Chapter 5

Optimal Receiver of Digital Communication Systems

From earlier chapters, we recall the performance of digital communication systems to be evaluated by ***bit error rate*** (BER). Influenced by the development of modern probability theory and statistical decision theory, communication theory has been developed to evaluate the performance of digital communication systems and to design more effective systems meeting the need from engineering. The first subject that we want to know is how to design an optimal receiver for a digital communication system given a signaling method being determined in advance. In other words, we shall develop a systematic way to design a mechanism of digital communication receiver that can determine whether "1" or "0" to be transmitted.

This design must be able to combat with at least two dimensions of challenges: noise from the channel (including the components used in transmitter and receiver), and the signal distortion through the channel. We may consider a phenomenon for baseband pulse transmission, after introducing pulse modulation in earlier chapter. Figure 5.1 illustrates a perfect pulse transmission as dot line drawing. The solid drawing shows pulses after transmission in channel to result in signal ***dispersion***, which can be understood by considering a pulse after non-perfect channel ***filtering***. Perfect channel has an impulse response like a delta function, but it is not in practical situations. We can consequently observe the "overlapping" of the pulses after transmission (i.e. at the receiver decision input), which results in unwanted situation that increase the difficulty for receiver to make a right decision between "1" and "0". This unwanted signal "overlapping" is generally known as ***inter-symbol interference*** (ISI), while a pulse may carry information more than a bit and it is considered as a ***symbol***. ISI may create signal distortion to influence the right decision of receiver, and may result in un-recoverable distortion at the receiver end. In this chapter, we would like to derive the optimal receiver based on mathematical structure for digital communications.

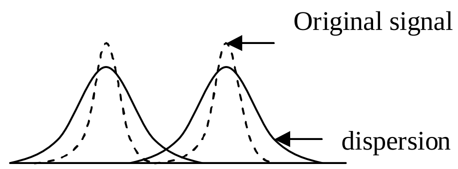


Figure 5.1: Signal dispersion

5.1 MATCHED FILTER

One of the best early description of matched filter might be the book *Introduction to Statistical Communication Theory* authored by John Thomas in 1960's. We pretty much follow the same way in the following development. We wish to identify a linear filter that can optimize the decision of baseband pulse transmission as Figure 5.2 showing, to correctly detect the pulse transmitted over channel with embedded noise.

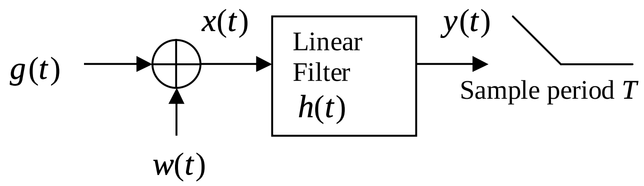


Figure 5.2: Matched filter for a linear receiver

Suppose the linear time-invariant filter having impulse response of $h(t)$. The filter input (or received waveform) $x(t)$ consists of the signal pulse $g(t)$ and additive noise $w(t)$ of 2-side power spectral density $N_0/2$.

$$x(t) = g(t) + w(t), \qquad 0 \le t \le T \tag{5.1}$$

where T is the observation interval and is usually the symbol period/time. The pulse binary signal may represent either "1" or "0" in a digital communication system. We wish to detect signal pulse and make a decision about "1" or "0" based on the received waveform, under the uncertainty of noise. We consequently wish to optimize filter design, in order to minimize noise effects and to optimize detection. Modern communication theory considers such a problem in a statistical way. The resulting filter output is

$$y(t) = g_o(t) + n(t) \tag{5.2}$$

where $g_o(t)$ and $n(t)$ are produced by signal and noise respectively. Please recall SNR to be a good index of system performance. We therefore maximize the *peak pulse signal-to-noise ratio* as

$$h = \frac{|g_o(T)|^2}{E[n^2(t)]} \tag{5.3}$$

In other words, we wish to design a linear filter to optimize its output SNR. Let $g(t) \leftrightarrow G(f), h(t) \leftrightarrow H(f)$. Via inverse Fourier transform,

$$g_o(t) = \int_{-\infty}^{\infty} H(f)G(f)\exp(j2p\, ft)df \tag{5.4}$$

When the filter output is sampled at time $t = T$, we have

$$|\, g_o(T)\,|^2 = \left| \int_{-\infty}^{\infty} H(f)G(f)\exp(j2p\, fT)df \right|^2 \tag{5.5}$$

Via the property in Chapter 2, the power spectral density of the output noise $n(t)$ is

$$S_N(f) = \frac{N_0}{2} |\, H(f)\,|^2 \tag{5.6}$$

The average power of the output noise $n(t)$ is

$$\begin{aligned} E[n^2(t)] &= \int_{-\infty}^{\infty} S_N(f)\,df \\ &= \frac{N_0}{2} \int_{-\infty}^{\infty} |H(f)|^2\, df \end{aligned} \tag{5.7}$$

We are ready to re-write the peak pulse SNR as

$$h = \frac{\left| \int_{-\infty}^{\infty} H(f)G(f)\exp(j2p\, fT)df \right|^2}{\frac{N_0}{2} \int_{-\infty}^{\infty} |H(f)|^2\, df} \tag{5.8}$$

Lemma 5.1: (Cauchy's Inequality)

$$(\sum x_i y_i)^2 \le (\sum x_i^{\ 2})(\sum y_i^{\ 2}) \tag{5.9}$$

Lemma 5.2: (Schwartz Inequality) If $\int_{-\infty}^{\infty} \left|f_1(x)\right|^2 dx < \infty$ and $\int_{-\infty}^{\infty} \left|f_2(x)\right|^2 dx < \infty$, then

$$|\int f_1(x) f_2(x) dx|^2 \le \int | f_1(x)|^2 dx \int | f_2(x)|^2 dx \tag{5.10}$$

Equality holds when

$$f_1(x) = k_1 f_2^*(x) \tag{5.11}$$

and k_1 is a constant.

Applying Schwartz inequality, we get

$$\left|\int_{-\infty}^{\infty} H(f) G(f) \exp(j2p\, fT) df\right|^2 \le \int_{-\infty}^{\infty} \left|H(f)\right|^2 df \int_{-\infty}^{\infty} \left|G(f)\right|^2 df \tag{5.12}$$

We can further get

$$h \le \frac{2}{N_0} \int_{-\infty}^{\infty} \left|G(f)\right|^2 df \tag{5.13}$$

It suggests that h reaches maximum when $H(f)$ is selected to hold the equality. We get (k is a scaling constant)

$$h_{\max} = \frac{2}{N_0} \int_{-\infty}^{\infty} \left|G(f)\right|^2 df \tag{5.14}$$

$$H_{opt}(f) = kG^*(f) \exp(-j2p\, fT) \tag{5.15}$$

It is interesting to note that the optimal filter is actually the complex conjugate of original input signal except the factor $k \exp(-j2p\, fT)$. Taking the inverse Fourier transform, the impulse response of optimal filter is

$$h_{opt}(t) = k\int_{-\infty}^{\infty} G^*(f)\exp[-j2\pi f(T-t)]df \tag{5.16}$$

For a real signal $g(t)$, we have $G^*(f) = G(-f)$, and can express

$$\begin{aligned} h_{opt}(t) &= k\int_{-\infty}^{\infty} G(-f)\exp[-j2\pi f(T-t)]df \\ &= k\int_{-\infty}^{\infty} G(f)\exp[j2\pi f(T-t)]df \\ &= k\,g(T-t) \end{aligned} \tag{5.17}$$

(5.17) tells us that the impulse response of the optimal filter is a time-reversed and delayed version of the input signal $g(t)$, except the scaling factor *k*. Or, we may interpret that $h_{opt}(t)$ is matched to $g(t)$ but in a reverse direction. That is the reason why we call such a filter as a ***matched filter***. Please also note that we only assume noise to be white, and no further assumption about its statistics. In frequency domain, ignoring the delay factor, the matched filter is the complex conjugate of $G(f)$.

Proposition 5.3: The matched filter optimizes output SNR under the structure of Figure 5.2 has time-domain and frequency domain response as

$$h_{opt}(t) = k\,g(T-t)$$

$$H_{opt}(f) = kG^*(f)\exp(-j2\pi fT)$$

Property 5.4: The peak signal-to-noise ratio of the matched filter depends solely on the ratio of signal energy to the power spectral density of white noise at the filter input.

►Proof: The Fourier transform of matched filter output $g_o(t)$ is

$$\begin{aligned} G_o(f) &= H_{opt}(f)G(f) \\ &= kG^*(f)G(f)\exp(-j2\pi fT) \\ &= k|G(f)|^2\exp(-j2\pi fT) \end{aligned} \tag{5.18}$$

Via inverse Fourier transform, we can derive the matched filter output at $t = T$ as

$$g_o(T) = \int_{-\infty}^{\infty} G_o(f)\exp(j2p\, fT)df$$
$$= k\int_{-\infty}^{\infty} |G(f)|^2 df$$

According to Rayleigh's energy theorem (i.e. total energy in time domain and in frequency domain is the same),

$$E = \int_{-\infty}^{\infty} g^2(t)dt = \int_{-\infty}^{\infty} |G(f)|^2 df$$

Then,

$$g_o(T) = kE \tag{5.19}$$

The average output noise power is

$$\begin{aligned} E[n^2(t)] &= \frac{k^2 N_o}{2}\int_{-\infty}^{\infty} |G(f)|^2 df \\ &= k^2 N_o E / 2 \end{aligned} \tag{5.20}$$

Consequently, the peak SNR reaches its maximum

$$h_{\max} = \frac{(kE)^2}{(k^2 N_0 E/2)} = \frac{2E}{N_0} \tag{5.21}$$

Q.E.D.◄

Example 5.1:

►Let us consider operation of matched filter for a typical signaling, $g(t)$ with a rectangular pulse of amplitude A and duration T . Figure 5.3 shows the waveform and operating waveform from the matched filter. If we use the integrate-and-dump circuit as Figure 5.4, we can get the matched filter output at $t = T$ as Figure 5.3(c).◄

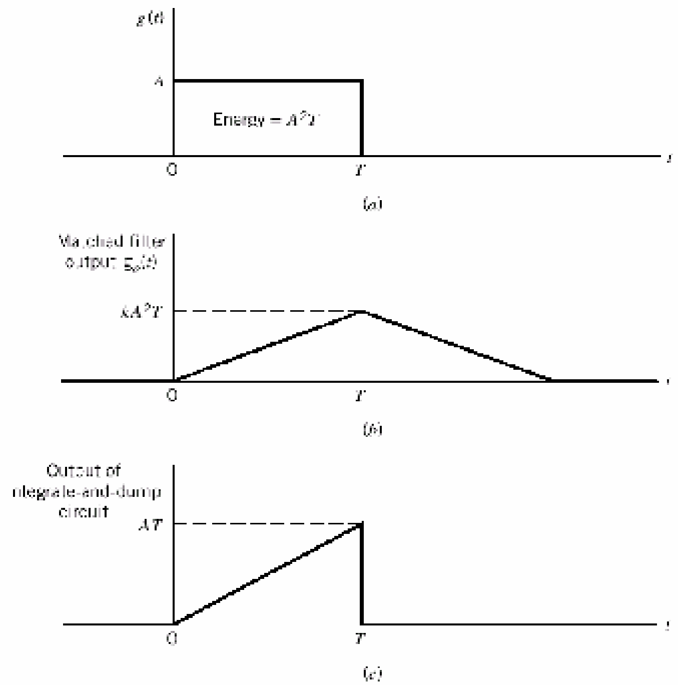


Figure 5.3: (a) Signal (b) Matched Filter Output (c) Integrator Output Waveform

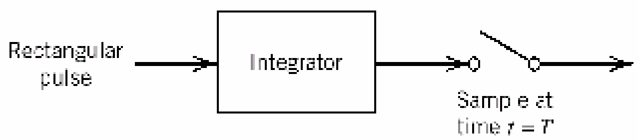


Figure 5.4: Integrate-and-dump Circuit

We are now ready to evaluate the error rate of detection pulse modulation, which is the most important performance index of a digital communication system. Before we develop a general signal space methodology in later section, we introduce statistical decision concept to design the receiver and thus to evaluate error rate.

Suppose we consider a binary pulse modulation using non-return-to-zero (NRZ) signaling format with bit period T_b and amplitude A. The channel is embedded an AWGN with zero mean and power spectral density $N_0/2$. The received waveform can be modeled as

$$x(t)=\begin{cases} +A+w(t), & \text{symbol "1" was sent (i.e. } \mathrm{H}_1) \\ -A+w(t), & \text{symbol "0" was sent (i.e. } \mathrm{H}_0) \end{cases} \tag{5.22}$$

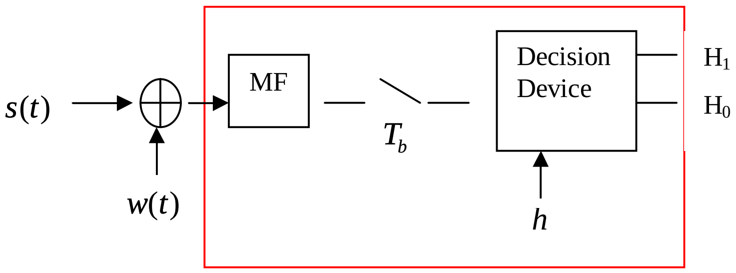


Figure 5.5: Receiver with matched filter

The mechanism decides H_1 or H_0 based on the threshold h . There are two possible kinds of error:

- Symbol "1" is determined when "0" is actually transmitted. It is the error of the first kind. We may also call such an error as *false alarm*.
- Symbol "0" is determined when "1" is actually transmitted. It is the error of the second kind. We may also call such an error as *missing*.

It is obvious for the importance of the selection of threshold. For (5.22), if we consider ***equally probable*** signaling (that is, ***a prior*** probabilities for "1" and "0" are equal or 1/2), we may observe from the following figure to see the conditional distributions for H_1 or H_0 . The threshold can be easily decided as $h = 0$ by symmetry.

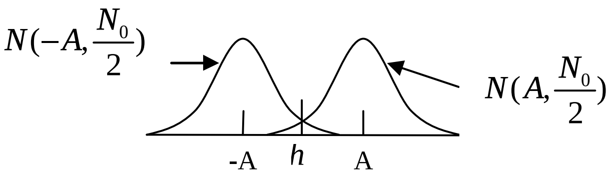


Figure 5.6 Conditional Distributions of H_1 or H_0

Before we calculate the error rate (or the probability of error), we introduce the Gaussian tail function as follows.

$$Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2p}} \exp(-\frac{x^2}{2}) dx \tag{5.23}$$

Suppose we first consider the case "0" is sent. The received signal is

$$x(t) = -A + w(t), \qquad 0 \le t \le T_b \tag{5.24}$$

The sample at $t = T_b$ of matched filter output is

$$\begin{aligned} y &= \int_0^{T_b} x(t)dt \\ &= -A + \frac{1}{T_b}\int_0^{T_b} w(t)dt \end{aligned} \tag{5.25}$$

It represents a sample value from a random variable *Y*. Since $w(t)$ is white and Gaussian, we can characterize *Y* as

- *Y* is Gaussian distributed with mean –*A*.
- The variance of *Y* is

$$\begin{aligned} \sigma_Y^2 &= E[(Y+A)^2] \\ &= \frac{1}{T_b^2} E\left[\int_0^{T_b}\int_0^{T_b} w(t)w(u)\,dt\,du\right] \\ &= \frac{1}{T_b^2}\int_0^{T_b}\int_0^{T_b} E\left[w(t)w(u)\right]dt\,du \\ &= \frac{1}{T_b^2}\int_0^{T_b}\int_0^{T_b} R_W(t,u)\,dt\,du \end{aligned} \tag{5.26}$$

where $R_W(t,u)$ is the autocorrelation of $w(t)$ that is white with power spectral density $N_0/2$.

$$R_W(t,u) = \frac{N_0}{2}\delta(t-u) \tag{5.27}$$

Consequently,

$$\begin{aligned} \sigma_Y^2 &= \frac{1}{T_b^2}\int_0^{T_b}\int_0^{T_b} \frac{N_0}{2}\delta(t-u)\,dt\,du \\ &= \frac{N_0}{2T_b} \end{aligned} \tag{5.28}$$

The conditional probability density function of *Y* given symbol “0” being sent is therefore

$$f_Y(y\,|\,0) = \frac{1}{\sqrt{\pi N_0/T_b}}\exp\left(-\frac{(y+A)^2}{N_0/T_b}\right) \tag{5.29}$$

We denote p_{10} as the conditional probability of error given “0” being sent as Figure 5.7.

$$
\begin{aligned}
p_{10} &= P(y > h \mid \text{symbol 0 was sent}) \\
&= \int_h^{\infty} f_Y(y \mid 0)\, dy \\
&= \frac{1}{\sqrt{p\ N_0 / T_b}} \int_h^{\infty} \exp\left(-\frac{(y+A)^2}{N_0 / T_b} \right) dy \\
&= Q\left(\frac{A+h}{\sqrt{N_0 / 2T_b}} \right)
\end{aligned}
\tag{5.30}
$$

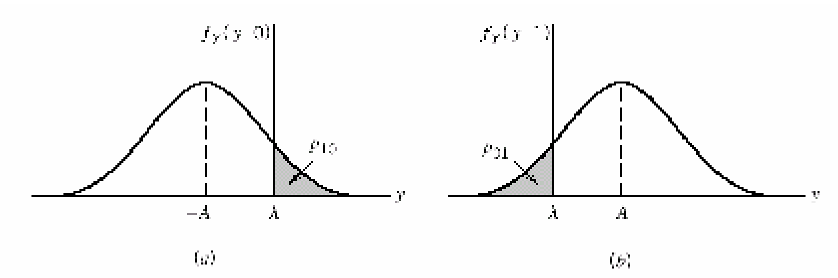


Figure 5.7 Graphic Conditional Distributions

Similarly, via conditional distribution

$$
f_Y(y \mid 1) = \frac{1}{\sqrt{p\ N_0 / T_b}} \exp\left(-\frac{(y-A)^2}{N_0 / T_b} \right) \tag{5.31}
$$

We can obtain the conditional probability error p_{01} as

$$
\begin{aligned}
p_{01} &= P(y < h \mid \text{symbol 1 was sent}) \\
&= \int_{-\infty}^{h} f_Y(y \mid 1)\, dy \\
&= \frac{1}{\sqrt{p\ N_0 / T_b}} \int_{-\infty}^{h} \exp\left(-\frac{(y-A)^2}{N_0 / T_b} \right) dy \\
&= Q\left(\frac{A-h}{\sqrt{N_0 / 2T_b}} \right)
\end{aligned}
\tag{5.32}
$$

To evaluate the performance, we can derive the ***average probability of (symbol) error***, P_e , via conditional probability of error weighted by ***a priori*** probability distribution of "1" and "0".

$$\begin{aligned} P_e &= p_0 p_{10} + p_1 p_{01} \\ &= p_0 \mathrm{Q}\left(\frac{A+h}{\sqrt{N_0/2T_b}}\right) + p_1 \mathrm{Q}\left(\frac{A-h}{\sqrt{N_0/2T_b}}\right) \end{aligned} \tag{5.33}$$

Lemma 5.5: (Leibniz's Rule)

Consider the integral $\int_{a(u)}^{b(u)} f(z,u)\,dz$, the derivative of this integral with respect to u is

$$\begin{aligned} &\frac{d}{du}\int_{a(u)}^{b(u)} f(z,u)\,dz \\ &= f(b(u),u)\frac{db(u)}{du} - f(a(u),u)\frac{da(u)}{du} + \int_{a(u)}^{b(u)} \frac{\partial f(z,u)}{\partial u}dz \end{aligned}$$

Differentiating (5.33) with respect to h using Lemma 5.1, the optimal threshold can be derived as

$$h_{\text{opt}} = \frac{N_0/2}{2AT_b}\log\left(\frac{p_0}{p_1}\right) \tag{5.34}$$

For the most interesting case of equally probable signaling $p_1 = p_0 = \frac{1}{2}$, we can have $h_{\text{opt}} = 0$, and thus $p_{01} = p_{10}$. The resulting average probability of error is

$$P_e = Q\left(\frac{A}{\sqrt{N_0/2T_b}}\right) \tag{5.35}$$

Please recall the transmitted energy per bit/symbol is defined as

$$E_b = A^2 T_b \tag{5.36}$$

The average probability of error is therefore a sole function of E_b / N_0 (the ratio of transmitted energy per bit to noise spectral density). A typical plot of error rate is shown in Figure 5.8. Please note the figure is plotted in logarithm scale, and it suggests a very fast decreasing of error rate (more than exponential).

$$P_e = Q\left(\sqrt{\frac{2E_b}{N_0}}\right) \tag{5.37}$$

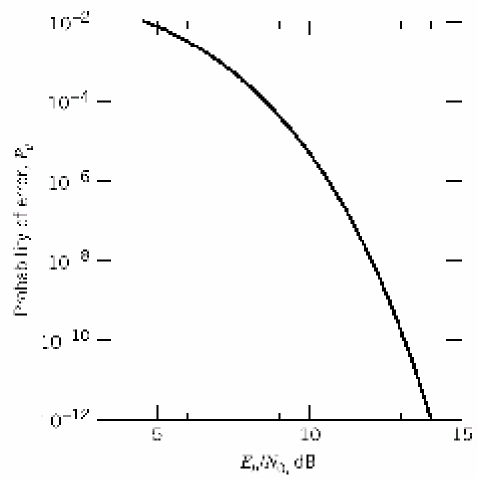


Figure 5.8 Probability of Error for Binary Equally Probable Signalling

►Example 5.2: (4-Level PAM) In above discussion, we consider binary PAM signaling. We may extend into 4-level PAM with signal constellations at $-3A, -A, +A, +3A$ of equally probable distribution. Please note that each pulse (symbol) can carry 2-bit information in this case. The decision thresholds are set at $-2A, 0, 2A$. Assuming normalized symbol duration and using the same receiver structure with above decision thresholds, the symbol error rate, p_e, (not bit error rate) is

$$p_e = (\frac{1}{4}+\frac{1}{4})Q(\frac{A}{\sqrt{N_0/2}}) + (\frac{1}{4}+\frac{1}{4})2Q(\frac{A}{\sqrt{N_0/2}}) = \frac{3}{2}Q(\frac{A}{\sqrt{N_0/2}})$$

Then, we may find the average energy per bit as $E_b = 5A^2$. The average symbol probability of error (symbol error rate) is therefore

$$p_e = \frac{3}{2} Q\left(\sqrt{\frac{E_b/5}{N_0/2}}\right)$$

In order to achieve the same error rate, 4-level PAM uses much more power for transmission than binary PAM.
◄

5.2 LINEAR RECEIVER

As we know from earlier description, signal dispersion can create inter-symbol interference (ISI) in baseband pulse transmission systems. Among so many pulse modulations, we can use pulse amplitude modulation (PAM) as the representative to study such a problem. We may consider a generic binary baseband PAM system as Figure 5.9. The input binary data sequence $\{b_k\}$ consists of symbols "1" and "0" with symbol/bit duration T_b. The PAM modulator translates this binary sequence into a sequence of pulses whose amplitude in polar form

$$a_k = \begin{cases} +1 & \text{if symbol } b_k \text{ is } 1 \\ -1 & \text{if symbol } b_k \text{ is } 0 \end{cases} \tag{5.38}$$

The transmit filter of impulse response $g(t)$ conducting *waveform shaping* delivers the transmitted signal waveform

$$s(t) = \sum_k a_k g(t - kT_b) \tag{5.39}$$

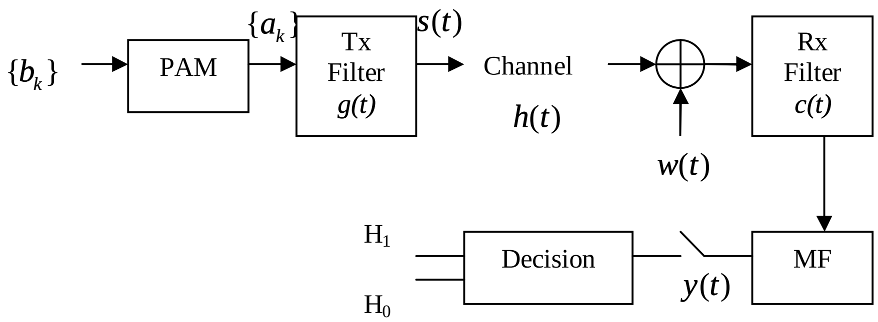


Figure 5.9 Binary Baseband Data Transmission System

The channel is also modeled as a filter of impulse response $h(t)$, with embedded additive noise $w(t)$ that is usually modeled as AWGN. The noisy signal after the channel $x(t)$ is then passed through a receive filter of impulse response $c(t)$. The resulting output is fed into the matched filter whose output $y(t)$ is synchronously sampled with the transmitter. Such *synchronization* is usually achieved by extracting timing information from the received waveform. The decision device has a threshold h to determine whether "1" or "0" is received. The receive filter output is

$$y(t) = m\sum_k a_k p(t - kT_b - t_0) + n(t) \tag{5.39}$$

where m is a scaling factor and t_0 can be considered as zero (perfect synchronization) without loss of generality. The scaled pulse $mp(t)$ with normalization $p(0) = 1$ can be expressed in both time domain and frequency domain as

$$m\,p(t) = g(t) * h(t) * c(t) \tag{5.40}$$

$$mP(f) = G(f)H(f)C(f) \tag{5.41}$$

Where $p(t) \leftrightarrow P(f), h(t) \leftrightarrow H(f), g(t) \leftrightarrow G(f), c(t) \leftrightarrow C(f)$. $n(t)$ is the noise term from the receive filter output. The receive filter output $y(t)$ is sampled at $t_i = iT_b$ to yield

$$\begin{aligned} y(t_i) &= m \sum_{k=-\infty}^{\infty} a_k p[(i-k)T_b] + n(t_i) \\ &= m a_i + m \sum_{\substack{k=-\infty \\ k \neq i}}^{\infty} a_k p[(i-k)T_b] + n(t_i) \end{aligned} \tag{5.42}$$

The first term is contributed from the desired *i*th transmitted bit; the second term is the residual effect of all other transmitted bits before/after the sampling instant and is called ISI; the third term is obviously noise contribution. Without noise and ISI, the desired output is $y(t_i) = m a_i$.

Since the frequency response of the channel and the transmitted pulse shaping are typically specified. The response of transmit and receive filters is determined from the reconstruction of data sequence $\{b_k\}$. The receiver has to estimate (or to decode) $\{a_k\}$ from output $y(t)$. Such a decoding requires the weighted pulse contribution free from ISI, which suggests the control of overall pulse $p(t)$ as

$$p(iT_b - kT_b) = \begin{cases} 1, & i = k \\ 0, & i \neq k \end{cases} \tag{5.43}$$

If $p(t)$ satisfies (5.43), $y(t_i) = m a_i \quad \forall i$ to ensure perfect reception in absence of noise. To transform (5.43) into frequency domain, recalling sampling process, we can have

$$P_d(f) = R_b \sum_{n=-\infty}^{\infty} P(f - nR_b) \tag{5.44}$$

Where $R_b = 1/T_b$ is the bit (symbol) rate; $P_d(f)$ is the Fourier transform of an infinite periodic sequence of delta functions of period T_b. That is,

$$P_d(f) = \int_{-\infty}^{\infty} \sum_{m=-\infty}^{\infty} [p(mT_b)\delta(t - mT_b)]\exp(-j2p\ f\,t)\,dt \tag{5.45}$$

Imposing (5.43) into (5.45), we get

$$\begin{aligned} P_d(f) &= \int_{-\infty}^{\infty} p(0)\delta(t)\exp(-j2p\ f\,t)\,dt \\ &= p(0) \end{aligned} \tag{5.46}$$

By normalization ($p(0) = 1$), we can reach the *Nyquist criterion* for distortionless baseband transmission on absence of noise.

> **Theorem 5.6:** (Nyquist Criterion)
>
> The condition of zero observed ISI is satisfied if
>
> $$\sum_{n=-\infty}^{\infty} P(f - nR_b) = T_b \tag{5.47}$$

The folded frequency function $P(f)$ eliminates inter-symbol interference (ISI) for samples taken at interval T_b provided (5.47) is satisfied. Please note that the folded spectrum in (5.47) and only for samples taken at interval T_b.

The simplest way to facilitate (5.47) is to specify frequency spectrum in the form of a rectangular function

$$\begin{aligned} P(f) &= \begin{cases} \dfrac{1}{2W}, & -W < f < W \\ 0, & |f| > W \end{cases} \\ &= \frac{1}{2W}\mathrm{rect}\left(\frac{f}{2W}\right) \end{aligned} \tag{5.48}$$

The system bandwidth is

$$W = \frac{R_b}{2} = \frac{1}{2T_b} \tag{5.49}$$

The desired pulse function without ISI is thus

$$\begin{aligned} p(t) &= \frac{\sin(2p\,Wt)}{2p\,Wt} \\ &= \mathrm{sinc}(2Wt) \end{aligned} \tag{5.50}$$

$R_b = 2W$ is called *Nyquist rate* and W is the Nyquist bandwidth. (5.48) and (5.50) describe an *ideal Nyquist channel*. There are two

To evaluate the effect of timing error (denoted by Δt), in the absence of noise, we have

$$\begin{aligned} y(\Delta t) &= m\sum_k a_k\, p(\Delta t - kT_b) \\ &= m\sum_k a_k \frac{\sin[2p\,W(\Delta t - kT_b)]}{2p\,W(\Delta t - kT_b)} \end{aligned} \tag{5.51}$$

Since $2WT_b = 1$,

$$y(\Delta t) = m\,a_0 \mathrm{sinc}(2W\Delta t) + \frac{m\sin(2p\,W\Delta t)}{p}\sum_{\substack{k \\ k\neq 0}} \frac{(-1)^k a_k}{(2W\Delta t - k)} \tag{5.52}$$

It is the desired signal plus ISI cause by timing error.

When we are conducting communication system design, zero-forcing technique can be used by meaning to force zero ISI at the sampling instants typically using Nyquist criterion.

5.3 OPTIMAL LINEAR RECEIVER

Up to this point, we treat noise and ISI separately. However, in practical applications, we have to deal both at the same time. A common approach to design a linear receiver consists of a zero-forcing equalizer followed by a decision device. The objective of equalizer is to keep ISI forced to zero at sampling time instants. The *zero-forcing equalizer* is relatively easy to design when ignoring the channel noise, at the price of *noise enhancement*. Another commonly used approach is based on the *minimum mean-squared error* (MMSE) criterion.

Again, considering the scenario of Figure 5.9, the received waveform after receive filter has the following response

$$y(t) = \int_{-\infty}^{\infty} c(t)x(t-t)dt \tag{5.53}$$

The channel output is

$$x(t) = \sum_k a_k q(t-kT_b) + w(t) \tag{5.54}$$

where a_k is the symbol transmitted at time $t = kT_b$ and $w(t)$ is the additive channel noise. $q(t) = g(t) * h(t)$, which can be considered as the pulse shape function after the channel. We can rewrite (5.53) and sample the resulting output waveform $y(t)$ at $t = iT_b$.

$$y(iT_b) = x_i + n_i \tag{5.55}$$

where x_i is contributed from signal and n_i is contributed from noise as

$$x_i = \sum_k a_k \int_{-\infty}^{\infty} c(t)q(iT_b - kT_b - t)dt \tag{5.56}$$

$$n_i = \int_{-\infty}^{\infty} c(t)w(iT_b - t)dt \tag{5.57}$$

We define the error signal is the difference between estimated signal and true signal to result in

$$\begin{aligned} e_i &= y(iT_b) - a_i \\ &= x_i + n_i - a_i \end{aligned} \tag{5.58}$$

The mean squared-error (MSE) is defined as

$$J = \frac{1}{2} E\left[e_i^2\right] \tag{5.59}$$

where 1/2 is introduced as a scaling factor. Substituting (5.58) into (5.59) gives

$$J = \frac{1}{2} E\left[x_i^2\right] + \frac{1}{2} E\left[n_i^2\right] + \frac{1}{2} E\left[a_i^2\right] + E\left[x_i n_i\right] - E\left[n_i a_i\right] - E\left[x_i a_i\right] \tag{5.60}$$

There are 6 terms in above equation, and we evaluate one by one.

- In a stationary environment, $E[x_i^2]$ is independent of sampling time instants, and thus

$$E\left[x_i^2\right] = \sum_l \sum_k E[a_l a_k] \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} c(t_1) c(t_2) q(lT_b - t_1) q(kT_b - t_2) \, dt_1 \, dt_2$$

We first assume binary symbol $a_k = \pm 1$, and the transmitted symbols are statistically independent. Then,

$$E[a_l a_k] = \begin{cases} 1 & \text{for } l = k \\ 0 & \text{otherwise} \end{cases} \tag{5.61}$$

and

$$E\left[x_i^2\right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_q(t_1, t_2) c(t_1) c(t_2) \, dt_1 \, dt_2 \tag{5.62}$$

where the temporal autocorrelation function of the sequence $\{q(kT_b)\}$ is

$$R_q(t_1, t_2) = \sum_k q(kT_b - t_1) q(kT_b - t_2) \tag{5.63}$$

and the stationary sequence means

$$R_q(t_1, t_2) = R_q(t_2 - t_1) = R_q(t_1 - t_2)$$

- The mean-squared noise term is

$$\begin{aligned} E\left[n_i^2\right] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} c(t_1)c(t_2)\; E[w(iT_b - t_1)w(iT_b - t_2)]dt_1\,dt_2 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} c(t_1)c(t_2)\; R_W(t_2 - t_1)dt_1\,dt_2 \end{aligned} \tag{5.64}$$

If we assume $w(t)$ to be white with power spectral density $N_0/2$, the ensemble-averaged autocorrelation function of channel noise $w(t)$ is

$$R_W(t_2 - t_1) = \frac{N_0}{2} d(t_2 - t_1) \tag{5.65}$$

Therefore,

$$E\left[n_i^2\right] = \frac{N_0}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} c(t_1)c(t_2)\; d(t_2 - t_1)dt_1\,dt_2 \tag{5.66}$$

- Since $a_i = \pm 1$, we can easily get

$$E[a_i^2] = 1 \quad \text{for all } i \tag{5.67}$$

- Since x_i and n_i are independent, and n_i has zero mean due to channel noise $w(t)$ zero mean, we have

$$E\left[x_i n_i\right] = 0 \quad \text{for all } i \tag{5.68}$$

- For similar reasons, we have

$$E\left[n_i a_i\right] = 0 \quad \text{for all } i \tag{5.69}$$

- Using (5.56), we have

$$E\left[x_i a_i\right] = \sum_k E\left[a_k a_i\right]\int_{-\infty}^{\infty} c(t)\, q(iT_b - kT_b - t)\, dt \tag{5.70}$$

As earlier independence assumption for transmitted symbols of (5.51), we reduce (5.70) to

$$\left(S_q(f) + \frac{N_0}{2}\right)C(f) = Q^*(f) \tag{5.71}$$

Above derivations result in MSE J for binary data transmission system as

$$J = \frac{1}{2} + \frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(R_q(t-t) + \frac{N_0}{2}d(t-t)\right)c(t)c(t)\,dt\,dt - \int_{-\infty}^{\infty}c(t)q(-t)\,dt \quad (5.72)$$

To design the receiver filter $c(t)$ in Figure 5.9, we differentiate (5.72) with respect to impulse response of receiver filter $c(t)$ and set to zero (necessary condition of optimality). We have

$$\int_{-\infty}^{\infty}\left(R_q(t-t) + \frac{N_0}{2}d(t-t)\right)c(t)\,dt = q(-t) \quad (5.73)$$

Based on (5.73), we can identify $c(t)$ of the equalizer optimized in MSE sense, and such an equalizer is known as ***minimum mean-square error (MMSE) equalizer***. Taking the Fourier transform ($c(t) \leftrightarrow C(f), q(t) \leftrightarrow Q(f), R_q(t) \leftrightarrow S_q(f)$), we obtain

$$\left(S_q(f) + \frac{N_0}{2}\right)C(f) = Q^*(f) \quad (5.74)$$

That is,

$$C(f) = \frac{Q^*(f)}{S_q(f) + \frac{N_0}{2}} \quad (5.75)$$

We can show that the power spectral density of $\{q(kT_b)\}$ can be expressed as

$$S_q(f) = \frac{1}{T_b}\sum_k \left|Q\left(f + \frac{k}{T_b}\right)\right|^2 \quad (5.76)$$

It suggests that the frequency response $C(f)$ of the optimum linear receiver is periodic with period $1/T_b$.

Proposition 5.7: Assuming linear $g(t)$ and $c(t)$, the optimum linear receiver for Figure 5.9 consists of cascading two filters

(a) A matched filter with impulse response $q(-t)$, while $q(t) = g(t) * h(t)$

(b) A transversal tap-delay-line equalizer with frequency response of the inverse of periodic function $S_q(f) + \frac{N_0}{2}$

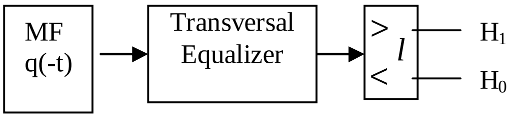


Figure 5.10: Optimum Linear Receiver

To deal with noise and ISI at the same time, we introduce the concept of equalizer. Equalizer can be realized by zero-forcing (ZF) criterion at the price of noise enhancement. We can also use MMSE criterion to strike balanced optimization or both ISI and noise. Given the optimal receiver is synchronous (i.e., perfectly recovered time, frequency, and phase), the optimal receiver is a cascaded filtering of matched filter, equalizer, and decision device (demodulator).

The tap-delay-line equalization/matched filter is usually implemented with delay less than a symbol (due to over-sampling). A special case but very common to use is spacing of adjacent taps at $T/2$ (i.e. half symbol period), and such a structure is named as *fractionally spaced equalizer* (FSE).

5.3.1 Adaptive Equalization

Ideally, if we know the channel response, a straightforward equalizer can be $\frac{1}{H(f)}$. But, $\frac{1}{H(f)}$ may not be realizable. ZF and MMSE provide general design approaches. Unfortunately, the receiver may not know the channel characteristics. Adaptive equalizer invented by Robert Lucky was introduced and has been applied even to modern digital communication systems. Figure 5.11 depicts an adaptive equalizer. Prior to data transmission, a ***training sequence*** known by transmitter and receiver is transmitted first so that the filter coefficients can be adjusted to appropriate values for later data communication. This stage is known as training mode. Then, it is switched to decision-directed mode by using coefficients obtained earlier, and keeps updating based on new decision to result in error signal $e[n]=\hat{a}[n]-y[n]$ driving further update.

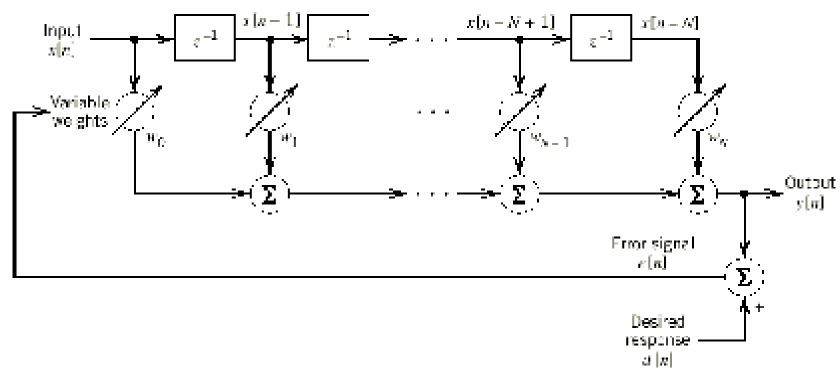


Figure 5.11: Adaptive Equalizer

The adjustment of adaptive equalizer coefficients usually relies on the steep-descent algorithms, and interesting readers can find in literatures about various algorithms and their behaviors.

One popular structure of adaptive filter is ***decision feedback equalizer*** (DFE) as Figure 5.12. The idea is based on

$$\begin{aligned} y[n] &= \sum_k h[k]x[n-k] \\ &= h[0]x[n] + \sum_{k<0} h[k]x[n-k] + \sum_{k>0} h[k]x[n-k] \end{aligned} \tag{5.77}$$

The second term is due to the *precursors* of the channel impulse response and the third term is due to *postcursors*. DFE is widely used in digital communication systems.

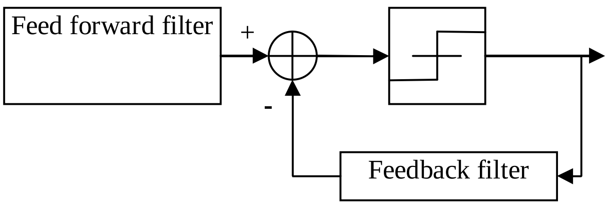


Figure 5.12: Decision Feedback Equalizer

When we observe the ISI from the instrument, we may overlap waveform to see so-called *eye pattern* as Figure 5.13 explains. The widen eye patterns mean little ISI.

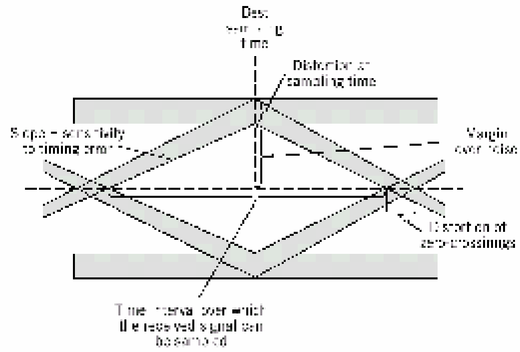
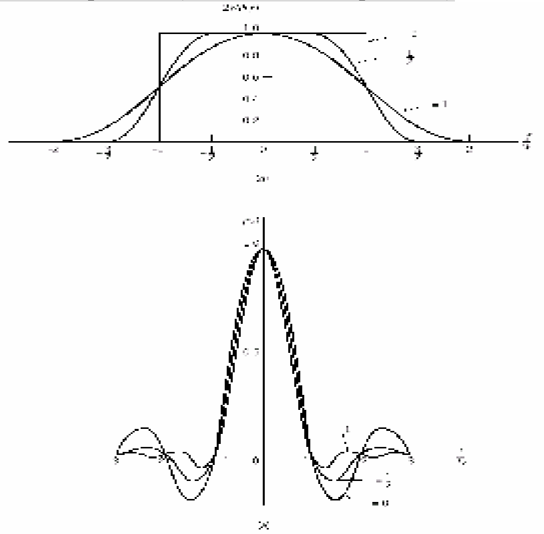


Figure 5.13: Eye Pattern

5.3.2 Line Coding

In addition to receive filter, transmit filter design is another important subject. We generally consider as waveform shaping. The first part is because no ideal Nquist channel existing, and realistic signaling has an excess bandwidth as $B_T = W(1+a)$, where a is the *roll-off factor* to indicate excess bandwidth over the ideal Nyquist bandwidth.

►Example 5.3: (Raised Cosine Spectrum)



◄

Up to this point, we treat ISI as something undesirable. However, a different point of view is to control the ISI to achieve signaling at Nyquist rate. Such a methodology is called *correlative-level coding* or *line coding*, which results in *partial response* signaling. Partial response signaling implies the signaling function over one symbol period, as a counterpart of full response signaling in typical cases. It is common to consider partial response signaling in cable communication and memory storage channels.

The basic idea of correlative coding can be illustrated by the *duo-binary signaling* that is also known as *class I partial response*. Consider a binary input sequence $\{b_k\}$ consisting of uncorrelated binary symbols "1" and "0" with symbol duration T_b. Binary PAM modulator gives

$$a_k = \begin{cases} +1 & \text{if } b_k = 1 \\ -1 & \text{if } b_k = 0 \end{cases} \tag{5.78}$$

The sequence then is applied to *duo-binary encoder*, which converts to a 3-level output, say -2, 0, and +2. Figure 5.15 shows the entire duo-binary signaling mechanism. The encoder output

$$c_k = a_k + a_{k-1} \tag{5.79}$$

Partial response is clear in this case as c_k can not be determined solely by a_k. The frequency response of duo-binary signaling can be found in Figure 5.16 and its merits can be easily observed at the nulls of $\pm 1/2T_b$. However, its drawback is also obvious due to its memory. If there exists any error, it can create further errors at later time, which is known as *error propagation*.

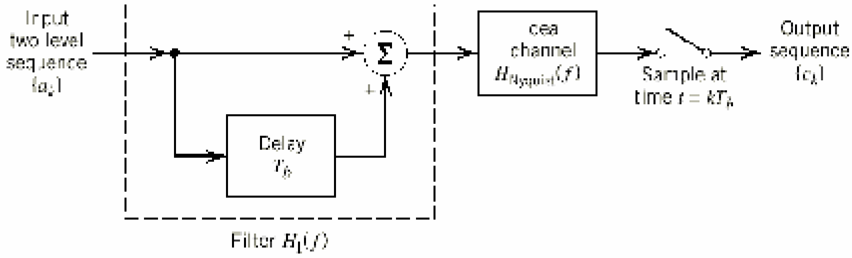
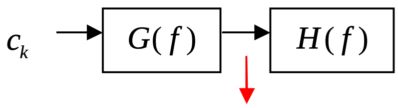


Figure 5.15 Duo-Binary Signalling



$$H_I(f) = H(f)[1 + \exp(-j2\pi fT_b)]$$

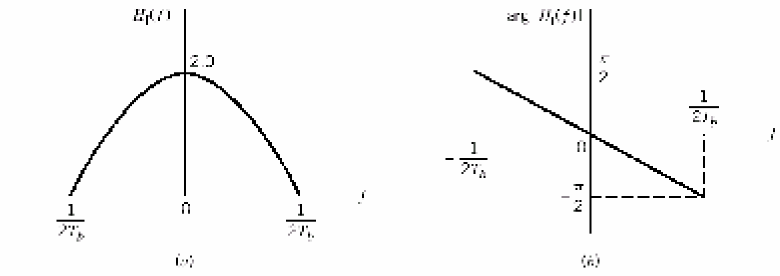


Figure 5.16 Frequency response

The frequency response can be derived as $H_I(f) = 2\cos(\pi fT_b)\exp(-j\pi fT_b)$, which suggests the maximum at DC term and implies waste of energy.

To avoid the error propagation, a brilliant idea named as *precoding* is used as Figure 5.17 showing, while

$$d_k = b_k \oplus d_{k-1} \tag{5.80}$$

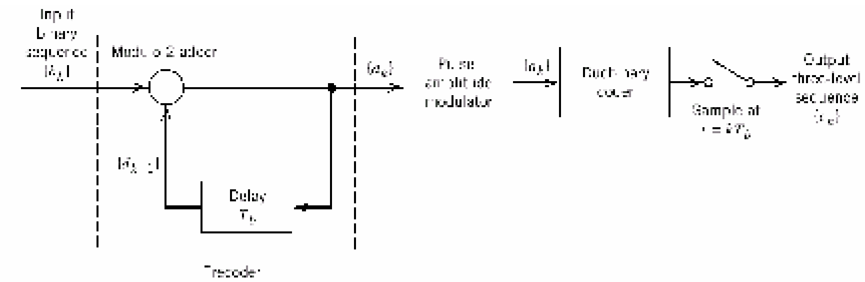


Figure 5.17 Precoding

5.3.3 M-ary PAM and Applications to Telecommunications

Baseband M-ary PAM:

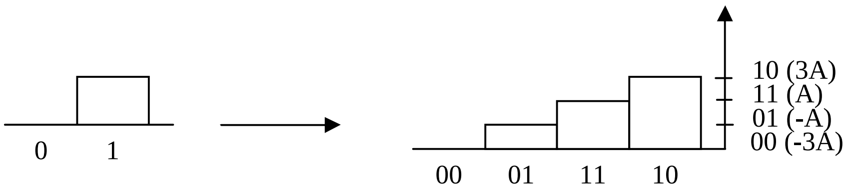


Figure 4.3.1: M-ary PAM

More applications can be summarized:

- Another variation of PAM is PPM that is widely used in optical fiber transmission.
- Subscriber Line (SL): plug-in telephone line
- DSL: transmit digital signal, e.g. modem.
- xDSL: replace the interface at end points, using the same SL in between.

5.4 SIGNAL SPACE ANALYSIS

We have described optimal design of receiver for digital communication system based on pulse transmission. We also conducted rather intuitive error rate analysis. In the following, we shall develop a general treatment of communication theory for digital communications.

Figure 5.18:

To deal with the uncertainty, we usually model phenomenon through statistics, probability, and stochastic process. Modern communication theory is based on the statistical decision theory.

Based on Figure 5.19, we summarize the following assumptions:

- Assuming equally probable inputs, we have

$$\begin{aligned} p_i &= P(m_i) \\ &= \frac{1}{M} \quad \text{for } i = 1,2,\mathbf{K},M \end{aligned} \tag{5.81}$$

- Assuming equal energy for (full-response) signaling, we have

$$\int_0^T s_i^2(t)dt = E_i = E \tag{5.82}$$

- Assuming LTI channel filter.
- Assuming AWGN

Figure 5.19

The received signal is

$$x(t) = s_i(t) + w(t), \qquad \begin{cases} 0 \le t \le T \\ i = 1,2,\ldots,M \end{cases} \tag{5.83}$$

The performance measure, average probability of error is

$$P_e = \sum_{i=1}^{M} p_i P(\hat{m} \ne m_i \,|\, m_i) \tag{5.84}$$

To develop a general theoretical framework, we use orthonormal function expansions

$$s_i(t) = \sum_{j=1}^{N} s_{ij} f_j(t), \qquad \begin{cases} 0 \le t \le T \\ i = 1,2,\ldots,M \end{cases} \tag{5.85}$$

$$s_{ij} = \int_0^T s_i(t) f_j(t)\,dt, \qquad \begin{cases} i = 1,2,\ldots,M \\ j = 1,2,\ldots,N \end{cases} \tag{5.86}$$

$$\int_0^T f_i(t) f_j(t)\,dt = d_{ij} = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \ne j \end{cases} \tag{5.87}$$

We may treat the N-dimensional vector $\mathbf{s}_i^T = (s_{i1}, s_{i2}, \cdots, s_{iN})$ and it corresponds to $s_i(t)$, for $i = 1,2,\cdots,M$, while T denotes transpose. Or, we may represent as

$$\mathbf{s}_i = \begin{bmatrix} s_{i1} \\ s_{i2} \\ \vdots \\ s_{iN} \end{bmatrix}, \qquad i = 1,2,\ldots,M \tag{5.88}$$

which is known as a signal vector. We may further visualize a set of M signal constellation points in N-dimensional space that is called ***signal space***. It is easy to imagine such a picture under Euclidean space.

Under the concept of signal space, we want to inspect the behaviors of noise process via the following Lemma.

Lemma 5.8: (White Gaussian Noise Characterization)

Suppose $n(t)$ is a white Gaussian noise process with zero mean and power spectral density $N_0/2$. $\{f_i(t)\}_{i=1}^{\infty}$ form an orthonormal basis.

$$n_i = \int_0^T n(t) f_i(t) dt$$

(a) $\{n_i\}_{i=1}^{\infty}$ are jointly Gaussian

(b) $\{n_i\}_{i=1}^{\infty}$ are mutually uncorrelated and thus mutually independent.

(c) $\{n_i\}_{i=1}^{\infty}$ are i.i.d. Gaussian with zero mean and spectral variance $N_0/2$.

We generally treat the signal space as an *N*-dimensional vector space. The squared-length (or squared-magnitude) of a signal vector $\mathbf{s}_i$ is

$$\begin{aligned} \|\mathbf{s}_i\|^2 &= \mathbf{s}_i^{\mathrm{T}} \mathbf{s}_i \\ &= \sum_{j=1}^{N} s_{ij}^2, \qquad i = 1, 2, \mathbf{K}, M \end{aligned} \tag{5.89}$$

The signal energy of $s_i(t)$ with duration T is

$$E_i = \int_0^T s_i^2(t)\, dt \tag{5.90}$$

Its equivalent form is

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

By interchanging the order of summation and integration, we have

$$E_i = \sum_{j=1}^{N} \sum_{k=1}^{N} s_{ij} s_{ik} \int_0^T f_j(t) f_k(t)\, dt \tag{5.91}$$

Since the orthonormal property, (5.91) reduces to

$$
\begin{aligned}
E_i &= \sum_{j=1}^{N} s_{ij}^2 \\
&= \|\mathbf{s}_i\|^2
\end{aligned}
\tag{5.92}
$$

For a pair of signals $s_i(t)$ and $s_k(t)$,

$$
\int_0^T s_i(t)\, s_k(t)\, dt = \mathbf{s}_i^{\mathrm{T}} \mathbf{s}_k \tag{5.93}
$$

It states the equivalence of inner product between time-domain representation and signal space vector form. The squared Euclidean distance d_{ik}^2 is described as

$$
\begin{aligned}
\|\mathbf{s}_i - \mathbf{s}_k\|^2 &= \sum_{j=1}^{N} (s_{ij} - s_{kj})^2 \\
&= \int_0^T (s_i(t) - s_k(t))^2\, dt
\end{aligned}
\tag{5.94}
$$

The angle between two signal vectors is

$$
\cos q_{ik} = \frac{\mathbf{s}_i^{\mathrm{T}} \mathbf{s}_k}{\|\mathbf{s}_i\|\,\|\mathbf{s}_k\|} \tag{5.95}
$$

►Example 5.3: (Schwartz Inequality) Considering a pair of signals, Schwartz inequality states

$$
\left(\int_{-\infty}^{\infty} s_1(t)\, s_2(t)\, dt \right)^2 \le \left(\int_{-\infty}^{\infty} s_1^2(t)\, dt \right) \left(\int_{-\infty}^{\infty} s_2^2(t)\, dt \right) \tag{5.96}
$$

with equality holds as $s_2(t) = c s_1(t)$ and c is a constant. We can prove by expressing both signals in terms of two orthogonal basis functions.

$$
\begin{aligned}
s_1(t) &= s_{11} f_1(t) + s_{12} f_2(t) \\
s_2(t) &= s_{21} f_1(t) + s_{22} f_2(t)
\end{aligned}
$$

where

$$
\int_{-\infty}^{\infty} f_i(t) f_j(t)\, dt = d_{ij} = \begin{cases} 1 & \text{for } j = i \\ 0 & \text{otherwise} \end{cases}
$$

The vector form is

$$
\mathbf{s}_1 = \begin{bmatrix} s_{11} \\ s_{12} \end{bmatrix}
$$

$$\mathbf{s}_2 = \begin{bmatrix} s_{21} \\ s_{22} \end{bmatrix}$$

The angle is

$$\begin{aligned} \cos q &= \frac{\mathbf{s}_1^{\mathrm{T}} \mathbf{s}_2}{\|\mathbf{s}_1\| \, \|\mathbf{s}_2\|} \\ &= \frac{\int_{-\infty}^{\infty} s_1(t) \, s_2(t) \, dt}{\left(\int_{-\infty}^{\infty} s_1^2(t) \, dt \right)^{1/2} \left(\int_{-\infty}^{\infty} s_2^2(t) \, dt \right)^{1/2}} \end{aligned} \tag{5.97}$$

$|\cos q| \leq 1$ with equality holds if and only if $\mathbf{s_2} = c\mathbf{s_1}$ and we have

$$\left| \int_{-\infty}^{\infty} s_1(t) \, s_2^*(t) \, dt \right|^2 \leq \left(\int_{-\infty}^{\infty} |s_1(t)|^2 \, dt \right) \left(\int_{-\infty}^{\infty} |s_2(t)|^2 \, dt \right) \tag{5.98}$$

◄

Above example shows the beauty of geometric representation of signals in vector space. The next challenge would be how to determine the appropriate orthonormal basis functions. From vector analysis, this can be facilitated via ***Gram-Schmidt (orthogonalization) procedure***. Suppose we have a set of M signals denoted by $s_1(t), \cdots, s_M(t)$. We arbitrarily select $s_1(t)$ and define the first basis function as

$$f_1(t) = \frac{s_1(t)}{\sqrt{E_1}} \tag{5.99}$$

where E_1 is the energy of signal $s_1(t)$. We therefore have

$$\begin{aligned} s_1(t) &= \sqrt{E_1} f_1(t) \\ &= s_{11} \, f_1(t) \end{aligned} \tag{5.100}$$

$f_1(t)$ clearly has unit energy. Next, we define coefficient s_{21} using signal $s_2(t)$

$$s_{21} = \int_0^T s_2(t) f_1(t) \, dt \tag{5.101}$$

Introducing a new intermediate function orthogonal to $f_1(t)$,

$$g_2(t) = s_2(t) - s_{21} f_1(t) \tag{5.102}$$

The second basis function can be defined as

$$f_2(t) = \frac{g_2(t)}{\sqrt{\int_0^T g_2^2(t)\,dt}} \tag{5.103}$$

By further simplifying and using E_2 as the energy of $s_2(t)$, we get

$$f_2(t) = \frac{s_2(t) - s_{21} f_1(t)}{\sqrt{E_2 - s_{21}^2}} \tag{5.104}$$

It is clear that

$$\int_0^T f_2^2(t)\,dt = 1$$

$$\int_0^T f_1(t) f_2(t)\,dt = 0$$

That is, $f_1(t)$ and $f_2(t)$ form an orthonormal pair. Continuing the same process, we generally define

$$g_i(t) = s_i(t) - \sum_{j=1}^{i-1} s_{ij} f_j(t) \tag{5.105}$$

$$s_{ij} = \int_0^T s_i(t) f_j(t)\,dt, \qquad j = 1,2,\mathbf{K}, i-1 \tag{5.106}$$

The set of basis functions

$$f_i(t) = \frac{g_i(t)}{\sqrt{\int_0^T g_i^2(t)\,dt}}, \qquad i = 1,2,\mathbf{K}, N \tag{5.107}$$

form an orthonormal set. The signal $s_1(t),\cdot\cdot, s_M(t)$ form a linearly independent set for $N = M$. The special case for $N<M$ results in $g_i(t) = 0, i > N$ and signals are not linearly independent.

Then we are ready to convert continuous-time AWGN channel into a vector channel form for in-depth development of mathematical framework. The AWGN channel output is fed into a bank of N correlators as Figure 5.20.

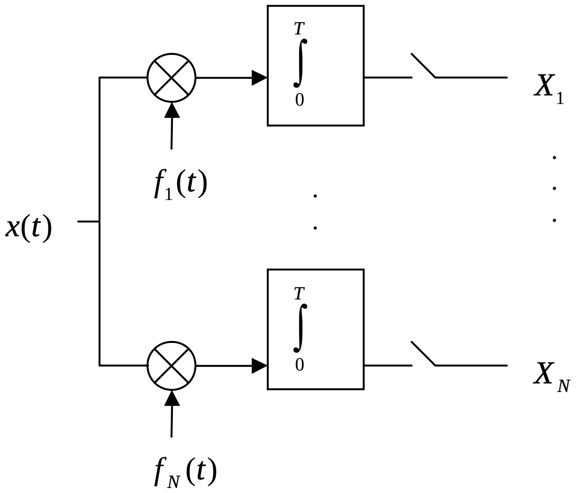


Figure 5.20: Bank of Correlators

$$x(t) = s_i(t) + w(t), \qquad \begin{cases} 0 \le t \le T \\ i = 1,2,\ldots,M \end{cases} \tag{5.108}$$

where $w(t)$ is the sampled function of a white Gaussian noise process $W(t)$ with zero mean and power spectral density $N_0/2$. The output of correlator j is the sampled value of a random variable X_j as

$$\begin{aligned} x_j &= \int_0^T x(t) f_j(t)\,dt \\ &= s_{ij} + w_j, \qquad j = 1,2,\ldots,N \end{aligned} \tag{5.109}$$

Contributed from signal and noise respectively, both terms in (5.109) are also

$$s_{ij} = \int_0^T s_i(t) f_j(t)\,dt \tag{5.110}$$

$$w_j = \int_0^T w(t) f_j(t)\,dt \tag{5.111}$$

We now consider a new random process $X'(t)$ whose sample function is $x'(t)$ with the following relationship.

$$x'(t) = x(t) - \sum_{j=1}^{N} x_j f_j(t) \tag{5.112}$$

Substituting (5.108) and (5.109) into (5.112), with expansion, we have

$$\begin{aligned} x'(t) &= s_i(t) + w(t) - \sum_{j=1}^{N} (s_{ij} + w_j) f_j(t) \\ &= w(t) - \sum_{j=1}^{N} w_j f_j(t) \\ &= w'(t) \end{aligned} \tag{5.113}$$

The received signal is consequently

$$\begin{aligned} x(t) &= \sum_{j=1}^{N} x_j f_j(t) + x'(t) \\ &= \sum_{j=1}^{N} x_j f_j(t) + w'(t) \end{aligned} \tag{5.114}$$

The next step is to characterize the N correlator outputs. Let $X(t)$ denote the random process whose sampled function is $x(t)$. $X(t)$ is a Gaussian process. As Chapter 2, X_j are Gaussian random variables, $j = 1, 2, \cdots, N$. To determine its mean and variance to characterize outputs,

$$\begin{aligned} m_{X_j} &= E[X_j] \\ &= E\left[s_{ij} + W_j\right] \\ &= s_{ij} + E\left[W_j\right] \\ &= s_{ij} \end{aligned} \tag{5.115}$$

and

$$\begin{aligned} s_{X_j}^2 &= \operatorname{var}\left[X_j\right] \\ &= E\left[(X_j - s_{ij})^2\right] \\ &= E\left[W_j^2\right] \end{aligned} \tag{5.116}$$

From (1.111),

$$W_j = \int_0^T W(t) f_j(t)\, dt$$

By expansion,

$$\begin{aligned} s_{X_j}^2 &= E\left[\int_0^T W(t) f_j(t)\, dt \int_0^T W(u) f_j(u)\, du\right] \\ &= E\left[\int_0^T \int_0^T f_j(t)\, f_j(u) W(t) W(u)\, dt\, du\right] \end{aligned} \tag{5.117}$$

Interchanging expectation
and integration,

$$\begin{aligned} s_{X_j}^2 &= \int_0^T \int_0^T f_j(t)\, f_j(u)\, E[W(t)W(u)] dt\, du \\ &= \int_0^T \int_0^T f_j(t)\, f_j(u)\, R_W(t,u)\, dt\, du \end{aligned} \tag{5.118}$$

where $R_W(t,u)$ is the autocorrelation function of noise process. Due to white noise property,

$$R_W(t,u) = \frac{N_0}{2} d(t-u) \tag{5.119}$$

The variance is

$$\begin{aligned} s_{X_j}^2 &= \frac{N_0}{2} \int_0^T \int_0^T f_j(t)\, f_j(u) d(t-u)\, dt\, du \\ &= \frac{N_0}{2} \int_0^T f_j^2(t)\, dt \end{aligned} \tag{5.120}$$

Due to the unit energy property of basis function, we have

$$s_{X_j}^2 = \frac{N_0}{2} \quad \text{for all } j \tag{5.121}$$

It suggests that correlator outputs have the same variance as noise process. Furthermore, since $\{f(t)\}$ form an orthogonal set, X_j are mutually uncorrelated as

$$
\begin{aligned}
\operatorname{cov}[X_j X_k] &= E[(X_j - m_{X_j})(X_k - m_{X_k})] \\
&= E\left[(X_j - s_{ij})(X_k - s_{ik})\right] \\
&= E[W_j W_k] \\
&= E\left[\int_0^T W(t) f_j(t)\,dt \int_0^T W(u) f_k(u)\,du\right] \\
&= \int_0^T \int_0^T f_j(t)\, f_k(u) R_W(t,u)\,dt\,du \qquad (5.122) \\
&= \frac{N_0}{2} \int_0^T \int_0^T f_j(t)\, f_k(u) \delta(t-u)\,dt\,du \\
&= \frac{N_0}{2} \int_0^T f_j(t) f_k(t)\,dt \\
&= 0, \qquad j \neq k
\end{aligned}
$$

For Gaussian random variables, uncorrelated property suggests independence.

Define the vector of N random variables as

$$
\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} \qquad (5.123)
$$

The elements of (5.123) are independent Gaussian random variables with mean s_{ij} and variance $N_0/2$. The conditional probability density function of $\mathbf{X}$ given that $s_i(t)$ (or symbol m_i) is transmitted is

$$
f_{\mathrm{X}}(\mathbf{x}|m_i) = \prod_{j=1}^{N} f_{\mathrm{X_j}}(x_j|m_i), \qquad i = 1,2,\ldots,M \qquad (5.124)
$$

where the product form comes from the independence. The vector $\mathbf{x}$ is the *observation* vector. The channel satisfies (5.124) is a *memoryless* channel. Using Gaussian mean and variance, we have

$$
f_{\mathrm{X_j}}(x_j|m_i) = \frac{1}{\sqrt{\pi N_0}} \exp\left[-\frac{1}{N_0}(x_j - s_{ij})^2\right], \quad \begin{array}{l} j = 1,2,\ldots,N \\ i = 1,2,\ldots,M \end{array} \qquad (5.125)
$$

The resulting conditional probability is

$$f_X(\mathbf{x}|m_i) = (pN_0)^{-N/2} \exp\left[-\frac{1}{N_0}\sum_{j=1}^{N}(x_j - s_{ij})^2\right], \qquad i = 1,2,\mathbf{K},M \quad (5.126)$$

Since the noise process $W(t)$ is Gaussian with zero mean, $W'(t)$ represented by the sample function $w'(t)$ is also a zero-mean Gaussian process. Actually, $W'(t_k)$ sampling at any time t_k is independent of $\{X_j\}$. That is,

$$E[X_j W'(t_k)] = 0, \qquad \begin{cases} j = 1,2,\mathbf{K},N \\ 0 \le t_k \le T \end{cases} \quad (5.127)$$

(5.127) suggests that the random variable $W'(t_k)$ is irrelevant to the decision of transmitted signal. In other words, the correlator outputs determined by the received waveform $x(t)$ are the only data useful for decision-making, and we call *sufficient statistics* for the decision.

Theorem 5.9: (Theorem of Irrelevance) For signal detection in AWGN, only the projected terms onto basis functions of signal $\{s_i(t)\}_{i=1}^{M}$ provides sufficient statistics of the decision problem. All other irrelevant information can not help detection at all.

Corollary 5.10: AWGN channel of Figure 5.21 is equivalent to an N-dimensional vector channel described by the observation vector

$$\mathbf{x} = \mathbf{s}_i + \mathbf{w}, \qquad i = 1,2,\mathbf{K},M \quad (5.128)$$

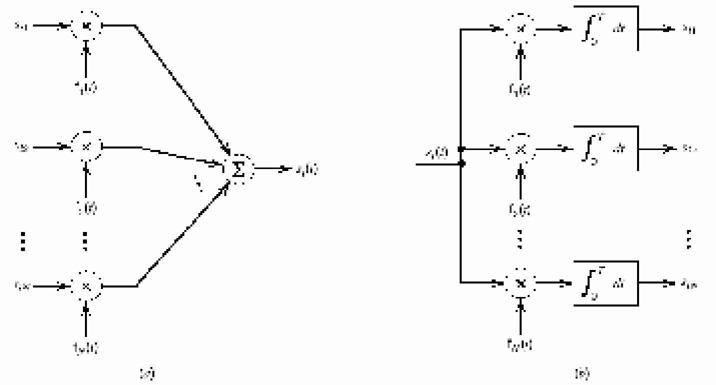


Figure 5.21 AWGN Channel by Functional Expansion

To design the optimum receiver, we have to consider the problem: Given observation vector $\mathbf{x}$, we would like to estimate the message symbol m_i generating $\mathbf{x}$. As we have obtained conditional probability density function $f_X(\mathbf{x}|m_i), i = 1,2,\mathbf{K},M$ to characterize AWGN, *likelihood function* denoted by $L(m_i)$ is introduced to design the receiver.

$$L(m_i) = f_X(\mathbf{x}|m_i), \qquad i = 1,2,\mathbf{K},M \tag{5.129}$$

In many cases, we find more convenient to use *log-likelihood function* as

$$l(m_i) = \log L(m_i), \qquad i = 1,2,\mathbf{K},M \tag{5.130}$$

Since the probability function is always non-negative and logarithm function is monotonically increasing, log-likelihood function is one-to-one mapping with likelihood function. For AWGN, by ignoring the constant term, we have

$$l(m_i) = -\frac{1}{N_0}\sum_{j=1}^{N}(x_j - s_{ij})^2, \qquad i = 1,2,\mathbf{K},M \tag{5.131}$$

Please note the engineering meaning of (3.131), which is the squared Euclidean distance normalized by noise spectral density, just like E_b / N_0 or SNR.

5.5 CORRELATION RECEIVER

IN the following, we will use concept of the signal space to develop *maximum likelihood* detection/decoding. Suppose in a symbol duration *T*, one of the possible signals $s_1(t),\cdot\cdot,s_M(t)$ with equal *a priori* probability $1/M$. By a bank of correlators, signal vector $\mathbf{s_i}$ is equivalent to $s_i(t)$ in such signal detection as long as $N \le M$. The set of message points corresponding to $\{s_i(t)\}_{i=1}^{M}$ is called the *signal constellation*.

> **Definition 5.11:** (Signal Detection Problem) Given the observation vector $\mathbf{x}$, we determine a *decision rule* (actually a mapping) to obtain estimate $\hat{m}$ of the transmitted symbol m_i, so that the probability of error can be minimized.

Given $\mathbf{x}$, we decide $\hat{m} = m_i$. The probability of error is

$$P_e(m_i|\mathbf{x}) = P(m_i \text{ not sent}|\mathbf{x}) \\ = 1 - P(m_i \text{ sent}|\mathbf{x}) \tag{5.132}$$

The decision criterion is to minimize probability of error in mapping each given observation vector $\mathbf{x}$ to a decision. The *optimum decision rule* is stated as follows.

$$\text{Set } \hat{m} = m_i \text{ if} \\ P(m_i \text{ sent}|\mathbf{x}) \geq P(m_k \text{ sent}|\mathbf{x}) \quad k = 1,2,\cdots,M \text{ and } k \neq i \tag{5.133}$$

This decision rule is also known as *maximum a posteriori probability (MAP)* rule. Using Baye's rule and signal *a priori* probability p_k to transmit m_k, MAP rule can be re-stated as

$$\text{Set } \hat{m} = m_i \text{ if} \\ \frac{p_k f_X(\mathbf{x}|m_k)}{f_X(\mathbf{x})} \text{ is maximum for } k = i \tag{5.134}$$

Since $f_X(\mathbf{x})$ is independent of transmitted signal and equal *a priori* probability p_k, the conditional probability density function having one-to-one relationship to log-likelihood function gives

$$\text{Set } \hat{m} = m_i \text{ if} \\ l(m_k) \text{ is maximum for } k = i \tag{5.135}$$

(5.135) represents the *maximum likelihood (ML)* rule. The mechanism to implement this rule is a maximum likelihood decoder/detector. Please note that we use equal *a priori* probability to get ML detector from MAP detector.

Leveraging signal space concept, we can graphically explain the decision rule. Let Z denote the N-dimensional space of all possible observation vectors $\mathbf{x}$, which forms the *observation space*. The total observation space Z is partitioned into M decision regions corresponding to possible signals, and these decision regions are denoted by $Z_1,\cdots,Z_M$. The decision rule is now

$$\text{Observation vector } \mathbf{x} \text{ lies in region } Z_i \text{ if} \\ l(m_k) \text{ is maximum for } k = i \tag{5.136}$$

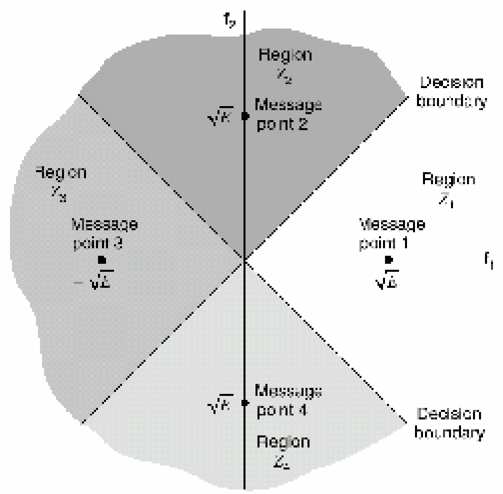


Figure 5.22 Decision Regions in Signal Space

From (5.131) for AWGN, $l(m_k)$ attains its maximum when

$$\sum_{j=1}^{N}(x_j - s_{kj})^2$$

is minimized by selecting $k = i$. The ML decision rule for AWGN is

$$\begin{gathered}\text{Observation vector } \mathbf{x} \text{ lies in region } Z_i \text{ if} \\ \sum_{j=1}^{N}(x_j - s_{kj})^2 \text{ is minimum for } k = i\end{gathered} \tag{5.137}$$

We can use the definition of Euclidean distance to get

$$\sum_{j=1}^{N}(x_j - s_{kj})^2 = \|\mathbf{x} - \mathbf{s}_k\|^2 \tag{5.138}$$

The decision rule is now as

$$\begin{gathered}\text{Observation vector } \mathbf{x} \text{ lies in region } Z_i \text{ if} \\ \text{the Euclidean distance } \|\mathbf{x} - \mathbf{s}_k\| \text{ is minimum for } k = i\end{gathered} \tag{5.139}$$

> **Theorem 5.12:** The ML decision rule is simply to select the message point closest (w.r.t. Euclidean distance) to the received signal.

Straightforward algebra gives

$$\sum_{j=1}^{N}(x_j - s_{kj})^2 = \sum_{j=1}^{N} x_j^2 - 2\sum_{j=1}^{N} x_j s_{kj} + \sum_{j=1}^{N} s_{kj}^2 \tag{5.140}$$

The first term is independent of k and can be ignored. The second term is the inner product of the observation vector $\mathbf{x}$ and signal vector $\mathbf{s_k}$. The third term is the energy of the transmitted signal. Consequently, the decision rule is

$$\begin{aligned}&\text{Observation vector } \mathbf{x} \text{ lies in region } \mathrm{Z_i} \text{ if}\\ &\sum_{j=1}^{N} x_j s_{kj} - \frac{1}{2}E_k \text{ is maximum for } k = i\end{aligned} \tag{5.141}$$

where

$$E_k = \sum_{j=1}^{N} s_{kj}^2 \tag{5.142}$$

Based on above development, the optimum receiver for AWGN channel involves two sub-systems as Figure 5.23:

(a) A bank of correlators (product-integrator) corresponding to a set of orthonormal basis functions.

(b) A correlation metric calculator to compute (5.141) for each possible signal

Such a structure is known as *correlation receiver*.

> **Theorem 5.13:** The correlation receiver is equivalent to the matched filter in AWGN channels.

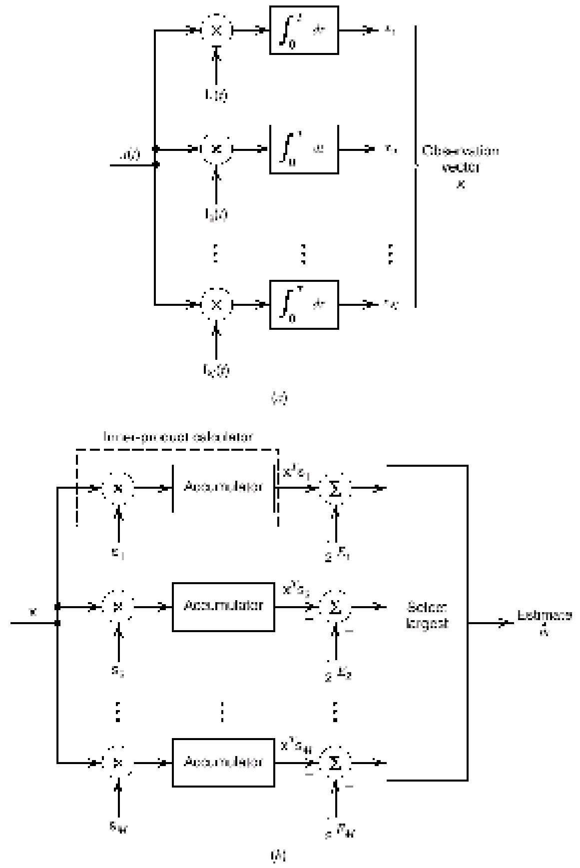


Figure 5.23 (a) Waveform Projection (b) Correlation Metric Calculator

Proof: With received waveform as input, the filter output is

$$y_j(t)=\int_{-\infty}^{\infty} x(t)h_j(t-t)\,dt \tag{5.143}$$

The matched filter is thus

$$h_j(t)=f_j(T-t) \tag{5.144}$$

The resulting filter output is

$$y_j(t)=\int_{-\infty}^{\infty} x(t)f_j(T-t+t)\,dt \tag{5.145}$$

At time $t=T$,

$$y_j(T)=\int_{-\infty}^{\infty} x(t)f_j(t)\,dt$$

Then,

$$y_j(T) = \int_0^T x(t) f_j(t)\, dt \tag{5.146}$$

■

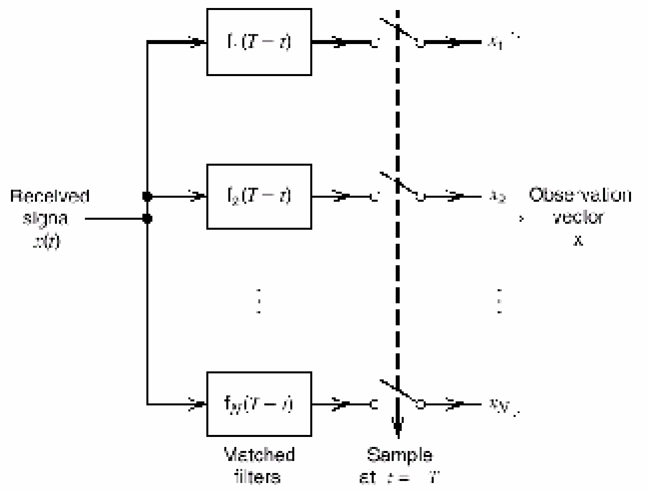


Figure 5.24 Matched Filter Receiver

To learn the effectiveness of receiver, we shall evaluate the probability of error associated with the receiver. We continue the concept of partitioning observation space Z into M regions $\{Z_i\}_{i=1}^{M}$, following the maximum likelihood decision rule. Suppose the signal m_i is transmitted and the observation vector $\mathbf{x}$ is received from channel. The error occurs when $\mathbf{x}$ does not fall into Z_i associated with m_i.

$$\begin{aligned} P_e &= \sum_{i=1}^{M} p_i\, P(\mathbf{x} \text{ does not lie in } Z_i | m_i \text{ sent}) \\ &= \frac{1}{M}\sum_{i=1}^{M} P(\mathbf{x} \text{ does not lie in } Z_i | m_i \text{ sent}) \\ &= 1 - \frac{1}{M}\sum_{i=1}^{M} P(\mathbf{x} \text{ lies in } Z_i | m_i \text{ sent}) \end{aligned} \tag{5.147}$$

We can rewrite the equation in terms of likelihood function as the following (likewise N-dimensional) integral equation.

$$P_e = 1 - \frac{1}{M}\sum_{i=1}^{M} \int_{Z_i} f_{\mathrm{X}}(\mathbf{x}|m_i)\, d\mathbf{x} \tag{5.148}$$

As a summary, via partitioning observation space into regions $Z_1, Z_2, \cdots, Z_M$, the ML detection of a signal in AWGN is uniquely defined by the signal constellation, which has the error rate derived as above. We may conclude as follows.

Proposition 5.14: In ML detection of signal in AWGN, the probability of (symbol) error solely depends on the Euclidean distance between the message signal constellation points.

Lemma 5.15: AWGN is spherically symmetric in signal space. The rotated noise vector is also Gaussian with zero mean. Furthermore, the components of rotated noise vector are uncorrelated and thus independent.

Proof: Let $\mathbf{I}$ denote an identity matrix. The N-dimensional signal vector $\mathbf{s}_i$ multiplying an $N \times N$ orthonormal (rotating) matrix $\mathbf{Q}$ means an rotation operation, where

$$\mathbf{Q}\mathbf{Q}^T = \mathbf{I} \tag{5.149}$$

The rotated signal vector is

$$\mathbf{s}_{i,\,\text{rotate}} = \mathbf{Q}\mathbf{s}_i, \qquad i = 1,2,\ldots,M \tag{5.150}$$

The corresponding noise vector $\mathbf{w}$ after rotation is

$$\mathbf{w}_{\text{rotate}} = \mathbf{Q}\mathbf{w} \tag{5.151}$$

Since linear transformation (or combination) of Gaussian random variables is jointly Gaussian, $\mathbf{w}_{\text{rotate}}$ is Gaussian with mean

$$\begin{aligned} E[\mathbf{w}_{\text{rotate}}] &= E[\mathbf{Q}\mathbf{w}] \\ &= \mathbf{Q}E[\mathbf{w}] \\ &= \mathbf{0} \end{aligned} \tag{5.152}$$

The covariance matrix of the noise vector is

$$E[\mathbf{w}\mathbf{w}^T] = \frac{N_0}{2}\mathbf{I} \tag{5.153}$$

Consequently, the covariance matrix of the rotated noise vector $\mathbf{w}_{\text{rotate}}$ is

$$\begin{aligned} E[\mathbf{w}_{\text{rotate}}\mathbf{w}_{\text{rotate}}^T] &= E[\mathbf{Q}\mathbf{w}(\mathbf{Q}\mathbf{w})^T] \\ &= E[\mathbf{Q}\mathbf{w}\mathbf{w}^T\mathbf{Q}^T] \\ &= \mathbf{Q}E[\mathbf{w}\mathbf{w}^T]\mathbf{Q}^T \\ &= \frac{N_0}{2}\mathbf{Q}\mathbf{Q}^T \\ &= \frac{N_0}{2}\mathbf{I} \end{aligned} \tag{5.154}$$

■

The observation vector for the rotated signal constellation can be expressed as

$$\mathbf{x}_{\text{rotate}} = \mathbf{Q}\mathbf{s}_i + \mathbf{w}, \qquad i = 1,2,...,M \tag{5.155}$$

As the decision rule of ML detection depend solely on the Euclidean distance between $\mathbf{x}_{\text{rotate}}$ and $\mathbf{s}_{i,\text{ rotate}}$, it is straightforward to conclude

$$\left\|\mathbf{x}_{\text{rotate}} - \mathbf{s}_{i,\text{ rotate}}\right\| = \left\|\mathbf{x} - \mathbf{s}_i\right\| \qquad \text{for all } i \tag{5.156}$$

Proposition 5.16: (Principle of Rotational Invariance) If a signal constellation is rotated by an orthonormal transformation $\mathbf{s}_{i,\text{ rotate}} = \mathbf{Q}\mathbf{s}_i, i = 1,2,...,M$, the probability of error associated with ML detection in AWGN remains the same.

Proposition 5.17: (Principle of Translation Invariance) If a signal constellation is translated by a constant vector, the probability of error associated with ML detection in AWGN remains the same.

Proof: Suppose all message points in a signal constellation are translated by a constant vector $\mathbf{c}$. That is,

$$\mathbf{s}_{i,\text{ translate}} = \mathbf{s}_i - \mathbf{c}, \qquad i = 1,2,...,M \tag{5.157}$$

The observation vector is correspondingly translated by the same constant vector as

$$\mathbf{x}_{\text{translate}} = \mathbf{x} - \mathbf{c} \tag{5.158}$$

It is obvious to reach

$$\left\|\mathbf{x}_{\text{translate}} - \mathbf{s}_{i,\,\text{translate}}\right\| = \left\|\mathbf{x} - \mathbf{s}_i\right\| \qquad \text{for all } i \tag{5.159}$$

■

►Example 5.4: Figure 5.25 illustrates a case of rotational invariance. Figure 5.26 depicts another case of translational invariance.

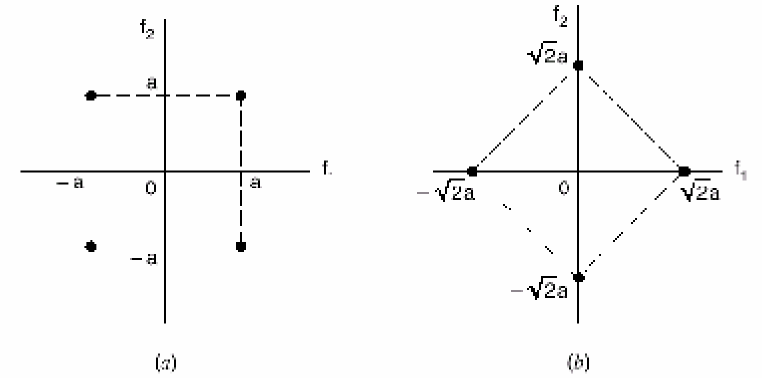


Figure 5.25

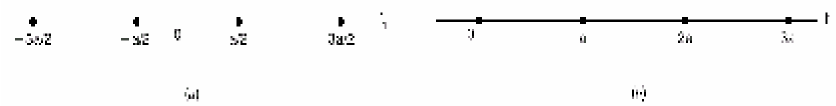


Figure 5.26
◄

As Chapter 1 mentioned, many challenges in communication system design lie in power-limited channel, and thus design of the minimum energy signals under the same probability of error (using Propositions 5.16 and 5.17) is very often an important task. To inspect this issue, we again consider a set of signals $m_1, m_2, \cdots, m_M$ represented by the signal vectors $\mathbf{s}_1, \mathbf{s}_{2,\cdots,}\mathbf{s}_M$. The average energy of this signal constellation translated by a vector $\mathbf{a}$ with p_i as *a priori* probability for m_i is

$$\mathcal{E}_{\text{translate}} = \sum_{i=1}^{M} \|\mathbf{s}_i - \mathbf{a}\|^2 p_i \tag{5.160}$$

The squared Euclidean distance between $\mathbf{s}_i$ and $\mathbf{a}$ is

$$\|\mathbf{s}_i - \mathbf{a}\|^2 = \|\mathbf{s}_i\|^2 - 2\mathbf{a}^T \mathbf{s}_i + \|\mathbf{a}\|^2$$

Therefore, the average energy is

$$\begin{aligned}\mathcal{E}_{\text{translate}} &= \sum_{i=1}^{M} \|\mathbf{s}_i\|^2 p_i - 2\sum_{i=1}^{M} \mathbf{a}^T \mathbf{s}_i p_i + \|\mathbf{a}\|^2 \sum_{i=1}^{M} p_i \\ &= \mathcal{E} - 2\mathbf{a}^T E[\mathbf{s}] + \|\mathbf{a}\|^2\end{aligned} \tag{5.161}$$

where $\mathcal{E} = \sum_{i=1}^{M} \|\mathbf{s}_i\|^2 p_i$ is the average energy of the original signal constellation, and

$$E[\mathbf{s}] = \sum_{i=1}^{M} \mathbf{s}_i p_i \tag{5.162}$$

The minimization of (5.161) gives the translation to yield minimum energy as

$$\mathbf{a}_{\min} = E[\mathbf{s}] \tag{5.163}$$

The resulting minimum energy is

$$\mathcal{E}_{\text{translate, min}} = \mathcal{E} - \|\mathbf{a}_{\min}\|^2 \tag{5.164}$$

Proposition 5.18: (Minimum Energy Translation) Given a signal constellation $\{s_i(t)\}_{i=1}^{M}$, the signal constellation achieves the minimum average energy by translating $\mathbf{s}_i$ to a constant $E[\mathbf{s}]$ as (5.163). If we further consider *a priori* probability $\{p_i\}_{i=1}^{M}$, it is equivalent to translation to the *center of (probability) mass*.

It is difficult to calculate the exact expressions of probability of error in many cases. However, we may still wish to obtain the knowledge regarding probability of error. The common approaches to resolve this dilemma are either to find a good approximation or to obtain the performance bounds (upper bound and lower bound). The upper bound of probability of error is of particular interests, because it provides the worst case scenario. In the following, we present a useful upper bound (known as *union bound*) for a set of M equally likely signals in AWGN channel.

Let A_{ik} $i,k = 1,2,\cdots,M$, denote the event that the observation vector $\mathbf{x}$ is closer to the signal vector $\mathbf{s}_k$ than to $\mathbf{s}_i$, given m_i is sent. The conditional probability of error given m_i , $P_e(m_i)$, is equal to the probability of the union of events $A_{i1}, A_{i2},\cdots, A_{iM}$ except A_{ii} . We may write

$$P_e(m_i) \le \sum_{\substack{k=1 \\ k\neq i}}^{M} P(A_{ik}), \qquad i = 1,2,\ldots,M \tag{5.165}$$

By introducing the pair-wise error probability $P_2(\mathbf{s}_i, \mathbf{s}_k)$,

$$P_e(m_i) \le \sum_{\substack{k=1 \\ k\neq i}}^{M} P_2(\mathbf{s}_i, \mathbf{s}_k), \qquad i = 1,2,\ldots,M \tag{5.166}$$

Please recall the probability of error solely depends on the Euclidean distance. By defining Euclidean distance between $\mathbf{s}_i$ and $\mathbf{s}_k$,

$$d_{ik} = \|\mathbf{s}_i - \mathbf{s}_k\| \tag{5.167}$$

Then,

$$\begin{aligned} P_2(\mathbf{s}_i, \mathbf{s}_k) &= P(\mathbf{x} \text{ is closer to } \mathbf{s}_k \text{ than } \mathbf{s}_i, \text{ when } \mathbf{s}_i \text{ is sent}) \\ &= \int_{d_{ik}/2}^{\infty} \frac{1}{\sqrt{pN_0}} \exp\left(-\frac{v^2}{N_0}\right) dv \end{aligned} \tag{5.168}$$

That is,

$$P_2(\mathbf{s}_i, \mathbf{s}_k) = Q\left(\frac{d_{ik}/2}{\sqrt{N_0/2}}\right) \tag{5.169}$$

We therefore have

$$P_e(m_i) \le \sum_{\substack{k=1 \\ k\ne i}}^{M} \mathrm{Q}(\frac{d_{ik}/2}{\sqrt{N_0/2}}), \qquad i = 1,2,\ldots,M \tag{5.170}$$

The average probability of error (over M symbols with ***a priori*** probability p_i) is upper bounded by

$$\begin{aligned} P_e &= \sum_{i=1}^{M} p_i P_e(m_i) \\ &\le \sum_{i=1}^{M} \sum_{\substack{k=1 \\ k\ne i}}^{M} p_i \ \mathrm{Q}(\frac{d_{ik}/2}{\sqrt{N_0/2}}) \end{aligned} \tag{5.171}$$

Two special cases are of particular interests in this general derivation of probability of error upper bound for general ML detection in AWGN:

(a) Suppose the signal constellation is circularly symmetric about the origin and equally probable (i.e. $p_i = 1/M$),

$$P_e \le \sum_{\substack{k=1 \\ k\ne i}}^{M} \mathrm{Q}(\frac{d_{ik}/2}{\sqrt{N_0/2}}) \ \ \forall i,k \tag{5.172}$$

(b) Define

$$d_{\min} = \min_{k\ne i} d_{ik} \quad \forall i,k \tag{5.173}$$

Then,

$$P_e \le (M-1)\mathrm{Q}(\frac{d_{\min}/2}{\sqrt{N_0/2}}) \tag{5.175}$$

Please also note that above general derivations are for probability of (symbol) error that is different from the probability of bit error. By assuming all symbol errors are equally likely,

$$BER = \frac{M/2}{M-1} P_e \tag{5.176}$$

5.6 COMMONLY APPLIED SIGNALING

PAM in AWGN or OOK:

The following signaling represents OOK (on-off keying) that is a special case of amplitude shifted keying (ASK)

$$\begin{cases} "1" & A\sqrt{2}\cos wt \\ "0" & 0 \end{cases} + n(t)$$

Assuming equal probable gives $p_e = \frac{1}{2}Q(\frac{A/2}{\sqrt{N_0/2}}) + \frac{1}{2}Q(\frac{A/2}{\sqrt{N_0/2}})$.

The following signaling represents frequency shifted keying (FSK):

$$\begin{cases} "1" & A\sqrt{2}\cos w_1 t \\ "0" & A\sqrt{2}\cos w_2 t \end{cases} + n(t)$$

Its signal space can be viewed as

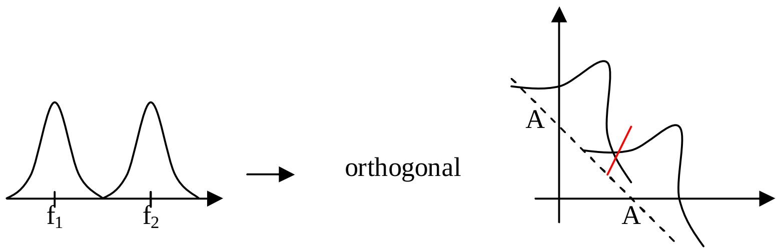


Figure 5.27: FSK signal space

$$p_e = Q(\frac{\sqrt{2}A/2}{\sqrt{N_0/2}}).$$

1. ASK with equal probable binary signaling, a.k.a. PAM à $p_e = Q(\frac{A/\sqrt{2}}{\sqrt{N_0/2}})$.

2. FSK: orthogonal, noise variance remains the same in both bases.

à $p_e = Q(\frac{A/\sqrt{2}}{\sqrt{N_0/2}})$, (coherent).

The next signaling is phase shifted keying (PSK):

$$\begin{cases} \text{"1"} \ A\sqrt{2}\cos wt \\ \text{"0"} \ A\sqrt{2}\cos(wt+t) \end{cases} \quad \text{à} \quad p_e = Q(\frac{A}{\sqrt{N_0/2}}).$$

- Performance: anti-podal > orthogonal > PAM.
- Anti-podal signaling is optimal in binary signaling with AWGN.

QPSK: $p_e = Q(\cancel{\frac{A/\sqrt{2}}{\sqrt{N_0/2}}}) = Q(\frac{A}{\sqrt{N_0/2}})$

Please note that QPSK is the superposition of I-channel and Q-channel transmission, which are orthogonal. It is equivalent to two parallel BPSK transmissions in the channel. It suggests $p_{e,QPSK} \equiv p_{e,BPSK}$.

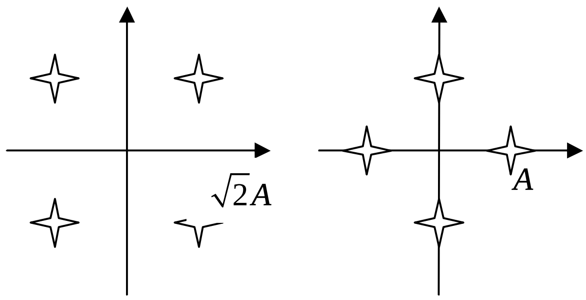


Figure 5.28: QPSK constellation

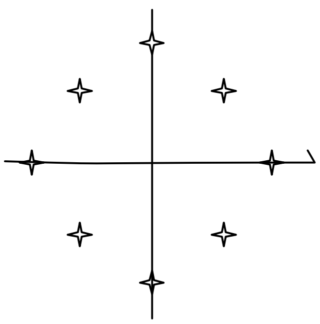


Figure 5.29: 8PSK constellation

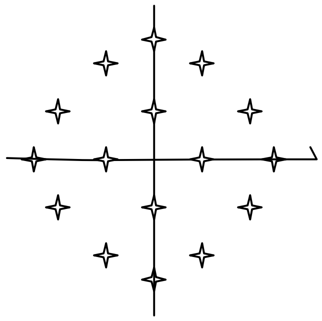


Figure 5.30: 16-APSK constellation

Although PSK has good performance, it requires good phase recovery, which is difficult when transmission speed is high (higher bandwidth is easier to result in severe signal facing), and we may use FSK instead.

FSK:

- Coherent: requires perfect recovery of frequency and phase.
- Non-coherent: determine max energy of frequency sub-band.

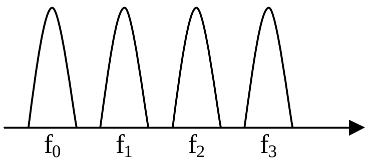


Figure 5.31: Frequency-domain PAM

Average prob. of error (symbol)

$$p_e = \sum_{i=1}^{M} p_i P(\aleph \notin z_i \mid m_i) = \frac{1}{M}\sum_{i=1}^{M} P(\aleph \notin z_i \mid m_i) = 1 - \frac{1}{M}\sum_{i=1}^{M} P(\aleph \in z_i \mid m_i)$$

4-level PAM, equal probable

1. $p_e = (\frac{1}{4}+\frac{1}{4})Q(\frac{A}{\sqrt{N_0/2}}) + (\frac{1}{4}+\frac{1}{4})2Q(\frac{A}{\sqrt{N_0/2}}) = \frac{3}{2}Q(\frac{A}{\sqrt{N_0/2}})$, with

roughly the same error rate, 4-level PAM uses much power for tx **à** use $\frac{E_b}{N_0}$ for comparison.

2. $E_b = 5A^2$ **à** $p_e = \frac{3}{2}Q(\sqrt{\frac{E_b/5}{N_0/2}})$, p_e : symbol error rate.

16-QAM, I, Q channel introduce i.i.d. noise.

1. $p_e = \frac{1}{4}2Q(\frac{A}{\sqrt{N_0/2}}) + \frac{1}{2}3Q(\frac{A}{\sqrt{N_0/2}}) + \frac{1}{4}4Q(\frac{A}{\sqrt{N_0/2}}) = 3Q(\frac{A}{\sqrt{N_0/2}})$

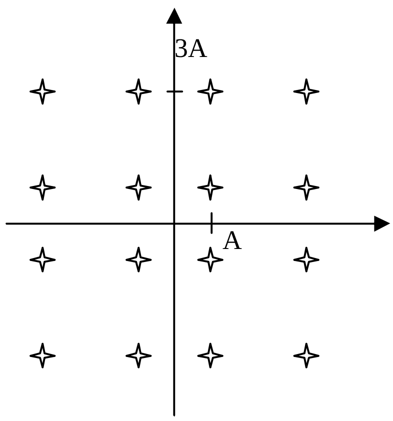


Figure 5.32: 16-QAM

$E_b = 10A^2$ **à** $p_e = 3Q(\sqrt{\frac{E_b/10}{N_0/2}})$

1. 2. 3.

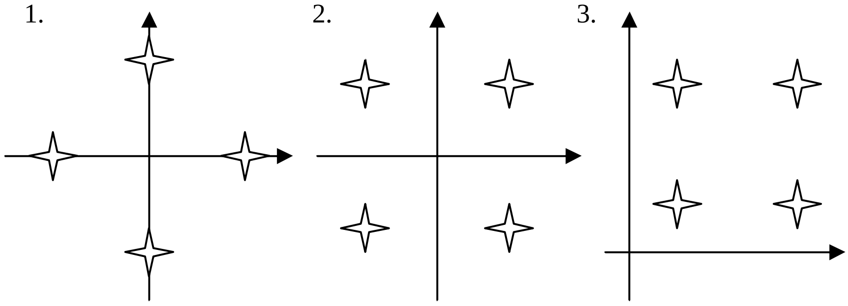


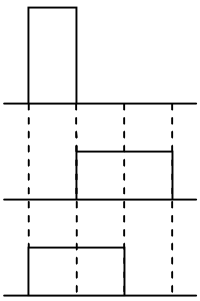
Figure 5.33: 1, 2, 3: equal error rates; 3: higher energy.

To minimize energy **à** $\sum_i r_i s_i = 0$.

Min $\|r-s\|^2$ **à** p_e rotationally invariant, translation invariant since p_e depends only on Euclidean distance in AWGN.

General Signaling: we have to transform time-domain/frequency-domain waveforms into signal space.

Q $\int_{-\infty}^{\infty} f_i(t) f_j(t) dt = d_{ij}$ **à** orthogonal waveform.



$s_1(t) = 2f_1(t) \rightarrow (2,0,0)$ $\quad (1,-\frac{2}{3},\frac{1}{3})$

$s_2(t) = f_2(t) + f_3(t) \rightarrow (0,1,1) \xrightarrow{\text{min energy}} (-1,\frac{1}{3},\frac{2}{3})$

$s_3(t) = f_1(t) + f_2(t) \rightarrow (1,1,0)$ $\quad (0,\frac{1}{3},-\frac{1}{3})$