Binary Detection

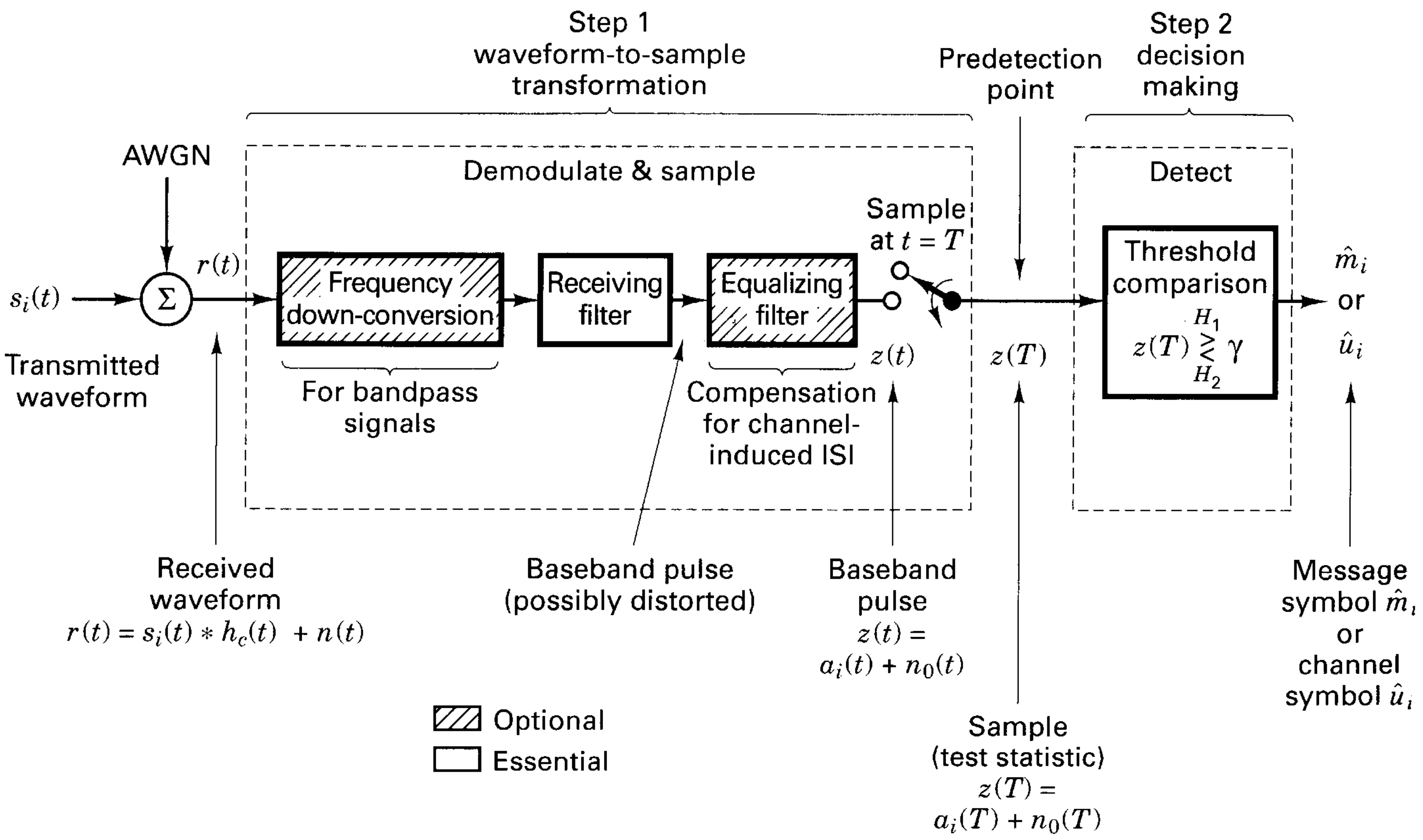
The transmitted signal is represented by

$$s_i(t) = \begin{cases} s_1(t) & \text{for binary 1} \\ s_2(t) & \text{for binary 0} \end{cases} \qquad 0 \le t \le T$$

The received waveform $r(t)$ is given by

$$r(t) = s_i(t) + n(t) \quad i = 1, 2 \quad 0 \le t \le T$$

where $n(t)$ is a zero-mean **additive white Gaussian noise** (AWGN) process.



Two basic steps in the demodulation/detection of digital signals.

Two separate steps are involved in the signal detection process:

- Reducing the received waveform $r(t)$, $0 \le t \le T$, to a single number, $Z(T) = a_i(T) + n_0(T)$, $i = 1, 2$, in which $a_i(T)$ and $n_0(T)$ are due to the signal and noise, respectively.
- Comparing the **test statistic** $Z(T)$ to a **threshold** level γ to decide which signal, $s_1(t)$ or $s_2(t)$, has been sent at the transmitter.

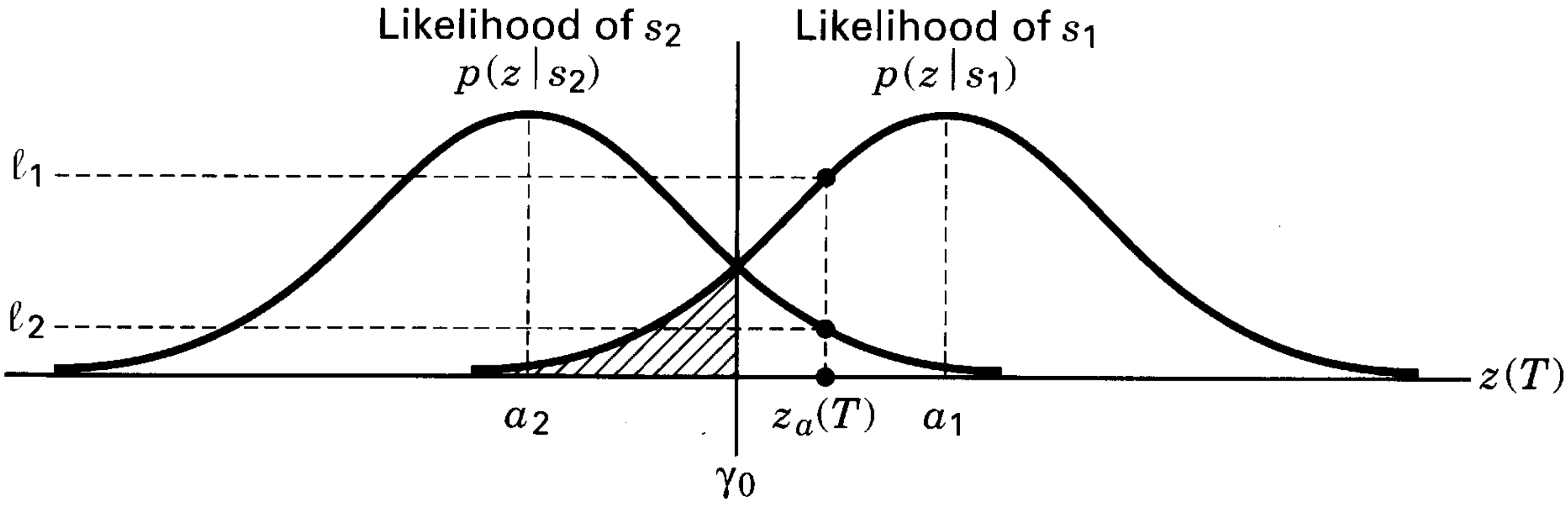
Test Statistic *Z*(*T*)

This operation can be performed by a linear filter followed by a sampler, as shown in Step 1. The noise component, $n_0(T)$, is a zero-mean Gaussian RV with pdf

$$p(n_0) = \frac{1}{\sigma_0\sqrt{2\pi}} \exp\left(-n_0^2 / 2\sigma_0^2\right)$$

Thus $Z(T)$ is a conditional Gaussian RV with mean either a_1 or a_2. The conditional pdf's are

$$p(z \mid s_i) = \frac{1}{\sigma_0\sqrt{2\pi}} \exp\left[-(z-a_i)^2 / 2\sigma_0^2\right], \quad i = 1, 2$$



Conditional probability density functions: $p(z|s_1)$ and $p(z|s_2)$.

In Step 2, detection is performed by choosing the hypothesis based on the outcome of

$$Z(T) \underset{H_2}{\overset{H_1}{\gtrless}} \gamma$$

where H_1 and H_2 are corresponding to $s_1(t)$ and $s_2(t)$, respectively.

Note that the filtering operation in Step 1 does not depend on the decision criterion in Step 2. Thus the choice of how best to implement Step 1 can be independent of the particular decision strategy (choice of the threshold setting, γ).

Maximum-Likelihood Receiver

In Step 2, a reasonable choice for choosing the threshold γ is based on minimizing the probability of error. To find the optimal value of $\gamma = \gamma_0$, we first perform the likelihood ratio test as

$$L(z) = \frac{p(z \mid s_1)}{p(z \mid s_2)} \underset{H_2}{\overset{H_1}{\gtrless}} \frac{\Pr(s_2)}{\Pr(s_1)} = \gamma$$

When the prior probabilities are the same, i.e., $\Pr(s_1) = \Pr(s_2)$, the above criterion becomes the **maximum-likelihood** (ML) **criterion**. That is,

$$L(z) = \frac{p(z|s_1)}{p(z|s_2)} \underset{H_2}{\overset{H_1}{\gtrless}} 1$$

If the conditional pdf's are symmetrical, the **likelihood ratio test** is equivalent to

$$z(T) \underset{H_2}{\overset{H_1}{\gtrless}} \gamma_0 = \frac{a_1 + a_2}{2}$$

where a_1 is the signal component of $z(T)$ when $s_1(t)$ is transmitted and a_2 is the signal component of $z(T)$ when $s_2(t)$ is transmitted. The threshold level $(a_1 + a_2)/2$ is the optimum threshold to minimize the probability of making an incorrect decision. Let us look at two special cases:

- If $a_2 = 0$, then

$$z(T) \underset{H_2}{\overset{H_1}{\gtrless}} \gamma_0 = \frac{a_1}{2}$$

- For antipodal signals, $s_1(t) = -s_2(t)$ and $a_1 = -a_2$. Therefore, the test is

$$z(T) \underset{H_2}{\overset{H_1}{\gtrless}} \gamma_0 = 0$$

Probability of Bit Error

For the ML binary decision, the probability of bit error is

$$P_B = \Pr(\rightarrow H_2 \mid H_1)\Pr(H_1) + \Pr(\rightarrow H_1 \mid H_2)\Pr(H_2)$$

where $\Pr(\rightarrow H_2|H_1)$ denotes the probability of choosing H_2, given that H_1 is actually correct. Assuming $\Pr(H_1) = \Pr(H_2) = 0.5$ in digital communications, P_B becomes

$$P_B = \frac{1}{2}\left[\Pr(\rightarrow H_2 \mid H_1) + \Pr(\rightarrow H_1 \mid H_2)\right]$$

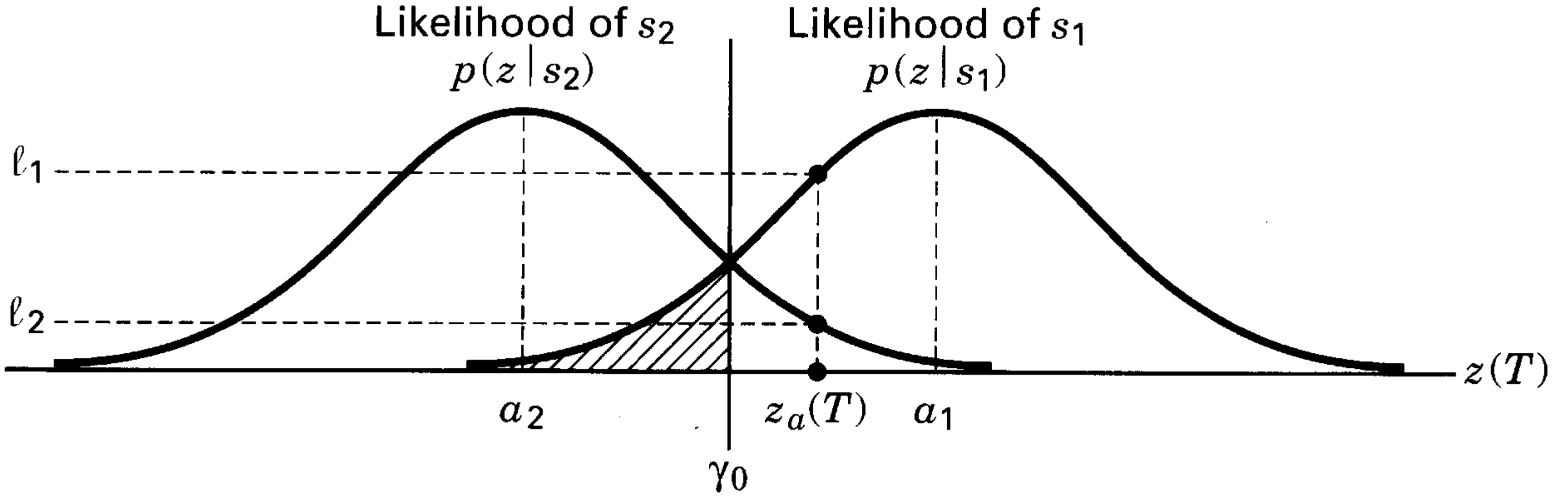
$$= \frac{1}{2}\left[\Pr(\rightarrow H_2 \mid s_1) + \Pr(\rightarrow H_1 \mid s_2)\right]$$

If the conditional pdf"s are symmetrical, we observe

$$\Pr(\rightarrow H_2 \mid s_1) = \Pr(\rightarrow H_1 \mid s_2)$$

Hence, the error probability is simplified to

$$P_B = \Pr(\rightarrow H_2 \mid s_1) = \Pr(\rightarrow H_1 \mid s_2)$$



Conditional probability density functions: $p(z|s_1)$ and $p(z|s_2)$.

Note that

$$\Pr(\rightarrow H_1 \mid s_2) = \Pr\left(Z > \gamma_0 \mid s_2\right)$$

$$= \int_{\gamma_0}^{\infty} p(z \mid s_2)\, dz$$

$$= \int_{\gamma_0}^{\infty} \frac{1}{\sigma_0 \sqrt{2\pi}} \exp\left[-\frac{(z-a_2)^2}{2\sigma_0^2}\right] dz$$

Let $u = (z - a_2)/\sigma_0$. Then

$$P_B = \int_{\frac{\gamma_0 - a_2}{\sigma_0}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du = Q\left(\frac{\gamma_0 - a_2}{\sigma_0}\right) = Q\left(\frac{a_1 - a_2}{2\sigma_0}\right)$$

where $Q(\cdot)$ has been defined previously.

The Matched Filter

The **matched filter** (MF) is a linear filter designed to provide the maximum signal-to-noise power ratio (SNR) at its output for a given transmitted symbol waveform.

Referring to the diagram on page 136, we have

$$Z(T) = a_i + n_0$$

where a_i is the signal component and n_0 is the noise component. Hence,

$$\left(S/N\right)_T = \frac{a_i^2}{\sigma_0^2}$$

where σ_0^2 is the variance of n_0. We wish to find the filter with transfer function $H_0(f)$ that **maximizes** $\left(S/N\right)_T$. In the subsequent analysis, we drop the subscript of $a_i(t)$ for the purpose of convenience.

Since the input noise is AWGN with mean zero and two-sided PSD $N_0/2$,

$$\sigma_0^2 = \frac{N_0}{2}\int_{-\infty}^{\infty} \left|H(f)\right|^2 df$$

We can also express $a(t)$ at the filter output in terms of $H(f)$ and the Fourier transform of the input signal, $s(t)$, as

$$a(t) = \mathrm{F}^{-1}\left\{H(f)S(f)\right\}$$

$$= \int_{-\infty}^{\infty} H(f)S(f)e^{j2\pi ft}\,df$$

Setting $t = T$, we obtain

$$(S/N)_T = \frac{\left|\int_{-\infty}^{\infty} H(f)S(f)e^{j2\pi fT}\,df\right|^2}{\frac{N_0}{2}\int_{-\infty}^{\infty}\left|H(f)\right|^2 df}$$

To maximize this equation with respect to $H(f)$, we employ **Schwarz's inequality**.

Schwarz's Inequality

$$\left| \int_{-\infty}^{\infty} f_1(x) f_2(x)\, dx \right|^2$$

$$\leq \int_{-\infty}^{\infty} | f_1(x)|^2\, dx \int_{-\infty}^{\infty} | f_2(x)|^2\, dx$$

The equality holds if and only if $f_1(x) = k\, f_2^*(x)$, where k is an arbitrary constant and * indicates complex conjugate.

If we identify $H(f)$ with $f_1(x)$ and $S(f)e^{j2\pi fT}$ with $f_2(x)$, we can write

$$\left|\int_{-\infty}^{\infty} H(f)S(f)e^{j2\pi fT}\,df\right|^2$$

$$\leq \int_{-\infty}^{\infty}|H(f)|^2\,df \int_{-\infty}^{\infty}|S(f)|^2\,df$$

It turns out that

$$\left(S/N\right)_T \leq \frac{2}{N_0}\int_{-\infty}^{\infty}|S(f)|^2\,df = \frac{2E}{N_0}$$

where

$$E = \int_{-\infty}^{\infty} s^2(t)\,dt = \int_{-\infty}^{\infty}|S(f)|^2\,df$$

is the input signal energy. Note that the maximum value of $(S/N)_T$ depends on the **input signal energy** E and the PSD of the noise, **not on the particular shape** of the waveform that is used.

The transfer function of the optimized (matched) filter is

$$H_0(f) = k\,S^*(f)\,e^{-j2\pi fT}$$

For a real-valued signal $s(t)$,

$$S^*(f) = S(-f).$$

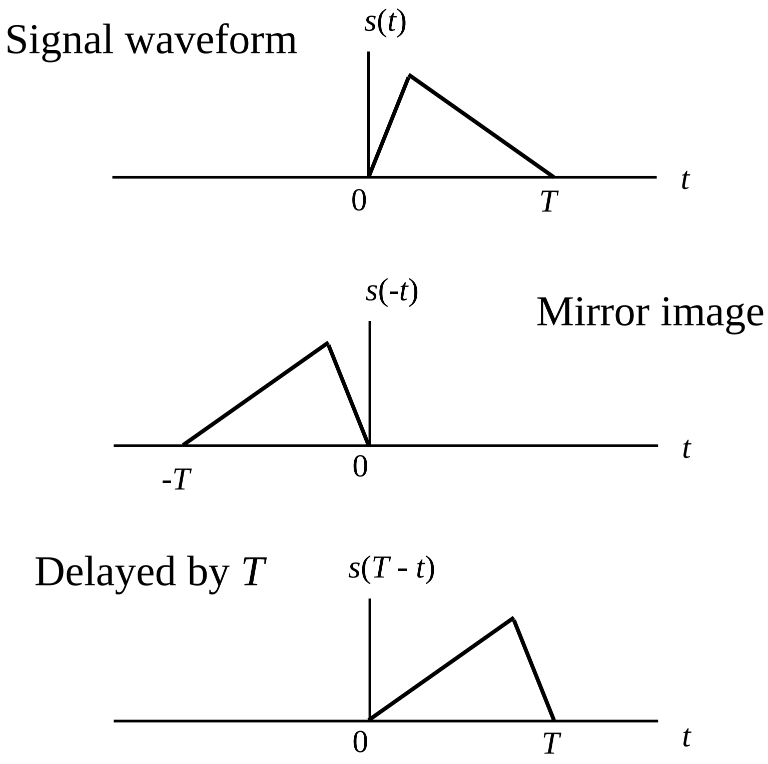
Hence, $H_0(f) = k\,S(-f)\,e^{-j2\pi fT}$ and

$$
\begin{aligned}
h_0(t) &= \mathbf{F}^{-1}\left[H_0(f)\right] \\
&= \int_{-\infty}^{\infty} k\, S(-f)\, e^{-j2\pi f(T-t)}\, df \\
&= \int_{-\infty}^{\infty} k\, S(\lambda)\, e^{j2\pi\lambda(T-t)}\, d\lambda \\
&= k\, s(T-t)
\end{aligned}
$$

Thus the impulse response, $h_0(t)$, of the MF that maximizes the output signal-to-noise power ratio is the mirror image of the input signal, $s(t)$, delayed by the symbol duration T. Note that $h_0(t)$ is a **causal** filter.

Impulse Response of Matched Filter

$K = 1$



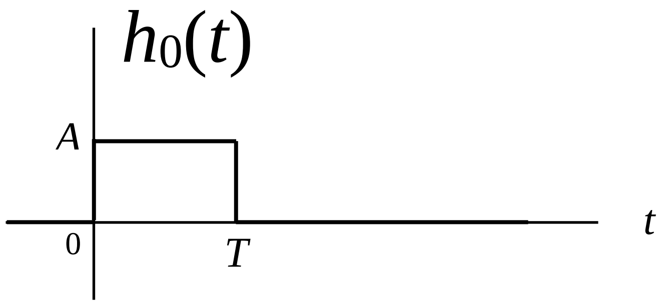
Example

Consider the pulse signal

$$s(t) = \begin{cases} A, & 0 \le t \le T \\ 0, & \text{otherwise} \end{cases}$$

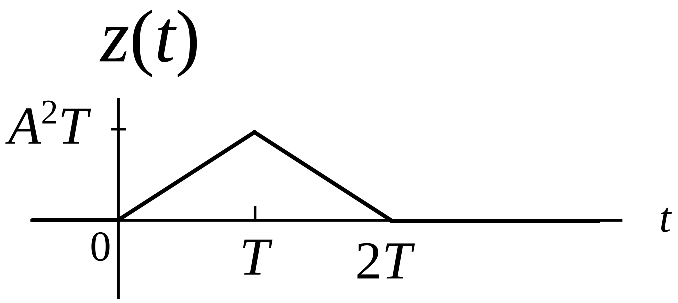
The MF for $s(t)$ is

$$h_0(t) = s(T-t)$$



The response $z(t)$ of the matched filter to $s(t)$ is

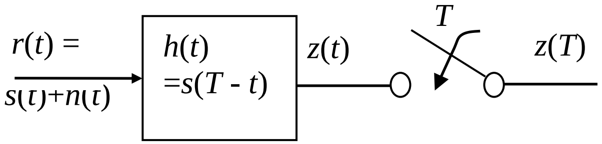
$$z(t) = h_0(t) \otimes s(t)$$



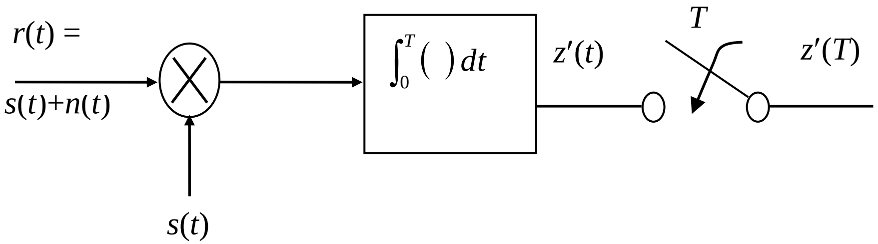
We notice that the peak output signal occurs at $t = T$, which is also the time instant of peak signal-to-noise power ratio.

Correlator Receiver

An alternative receiver structure of the MF is a correlator.



(a) ***Matched filter***



(b) ***Correlator***

To show that the operations given in (a) and (b) are equivalent, we will show that $z(T)$ in (a) is equal to $z'(T)$ in (b). The output of the MF in (a) is

$$\begin{aligned} z(t) &= h(t) \otimes r(t) \\ &= \int_0^t r(\tau) h(t-\tau)\, d\tau \\ &= \int_0^t r(\tau) s(T-t+\tau)\, d\tau \end{aligned}$$

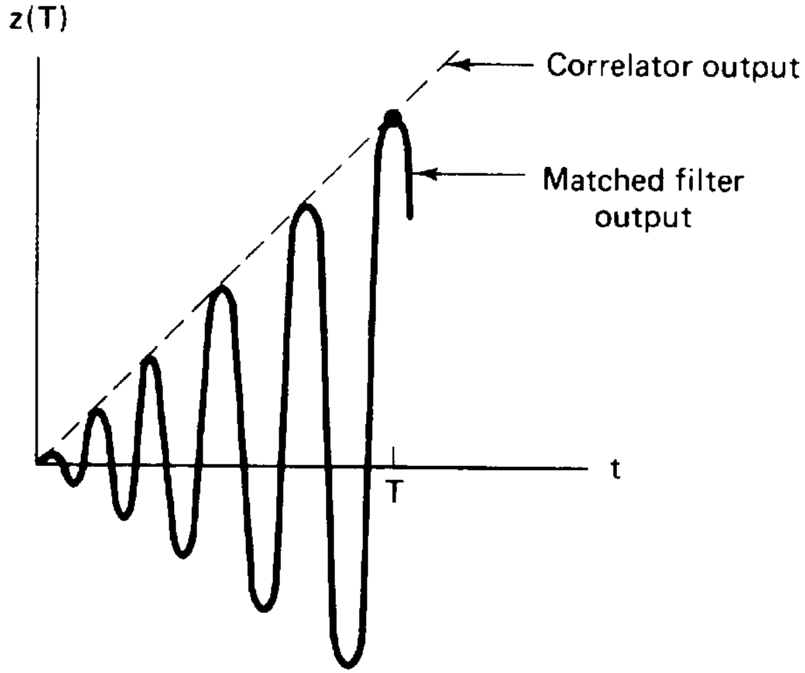
which follows because $h(t) = s(T-t)$ for $0 \leq t \leq T$ and is zero otherwise. When $t = T$, we obtain

$$z(T) = \int_0^T r(\tau) s(\tau)\, d\tau$$

Considering next the output of the correlator in (b), we find

$$z'(T) = \int_0^T r(\tau)s(\tau)\, d\tau$$

which is identical to $z(T)$. In general, their outputs are NOT the same at all time.

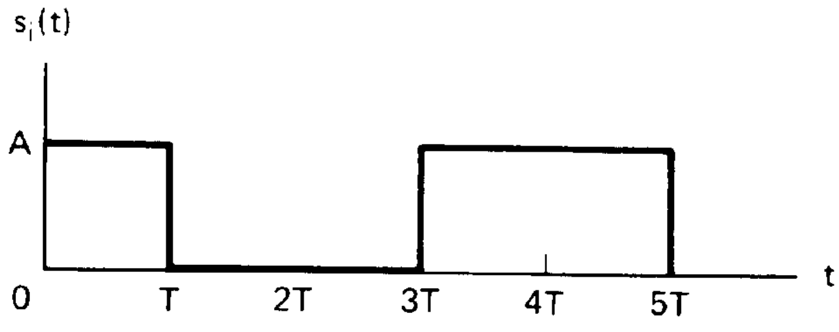


BER of Unipolar Signaling

Suppose unipolar or on-off signaling is used to represent binary information as

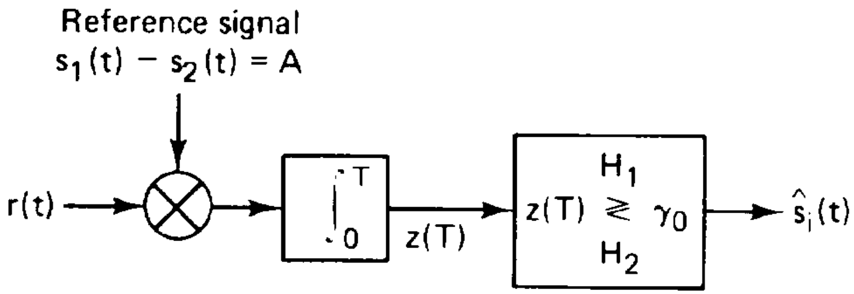
$$s_1(t) = A \quad 0 \le t \le T \quad \text{for binary 1}$$
$$s_2(t) = 0 \quad 0 \le t \le T \quad \text{for binary 0}$$

Assume that the unipolar signal plus AWGN is present at the input of a correlator. The correlator multiples the incoming signal, $r(t)$, with the difference of the prototype signal $d(t) = s_1(t) - s_2(t) = A$, and integrates over a symbol duration, T, to obtain $Z(T)$. The result $Z(T)$ is compared with the threshold γ_0.



$s_i(t)$
A
0
T
2T
3T
4T
5T
t

(a)



Reference signal
$s_1(t) - s_2(t) = A$
$r(t)$
$\int_0^T$
$z(T)$
H_1
$z(T) \gtrless \gamma_0$
H_2
$\hat{s}_i(t)$

(b)

When $r(t) = s_1(t) + n(t)$, the signal component $Z(T)$ at the output due to $s_1(t)$ is given by

$$a_1 = \int_0^T A \cdot A \, dt = A^2 T$$

Similarly, when $r(t) = s_2(t) + n(t)$,

$$a_2 = \int_0^T 0 \cdot A \, dt = 0$$

Thus the optimum threshold is

$$\gamma_0 = \frac{1}{2}\left[a_1 + a_2\right] = \frac{1}{2} A^2 T$$

Note that the output noise is

$$n_0(T) = \int_0^T n(t) \cdot A \; dt$$

Its variance is thus $\sigma_0^2 = \mathrm{var}\left[n_0(T)\right] = A^2 T \left(N_0 / 2\right)$.

The BER is thus

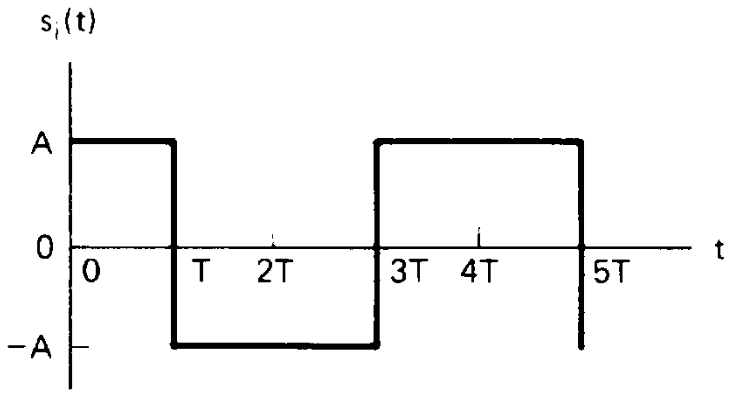
$$P_B = Q\left(\frac{\gamma_0 - a_2}{\sigma_0}\right) = Q\left(\sqrt{\frac{E_b}{N_0}}\right)$$

where $E_b = A^2 T/2$ is the average bit energy.

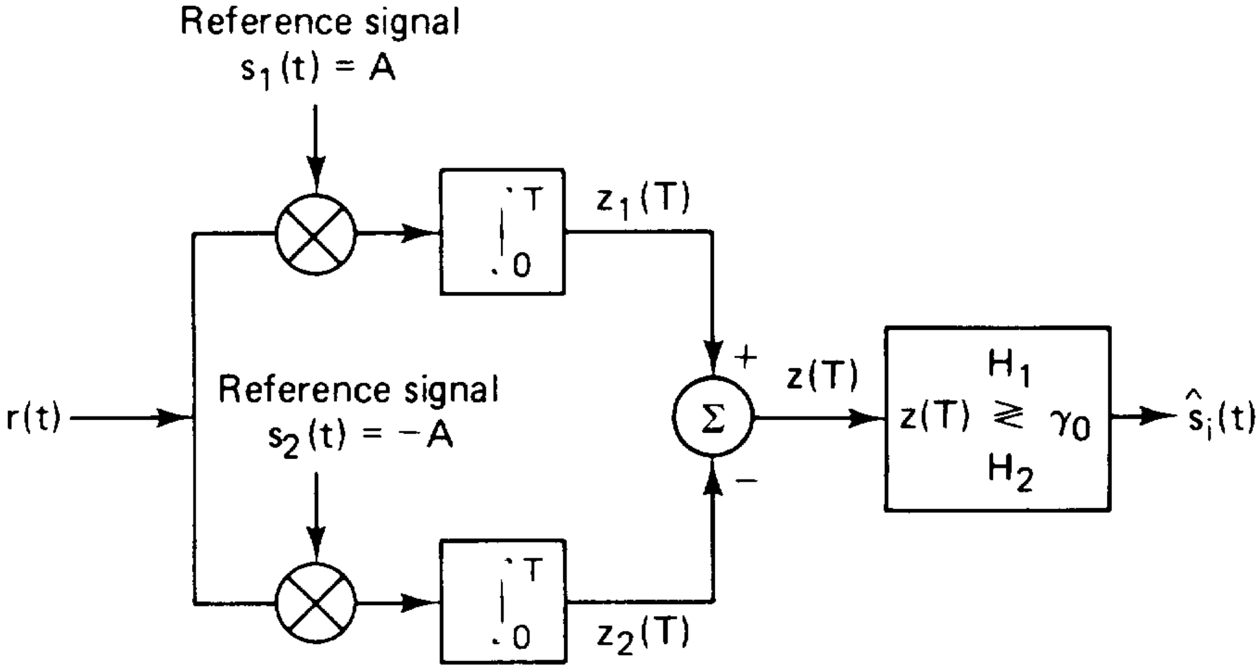
BER of Bipolar Signaling

Suppose bipolar signaling or polar signaling is used to represent binary information as

$$s_1(t) = +A \quad \text{for binary 1} \quad 0 \le t \le T$$
$$s_2(t) = -A \quad \text{for binary 0} \quad 0 \le t \le T$$



Since the signals are antipodal, i.e., $s_1(t) = -s_2(t)$, the optimum threshold value is 0. The correlator receiver is shown below.



When $r(t) = s_1(t) + n(t)$, the signal component at the output $Z(T)$ due to $s_1(t)$ is given by

$$a_1 = \int_0^T s_1(t)\left[s_1(t) - s_2(t)\right] \, dt = 2A^2T$$

Similarly, $a_2 = -2A^2T$. Accordingly, the optimum threshold is $\gamma_0 = (a_1 + a_2)/2 = 0$. Note that the output noise is given by

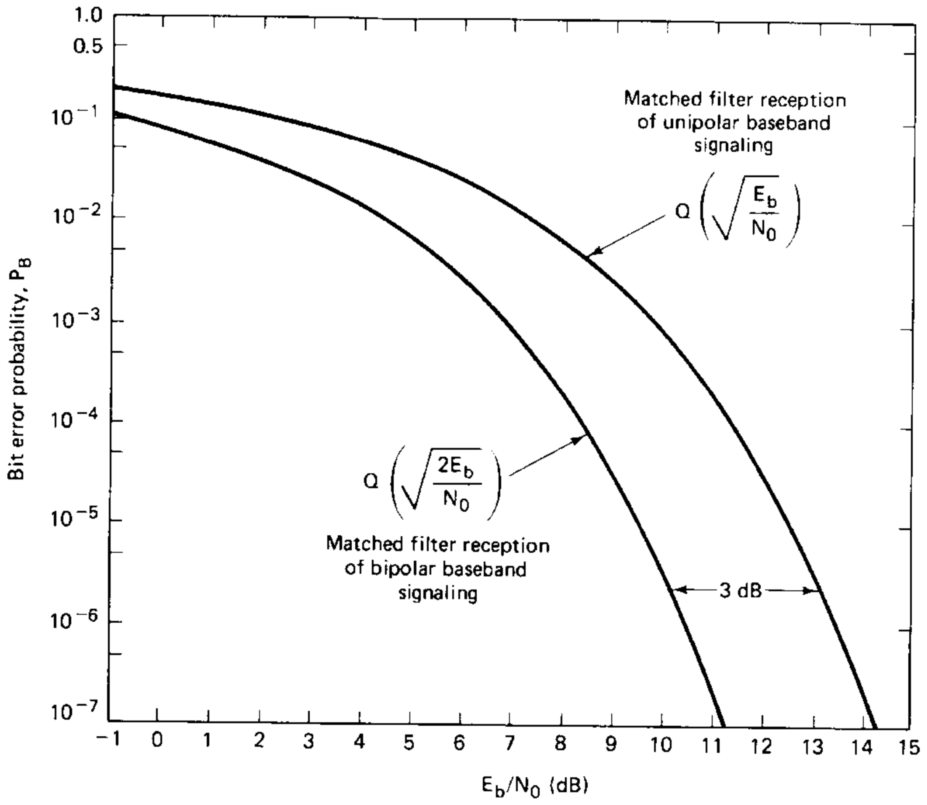
$$n_0(T) = \int_0^T n(t) \cdot \left[s_1(t) - s_2(t)\right] \, dt$$

with $\operatorname{var}\left[n_0(T)\right] = \sigma_0^2 = 2A^2TN_0$. Hence, the BER is given by

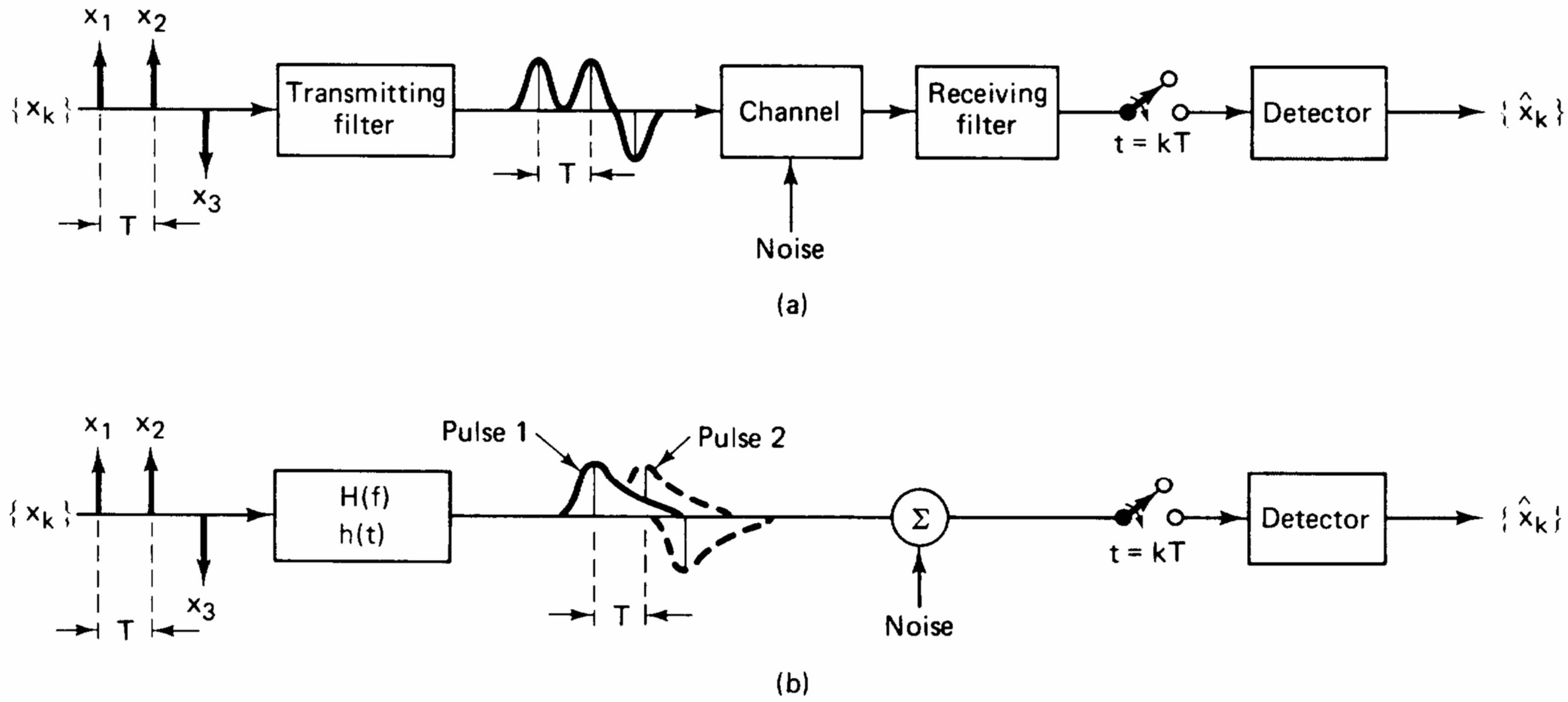
$$P_B = Q\left(\frac{\gamma_0 - a_2}{\sigma_0}\right) = Q\left(\sqrt{\frac{2E_b}{N_0}}\right)$$

where $E_b = A^2T$ is the average energy per bit.

As we shall discuss later, the BER results presented here are also valid for bandpass signals such as ASK and PSK signaling.



Intersymbol Interference (ISI)



Intersymbol interference in the detection process. (a) Typical baseband digital system. (b) Equivalent model.

In the transmitter the impulses at the input are low-pass filtered to confine to some desired bandwidth. Channel impairment can cause amplitude distortion and phase distortion to the transmitted pulses. At the receiver, an equalizing filter, $H_r(f)$, is often employed to compensate for the distortion caused by the transmitting filter, $H_t(f)$, and the channel transfer function, $H_c(f)$. For the case of binary bipolar signaling, the detector makes symbol decisions by comparing the received bipolar pulses to a zero threshold. Fig. (b) on the last page illustrates a convenient model for the system, lumping all the filtering effects into one overall equivalent system transfer function

$$H(f) = H_t(f)\,H_c(f)\,H_r(f)$$

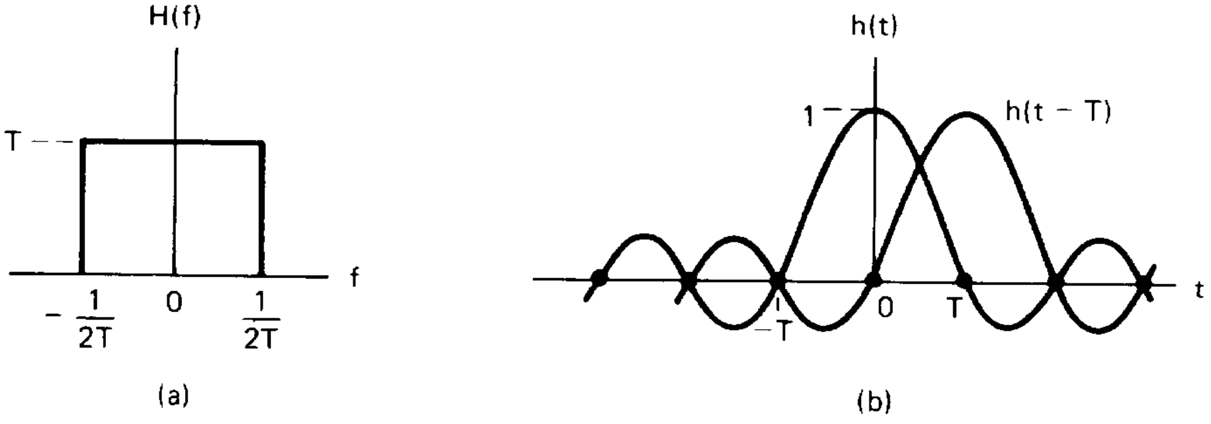
Due to the effects of $H(f)$, the tail of one pulse "smears" into adjacent symbol intervals and affects the accuracy of detection process. This type of interference is called **intersymbol interference** (**ISI**). Even in the absence of noise, imperfect filtering and insufficient system bandwidth lead to ISI. In practice, the channel transfer function $H_c(f)$ is usually specified *a priori*, and the problem remains to find $H_t(f)$ and $H_r(f)$ such that the ISI of the pulses are minimized at the output of $H_r(f)$.

Zero ISI

Nyquist's first method for eliminating ISI is to use

$$H(f)=\begin{cases} T & |f| \le 1/(2T) \\ 0 & \text{otherwise} \end{cases}$$

The corresponding impulse response is $h(t)=\mathrm{sinc}(t/T)$.



Nyquist channels for zero ISI. (a) Rectangular system transfer function $H(f)$. (b) Received pulse shape $h(t) = \mathrm{sinc}\,(t/T)$.

If each pulse of a received sequence is of the form $h(t)$, then the pulses can be detected without ISI. In other words, a system with bandwidth $W = 1/(2T) = R_s / 2$ Hz can support a maximum symbol rate of $R_s = 1/T$ symbols per second. Hence, without ISI, the maximum symbol transmission rate per hertz is $R_s/W = 2$ symbols/s/Hz.

Note that there is no restriction on bits/s/Hz. If each symbol represents $M = 2^6 = 64$ levels, then the **bandwidth efficiency** of the system is 12 bits/s/Hz.

Practical Difficulties of Realizing $H(f)$

The pulse shape of $h(t)$ has two practical difficulties.

- The impulse response $h(t)$ is not causal. Hence, it is not physically realizable.
- The detection process would be very sensitive to small errors. The pulse $h(t)$ has zero value in adjacent pulse times only when the sampling is performed at exactly the correct sampling time. Timing error will produce ISI.

Raised Cosine Filter

Because of the above difficulties, we need to consider other pulse shapes that have slightly wider bandwidth. One commonly used in practice is the raised cosine filter

$$H(f)=\begin{cases} 1 & |f|<2W_0-W \\ \cos^2\left(\dfrac{\pi}{4}\dfrac{|f|+W-2W_0}{W-W_0}\right) & 2W_0-W<|f|<W \\ 0 & |f|>W \end{cases}$$

where W is the absolute bandwidth, W_0 represents the minimum Nyquist bandwidth. The difference between $(W-W_0)$ is called the **excess bandwidth**.

The **roll-off factor** is

$$r = \left(W - W_0\right)/W_0$$

The impulse response for $H(f)$ is

$$h(t) = 2W_0 \,\mathrm{sinc}(2W_0 t)\,\frac{\cos\left[2\pi(W - W_0)t\right]}{1 - 4(W - W_0)t^2}$$

The total bandwidth is

$$W = \left(1 + r\right)W_0 = \frac{1}{2}\left(1 + r\right)R_s \qquad (\because W_0 = R_s / 2)$$

For basspass-modulated signals such as ASK and PSK, the required BW is $W_{\mathrm{DSB}} = \left(1 + r\right)R_s$.

Example

(a) Find the minimum required bandwidth for the baseband transmission of a four-level PCM pulse sequence having a data rate of R = 2400 bits/s if the system transfer characteristic consists of a raised cosine spectrum with 100% excess BW (r = 1).

(b) The same PCM sequence is modulated onto a carrier wave, so that the baseband spectrum is shifted and centered at frequency f_0. Find the minimum required DSB bandwidth for transmitting the modulated PCM sequence. Assume that the system transfer characteristic is the same as in part (a).

Solution

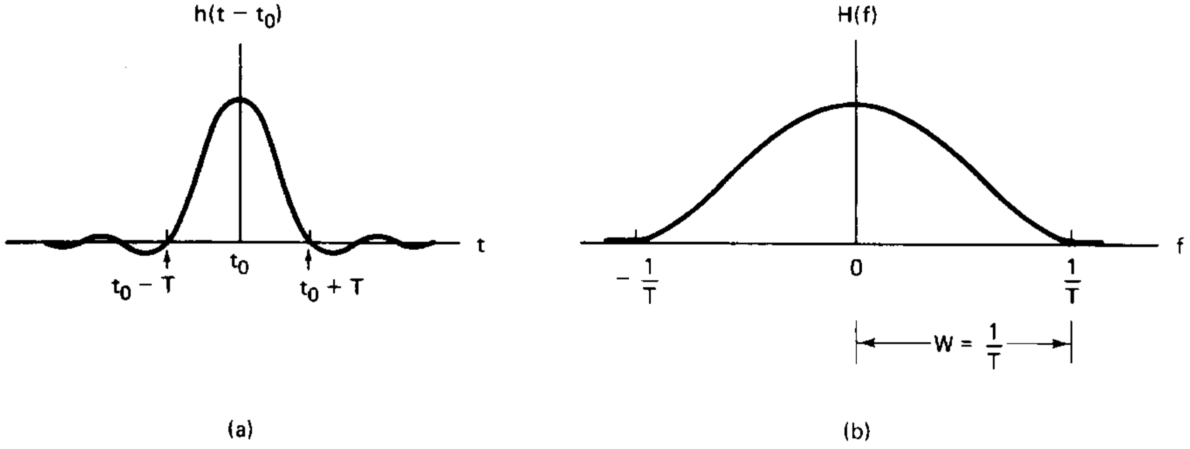
(a) $M = 2^k = 4$, where $k = 2$

symbol rate is

$$R_s = R/k$$
$$= 2400/2$$
$$= 1200 \text{ symbols/s}$$

min. BW is

$$W = 0.5(1+r)\, R_s$$
$$= 0.5(1+1)(1200)$$
$$= 1200 \text{ Hz}$$



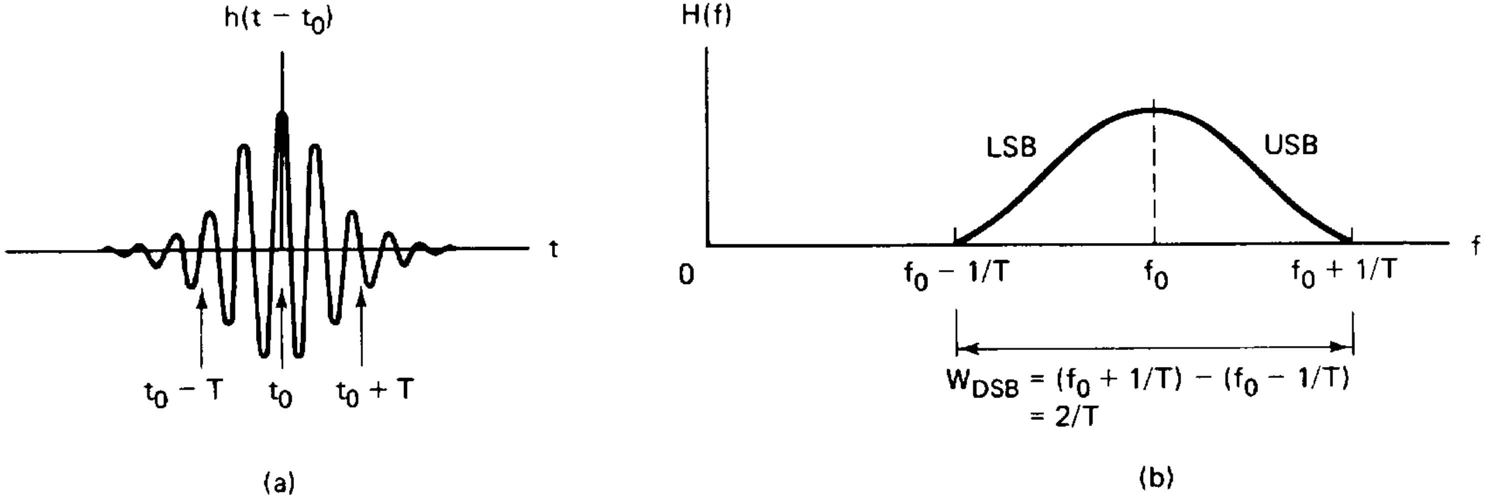
(a) Shaped pulse. (b) Baseband raised cosine spectrum.

(b) As in part (a),

R_s = 1200 symbols/s

$$
\begin{aligned}
W_{\mathrm{DSB}} &= (1+r)\,R_s \\
&= 2(1200) \\
&= 2400 \text{ Hz}
\end{aligned}
$$

When the spectrum is shifted up in frequency, the negative and positive halves of the baseband spectrum are shifted up in frequency, thereby doubling the required transmission BW. They are referred to as the upper sideband (USB) and lower sideband (LSB), as labeled in the figure below.



(a) Modulated shaped pulse. (b) DSB-modulated raised cosine spectrum.