right] \ ,\ \mathbf{s}_3 = \left[0,\ -\sqrt{2},\ 1\right] \ ,\ \mathbf{s}_4 = \left[\sqrt{2},\ 0,\ 1\right]$$

$$d_{12} = |\mathbf{s}_1 - \mathbf{s}_2| = \left[\left(\sqrt{2}-0\right)^2 + \left(0-\sqrt{2}\right)^2 + (0-0)^2\right]^{0.5} = 2\ ,\ d_{13} = |\mathbf{s}_1 - \mathbf{s}_3| = \sqrt{5}$$

$$d_{14} = |\mathbf{s}_1 - \mathbf{s}_4| = 1\ ,\ d_{23} = |\mathbf{s}_2 - \mathbf{s}_3| = 3\ ,\ d_{24} = |\mathbf{s}_2 - \mathbf{s}_4| = \sqrt{5}\ ,\ d_{34} = |\mathbf{s}_3 - \mathbf{s}_4| = 2 \tag{1.17}$$



a) Signal waveforms $s_1(t) \cdots s_4(t)$



b) Orthonormal waveforms (basis function) $\psi_1(t)\cdots\psi_3(t)$

Fig. 1.1 Signals waveforms and orthonormalized basis functions for Example 1.1

We do not always have to go to such lengths of finding orthonormalized basis functions by the use of **Gram-Schmidt Orthogonalization Procedure**. As an alternative, we can use intuition and eye inspection to arrive at a set of orthonormalized basis functions. Here the rule is that orthonormalized basis functions $\psi_1(t)\cdots\psi_N(t)$ should satisfy three requirements

1. $\psi_1(t)\cdots\psi_N(t)$ should be orthogonal amongst themselves, this means

$$\int_{-\infty}^{\infty} \psi_i(t)\ \psi_k(t)dt = 0 \quad \text{for } i \neq k\ ,\ i = 1\cdots N\ ,\ k = 1\cdots N \tag{1.18}$$

2. $\psi_1(t)\cdots\psi_N(t)$ should have unit energy, this means

$$\int_{-\infty}^{\infty} \psi_n^2(t)\ dt = 1\ ,\ n = 1\cdots N \tag{1.19}$$

3. $\psi_1(t)\cdots\psi_N(t)$ should be able to represent (or span) all the signals in the set of $s_1(t)\cdots s_M(t)$ in an *N* dimensional signal space, or alternatively we should be able to express all of the signal in the set of $s_1(t)\cdots s_M(t)$ in terms of $\psi_1(t)\cdots\psi_N(t)$.

So bearing in mind these requirements and particularly paying attention to time slicing in wave forms of $s_1(t)\cdots s_4(t)$, we can deduce an alternative set of orthonormalized basis functions of $\psi_1^a(t)\cdots\psi_3^a(t)$, as shown in Fig. 1.3.

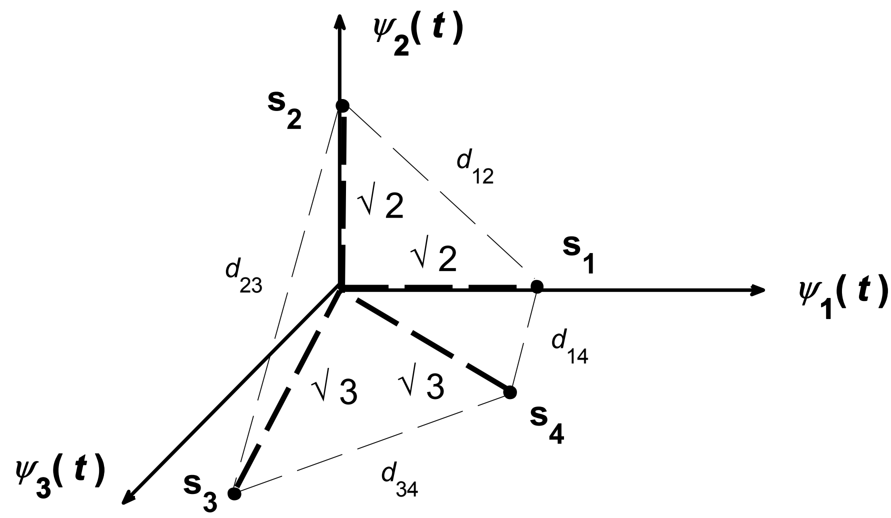


Fig. 1.2 Signal space diagram of signal waveforms $s_1(t)\cdots s_4(t)$ in Example 1.1.

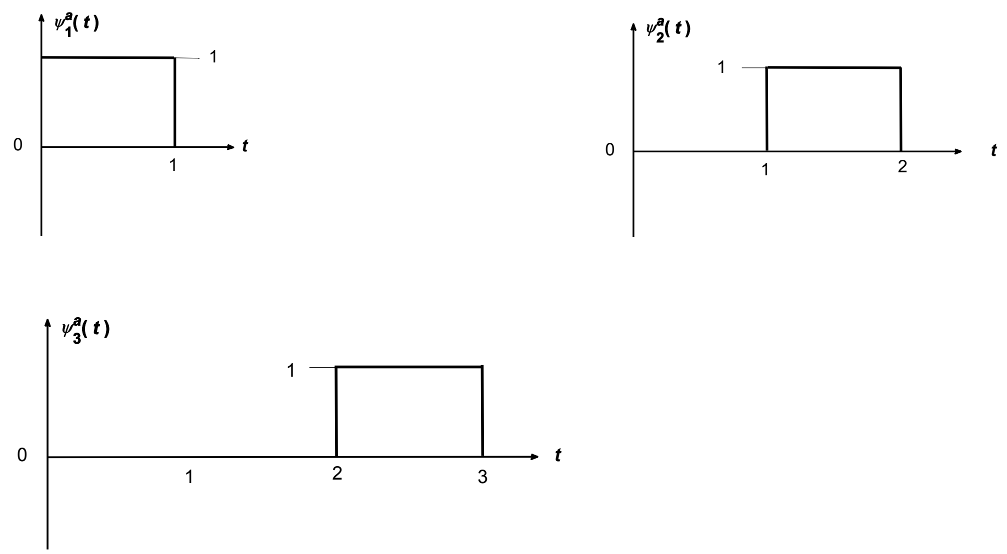


Fig. 1.3 Alternative set of orthonormalized basis functions of $\psi_1^a(t)\cdots\psi_3^a(t)$ for Example 1.1.

Exercise 1.1 : Find the representation of the signal set $s_1(t)\cdots s_4(t)$ in terms of $\psi_1^a(t)\cdots\psi_3^a(t)$. Compare your new signal space diagram, signal vectors and the distances between their ends to those results of $\psi_1(t)\cdots\psi_3(t)$. Make your comments by plotting and writing for the relevant mathematical expressions.

Exercise 1.2 : By using the Matlab file ECE632_GSOrthogonalWaveforms_Exp1.m (available on course webpage), find orthonormalized basis functions $\psi_1(t)\cdots\psi_N(t)$ for the signal waveforms given in the Fig 1.4. Verify your results by hand derivation using Gram-Schmidt Orthogonalization Procedure or by eye inspection. Note that number of dimensions N cannot be greater than time slicing along t axis. Write your $s_1(t) \cdots s_M(t)$ in terms of $\psi_1(t) \cdots \psi_N(t)$, draw the signal space diagram for this case.

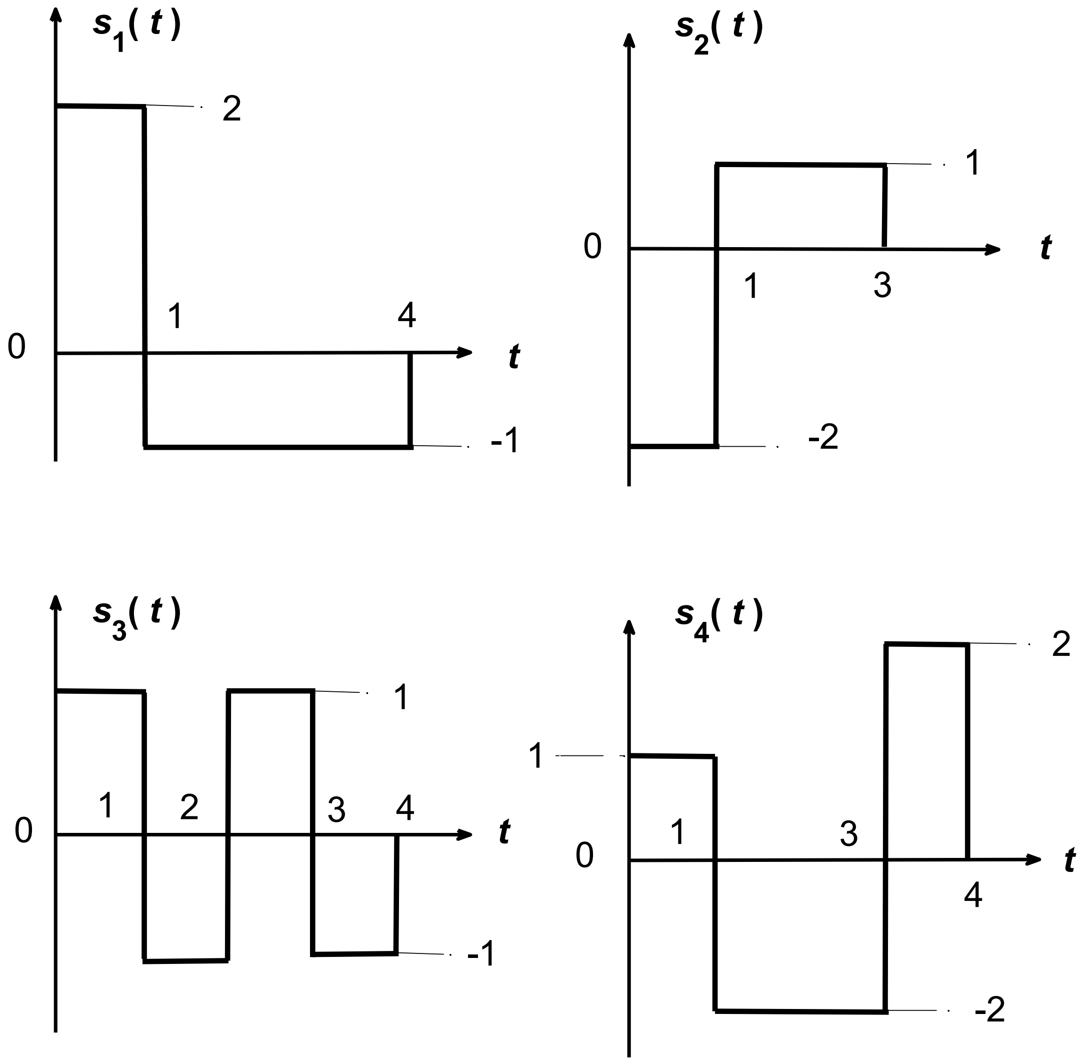


Fig. 1.4 Signal waveforms $s_1(t) \cdots s_4(t)$ for Exercise 1.2.

Before we start to investigate different modulation types, it is instructive to show the simplified block diagram of the modulator we are referring to. This model is shown below in Fig. 1.5. According to this figure, modulation means taking unmodulated input of binary waveforms in groups of $k = \log_2 M$ and converting (mostly called mapping) them into modulated output of *M* ary signals $s_1(t) \cdots s_M(t)$. The way that this modulator functions will correspond to different types of modulations examined next. Note that apart from the case of $M = 2$, it is always the case that $\text{(symbol duration) } T > T_b \text{(binary waveform duration)}$.

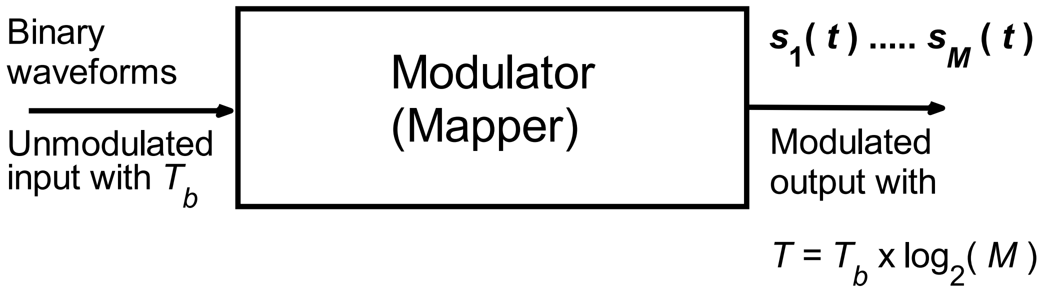
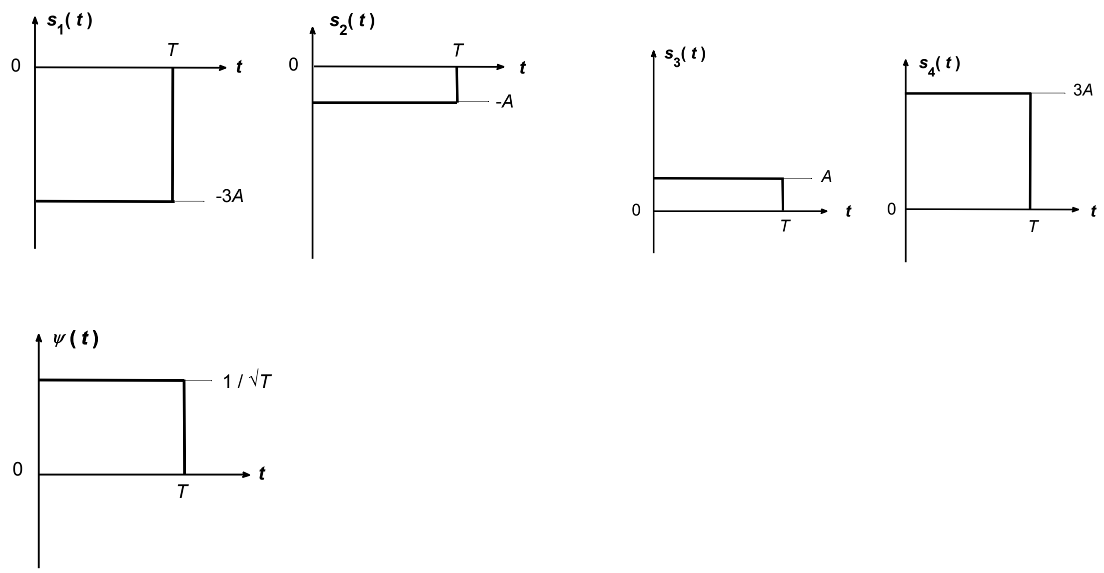


Fig. 1.5 Simplified block diagram of modulator.

2. Amplitude Shift Keying (ASK) or Pulse Amplitude Modulation (PAM)

Now we examine different modulation types with the perspective of dimensionality of signal. The first and the simplest one is ASK (or PAM). In this modulation type, $N=1$, so we say that ASK is one dimensional, thus we need a single orthonormalized function $\psi(t)$ which is usually drawn as a horizontal line, where the signal vector are placed according to their respective energies. In this sense ASK signals will be differentiated by energy differences (essentially amplitude differences) and their respective orientation to the left (negative pulse) and to the right (positive pulse).

By taking a symbol duration of *T*, an ASK example of $M = 4$ is given in Fig. 2.1



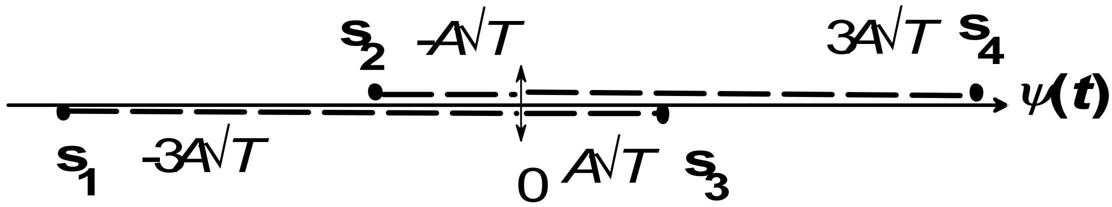


Fig. 2.1 An example of ASK signals, orthonormalized basis function and signal space diagram, for $M = 4$.

By looking at Fig. 2.1 and benefitting from the formulations given in section 1., it is possible to write the following expressions for signal waveforms $s_1(t) \cdots s_4(t)$, signal vectors $\mathbf{s}_1 \cdots \mathbf{s}_4$ and their respective energies. This way $s_1(t)$ and $s_4(t)$ will have greater energies than $s_2(t)$ and $s_3(t)$

$$s_1(t) = -3A \quad , \quad s_2(t) = -A \quad , \quad s_3(t) = A \quad , \quad s_4(t) = 3A \quad , \quad \psi(t) = 1/\sqrt{T} \qquad 0 \le t \le T$$
$$s_1(t) = -3A\sqrt{T}\psi(t) \quad , \quad s_2(t) = -A\sqrt{T}\psi(t) \quad , \quad s_3(t) = A\sqrt{T}\psi(t) \quad , \quad s_4(t) = 3A\sqrt{T}\psi(t)$$
$$\mathbf{s}_1 = \left[-3A\sqrt{T}\right] \ , \ \mathbf{s}_2 = \left[-A\sqrt{T}\right] \ , \ \mathbf{s}_3 = \left[A\sqrt{T}\right] \ , \ \mathbf{s}_4 = \left[3A\sqrt{T}\right]$$
$$\varepsilon_1 = \|\mathbf{s}_1\|^2 = 9A^2T \quad , \quad \varepsilon_2 = \|\mathbf{s}_2\|^2 = A^2T \quad , \quad \varepsilon_3 = \|\mathbf{s}_3\|^2 = A^2T \quad , \quad \varepsilon_4 = \|\mathbf{s}_4\|^2 = 9A^2T \qquad (2.1)$$

Mostly to conserve bandwidth and to avoid intersymbol interference that will occur transmission, we will not use rectangular waveforms, but instead use a shaping waveform called $g_T(t)$.

3. Phase Shift Keying (PSK) and Quadrature Amplitude Modulation (QAM) – Two Dimensional Signals

To establish two dimensions, it is natural to use two (orthogonal) axes which will be $\psi_1(t)$ and $\psi_2(t)$. By taking a symbol duration of *T*, it is possible to generate two types of orthogonal set of $\psi_1(t)$ and $\psi_2(t)$ as illustrated in Fig. 3.1.

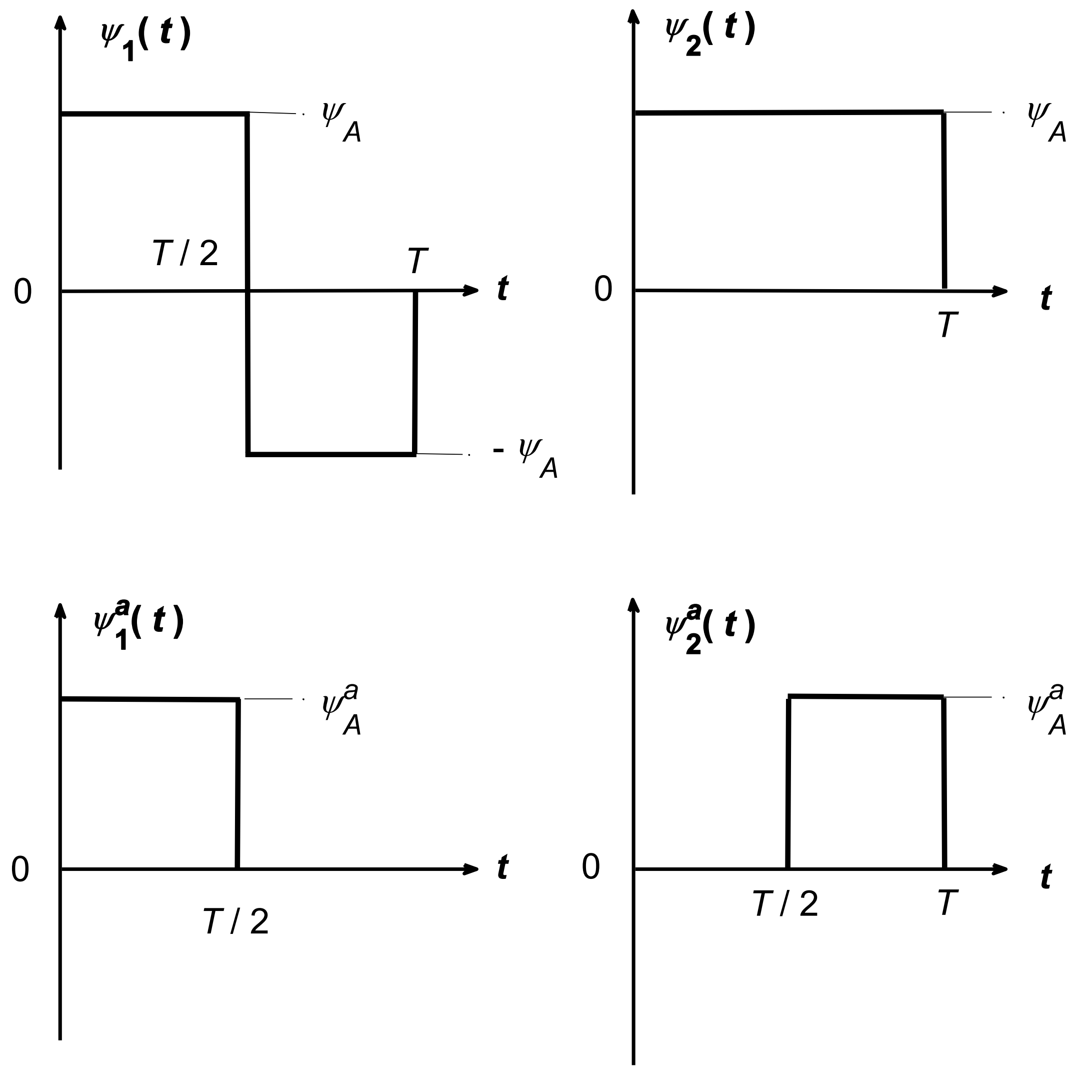


Fig. 3.1 Two sets orthogonal functions $\psi_1(t)$ and $\psi_2(t)$, $\psi_1^a(t)$ and $\psi_2^a(t)$.

We can write the mathematical expressions of $\psi_1(t)$ and $\psi_2(t)$, $\psi_1^a(t)$ and $\psi_2^a(t)$ as

$\psi_1(t)=\psi_A$ for $0\le t\le T/2$, $\psi_1(t)=-\psi_A$ for $T/2\le t\le T$, $\psi_1(t)=0$ for $t<0$ or $t>T$
$\psi_2(t)=\psi_A$ for $0\le t\le T$, $\psi_2(t)=0$ for $t<0$ or $t>T$
$\psi_1^a(t)=\psi_A^a$ for $0\le t\le T/2$, $\psi_1^a(t)=0$ for $t<0$ or $t>T/2$
$\psi_2^a(t)=\psi_A^a$ for $T/2\le t\le T$, $\psi_2^a(t)=0$ for $t<T/2$ or $t>T$ (3.1)

It is quite easy to see that

$$\int_{-\infty}^{\infty}\psi_1(t)\ \psi_2(t)dt=0 \quad , \quad \int_{-\infty}^{\infty}\psi_1^a(t)\ \psi_2^a(t)dt=0 \tag{3.2}$$

So both $\psi_1(t)$ and $\psi_2(t)$ and $\psi_1^a(t)$ and $\psi_2^a(t)$ are orthogonal amoung themselves. Note that $\psi_1^a(t)$ and $\psi_2^a(t)$ achieves orthogonolization by nonoverlapping along time axis, while $\psi_1(t)$ and $\psi_2(t)$ are overlapping. To establish that $\psi_1(t)$ and $\psi_2(t)$ and $\psi_1^a(t)$ and $\psi_2^a(t)$ are orthonormalized as well, we demand that energies are unity such that

$$\int_{-\infty}^{\infty}\psi_1^2(t)\ dt=1 \quad , \quad \int_{-\infty}^{\infty}\psi_2^2(t)\ dt=1 \quad , \quad \int_{-\infty}^{\infty}\left[\psi_1^a(t)\right]^2\ dt=1 \quad , \quad \int_{-\infty}^{\infty}\left[\psi_2^a(t)\right]^2\ dt=1 \tag{3.3}$$

From the evaluations in (3.2), we get $\psi_A=\dfrac{1}{\sqrt{T}}$ and $\psi_A^a=\sqrt{\dfrac{2}{T}}$. So now $\psi_1(t)$ and $\psi_2(t)$ and $\psi_1^a(t)$ and $\psi_2^a(t)$ are both orthogonal and orthonormalized.

Now we choose two time signal waveforms $s_1(t)$ and $s_2(t)$, similar to $\psi_1(t)$ and $\psi_2(t)$ as displayed in Fig. 3.2.

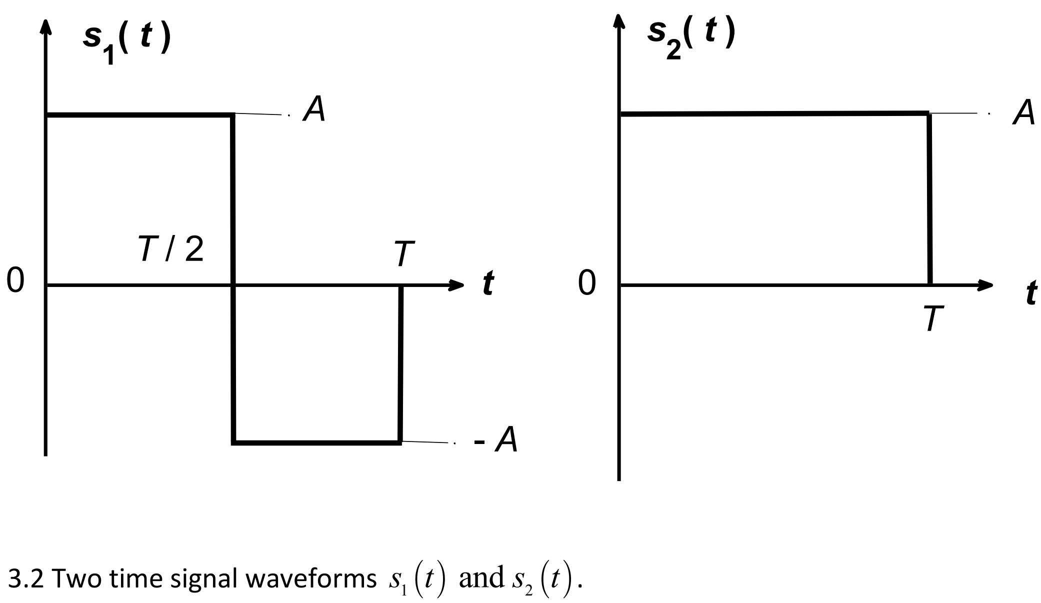


Fig. 3.2 Two time signal waveforms $s_1(t)$ and $s_2(t)$.

If $\psi_1^a(t)$ and $\psi_2^a(t)$ are selected to represent $s_1(t)$ and $s_2(t)$, then it is possible to write $s_1(t)$ and $s_2(t)$ in terms of the orthonormalized basis functions $\psi_1^a(t)$ and $\psi_2^a(t)$ as

$$s_1(t) = A \text{ for } 0 \le t \le T/2 \,,\, s_1(t) = -A \text{ for } T/2 \le t \le T \,,\, s_1(t) = 0 \text{ for } t < 0 \text{ or } t > T$$
$$s_2(t) = A \text{ for } 0 \le t \le T \;,\; s_2(t) = 0 \text{ for } t < 0 \text{ or } t > T$$
$$s_1(t) = A\sqrt{\frac{T}{2}}\psi_1^a(t) - A\sqrt{\frac{T}{2}}\psi_2^a(t) \text{ for } 0 \le t \le T \;,\; s_1(t) = 0 \text{ for } t < 0 \text{ or } t > T$$
$$s_2(t) = A\sqrt{\frac{T}{2}}\psi_1^a(t) + A\sqrt{\frac{T}{2}}\psi_2^a(t) \text{ for } 0 \le t \le T \;,\; s_2(t) = 0 \text{ for } t < 0 \text{ or } t > T$$
$$\mathbf{s}_1 = [s_{11},\, s_{12}] = \left[A\sqrt{\frac{T}{2}},\; -A\sqrt{\frac{T}{2}}\right] \;,\; \mathbf{s}_2 = [s_{21},\, s_{22}] = \left[A\sqrt{\frac{T}{2}},\; A\sqrt{\frac{T}{2}}\right] \qquad (3.4)$$

Note that on the first two lines of (3.4), we have intentionally written for $s_1(t)$ and $s_2(t)$ as they are seen from $s_1(t)$ and $s_2(t)$ (without $\psi_1^a(t)$ and $\psi_2^a(t)$). On the third and fourth lines of (3.4), where there are the expressions of $s_1(t)$ and $s_2(t)$ in terms of $\psi_1^a(t)$ and $\psi_2^a(t)$, we do not actually need the time range specifications given at the end of lines, since these time ranges are readily built into $\psi_1^a(t)$ and $\psi_2^a(t)$. On the last line of (3.4) we have the vectorial representation of $s_1(t)$ and $s_2(t)$, i.e. $\mathbf{s}_1 = [s_{11},\, s_{12}]$ and $\mathbf{s}_2 = [s_{21},\, s_{22}]$ whose vectorial coefficients can be calculated either using the integral in (1.13) or by eye inspection. With these vectorial coefficients, it is now possible to construct the signal space diagram as shown in Fig. 3.3.

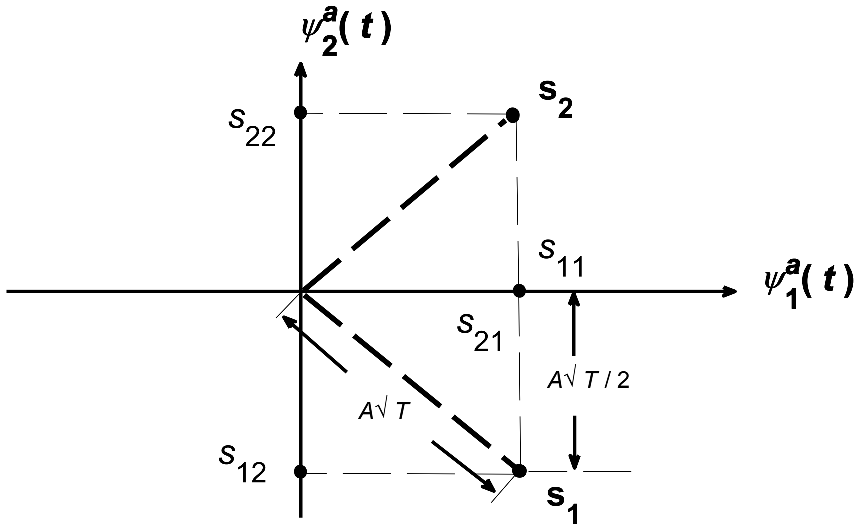


Fig. 3.3 Signal space diagram for $s_1(t)$ and $s_2(t)$ of Fig. 3.2.

As seen from Fig. 3.3. the two signal vectors are placed at 90^0 with respect to each other which is not surprising since $s_1(t)$ and $s_2(t)$ are orthogonal to each other. Additionally the angles $s_1(t)$ and $s_2(t)$ make with $\psi_1^a(t)$ is 45^0, since $|s_{11}| = |s_{12}| = A\sqrt{T/2}$, $|s_{21}| = |s_{22}| = A\sqrt{T/2}$.

Furthermore, the energies in $s_1(t)$ and $s_2(t)$ can be calculated from the lengths of $\mathbf{s}_1$ and $\mathbf{s}_2$ as well as from the time waveforms as shown in (3.5). Of course in both cases we arrive at identical results.

$$\varepsilon_1 = \|\mathbf{s}_1\|^2 = s_{11}^2 + s_{12}^2 = \int_{-\infty}^{\infty} s_1^2(t)\,dt = A^2T$$
$$\varepsilon_2 = \|\mathbf{s}_2\|^2 = s_{21}^2 + s_{22}^2 = \int_{-\infty}^{\infty} s_2^2(t)\,dt = A^2T \ , \ \varepsilon_1 = \varepsilon_2 = \varepsilon_s \quad (3.5)$$

Exercise 3.1 : Find $s_1(t)$ and $s_2(t)$ in terms of $\psi_1(t)$ and $\psi_2(t)$, draw the related signal space diagram, signal vectors, energies and compare your results with the case of $\psi_1^a(t)$ and $\psi_2^a(t)$ and comment on them.

$s_1(t)$ and $s_2(t)$ time waveforms of Fig. 3.2 and the associated signal space diagram in Fig. 3.3. constitute what is called PSK, since here the energies of signals (thus the length of signal vectors) are the same (denoted commonly as ε_s) and the only differentiating factor is the respective angular location corresponding to phases in $s_1(t)$ and $s_2(t)$. Looking at Fig. 3.3, we see that the two dimensional signal space is not used efficiently and we can place two more signals, namely $s_3(t)$ and $s_4(t)$ such that $s_1(t) = -s_3(t)$ and $s_4(t) = -s_2(t)$, hence the vectors $\mathbf{s}_3$ and $\mathbf{s}_4$ will have a rotation of 180^0 with respect to $\mathbf{s}_1$ and $\mathbf{s}_2$. The new signal space diagram comprising $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3$ and $\mathbf{s}_4$ is given in Fig. 3.4. Note that here we have reverted from $\psi_1^a(t)$ and $\psi_2^a(t)$ to $\psi_1(t)$ and $\psi_2(t)$.

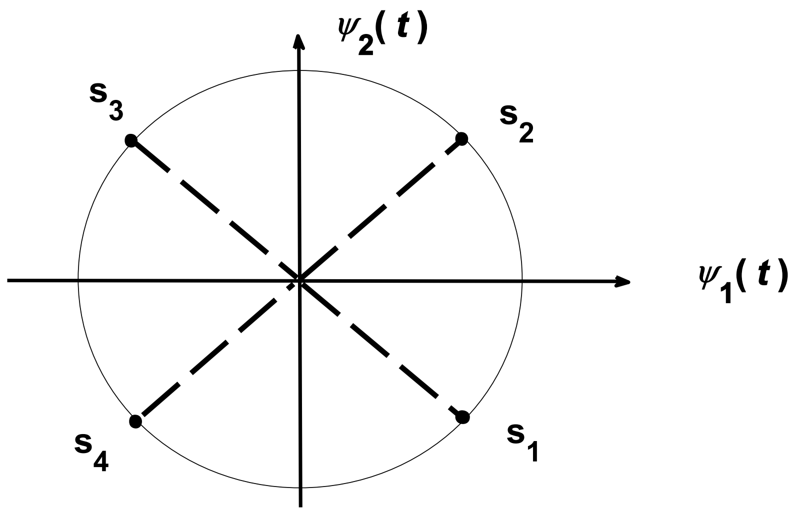


Fig. 3.4 Signal space diagram for 4 PSK signals $s_1(t) \cdots s_4(t)$.

In Fig. 3.4, $M = 4$ (corresponding to 4 level signaling), this PSK scheme is also known as quadrature PSK. In the signal space diagram of Fig. 3.3, we had $M = 2$ (binary). Since in PSK, all signal vectors are of same length (and same energies), it is customary to draw a circle passing through signal end points as indicated in Fig. 3.4. It is of course possible to add more signals to the two dimensional signal space of PSK. For instance, the case of $M = 8$ is shown in Fig. 3.5 where we have removed the connections of signal vector ends to the origin for clarity. It is also possible to go to higher M values. This way, the appearance in the signal space diagrams will turn more into a constellation of stars. For this reason, signal space diagram is also called constellation diagram. For a general M , the *m*th signal waveform $s_m(t)$ and signal vector $\mathbf{s}_m$ from the signal set of $s_1(t) \cdots s_m(t) \cdots s_M(t)$ can be formulated as

$$
\begin{aligned}
& s_m(t) = A\sqrt{T}\left[C_c\psi_1(t) + C_s\psi_2(t)\right] \ , \ C_c = \cos\left[2\pi(m-1)/M\right] \ , \ C_s = \sin\left[2\pi(m-1)/M\right] \\
& \psi_1(t) = \begin{cases} \sqrt{2/T} & \text{for} \quad 0 \le t \le T/2 \\ 0 & \text{elsewhere} \end{cases} \quad \psi_2(t) = \begin{cases} \sqrt{2/T} & \text{for} \quad T/2 \le t \le T \\ 0 & \text{elsewhere} \end{cases} \\
& \mathbf{s}_m = \left\{A\sqrt{T}\cos\left[2\pi(m-1)/M\right] , A\sqrt{T}\sin\left[2\pi(m-1)/M\right]\right\} \ , \ m = 1 \cdots M
\end{aligned} \tag{3.6}
$$

Note that the orthonormalized basis functions defined in (3.6) are the same as $\psi_1^a(t)$ and $\psi_2^a(t)$ of Fig. 3.1.

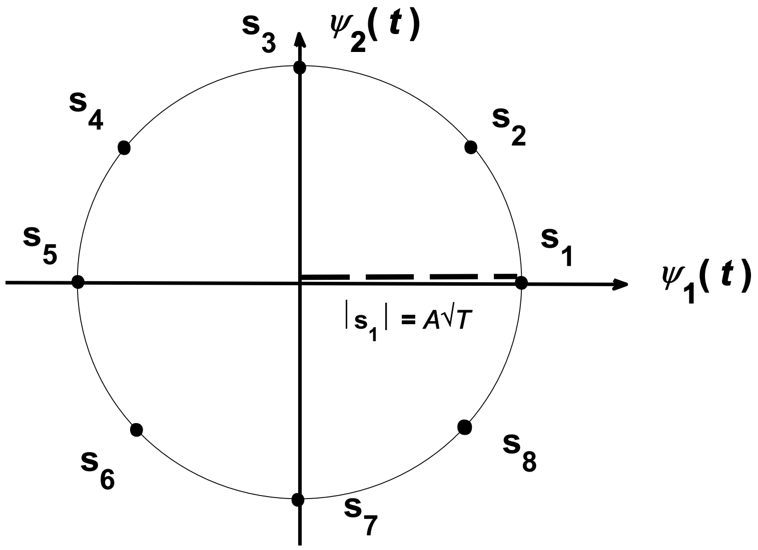


Fig. 3.5 Signal space diagram for 8 PSK signals $s_1(t) \cdots s_8(t)$.

Exercise 3.2 : From (3.6) write for the time signals of $s_1(t) \cdots s_8(t)$, signal vectors of $\mathbf{s}_1 \cdots \mathbf{s}_8$ and see if they agree with the signal constellation of Fig. 3.5. Find the length and the energies of the signal vectors $\mathbf{s}_1 \cdots \mathbf{s}_8$.

An interesting feature of PSK (and also QAM) constellation is that positions of signal vectors can also be represented on a complex plane, since a complex plane is two dimensional as well. With this arrangement, $\psi_1(t)$ will be replaced by the real part of the complex exponential and $\psi_2(t)$ will be replaced by the imaginary part of the same exponential. Thus the signal vector $\mathbf{s}_m$ of *M* ary PSK will become

$$\mathbf{s}_m = \left\{ A\sqrt{T} \exp\left[2\pi j \left(m-1\right)/M\right]\right\} \ , \ m = 1 \cdots M \tag{3.7}$$

It is worthwhile to note that Matlab uses the notation expressed in (3.7).

It is easy to see that even with increasing M , the two dimensional signal space is still not efficiently used. One solution would be to create energy variations as well as phase variations in signal vectors, this way achieve a combination of ASK plus PSK. This combination will be called Quadrature Amplitude Modulation (QAM). An example of 16 QAM is shown in Fig. 3.6. As seen here, QAM constellations are usually arranged in the form of rectangles, although from probability of error view point, this is not the best placement of signal vectors, there is little difference between the rectangular arrangements and the optimum ones. Fig. 3.6 proves that QAM is indeed a combination of ASK and PSK. For instance, in the given constellation of Fig. 3.6, the collection of signal vectors $\mathbf{s}_7,\ \mathbf{s}_6,\ \mathbf{s}_3$ and $\mathbf{s}_2$ constitute 4 ASK, whereas the collection $\mathbf{s}_4,\ \mathbf{s}_5,\ \mathbf{s}_{10}$ and $\mathbf{s}_{15}$ represents 4 PSK. It is equally possible to identify other similar groupings.

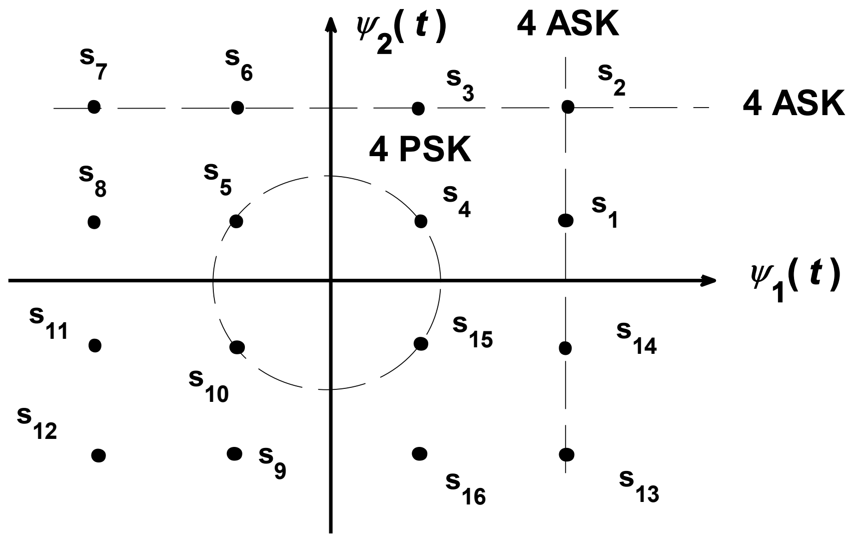


Fig. 3.6 Rectangular signal constellation for 16 QAM signals.

Arrangements other than the rectangular type are also possible in QAM. For instance placing of signal vectors on different circles within each other is another option as illustrated in Fig. 3.7. Of course, the objective here is to find the constellation (distribution of signal vectors) that will give the maximum distances between vector ends for the same total or average energy, since it this criteria which will determine the probability of error performance.

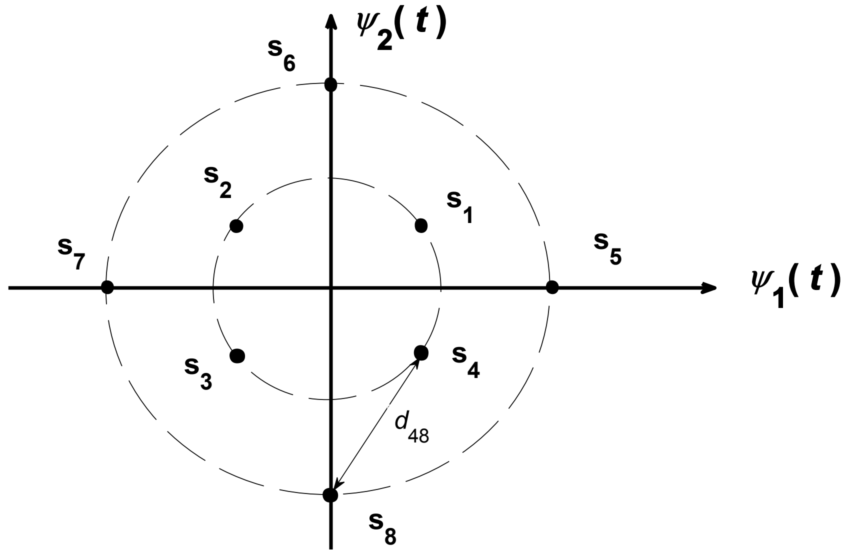


Fig. 3.7 Circular signal constellation for 8 QAM signals.

In practice QAM is used mostly in radio links.

4. Multidimensional Signals

With the context of multidimensional signals, here we will only study Frequency Shift Keying (FSK). Although the other modulation types can be represented both in baseband (without carrier) and bandpass (with carriers), FSK can only be written in terms of sinusoidal carriers. Assume that we choose $M = 2$ and assign frequencies f_1 and f_2 to our message signals of $s_1(t)$ and $s_2(t)$, then

$$s_1(t) = \sqrt{\frac{2\varepsilon_b}{T_b}} \cos(2\pi f_1 t) \;\; , \;\; 0 \le t \le T_b \;\; , \;\; T_b : \text{ Binary waveform duration}$$

$$s_2(t) = \sqrt{\frac{2\varepsilon_b}{T_b}} \cos(2\pi f_2 t) \;\; , \;\; 0 \le t \le T_b \;\; , \;\; \varepsilon_b = \int_0^{T_b} s_1^2(t)\, dt = \int_0^{T_b} s_2^2(t)\, dt \qquad (4.1)$$

Note that in the writing of signal waveforms, we have used slightly different notation than ASK and PSK. By setting $\Delta f = f_m - f_{m-1}$, adopting a starting frequency of f_c such that $f_m = f_c + (m-1)\Delta f$ we can write for the *m* th signal as follows

$$s_m(t) = \sqrt{\frac{2\varepsilon_s}{T}} \cos\left[2\pi f_c t + 2\pi\Delta f (m-1)t\right] \;\; , \;\; 0 \le t \le T$$

$$\varepsilon_s = \int_0^T s_m^2(t)\, dt = k\varepsilon_b = \varepsilon_b \log_2 M \;\; , \;\; m = 1 \cdots M \;\; , \;\; T = kT_b \qquad (4.2)$$

It is interesting to examine the variation of Δf (frequency separation) against T (symbol duration). To this end we define the correlation coefficient γ_{mn} as follows

$$
\begin{aligned}
\gamma_{mn} &= \frac{1}{\varepsilon_s}\int_0^T s_m(t)\, s_n(t)\ dt \\
&= \frac{1}{\varepsilon_s}\int_0^T \frac{2\varepsilon_s}{T}\cos\left[2\pi f_c t + 2\pi\Delta f\,(m-1)t\right]\ \cos\left[2\pi f_c t + 2\pi\Delta f\,(n-1)t\right]\ dt \\
&= \frac{1}{T}\int_0^T \cos\left[2\pi f_c t + 2\pi\Delta f\,(m-n)t\right]\ dt + \frac{1}{T}\int_0^T \cos\left[2\pi f_c t + 2\pi\Delta f\,(m+n-2)t\right]\ dt \\
&\simeq \frac{\sin\left[2\pi\Delta f\,(m-n)T\right]}{2\pi\Delta f\,(m-n)T}
\end{aligned} \tag{4.3}
$$

where on the second line, we have substitutions from (4.1) and the approximation on the line is due to $f_c \gg 1/T$. A plot of γ_{mn} against Δf is given in Fig. 4.1. As seen from this figure, γ_{mn} passes through zero at integer multiples of $1/2T$. It means at this values of Δf, the signals $s_m(t)$ and $s_n(t)$ are orthogonal. The minimum value of γ_{mn} is -0.217 and reached at $\Delta f = 0.715/T$. These markings are important and form the basis of Orthogonal Frequency Division Multiplexing (OFDM).

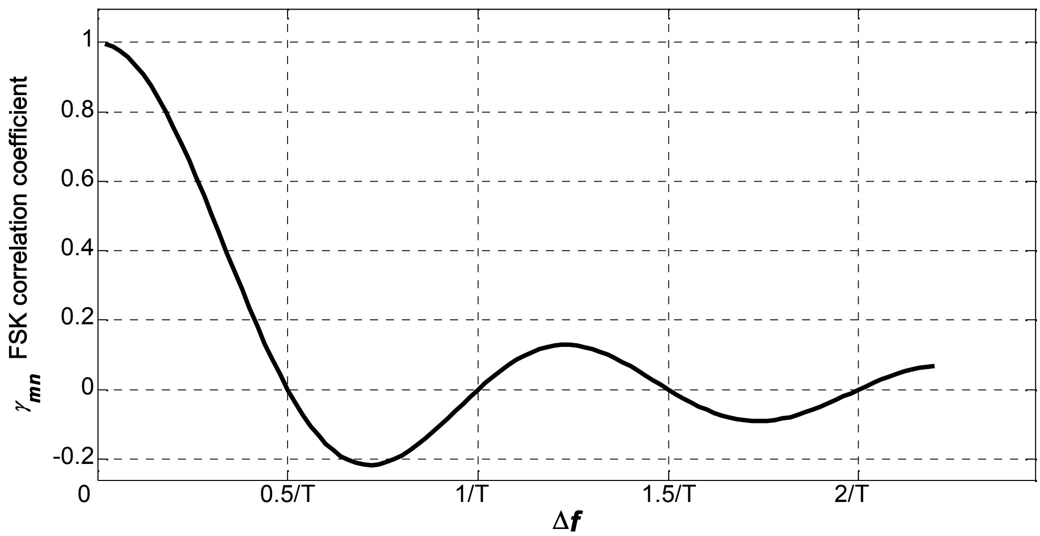


Fig. 4.1 The variation of FSK correlation coefficient γ_{mn} against Δf.

Exercise 4.1 : Write the mathematical expressions for FSK time signals and signal vectors for $M = 8$.

5. Detection of Signal in Presence of Additive White Gaussian Noise – Correlators and Matched Filters

We assume that within a time interval of $0 \le t \le T$, our transmitter randomly sends one of the $s_1(t) \cdots s_M(t)$ signals, namely $s_m(t)$ and in the communication channel, only additive white Gaussian noise (AWGN) is added to the signal, so that the received signal $r(t)$ is

$$r(t) = s_m(t) + n(t) \quad , \quad S_n(f) = \frac{N_0}{2} = \sigma_n^2 \tag{5.1}$$

where $S_n(f)$ is known as noise spectral density function, N_0 and σ_n^2 are noise spectral density level and noise variance. It is obvious that $S_n(f)$ is independent of frequency f, hence the White nature of Gaussian noise.

Such a channel model is known as (band unlimited) AWGN channel and depicted in Fig. 5.1. It is clear that in this channel, there is no band limitation, which means that

$$C(f) = 1 \text{ , } c(t) = \delta(t) \text{ , } \delta(t) \text{ : Time delta function} \tag{5.2}$$

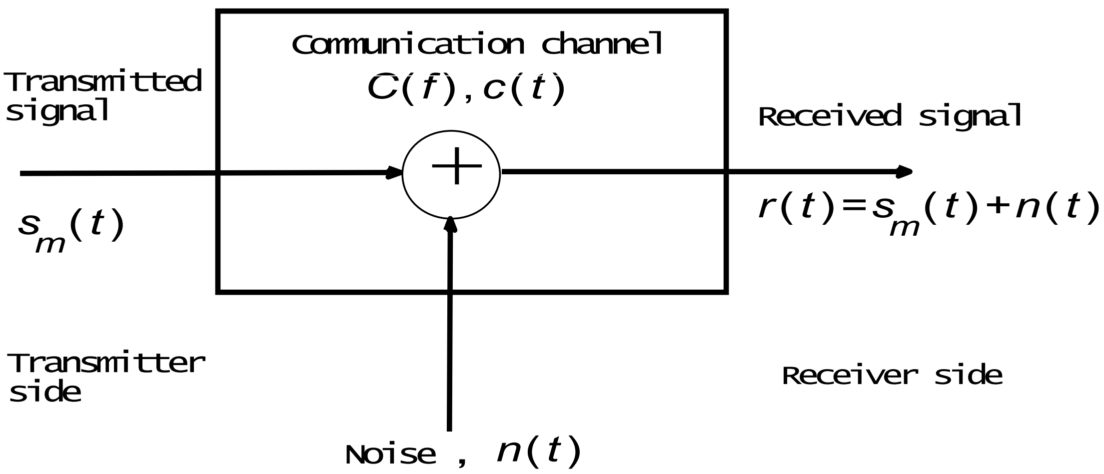


Fig. 5.1 AWGN channel model.

The receiver has a knowledge of modulation type and symbol duration, T that are employed at the transmitter. Furthermore the receiver also knows the set of signals $s_1(t) \cdots s_M(t)$, i.e. the alphabet used by the transmitter. Finally we assume that the receiver is able to extract the beginning of time interval $0 \le t \le T$, called synchronization. So the job of the receiver is to demodulate the incoming signal $r(t)$ and decide correctly which $s_m(t)$ was sent from the transmitter within the time interval $0 \le t \le T$. Note that since we are dealing with an unlimited channel, it is sufficient to consider any symbol interval. Here we choose, the interval, $0 \le t \le T$. To perform demodulation tasks, we pass the received signal through a correlator as shown in Fig. 5.2. Basically the operations performed in the correlator of Fig. 5.2 are feeding the received signal simultaneously to N branches, multiplying the received signal $r(t)$ on each branch by one of the orthonormal basis functions $\psi_1(t) \cdots \psi_N(t)$ (the same ones used at transmitter to construct the signal $s_m(t)$), integrating the resultant over one symbol duration, and sampling at the end of this duration, forming the vector array $\mathbf{r}$ by collecting the individual components $r_1 \cdots r_N$, eventually sending $\mathbf{r}$ to a detector to decide which $s_m(t)$ was sent from the transmitter. The operations carried out on the *n* th branch of this correlator can mathematically be described as

$$\int_0^T r(t)\ \psi_n(t)\ dt = \int_0^T \left[s_m(t)+n(t)\right]\ \psi_n(t)\ dt$$

$$r_n = s_{mn} + n_n \quad n = 1 \cdots N$$

$$s_{mn} = \int_0^T s_m(t)\ \psi_n(t) dt \quad , \quad n_n = \int_0^T n(t)\ \psi_n(t) dt \qquad (5.3)$$

Note that the definition of s_{mn} is equivalent to the one in (1.13).

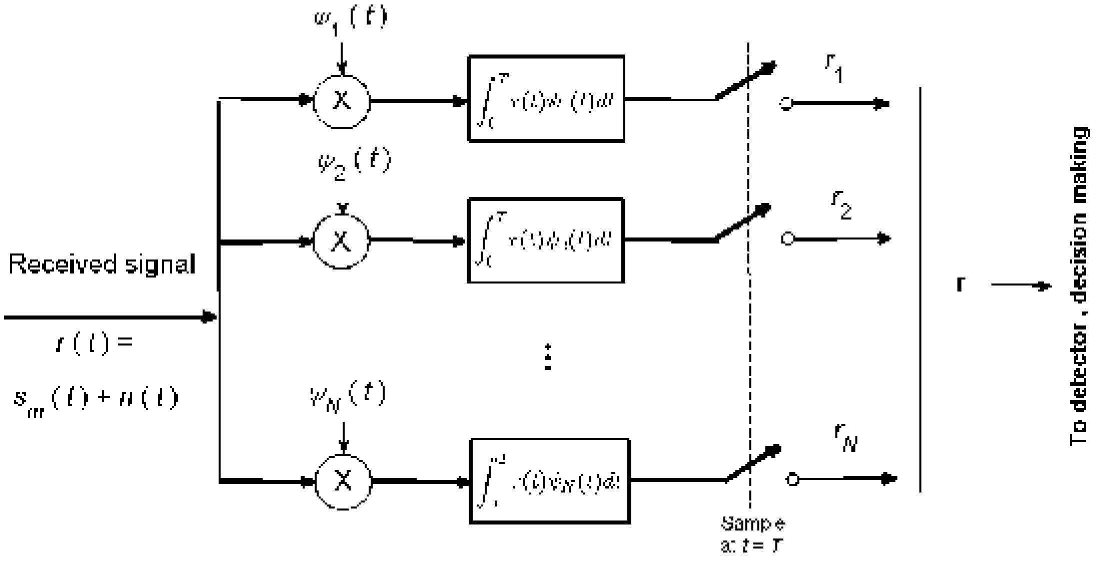


Fig. 5.2 Correlator type of demodulator.

Considering all branches in the correlator of Fig. 5.2, we get

$$\mathbf{r} = \mathbf{s}_m + \mathbf{n} \qquad (5.4)$$

(5.4) means the totality of the operations in the correlator can be treated as row arrays of $\mathbf{r}$, $\mathbf{s}_m$ and $\mathbf{n}$, where

$$\mathbf{r} = \left[r_1 \cdots r_n \cdots r_N\right] \quad , \quad \mathbf{s}_m = \left[s_{m1} \cdots s_{mn} \cdots s_{mN}\right] \quad , \quad \mathbf{n} = \left[n_1 \cdots n_n \cdots n_N\right] \qquad (5.5)$$

It is important to point out that $\mathbf{s}_m$ is deterministic in the sense that it will take upon one of the values from the set $\mathbf{s}_1 \cdots \mathbf{s}_M$, while $\mathbf{n}$ is random. The probability density function (pdf) for the amplitude distribution of one n_n sample of $\mathbf{n}$ is the same as input noise, hence

$$f(n_n) = \frac{1}{(\pi N_0)^{0.5}} \exp\left(-\frac{n_n^2}{N_0}\right) \quad , \quad N_0 = 2\sigma_n^2 \quad , \quad n = 1 \cdots N \qquad (5.6)$$

where $N_0 / 2$ is the noise spectral density and σ_n^2 is the variance. All n_n samples have zero mean and are uncorrelated, which means

$$E[n_n]=\int_0^T E[n(t)]\ \psi_n(t)dt=0$$

$$
\begin{aligned}
E[n_n n_m] &= \int_0^T\int_0^T E[n(t)n(\tau)]\ \psi_n(t)\psi_m(\tau)dtd\tau \\
&= \int_0^T\int_0^T \frac{N_0}{2}\delta(t-\tau)\ \psi_n(t)\psi_m(\tau)dt \\
&= \frac{N_0}{2}\int_0^T \psi_n(t)\psi_m(t)dt \\
&= \frac{N_0}{2}\delta_{mn}\ ,\ \ \delta_{mn}=0\ \text{ if }\ n\neq m\ \ ,\ \ \delta_{mn}=1\ \text{ if }\ n=m
\end{aligned}
\qquad (5.7)
$$

As a consequence of (5.5) and (5.6)

$$
\begin{aligned}
f(\mathbf{n}) &= \prod_{n=1}^{N} f(n_n)=\frac{1}{(\pi N_0)^{N/2}}\exp\left(-\sum_{n=1}^{N}\frac{n_n^2}{N_0}\right) \\
E[r_n] &= E[s_{mn}+n_n]=E[s_{mn}]+E[n_n]=s_{mn}+0=s_{mn} \\
f(\mathbf{r}\mid\mathbf{s}_m) &= \prod_{n=1}^{N} f(r_n\mid s_{mn})\qquad ,\qquad m=1\cdots M \\
f(r_n\mid s_{mn}) &= \frac{1}{(\pi N_0)^{0.5}}\exp\left[-(r_n-s_{mn})^2/N_0\right]\qquad ,\qquad n=1\cdots N \\
f(\mathbf{r}\mid\mathbf{s}_m) &= \frac{1}{(\pi N_0)^{N/2}}\exp\left[-\sum_{n=1}^{N}\frac{(r_n-s_{mn})^2}{N_0}\right] \\
&= \frac{1}{(\pi N_0)^{N/2}}\exp\left[-\frac{\|\mathbf{r}-\mathbf{s}_m\|^2}{N_0}\right]\qquad ,\qquad m=1\cdots M
\end{aligned}
\qquad (5.8)
$$

The development in (5.8) means that (vectorwise) when noise $\mathbf{n}$ is added to the incoming signal $\mathbf{s}_m$, then the received signal $\mathbf{r}$ becomes a Gaussian random variable as well. This way the received signal inherits all properties of noise, except that the previous zero mean is now shifted to $\mathbf{s}_m$. In a way, this is like adding a DC shift ($\mathbf{s}_m$) to an AC signal $\mathbf{n}$. Note that in (5.5) noise vector $\mathbf{n}$ is shown to be N dimensional. This is because the correlator of Fig. 5.2 takes the projections of noise signal $n(t)$ onto an N dimensional space. Prior to such an operation, noise signal $n(t)$ (i.e. the noise in nature) has infinite number of dimensions.

It is also possible to perform the demodulation via match filter type of demodulator. This is shown in Fig. 5.3. As seen in Fig. 5.3, the previous acts of multiplying the received signal by orthonormal basis functions and the integrating the resulting product are concentrated in the boxes known as matched filters (MFs). This can be proven mathematically by writing the output from MF *n* th branch on the as follows.

$$y_n(t) = \int_0^t r(\tau)\, h_n(t-\tau)\, d\tau$$
$$= \int_0^t r(\tau)\, \psi_n(T-t+\tau)\, d\tau \quad , \quad n = 1 \cdots N \qquad (5.9)$$

After sampling at $t = T$, we obtain

$$y_n(T) = \int_0^T r(\tau)\, \psi_n(\tau)\, d\tau \quad , \quad n = 1 \cdots N \qquad (5.10)$$

(5.10) will deliver the same result as the first line of the integral in (5.3).

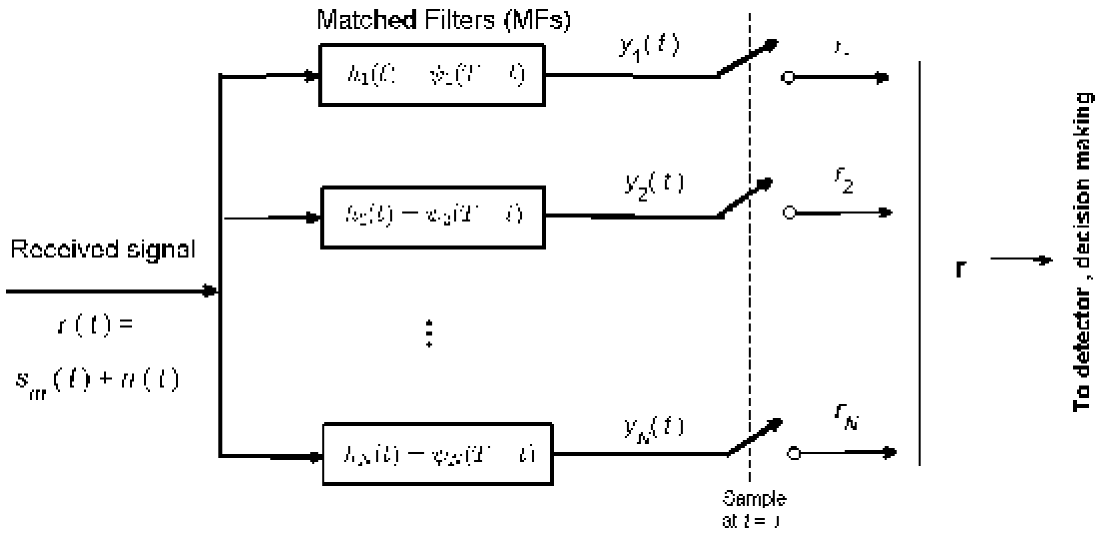
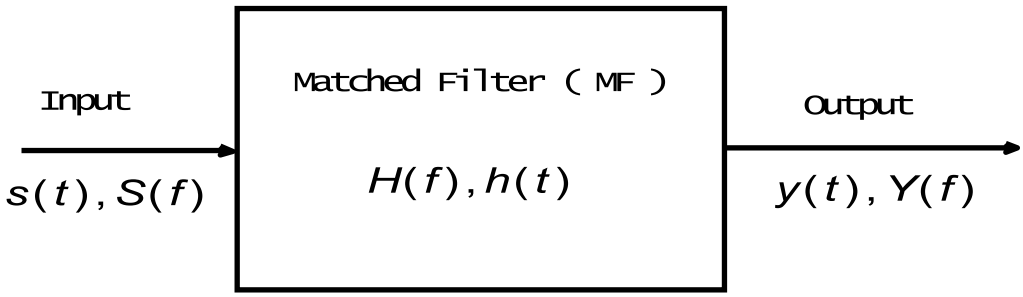


Fig. 5.3 Matched Filter (MF) type of demodulator.

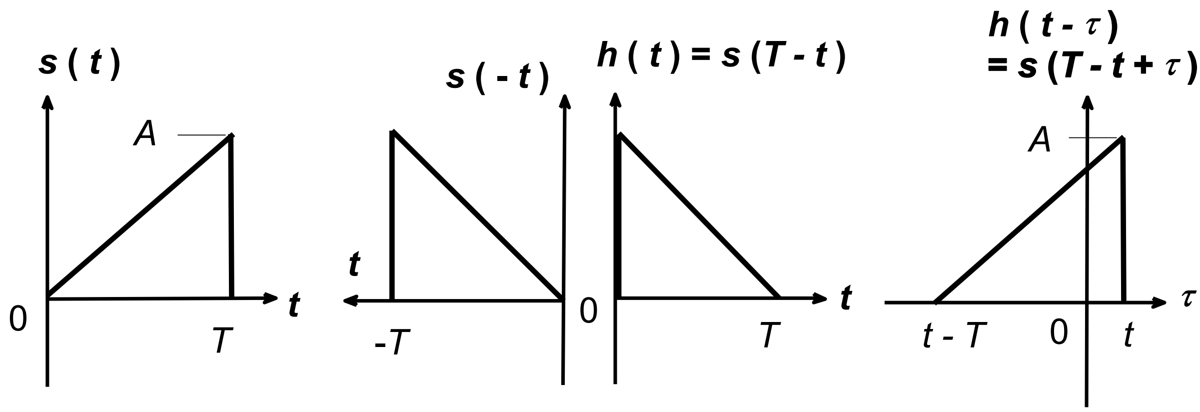
It is instructive to examine the time domain properties of MF. The impulse response of a filter matched to an input signal of $s(t)$ is given by $h(t) = s(T-t)$. Then the response from such a filter would be

$$y(t) = \int_0^t s(\tau)\, h(t-\tau)\, d\tau = \int_0^t s(\tau)\, s(T-t+\tau)\, d\tau \qquad (5.11)$$

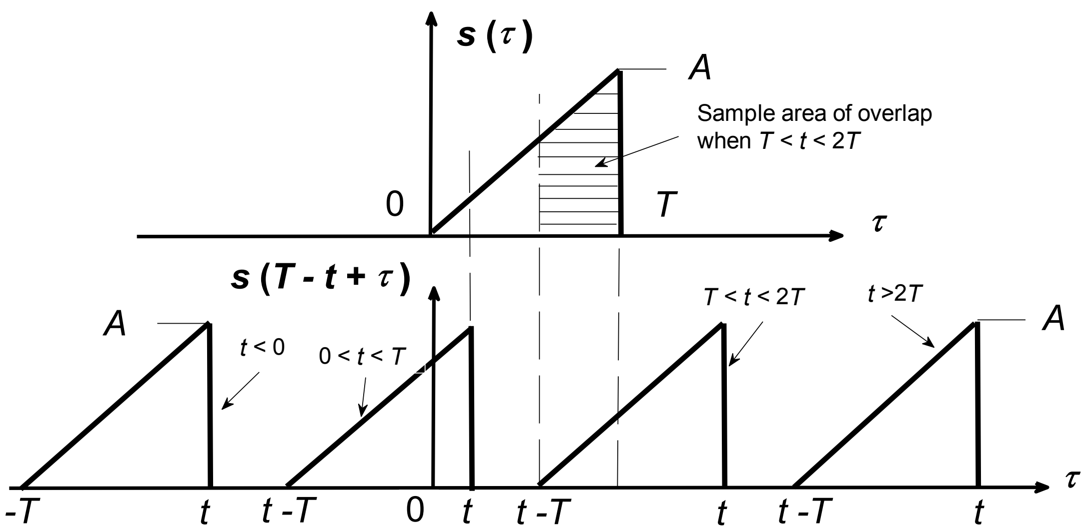
So the output from the matched filter can be interpreted as the time autocorrelation function of the input signal $s(t)$. An example input and output of MF are given in Fig. 5.4. As seen from Fig. 5.4c, the match filter response inside the convolution integral becomes oriented in the same direction as the input so that maximum similarity (correlation) is established at the output at the instance of $t = T$.



a) Block diagram of matched filter (MF)



b) Orientation of input signal to matched filter (MF)



c) Output from matched filter (MF) via convolution

Fig. 5.4 Block diagram, orientation of input signal and obtaining output signal from MF via convolution.

To find the output of MF for the input given in Fig. 5.4, we need the mathematical expressions of $s(\tau)$ and $s(T-t+\tau)$ as seen for the integration in (5.11). These are

$$s(\tau) = A\frac{\tau}{T} \quad , \quad s(T-t+\tau) = -A\frac{\tau - T}{T} + A \tag{5.12}$$

Again looking at Fig. 5.4 c), we identify four different regions of integration for the expression of (5.11) which are

$$y(t) = \begin{cases} y_1(t) = 0 & t < 0 \\ y_2(t) = \int\limits_0^t s(\tau)\, s(T-t+\tau)\, d\tau & 0 < t < T \\ y_3(t) = \int\limits_{t-T}^T s(\tau)\, s(T-t+\tau)\, d\tau & T < t < 2T \\ y_4(t) = 0 & t > 2T \end{cases} \tag{5.13}$$

Note that on the first and last lines of (5.13), the result is zero because the is no overlap between $s(\tau)$ and $s(T-t+\tau)$, while on the second and third lines, the integrand is the same as expected, but the integration limits are adjusted according to the areas of overlap. After using (5.12) in (5.13), we get

$$y(t) = \begin{cases} y_1(t) = 0 & t < 0 \\ y_2(t) = -\frac{A^2 t^3}{6T^2} + \frac{A^2 t^2}{2T} & 0 < t < T \\ y_3(t) = \frac{A^2}{6T^2}(t-T)^3 - \frac{A^2 t}{2} + \frac{5A^2 T}{6} & T < t < 2T \\ y_4(t) = 0 & t > 2T \end{cases} \tag{5.14}$$

For the results (meaning the second and third lines) in (5.14), the tests conducted at the check points $t = 0,\ t = T$ and $t = 2T$ are given in (5.15). Note that these tests, of course, do not guarantee the absolute correctness of the formulations given for $y_2(t)$ and $y_3(t)$ in (5.14). We show the plot of $y(t)$ in Fig. 5.5.

$$y_2\left(t=0\right)=-\frac{A^2\left(t=0\right)^3}{6T^2}+\frac{A^2\left(t=0\right)^2}{2T}=0 \qquad \text{Test for } t=0:\text{OK}$$

$$y_2\left(t=T\right)=-\frac{A^2\left(t=T\right)^3}{6T^2}+\frac{A^2\left(t=T\right)^2}{2T}=\frac{A^2T}{3}=\int_0^T s^2\left(t\right)\,dt=\varepsilon_s \qquad \text{Test for } t=T:\text{OK}$$

$$y_3\left(t=T\right)=\frac{A^2}{6T^2}\left(t=T-T\right)^3-\frac{A^2\left(t=T\right)}{2}+\frac{5A^2T}{6}=\frac{A^2T}{3}=\varepsilon_s \qquad \text{Test for } t=T:\text{OK}$$

$$y_3\left(t=2T\right)=\frac{A^2}{6T^2}\left(t=2T-T\right)^3-\frac{A^2\left(t=2T\right)}{2}+\frac{5A^2T}{6}=\frac{A^2T}{3}=0 \qquad \text{Test for } t=2T:\text{OK}$$

(5.15)

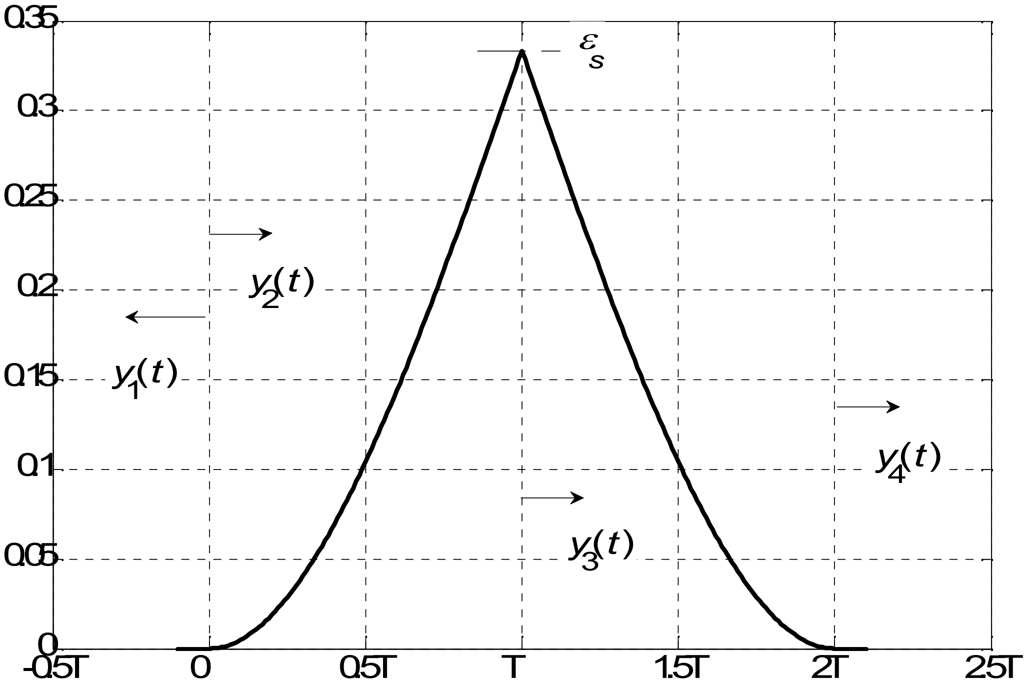


Fig. 5.5 The output from MF, when the input is as shown in Fig. 5.4.

As seen from Fig. 5.5, the peak of $y\left(t\right)$ occurs at $t=T$. This explains also why we have chosen the sampling instance $t=T$ for correlator in Fig. 5.2 and matched filter in Fig. 5.3. It is important to note that the output of MF comes out to be in units of energy, whereas we expect it to be in units of voltage (or current). To correct this, we have to divide the output of MF by the square root of energy of input.

Finally we examine the frequency domain interpretation of MF. To this end we assume that for a signal of $s\left(t\right)$ the time response of matched filter is given by $h\left(t\right)=s\left(T-t\right)$, thus the Fourier transform of $s\left(T-t\right)$ will be

$$
\begin{aligned}
H(f) &= \int_0^T s(T-t)\ \exp(-2j\pi ft)\ dt \\
&= \left[\int_0^T s(\tau)\ \exp(2j\pi f\tau)\ d\tau\right]\exp(-2j\pi fT) \\
&= S^*(f)\exp(-2j\pi fT) \qquad (5.16)
\end{aligned}
$$

The last line in (5.16) means that the frequency response of MF is equal to the multiplication of the complex conjugate of the frequency response of the input signal and phase factor $\exp(-2j\pi fT)$, representing the time delay of T in $s(T-t)$. The output from MF in will then be

$$
\begin{aligned}
y(t) &= \int_{-\infty}^{\infty} Y(f)\ \exp(2j\pi ft)\ df \\
&= \int_{-\infty}^{\infty} |S(f)|^2\ \exp(-2j\pi fT)\exp(2j\pi ft)\ df \\
y(t=T) &= \int_{-\infty}^{\infty} |S(f)|^2\ df = \int_{-\infty}^{\infty} s^2(t)\ dt = \varepsilon_s \qquad (5.17)
\end{aligned}
$$

where on the last line we have taken into account the sampling at the instance of $t=T$. Then we have used Perseval's relation to establish the energy equivalence of a (time limited) signal along time and frequency axis. Since it is a bit awkward to find the output of MF in units of energy, usually we scale the response of MF by the square root of energy. Note that such a scaling is already accounted for in the orthonormalized functions of $\psi_1(t)\cdots\psi_N(t)$.

The last line of (5.17) gives the signal output as amplitude, thus its square will give the output power, that is

$$
P_s = y^2(T) = \varepsilon_s^2 \qquad (5.18)
$$

The noise with a spectral density of $S_n(f) = N_0/2$, when fed to an MF whose frequency response is $H(f) = S^*(f)\exp(-2j\pi fT)$ will deliver spectral density output of

$$
S_0(f) = |H(f)|^2 S_n(f) = |S(f)|^2 N_0/2 \qquad (5.19)
$$

As a result, nose power at the output of MF will be

$$
P_n = \int_{-\infty}^{\infty} S_0(f)\ df = \frac{N_0}{2}\int_{-\infty}^{\infty} |S(f)|^2\ df = \frac{\varepsilon_s N_0}{2} \qquad (5.20)
$$

By using (5.18) and (5.20), we can calculate the signal to noise ratio (SNR) at the output as follows

$$
\mathrm{SNR} = \frac{P_s}{P_n} = \frac{\varepsilon_s^2}{\varepsilon_s N_0/2} = \frac{2\varepsilon_s}{N_0} \qquad (5.21)
$$

6. Optimum Detector

By optimum detector, we mean a detector that makes best use of the received (statistical) information and establishes a correct decision as far as possible. As seen from Figs. 5.2 and 5.3, the received to be used by the optimum detector is $\mathbf{r} = [r_1; \cdots ; r_N]$. From (5.8), we know that $\mathbf{r}$ is a Gaussian random variable (a property inherited from noise) with a mean of $\mathbf{s}_m$ (a property inherited from the transmitted signal). Vectorwise $\mathbf{r} = \mathbf{s}_m + \mathbf{n}$. So if the possible number of transmitted signals was $M = 4$ and the dimensionality of the signal space was $N = 3$, then the appearance of the signal space diagram would be something like that shown in Fig. 6.1 assuming $s_1(t)$ was transmitted. Here the cloud around each signalling point represents the spherical (since $N = 3$) noise observed after may receptions. At a given instance of time received vector $\mathbf{r}$ would be as shown.

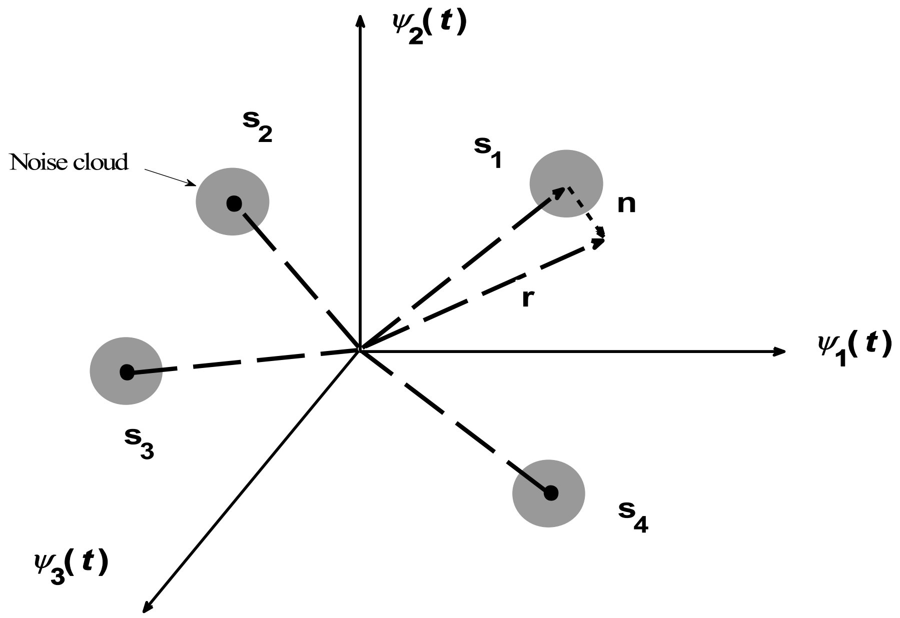


Fig. 6.1 The appearance of signal space after AWGN channel.

Now we aim for an optimum detector that will make a decision based on the computation of the posterior probability defined as

$$P(\text{signal } \mathbf{s}_m \text{was transmitted} | \mathrm{r}) = P(\mathbf{s}_m | \mathrm{r}) \tag{6.1}$$

Our criteria will be to find m value that will maximize $P(\mathbf{s}_m | \mathrm{r})$, when m ranges in $m = 1 \cdots M$. Upon finding the m value that has maximized $P(\mathbf{s}_m | \mathrm{r})$, we arrive at the decision that it is most likely that this particular $\mathbf{s}_m$ was transmitter. So our optimum decision rule boils down to evaluating $P(\mathbf{s}_m | \mathrm{r})$ and is named as maximum a posteriori probability (MAP) criterion.

Using Bayes rule, we can express $P\left(\mathbf{s}_m | \mathrm{r}\right)$ as

$$P\left(\mathbf{s}_m | \mathrm{r}\right) = \frac{f\left(\mathbf{r} | \mathbf{s}_m\right) P\left(\mathbf{s}_m\right)}{f\left(\mathbf{r}\right)} \tag{6.2}$$

where $f\left(\mathbf{r} | \mathbf{s}_m\right)$ is the conditional pdf of $\mathbf{r}$ given that $\mathbf{s}_m$ was transmitted. $P\left(\mathbf{s}_m\right)$ is the probability that signal $\mathbf{s}_m$ was transmitted. $f\left(\mathbf{r}\right)$ in the denominator of (6.2) is the pdf vector $\mathbf{r}$ and will be given by the following sum

$$f\left(\mathbf{r}\right) = \sum_{m=1}^{M} f\left(\mathbf{r} | \mathbf{s}_m\right) P\left(\mathbf{s}_m\right) \tag{6.3}$$

We can take the formulation of $f\left(\mathbf{r} | \mathbf{s}_m\right)$ from (5.8), but even then, it is not possible to arrive at a simplified expression of $P\left(\mathbf{s}_m | \mathrm{r}\right)$, since individual probability of signal sent from the transmitter may be different. If however the transmitter sends all $\mathbf{s}_m$ signals, $m = 1 \cdots M$ with equal probability, then

$$P\left(\mathbf{s}_m\right) = \frac{1}{M} \tag{6.4}$$

As a result

$$\mathrm{Max}\left[P\left(\mathbf{s}_m | \mathrm{r}\right)\right] = \mathrm{Max}\left[\frac{f\left(\mathbf{r} | \mathbf{s}_m\right)}{\sum_{m=1}^{M} f\left(\mathbf{r} | \mathbf{s}_m\right)}\right] \equiv \mathrm{Max}\left[f\left(\mathbf{r} | \mathbf{s}_m\right)\right] \tag{6.5}$$

The reason that we have been able to write the last expression in (6.5) is that the sum in the denominator of the middle expression remains the same whichever m is selected. Therefore this sum has no role in the determination of $\mathrm{Max}\left[P\left(\mathbf{s}_m | \mathrm{r}\right)\right]$. So finding $\mathrm{Max}\left[P\left(\mathbf{s}_m | \mathrm{r}\right)\right]$ is equivalent to finding $\mathrm{Max}\left[f\left(\mathbf{r} | \mathbf{s}_m\right)\right]$, such a reduced decision strategy is called maximum likelyhood (ML) criterion. From (5.8) we see that $f\left(\mathbf{r} | \mathbf{s}_m\right)$ contains a Gaussian exponential, thus it may be easier to work with the $\log_e$ (denoted by $\ln$), hence

$$\mathrm{Max}\left\{\ln\left[f\left(\mathbf{r} | \mathbf{s}_m\right)\right]\right\} = \mathrm{Max}\left[\frac{-N}{2}\ln\left(\pi N_0\right) - \frac{1}{N_0}\sum_{n=1}^{N}\left(r_n - s_{mn}\right)^2\right] \tag{6.6}$$

We note that terms that do not contain the index m are irrelevant in the maximizing process, therefore, we take the last term in (6.6) and set it to a distance metrics, $D\left(\mathbf{r}, \mathbf{s}_m\right)$

$$D\left(\mathbf{r}, \mathbf{s}_m\right) = \sum_{n=1}^{N}\left(r_n - s_{mn}\right)^2 \tag{6.7}$$

Because of the minus sign in front of the sum in (6.6), seeking $\mathrm{Max}\left[P\left(\mathbf{s}_m|\mathrm{r}\right)\right]$ will now be equivalent to $\mathrm{Min}\left[D\left(\mathbf{r},\mathbf{s}_m\right)\right]$. As the name implies and as also detected from (6.6), when the index m is run from 1 to M, the distance metrics $D\left(\mathbf{r},\mathbf{s}_m\right)$ will calculate one by one the distances of all signals that are likely to be sent from the transmitter to the received vector $\mathbf{r}$. In the end by selecting $\mathrm{Min}\left[D\left(\mathbf{r},\mathbf{s}_1\right),\ D\left(\mathbf{r},\mathbf{s}_2\right),\ \cdots D\left(\mathbf{r},\mathbf{s}_M\right)\right]$, we shall have arrived at the optimum decision based on MAP criterion. Such an operation is carried out for the sample constellation of Fig. 6.1 and illustrated in Fig. 6.2.

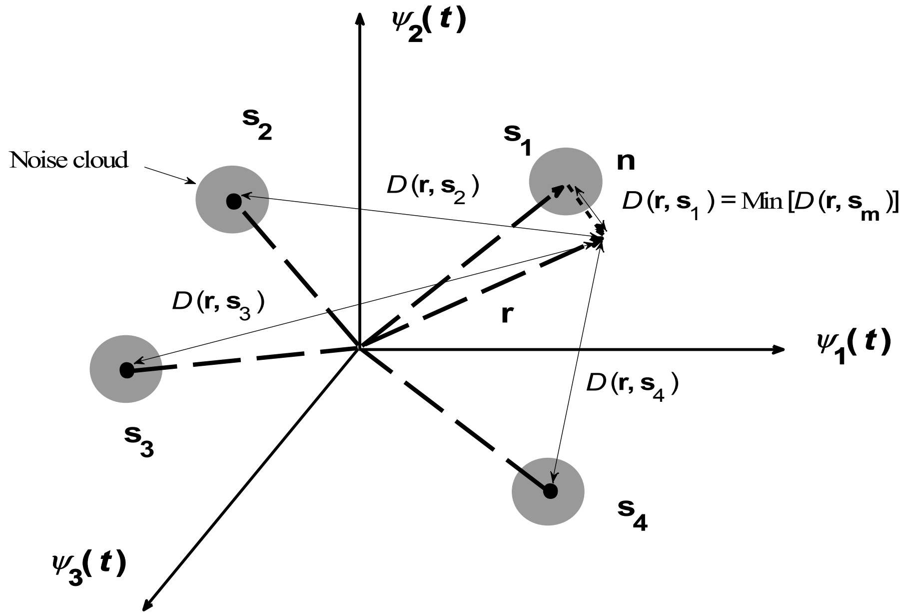


Fig. 6.2 Calculation of distance metrics for the sample constellation of Fig. 6.1.

As clearly seen from Fig. 6.2, this operation will definitely to the correct decision, so long as the noise $\mathbf{n}$ added to $\mathbf{s}_1$ (remember that we have already assumed that $s_1\left(t\right)$ was sent from the transmitter) is at the amplitude and angle as shown.

Expanding (6.7), we get

$$\begin{aligned} D\left(\mathbf{r},\mathbf{s}_m\right) &= \sum_{n=1}^{N} r_n^2 - 2\sum_{n=1}^{N} r_n s_{mn} + \sum_{n=1}^{N} s_{mn}^2 \\ &= \left\|\mathbf{r}\right\|^2 - 2\,\mathbf{r}\cdot\mathbf{s}_m + \left\|\mathbf{s}_m\right\|^2 \\ D^a\left(\mathbf{r},\mathbf{s}_m\right) &= -2\,\mathbf{r}\cdot\mathbf{s}_m + \left\|\mathbf{s}_m\right\|^2 \quad , \quad C\left(\mathbf{r},\mathbf{s}_m\right) = 2\,\mathbf{r}\cdot\mathbf{s}_m - \left\|\mathbf{s}_m\right\|^2 \end{aligned} \tag{6.8}$$

On the second line of (6.8), we have reverted to vectorial notation, on the third line in the definition of $D\left(\mathbf{r},\mathbf{s}_m\right)$, we have dropped $\left\|\mathbf{r}\right\|^2$, since it is common to all calculations of distance metrics, hence no effect on the result, thus have defined a new function called $D^a\left(\mathbf{r},\mathbf{s}_m\right)$. Finally on the third line of (6.8) we have introduced correlation metrics $C\left(\mathbf{r},\mathbf{s}_m\right)$ which is the negative $D^a\left(\mathbf{r},\mathbf{s}_m\right)$. Since in the search of $\mathrm{Max}\left[P\left(\mathbf{s}_m|\mathrm{r}\right)\right]$, we opted to seek $\mathrm{Min}\left[D(\mathbf{r},\mathbf{s}_m)\right]$ and $C(\mathbf{r},\mathbf{s}_m)$ is opposite sign to $D(\mathbf{r},\mathbf{s}_m)$ and consequently $D^a(\mathbf{r},\mathbf{s}_m)$, searching for $\mathrm{Max}\left[P\left(\mathbf{s}_m|\mathrm{r}\right)\right]$ (i.e. applying the rule of MAP criterion) must be equivalent to

$$\mathrm{Max}\left[P\left(\mathbf{s}_m|\mathrm{r}\right)\right]\equiv\mathrm{Min}\left[D(\mathbf{r},\mathbf{s}_m)\right]\equiv\mathrm{Min}\left[D^a(\mathbf{r},\mathbf{s}_m)\right]\equiv\mathrm{Max}\left[C(\mathbf{r},\mathbf{s}_m)\right] \quad (6.9)$$

(6.9) means that our optimum detection rule is simply finding the distances between the received vector $\mathbf{r}$ and all possible signals transmitted, $\mathbf{s}_1\cdots\mathbf{s}_M$ and deciding on $\mathbf{s}_m$ which gives the minimum distance, i.e. $\mathrm{Min}\left[D^a(\mathbf{r},\mathbf{s}_m)\right]$ or finding the correlation between the received vector $\mathbf{r}$ and all possible signals transmitted , $\mathbf{s}_1\cdots\mathbf{s}_M$ and deciding on the one which gives the maximum correlation, i.e. $\mathrm{Max}\left[C(\mathbf{r},\mathbf{s}_m)\right]$.

The above development is valid for the situation when all signals are sent from transmitter with equal probability. If this is not the case, then we go back to (6.2) and (6.3) and keep in mind that the pdf function $f(\mathbf{r})$ is a sum that remains constant whichever m is chosen, thus has no effect on maximization process. Under these circumstances, $\mathrm{Max}\left[P\left(\mathbf{s}_m|\mathrm{r}\right)\right]$ will become

$$\mathrm{Max}\left[P\left(\mathbf{s}_m|\mathrm{r}\right)\right]\equiv\mathrm{Max}\left[f\left(\mathbf{r}\mid\mathbf{s}_m\right)P\left(\mathbf{s}_m\right)\right] \quad (6.10)$$

It is clear that by the application of $D^a(\mathbf{r},\mathbf{s}_m)$ or $C(\mathbf{r},\mathbf{s}_m)$ we start to define an area of (correct) decision region for the signal $s_m(t)$ which we denote by R_m. Then, the probability of error $P_e(\mathbf{s}_m)$ for the signal $\mathbf{s}_m$ will be given by the integration of $f(\mathbf{r}\mid\mathbf{s}_m)$ over the entire area excluding the one belonging to R_m. This area is denoted by R_m^c. Then

$$P_e(\mathbf{s}_m)=\int_{R_m^c} f\left(\mathbf{r}\mid\mathbf{s}_m\right)d\mathbf{r} \quad (6.11)$$

The average probability of error over the total of M signals will be

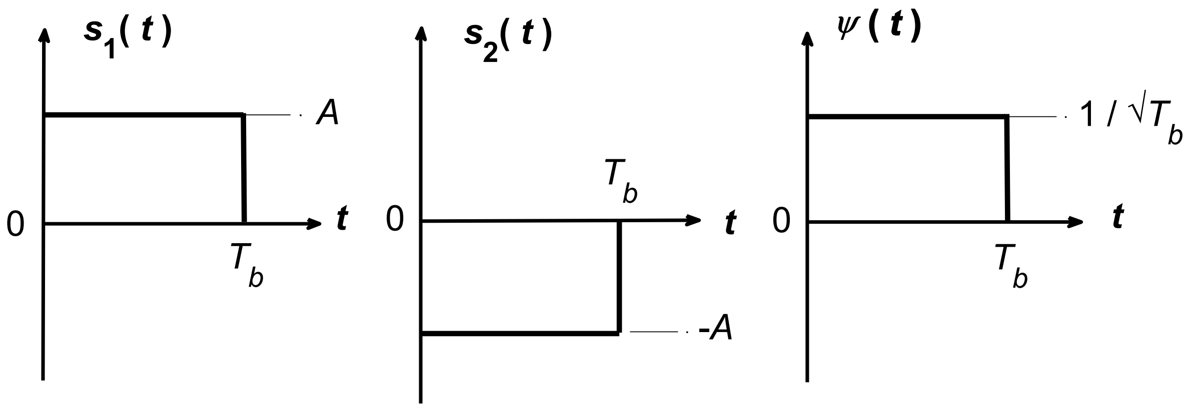
$$\begin{aligned}P_e&=\frac{1}{M}\sum_{m=1}^{M}P_e(\mathbf{s}_m)=\frac{1}{M}\sum_{m=1}^{M}\int_{R_m^c} f\left(\mathbf{r}\mid\mathbf{s}_m\right)d\mathbf{r}\\&=\frac{1}{M}\sum_{m=1}^{M}\left[1-\int_{R_m} f\left(\mathbf{r}\mid\mathbf{s}_m\right)d\mathbf{r}\right]\end{aligned} \quad (6.12)$$

In case the signals $\mathbf{s}_1\cdots\mathbf{s}_M$ are not sent with equal probability, i.e. if MAP criterion is valid, (6.12) turns into

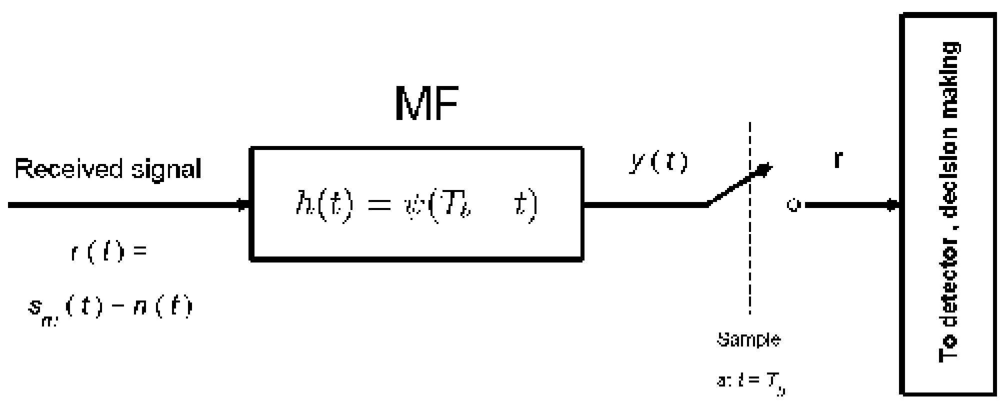
$$P_e=1-\sum_{m=1}^{M}P(\mathbf{s}_m)\int_{R_m} f\left(\mathbf{r}\mid\mathbf{s}_m\right)d\mathbf{r} \quad (6.13)$$

Now we solve two lengthy examples to illustrate the above points.

Example 6.1 : Consider an ASK system where $M = 2$ and the transmitter uses the signal set and the basis function, $s_1(t)$, $s_2(t)$ and $\psi(t)$ shown in Fig. 6.3a. $s_1(t)$ and $s_2(t)$ are transmitted with unequal probabilities of p and $1-p$ respectively. Determine the metrics, i.e., $\mathrm{Max}\left[P(\mathbf{s}_m|\mathrm{r})\right] \equiv \mathrm{Max}\left[f(\mathbf{r}\,|\mathbf{s}_m)P(\mathbf{s}_m)\right]$ for the MAP optimum detector.



a) Transmitted signal waveforms, the basis function



b) Block diagram of MF demodulator

Fig. 6.3 Transmitted signal waveforms, the basis function and the block diagram of MF demodulator for Example 6.1.

Solution : In Fig. 6.2b, the demodulator at the receiver side is shown as matched filter (MF). Accordingly, for the cases of $s_1(t)$ and $s_2(t)$, $y(t)$ after MF becomes the following

$$
\begin{aligned}
y_{s_1}(t) &= \int_0^t r(\tau)\ \psi(T_b - t + \tau)\ d\tau \\
&= \int_0^t s_1(\tau)\ \psi(T_b - t + \tau)\ d\tau + \int_0^t n(\tau)\ \psi(T_b - t + \tau)\ d\tau \\
y_{s_2}(t) &= \int_0^t r(\tau)\ \psi(T_b - t + \tau)\ d\tau \\
&= \int_0^t s_2(\tau)\ \psi(T_b - t + \tau)\ d\tau + \int_0^t n(\tau)\ \psi(T_b - t + \tau)\ d\tau
\end{aligned}
\tag{6.14}
$$

After sampling at $t = T_b$, (6.14) will be

$$
\begin{aligned}
y_{s_1}(T_b) &= \int_0^{T_b} s_1(\tau)\ \psi(\tau)\ d\tau + \int_0^{T_b} n(\tau)\ \psi(\tau)\ d\tau = A\sqrt{T_b} + n(T_b) = \sqrt{\varepsilon_b} + n(T_b) \\
y_{s_2}(t) &= \int_0^{T_b} s_2(\tau)\ \psi(\tau)\ d\tau + \int_0^{T_b} n(\tau)\ \psi(\tau)\ d\tau = -A\sqrt{T_b} + n(T_b) = -\sqrt{\varepsilon_b} + n(T_b) \\
n(T_b) &= \int_0^{T_b} n(\tau)\ \psi(\tau)\ d\tau \quad , \quad \varepsilon_b = \int_0^{T_b} s_1^2(t)\ dt = \int_0^{T_b} s_2^2(t)\ dt = A^2 T_b
\end{aligned}
\tag{6.15}
$$

$n(T_b)$ is the noise sample that has the same characteristics of $n(t)$ in the received signal. Thus $n(T_b)$ is Gaussian with zero mean and variance $\sigma_n^2 = N_0 / 2$. From (5.8) and (6.15), we work out the individual $f(r \mid s_1)$ and $f(r \mid s_2)$ as

$$
\begin{aligned}
f(r \mid s_1) &= \frac{1}{(\pi N_0)^{0.5}} \exp\left[-\left(r - \sqrt{\varepsilon_b}\right)^2 / N_0\right] \\
f(r \mid s_2) &= \frac{1}{(\pi N_0)^{0.5}} \exp\left[-\left(r + \sqrt{\varepsilon_b}\right)^2 / N_0\right]
\end{aligned}
\tag{6.16}
$$

For a detection strategy based on finding $\mathrm{Max}\left[P(\mathbf{s}_m | \mathrm{r})\right] \equiv \mathrm{Max}\left[f(\mathbf{r} \mid \mathbf{s}_m) P(\mathbf{s}_m)\right]$, we simply decide as follows

$$
\begin{aligned}
&\text{If} \quad f(r \mid s_1) P(s_1) > f(r \mid s_2) P(s_2) \quad \text{then decide } s_1(t) \text{ was transmitted} \\
&\text{If} \quad f(r \mid s_1) P(s_1) < f(r \mid s_2) P(s_2) \quad \text{then decide } s_2(t) \text{ was transmitted}
\end{aligned}
\tag{6.17}
$$

Substituting from (6.16) into (6.17)

$$
\begin{aligned}
&\text{If} \quad r > \frac{N_0}{4\sqrt{\varepsilon_b}} \ln\left(\frac{1-p}{p}\right) \quad \text{then decide } s_1(t) \text{ was transmitted} \\
&\text{If} \quad r < \frac{N_0}{4\sqrt{\varepsilon_b}} \ln\left(\frac{1-p}{p}\right) \quad \text{then decide } s_2(t) \text{ was transmitted}
\end{aligned}
\tag{6.18}
$$

So our decision requires a knowledge of $N_0, \ \varepsilon_b$ and p. It is quite possible that the transmitter sends information about the last two parameters, but N_0 has to be measured somehow. Note that if $p = 0.5$, meaning that $s_1(t)$ and $s_2(t)$ are sent with equal probabilities from the transmitter then then the decision rule of (6.18) will become independent of the parameters as shown below.

$$\begin{aligned} &\text{If} \quad r > 0 \quad \text{then decide } s_1(t) \text{ was transmitted} \\ &\text{If} \quad r < 0 \quad \text{then decide } s_2(t) \text{ was transmitted} \end{aligned} \tag{6.19}$$

Example 6.2 : Give at least two different sets of time waveforms, $s_1(t) \cdots s_4(t)$ for 4 PSK. For these waveforms, find appropriate orthonormalized basis functions, find the representation of $s_1(t) \cdots s_4(t)$ in terms of orthonormalized basis functions, find $\mathbf{s}_1 \cdots \mathbf{s}_4$ vectors, the distances between vector ends, draw constellation diagram and the diagram of demodulator comprising correlator and matched filter. Show that distance and correlation metrics function properly (i.e. give the correct decision) if $s_1(t)$ was sent from the transmitter and no noise is mixed with the signal at receiver. Give correct decision boundaries and find probability of error if all signals are sent from transmitter with equal probability.

Solution : Two possible sets of $s_1(t) \cdots s_4(t)$ and $\psi_1(t), \ \psi_2(t)$ are given in Figs. 6.4 and 6.5

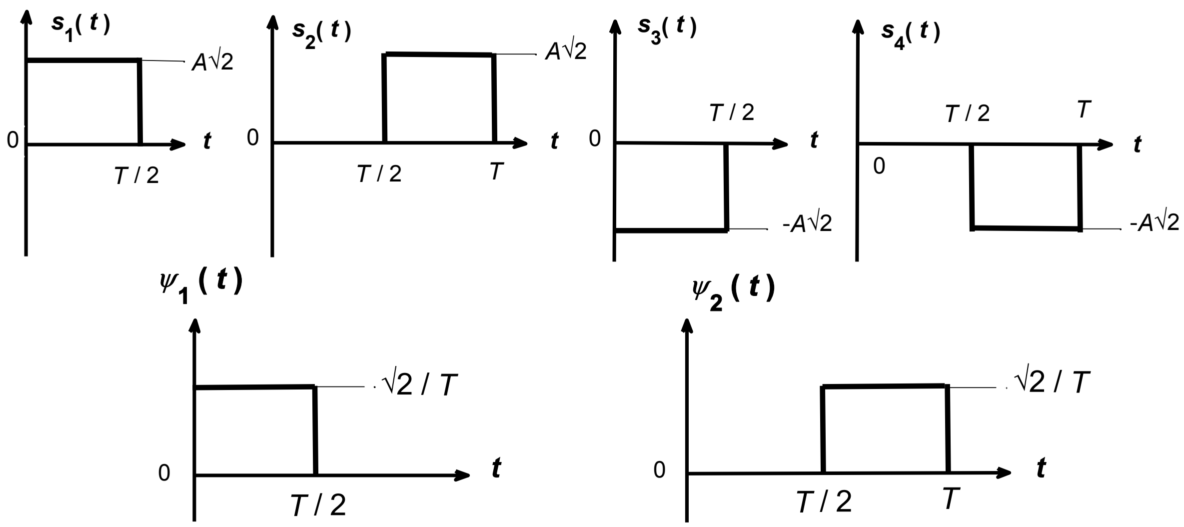


Fig. 6.4 First possible set of signal waveforms, $s_1(t) \cdots s_4(t)$ and orthonormalized basis functions, $\psi_1(t), \ \psi_2(t)$ for 4 PSK.

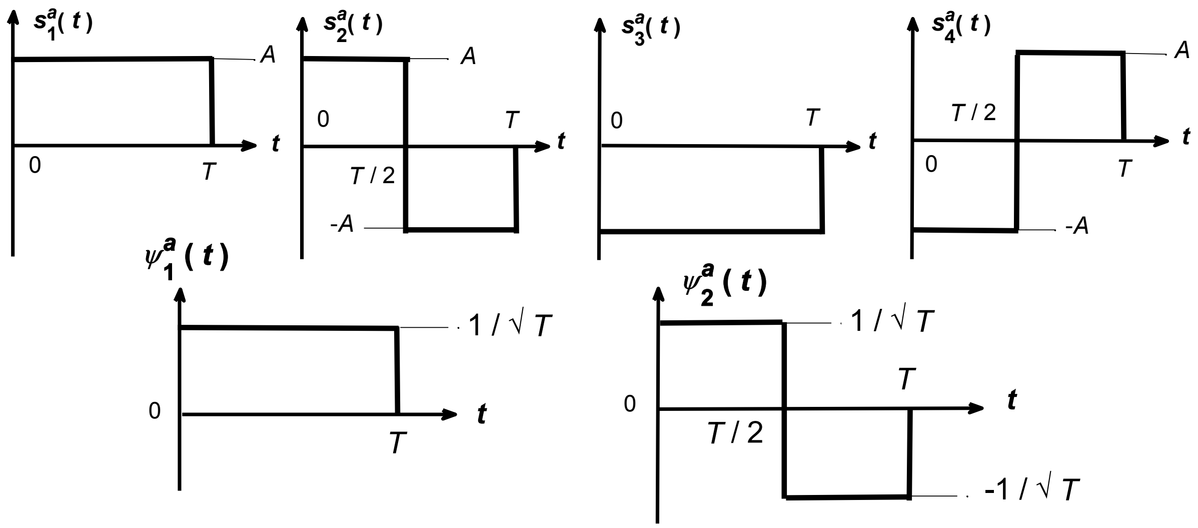


Fig. 6.5 First possible set of signal waveforms, $s_1^a(t) \cdots s_4^a(t)$ and orthonormalized basis functions, $\psi_1^a(t),\ \psi_2^a(t)$ for 4 PSK.

Note that signal waveforms and orthonormalized basis functions in Figs. 6.4 and 6.5 are interchangeable. Here we continue our solution with the set shown Fig. 6.4. Initially we write the time waveform expressions for $s_1(t) \cdots s_4(t)$ and $\psi_1(t),\ \psi_2(t)$

$$s_1(t) = \begin{cases} A\sqrt{2} & 0 \le t \le T/2 \\ 0 & \text{otherwise} \end{cases} \qquad s_2(t) = \begin{cases} A\sqrt{2} & T/2 \le t \le T \\ 0 & \text{otherwise} \end{cases}$$

$$s_3(t) = \begin{cases} -A\sqrt{2} & 0 \le t \le T/2 \\ 0 & \text{otherwise} \end{cases} \qquad s_4(t) = \begin{cases} -A\sqrt{2} & T/2 \le t \le T \\ 0 & \text{otherwise} \end{cases}$$

$$\psi_1(t) = \begin{cases} \sqrt{2/T} & 0 \le t \le T/2 \\ 0 & \text{otherwise} \end{cases} \qquad \psi_2(t) = \begin{cases} \sqrt{2/T} & T/2 \le t \le T \\ 0 & \text{otherwise} \end{cases} \tag{6.20}$$

Now either by eye inspection or by **Gram-Schmidt Orthogonalization Procedure**, we write $s_1(t) \cdots s_4(t)$ in terms of $\psi_1(t)$ and $\psi_2(t)$. Note that here there is no need to indicate the time intervals, since they are embedded in $\psi_1(t)$ and $\psi_2(t)$

$$s_1(t) = A\sqrt{T}\psi_1(t) \ ,\ s_2(t) = A\sqrt{T}\psi_2(t) \ ,\ s_3(t) = -A\sqrt{T}\psi_1(t) \ ,\ s_4(t) = -A\sqrt{T}\psi_2(t)$$

$$\mathbf{s}_1 = [s_{11},\ s_{12}] = \left[A\sqrt{T},\ 0\right] \ ,\ \mathbf{s}_2 = [s_{21},\ s_{22}] = \left[0,\ A\sqrt{T}\right]$$

$$\mathbf{s}_3 = [s_{31},\ s_{32}] = \left[-A\sqrt{T},\ 0\right] \ ,\ \mathbf{s}_4 = [s_{41},\ s_{42}] = \left[0,\ -A\sqrt{T}\right]$$

$$d_{12} = d_{14} = d_{23} = A\sqrt{2T} = \sqrt{2\varepsilon_s} \ ,\ d_{13} = d_{24} = 2A\sqrt{T} = 2\sqrt{\varepsilon_s}$$

$$|\mathbf{s}_1| = |\mathbf{s}_2| = |\mathbf{s}_3| = |\mathbf{s}_4| = A\sqrt{T} \ ,\ \varepsilon_1 = \varepsilon_2 = \varepsilon_3 = \varepsilon_4 = \varepsilon_s = A^2T \tag{6.21}$$

On the second and third lines of (6.21), we have included the vectorial representation of our signals, on the fourth line we have given the respective distance between vector ends, on the last line the length of vectors and the energies are given which can be calculated either from time signals or vectorial representations. As this is PSK, all vector lengths, thus the energies are equal. Now we can plot the constellation diagram of $\mathbf{s}_1 \cdots \mathbf{s}_4$. This is illustrated in Fig. 6.6

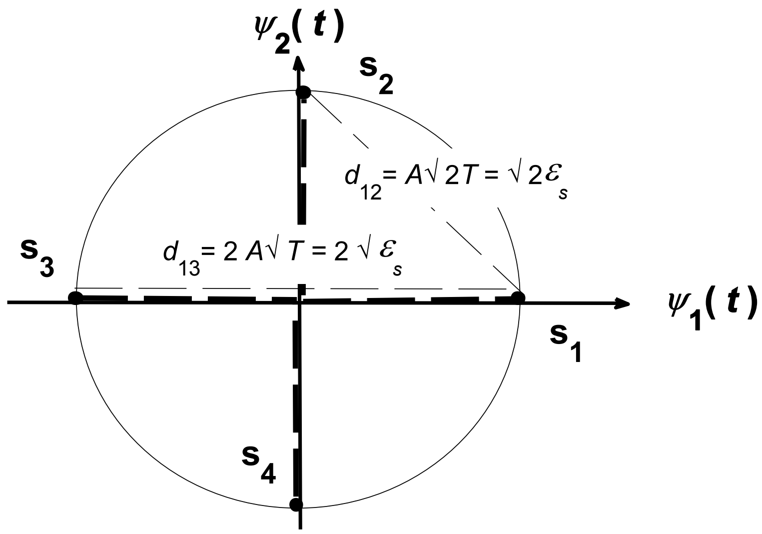
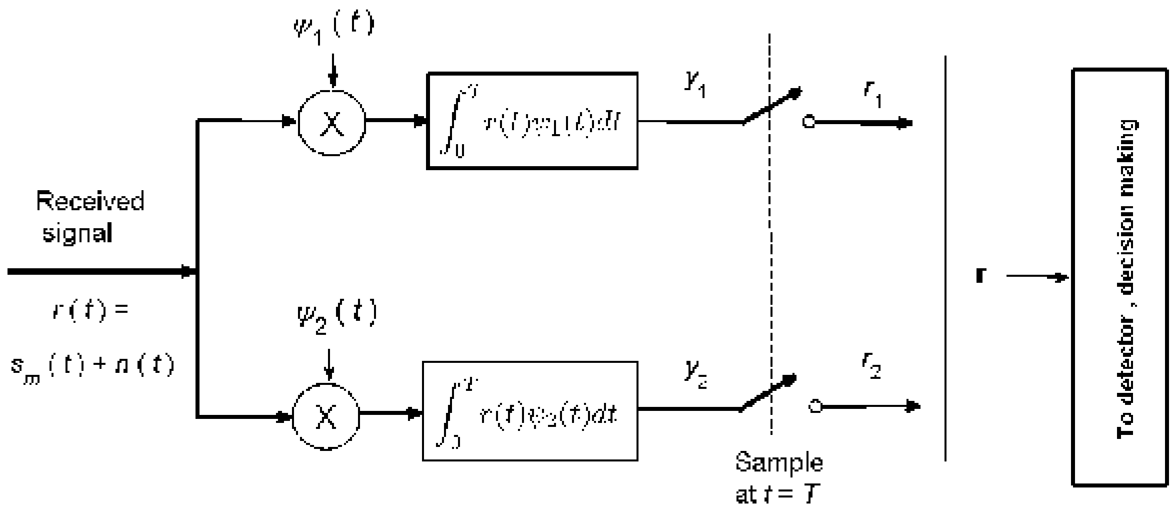
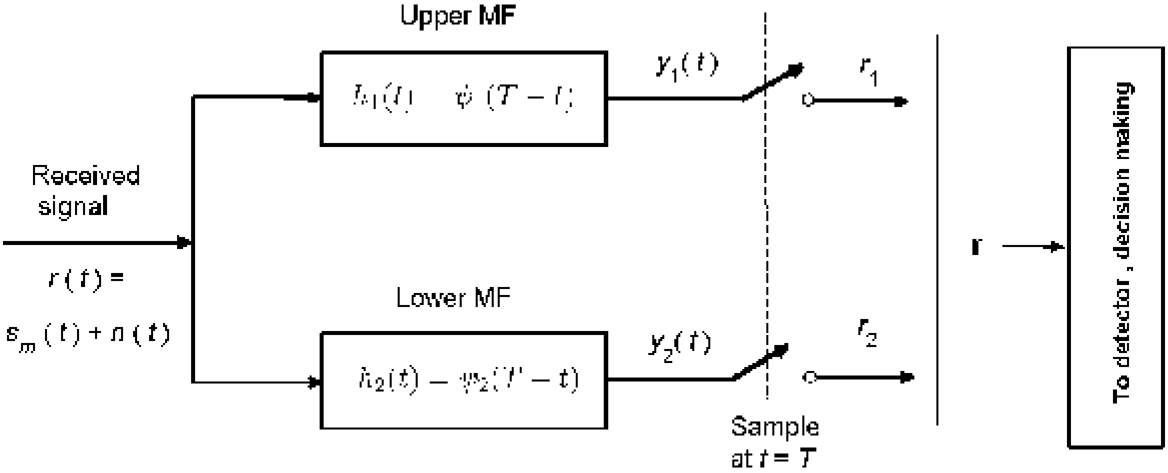


Fig. 6.6 Constellation diagram for the 4 PSK in Example 6.2.

Below, we show the block diagrams of correlator and matched filter type of demodulators.



a) Block diagram of correlator type of demodulator.



b) Block diagram of matched filter type of demodulator.

Fig. 6. 7 Block diagrams of correlator and matched filter type of demodulators for the 4 PSK in Example 6.2.

We first tackle the case of the correlator and assume that $s_1(t)$ was sent from the transmitter

$$
\begin{aligned}
y_1 = r_1 &= \int_0^T r(t)\ \psi_1(t)\ dt = \int_0^T s_1(t)\ \psi_1(t)\ dt + \int_0^T n(t)\ \psi_1(t)\ dt \\
&= A\sqrt{T} + n_1 \quad , \quad n_1 = \int_0^T n(t)\ \psi_1(t)\ dt \\
y_2 = r_2 &= \int_0^T r(t)\ \psi_2(t)\ dt = \int_0^T s_1(t)\ \psi_2(t)\ dt + \int_0^T n(t)\ \psi_2(t)\ dt \\
&= 0 + n_2 \quad , \quad n_2 = \int_0^T n(t)\ \psi_2(t)\ dt \quad , \quad \mathbf{r} = [r_1;\ r_2]
\end{aligned}
\tag{6.22}
$$

The reason that we have arranged vector $\mathbf{r}$ in the form of a column vector rather than a row vector, is because column arrangement facilitates metrics computation.

Doing the same for the matched filter case, we get

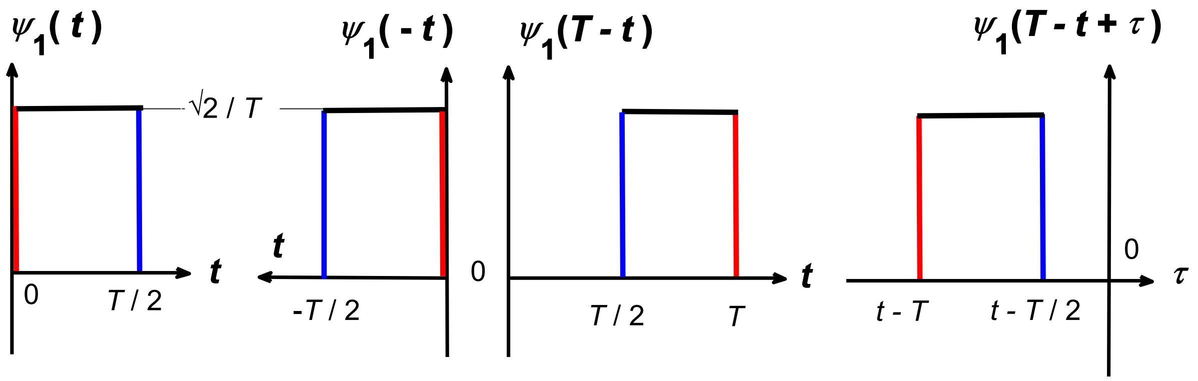
$$
\begin{aligned}
y_1(t) &= \int_0^t r(\tau)\, h_1(t-\tau)\ d\tau = \int_0^t r(\tau)\ \psi_1(T-t+\tau)\ d\tau \\
y_2(t) &= \int_0^t r(\tau)\, h_2(t-\tau)\ d\tau = \int_0^t r(\tau)\ \psi_2(T-t+\tau)\ d\tau
\end{aligned}
\tag{6.23}
$$

After sampling at $t = T$, (6.23) will become

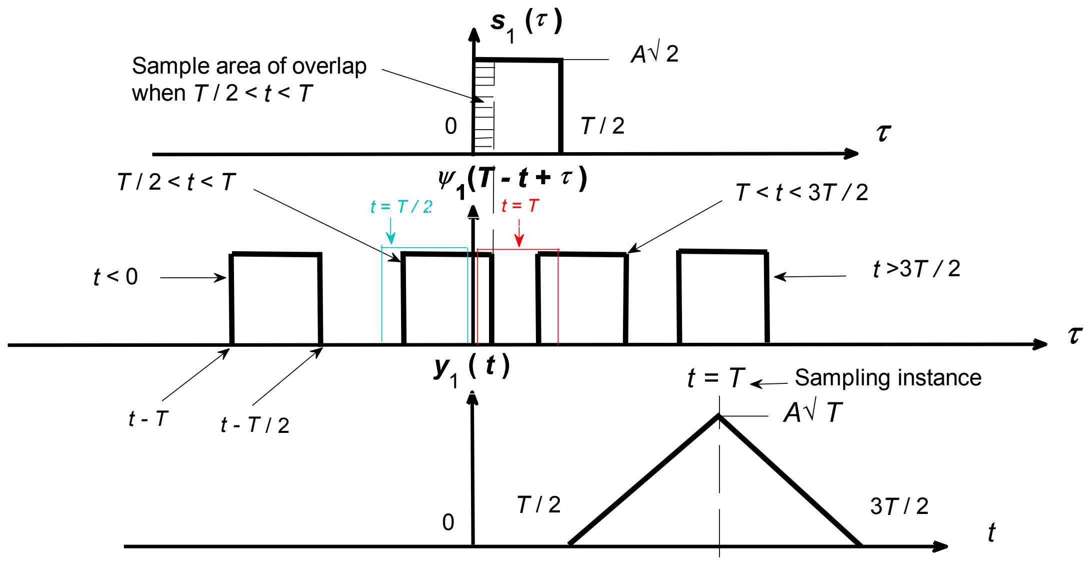
$$r_1 = y_1(T) = \int_0^T r(\tau)\ \psi_1(\tau)\ d\tau = A\sqrt{T} + n_1 \quad , \quad n_1 = \int_0^T n(t)\ \psi_1(t)\ dt$$

$$r_2 = y_2(T) = \int_0^T r(\tau)\ \psi_2(\tau)\ d\tau = n_2 \quad , \quad n_2 = \int_0^T n(t)\ \psi_2(t)\ dt \quad , \quad \mathbf{r} = [r_1;\ r_2] \qquad (6.24)$$

So the outputs, we obtain from correlator and matched filter are the same. For the case of matched filter demodulator, it is instructive to graphically illustrate the convolution operation carried out in $y_1(t)$ and $y_2(t)$ of (6.23) for the signal parts. For the upper and lower branches of Fig. 6.7 a, this is respectively done in Figs. 6.8 and 6.9.

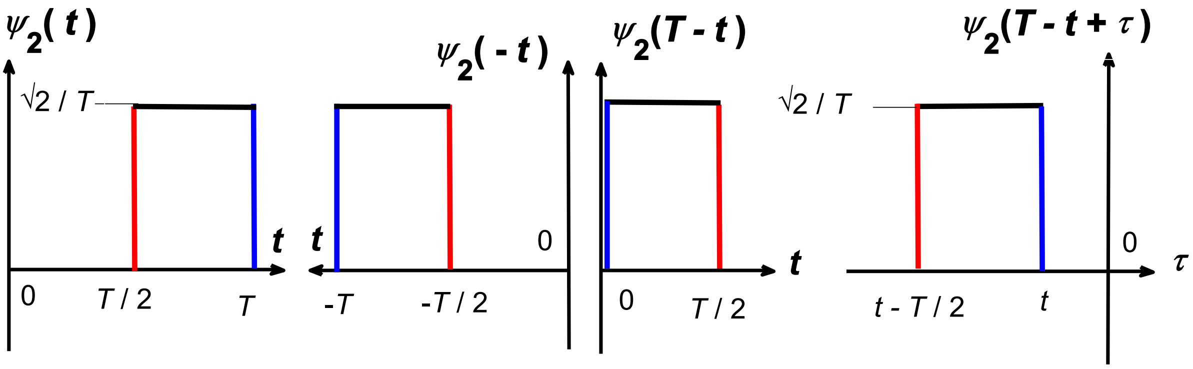


a) Orientation of matched filter (MF) $\psi_1(t)$ for convolution operation in Fig. 6.7b

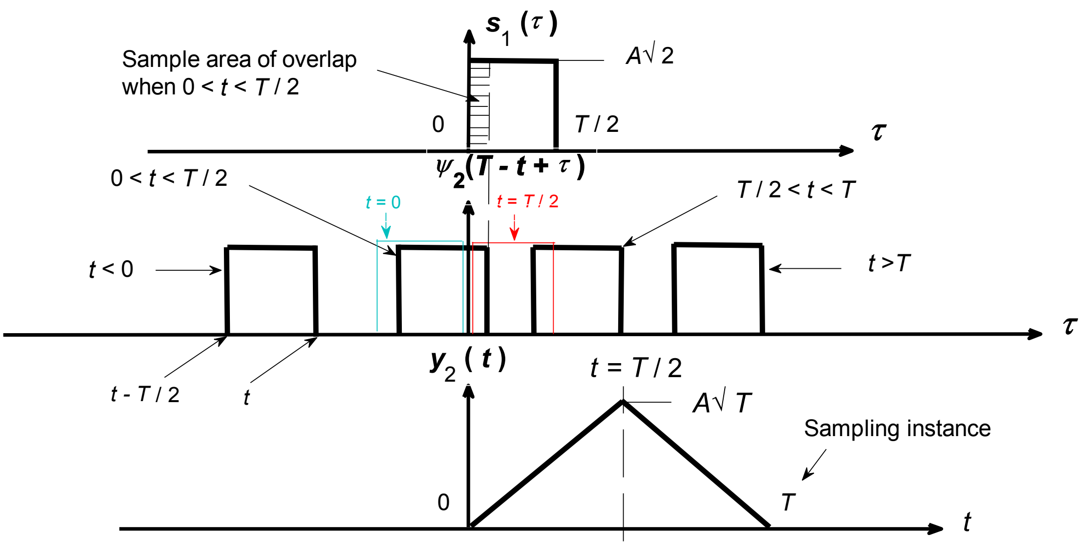


b) Output from matched filter (MF) $\psi_1(t)$ via convolution

Fig. 6. 8 Graphical illustration of the convolution operation implemented in the upper matched filter of Fig. 6.7b.



a) Orientation of matched filter (MF) $\psi_2(t)$ for convolution operation in Fig. 6.7b



b) Output from matched filter (MF) $\psi_1(t)$ via convolution

Fig. 6. 9 Graphical illustration of the convolution operation implemented in the lower matched filter of Fig. 6.7b.

Now we compute correlation metrics values by inserting $\mathbf{s}_1 \cdots \mathbf{s}_4$ into $C(\mathbf{r}, \mathbf{s}_m)$. From (6.8) and (6.22), we have

$$m=1\ ,\ C(\mathbf{r},\mathbf{s}_1)=2\,\mathbf{s}_1\cdot\mathbf{r}-\|\mathbf{s}_1\|^2=2\left[A\sqrt{T},\ 0\right]\begin{bmatrix}A\sqrt{T}+n_1\\ n_2\end{bmatrix}-A^2T=A^2T+2An_1\sqrt{T}$$
$$m=2\ ,\ C(\mathbf{r},\mathbf{s}_2)=2\,\mathbf{s}_2\cdot\mathbf{r}-\|\mathbf{s}_2\|^2=2\left[0,\ A\sqrt{T}\right]\begin{bmatrix}A\sqrt{T}+n_1\\ n_2\end{bmatrix}-A^2T=2An_2\sqrt{T}-A^2T$$
$$m=3\ ,\ C(\mathbf{r},\mathbf{s}_3)=2\,\mathbf{s}_3\cdot\mathbf{r}-\|\mathbf{s}_3\|^2=2\left[-A\sqrt{T},\ 0\right]\begin{bmatrix}A\sqrt{T}+n_1\\ n_2\end{bmatrix}-A^2T=-3A^2T-2An_1\sqrt{T}$$
$$m=4\ ,\ C(\mathbf{r},\mathbf{s}_4)=2\,\mathbf{s}_4\cdot\mathbf{r}-\|\mathbf{s}_4\|^2=2\left[0,\ -A\sqrt{T}\right]\begin{bmatrix}A\sqrt{T}+n_1\\ n_2\end{bmatrix}-A^2T=-2An_2\sqrt{T}-A^2T \quad (6.25)$$

It is clear from (6.25) that in the absence of noise i.e. $n_1=n_2=0$, $C(\mathbf{r},\mathbf{s}_1)$ becomes the largest in the set of $C(\mathbf{r},\mathbf{s}_m)$, $m=1\cdots4$. Under such circumstances, the detector correctly decides that $s_1(t)$ was transmitted. In the presence of noise, to arrive at a correct decision, it must be that

$$C(\mathbf{r},\mathbf{s}_1)>C(\mathbf{r},\mathbf{s}_2)\ ,\ C(\mathbf{r},\mathbf{s}_1)>C(\mathbf{r},\mathbf{s}_3)\ ,\ C(\mathbf{r},\mathbf{s}_1)>C(\mathbf{r},\mathbf{s}_4) \quad (6.26)$$

These three conditions correspond to

$$C(\mathbf{r},\mathbf{s}_1)>C(\mathbf{r},\mathbf{s}_2)\ :\ A^2T+2An_1\sqrt{T}>2An_2\sqrt{T}-A^2T\ \rightarrow A\sqrt{T}+n_1>n_2$$
$$C(\mathbf{r},\mathbf{s}_1)>C(\mathbf{r},\mathbf{s}_3)\ :\ A^2T+2An_1\sqrt{T}>-3A^2T-2An_1\sqrt{T}\ \rightarrow\ A\sqrt{T}>-n_1$$
$$C(\mathbf{r},\mathbf{s}_1)>C(\mathbf{r},\mathbf{s}_4)\ :\ A^2T+2An_1\sqrt{T}>-2An_2\sqrt{T}-A^2T\ \rightarrow\ A\sqrt{T}+n_1>-n_2 \quad (6.27)$$

From (6.24) and Fig. 6.6, we deduce that if $s_1(t)$ was transmitted and gets mixed with noise, then at receiver constellation diagram will look like the following

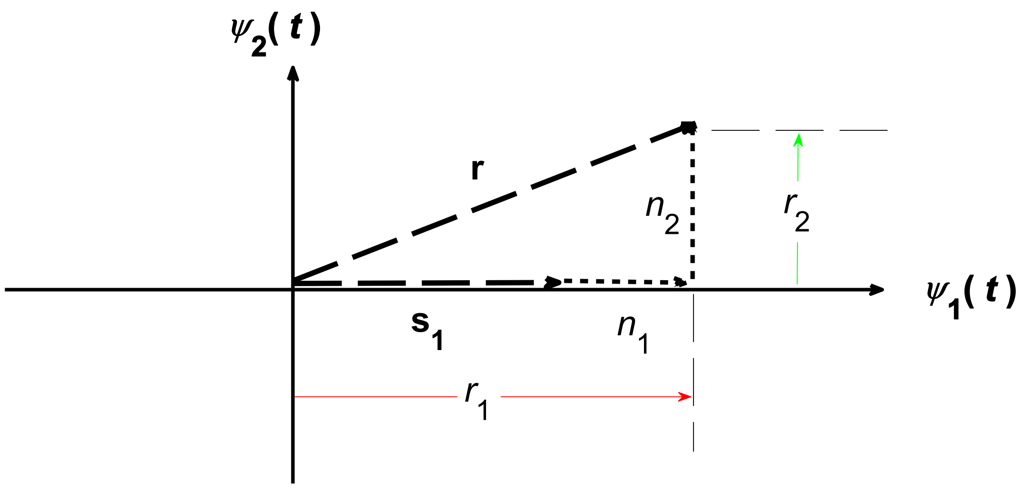


Fig. 6. 10 Constellation diagram at receiver if $s_1(t)$ was transmitted.

The three conditions in (6.27) individually define the regions shown in Fig. 6.11.

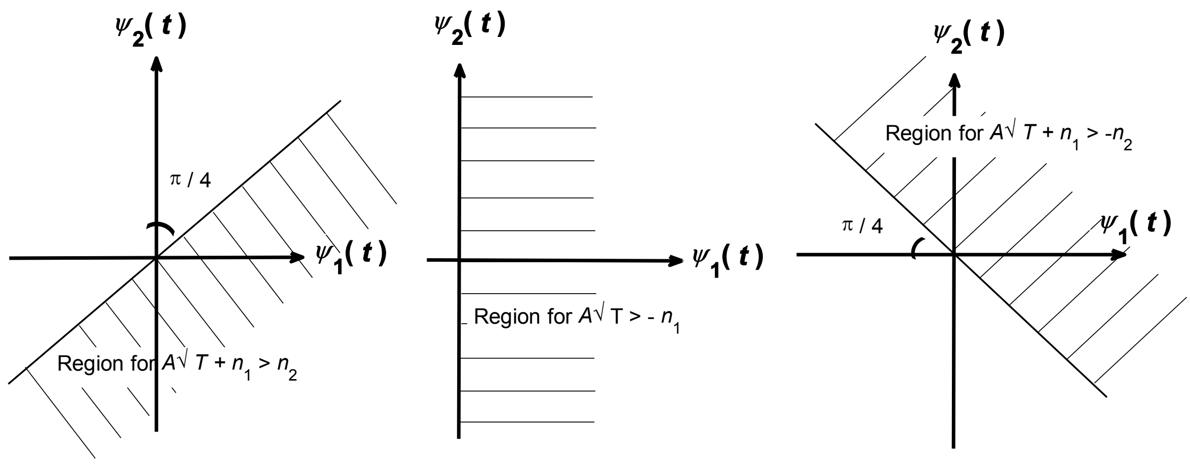


Fig. 6. 11 Regions for conditions in (6.27).

Since conditions in (6.27) are connected by AND relation, then the intersection of the three regions of Fig. 6.11 becomes

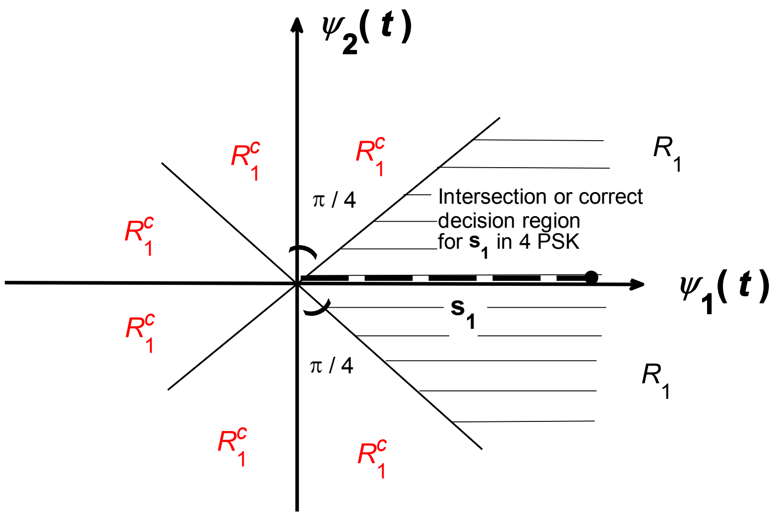


Fig. 6. 12 Intersection of correct decision region for conditions in (6.27).

With the view in Fig. 6.12, we are now in a position to estimate the probability of error for 4 PSK or *M* ary PSK in general. For simplicity, here we choose to find the probability of correct detection P_c and use it to find the probability of error P_e from the following relation

$$P_e = 1 - P_c \tag{6.28}$$

The conditions in (6.27) may be written in a different way, taking into account Fig. 6. 12, that is

$$-\frac{\pi}{4} < \tan^{-1}\left(\frac{n_2}{n_1+\sqrt{\varepsilon_s}}\right) < \frac{\pi}{4} \quad , \quad n_1 > -\sqrt{\varepsilon_s} \quad , \quad \varepsilon_s = A^2 T \tag{6.29}$$

where we have used the alternative definition of $\varepsilon_s = A^2 T$. The first inequality in (6.29) is a combination of first and third conditions from (6.27), while the second inequality of (6.29) is merely the second condition from (6.27) expressed in a different manner. Writing (6.29) purely in terms of n_1 and n_2, we get

$$-\left(n_1+\sqrt{\varepsilon_s}\right)\tan\left(\frac{\pi}{4}\right) < n_2 < \left(n_1+\sqrt{\varepsilon_s}\right)\tan\left(\frac{\pi}{4}\right) \quad , \quad n_1 > -\sqrt{\varepsilon_s} \tag{6.30}$$

So we can utilize the noise pdf definition given in (5.6) and set P_c as follows

$$P_c = \frac{1}{\left(\pi N_0\right)^{0.5}} \int_{-\sqrt{\varepsilon_s}}^{\infty} \exp\left(-\frac{n_1^2}{N_0}\right) dn_1 \frac{1}{\left(\pi N_0\right)^{0.5}} \int_{-\left(n_1+\sqrt{\varepsilon_s}\right)\tan\left(\frac{\pi}{4}\right)}^{\left(n_1+\sqrt{\varepsilon_s}\right)\tan\left(\frac{\pi}{4}\right)} \exp\left(-\frac{n_2^2}{N_0}\right) dn_2 \tag{6.31}$$

Note that in the general case of *M* ary PSK, (6.31) simply becomes

$$P_c = \frac{1}{\left(\pi N_0\right)^{0.5}} \int_{-\sqrt{\varepsilon_s}}^{\infty} \exp\left(-\frac{n_1^2}{N_0}\right) dn_1 \frac{1}{\left(\pi N_0\right)^{0.5}} \int_{-\left(n_1+\sqrt{\varepsilon_s}\right)\tan\left(\frac{\pi}{M}\right)}^{\left(n_1+\sqrt{\varepsilon_s}\right)\tan\left(\frac{\pi}{M}\right)} \exp\left(-\frac{n_2^2}{N_0}\right) dn_2 \tag{6.32}$$

The evaluation of P_e for the cases of $M = 2$ and general M are given below

$$\begin{aligned}
& M = 2 \text{ , } P_e = Q\left(\sqrt{\frac{2\varepsilon_s}{N_0}}\right) \quad , \quad \mathrm{SNR}_s = \frac{\varepsilon_s}{N_0} \quad , \quad \mathrm{SNR}_b = \frac{\varepsilon_s}{\log_2\left(M\right)N_0} = \frac{\varepsilon_b}{N_0} \\
& \text{Any } M \text{ , } P_e = 1 - \frac{1}{\pi^{0.5}}\exp\left(-\mathrm{SNR}_s\right)\int_0^{\infty}\exp\left(-z^2 + 2z\mathrm{SNR}_s^{0.5}\right)\mathrm{erf}\left[z\tan\left(\frac{\pi}{M}\right)\right] dz \\
& Q\left(x\right) = \frac{1}{\left(2\pi\right)^{0.5}}\int_x^{\infty}\exp\left(-\frac{z^2}{2}\right) dz \quad , \quad \Phi\left(x\right) = \frac{1}{\left(2\pi\right)^{0.5}}\int_{-\infty}^{x}\exp\left(-\frac{z^2}{2}\right) dz \\
& Q\left(x\right) + \Phi\left(x\right) = 1 \quad , \quad \Phi\left(x\right) = Q\left(-x\right) \quad , \quad \mathrm{erf}\left(x\right) + \mathrm{erfc}\left(x\right) = 1 \\
& \mathrm{erf}\left(x\right) = \frac{2}{\pi^{0.5}}\int_0^{x}\exp\left(-z^2\right) dz \quad , \quad \mathrm{erfc}\left(x\right) = \frac{2}{\pi^{0.5}}\int_x^{\infty}\exp\left(-z^2\right) dz \\
& Q\left(x\right) = 0.5\mathrm{erfc}\left(x/2^{0.5}\right) = 0.5 - 0.5\mathrm{erf}\left(x/2^{0.5}\right) \\
& \Phi\left(x\right) = 1 - \mathrm{erfc}\left(x/2^{0.5}\right) = 0.5 - 0.5\mathrm{erf}\left(x/2^{0.5}\right)
\end{aligned} \tag{6.33}$$

As seen from (6.33), it is possible to evaluate P_e analytically only in the case of $M = 2$, for $M > 2$, numeric evaluation is required. $\Phi\left(x\right)$ and $Q\left(x\right)$ are known as error function and complimentary error function respectively. Their definitions are slightly different in Matlab denoted as

$\text{erf}(x)$ and $\text{erfc}(x)$. The equivalences are given on the last but two lines of (6.33). SNR is the signal to noise ratio. As given in (6.33), it is signal energy divided by noise spectral density, sometimes it is taken as signal power divided by noise power. When signal part refers to symbol energy, we put the subscript *s*, but if it's the energy in the binary waveform, then we use the subscript *b*.

Finally we want to point out that the P_c and P_e derivations made above for the case of $\mathbf{s}_1$ actually represents the general case of any $\mathbf{s}_m$ and the total average P_c and P_e, since

$$
\begin{aligned}
P_c(\text{Average}) = & P(\mathbf{s}_1 \text{ being sent}) P(\mathbf{r} \text{ falling in correct decision of } \mathbf{s}_1) \\
& + P(\mathbf{s}_2 \text{ being sent}) P(\mathbf{r} \text{ falling in correct decision of } \mathbf{s}_2) \\
& \quad \vdots \\
& + P(\mathbf{s}_M \text{ being sent}) P(\mathbf{r} \text{ falling in correct decision of } \mathbf{s}_M) \\
= & P_c(\mathbf{r} \text{ falling in correct decision of } \mathbf{s}_m | \text{given } \mathbf{s}_m \text{ was sent from transmitter}) \qquad (6.34)
\end{aligned}
$$

The last line in (6.34) is due to the fact that

$$P(\mathbf{s}_1 \text{ being sent}) = P(\mathbf{s}_2 \text{ being sent}) = \cdots = P(\mathbf{s}_M \text{ being sent}) = \frac{1}{M} \qquad (6.35)$$

Exercise 6.1 : By using the signal set given in Fig. 6.5 $s_1^a(t) \cdots s_4^a(t)$, repeat the steps of Example 6.2.

Exercise 6.2 : On the course webpage, there is the MATLAB code MallPeMdl_2012.m to evaluate theoretical and experimental P_e for ASK, PSK, rectangular QAM using the model file modulators and demodulators called Allmodd and AllDmodd together with the associated m files of PeMaryPSK.m and quade.m. For the theoretical formulation, for PSK for instance, we use the second line of (6.30). To arrive at experimental P_e, we simply transmit a number of symbols of the related modulation type, then add noise to these symbols in the defined SNR ratio, then perform demodulation. In the end, the following ratio is calculated to get experimental P_e

$$(\text{experimental})\ P_e = \frac{\text{Number of symbols in error}}{\text{Total number of symbols transmitted } (n)} \qquad (6.36)$$

Here care must be taken to send sufficient number of symbols so we can represent the small P_e values properly. If this is not done, fluctuations occur on probability of error curves and small P_e values will be absent from the graph. Usually the golden rule is

$$n \geq 10 / P_e \qquad (6.37)$$

By using the MATLAB code and the model files, find the P_e curves at $M = 2,\ 4,\ 8,\ 32,\ 64,\ 128$ for ASK, PSK, QAM. Compare your results with Figs. 7.55, 7.57 and 7.62 of Proakis 2002. Make comments on the dependency of P_e on M and modulation type.

Exercise 6.3 : By using the analysis in Example 6.2, find by hand derivation, P_e for the two PSK constellations of $M = 2$. Explain the difference between the two cases.

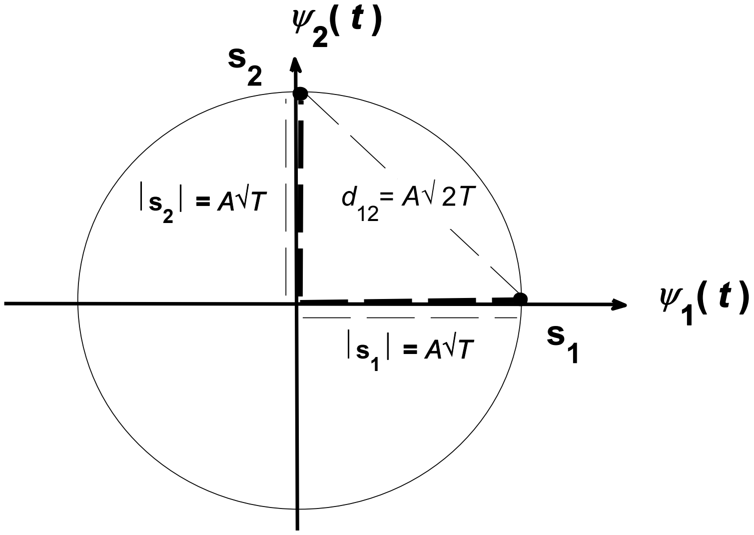


Fig. 6.13 Orthogonal PSK constellation of $M = 2$ for Exercise 6.2.

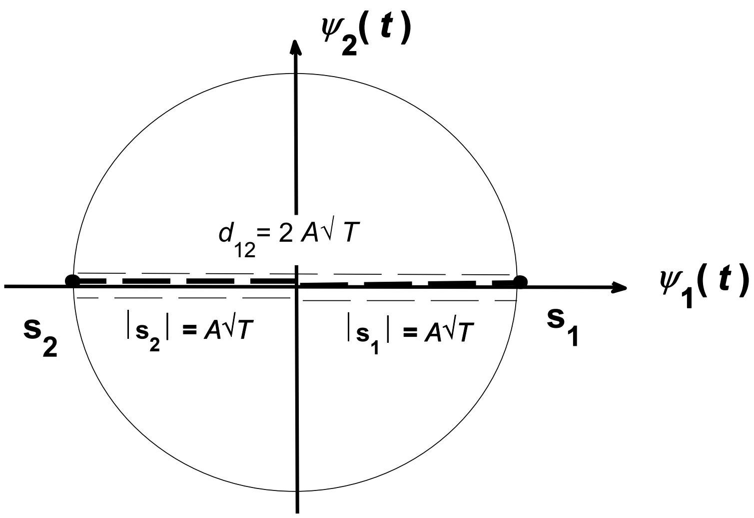


Fig. 6.13 Antipodal PSK constellation of $M = 2$ for Exercise 6.2.

The above text is based on

1) John G. Proakis, Masoud Salehi, “Communication Systems Engineering” 2nd Ed. 2002, ISBN : 0-13-061793-8.
2) My own lecture notes.