ECE 8700: Communication Systems Engineering

Module 2
Baseband Demodulation/Detection

Mojtaba Vaezi
http://www.ece.villanova.edu/ mvaezi

Villanova University

March 23, 2025

Contents

1 E_b/N_0–The Basic SNR Parameter for Digital Communication **3**

1.1 Definition of E_b/N_0 . . . 3

1.2 E_b/N_0 vs. SNR . . . 5

1.3 Why E_b/N_0 Is a Natural Figure of Merit? . . . 7

2 Demodulation and Detection **9**

2.1 Correlation Realization of the Matched Filter . . . 14

3 Binary Signal Detection **18**

3.1 Error Analysis for AWGN Channel . . . 19

3.2 Evaluating the Probability of Errors . . . 23

3.3 Energy and Correlation . . . 30

4 Filtering without Intersymbol Interference (ISI) **33**

4.1 Why Pulse Shaping? . . . 35

4.2 Practical Nyquist Filters . . . 41

5 The Nyquist Pulse Shaping (Filtering) Criterion **47**

5.1 Bandwidth Efficiency . . . 55

5.2 ISI and Pulse Shaping . . . 58

1 E_b/N_0–The Basic SNR Parameter for Digital Communication

- In analog communications, the commonly used figure of merit is the `average signal power to average noise power ratio` (S/N or SNR).
- In digital communications, often E_b/N_0 is used as a figure of merit.
- E_b/N_0 is the normalized version of SNR

1.1 Definition of E_b/N_0

The bit energy-to-noise power spectral density ratio ($\frac{E_b}{N_0}$) is defined as:

$$\frac{E_b}{N_0} = \frac{\text{average energy per bit}}{\text{noise power spectral density}}$$

Dimension of E_b/N_0

Just as SNR is a dimensionless ratio, so too is E_b/N_0.

– E_b**:** Represents the *energy per bit*.

$$\text{Dimension of } E_b = [\text{Energy}] = [\text{Joules}]$$

– N_0**:** Represents the *noise power spectral density*, which is the noise power per unit bandwidth.

$$\text{Dimension of } N_0 = \frac{[\text{Power}]}{[\text{Frequency}]} = \frac{[\text{Watts}]}{[\text{Hertz}]} = [\text{Joules}]$$

1.2 E_b/N_0 vs. SNR

- $T_b = 1/R_b$: **Bit time** T_b and the **bit rate** R_b are reciprocal ($T_b = 1/R_b$)
- E_b: **Energy per bit**, described as the signal power S multiplied by the bit time T_b.
- N_0: **Noise power spectral density**, which is the noise power N divided by the bandwidth W.

$$\frac{E_b}{N_0} = \frac{ST_b}{N/W} = \frac{S/R_b}{N/W} = \frac{S}{N} \cdot \frac{W}{R_b} \tag{1}$$

- In M-ary signaling

$$\frac{W}{R_b} \approx \frac{1}{\log_2 M} \implies \frac{E_b}{N_0} \approx \frac{S}{N} \cdot \frac{1}{\log_2 M}$$

- The E_b/N_0 ratio relates the S/N to the bandwidth-to-bit-rate ratio (W/R_b).
- How can we find the value of W/R_b?
  - In an ideal case, the minimum bandwidth required to transmit a signal depends on the Nyquist criterion:

$$W_{\min} = \frac{1}{2T} = \frac{R_s}{2}$$

This is the absolute lower bound for a baseband signal (no carrier modulation).

– However, for passband (RF) signals, the bandwidth is approximately:

$$W \approx R_s$$

This is because for practical bandpass signals

◇ The signal is modulated onto a carrier, occupying bandwidth around the carrier frequency.

◇ The actual occupied bandwidth depends on the pulse shape and modulation scheme.

◇ Common modulation techniques like BPSK, QPSK, and M-PSK spread the spectrum over approximately the symbol rate R_s.

◇ If no pulse shaping (ideal rectangular pulses) is applied, the main lobe width is roughly R_s.

◇ If pulse shaping (e.g., raised cosine) is applied, the occupied bandwidth is slightly larger than R_s (depending on the roll-off factor β):

$$W = R_s(1 + \alpha)$$

where α (roll-off factor) typically ranges from 0 to 1.

1.3 Why E_b/N_0 Is a Natural Figure of Merit?

- In digital communications, E_b/N_0 is a key figure of merit—a parameter used to assess system performance.
- E_b/N_0 is a measure of the SNR per bit and is crucial for understanding the energy efficiency and reliability of communication systems.
- Why E_b/N_0 is a natural figure of merit in digital communication systems? Why can't we continue to use SNR as a figure of merit in digital communications?

  There are **two main reasons**:

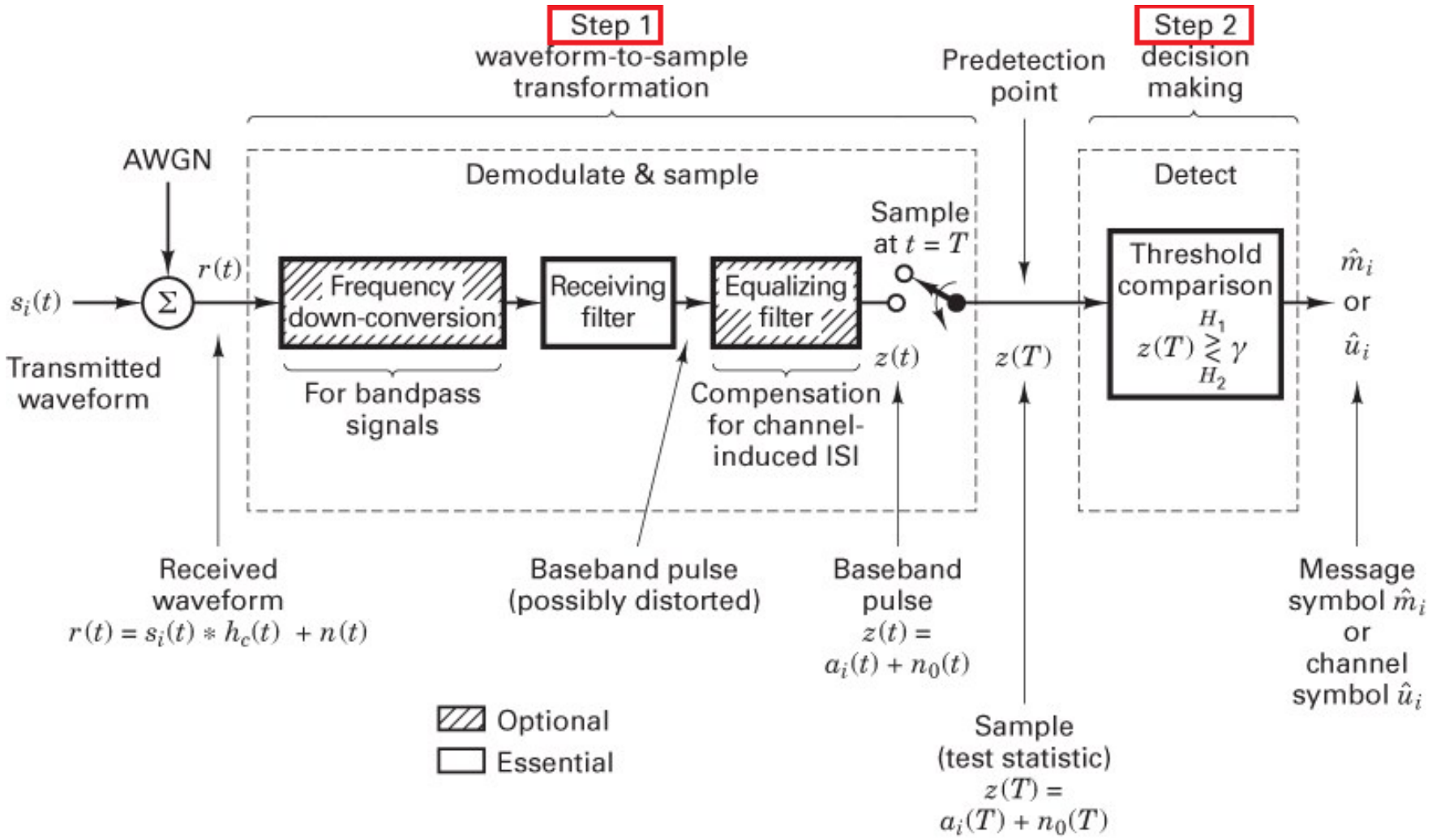
  1. A digital signal is best characterized by its received energy not its power, unlike an analog signal. This is because
     - Analog waveforms are classified as power signals since they typically have finite average power and infinite duration, resulting in infinite energy. Hence, energy is not a useful metric for analog waveforms; instead, power, which represents the rate of energy delivery, is more appropriate.
     - In digital communication, symbols are transmitted within a finite window of time (T). Over this limited duration, power averaged over all time ap-

proaches zero, making power an unsuitable metric for digital waveforms. Instead, symbol energy (power integrated over T) is a more meaningful parameter for characterizing digital signals.

2. A digital waveform can encode messages with varying levels of granularity, such as 1 bit (binary), 2 bits (4-ary) or even 10 bits (1024-ary). Unlike analog systems, which use continuous waveforms, digital systems require a metric that allows comparisons at the bit level, irrespective of the number of bits per symbol.

- To address these, E_b/N_0 is used which evaluates the energy efficiency of digital communication at the bit level, as it
  - provides a standardized metric that eliminates the need for adjustments based on the waveform's meaning or encoding structure.
  - directly impacts the bit error rate (BER) and overall system design.

2 Demodulation and Detection



Baseband Demodulation/Detection

- The core block is

$$r(t) = s(t) + n(t) \longrightarrow \boxed{\text{LTI filter of impulse response } h(t)} \longrightarrow z(t) = a(t) + n_0(t)$$

- The function of the detection is to detect $s(t)$ from the received signal $r(t)$

- Received signal (filter input): $r(t) = s(t) + n(t) \qquad 0 \le t \le T$
  - $s(t)$ is a general pulse representing 0 or 1
  - T is the duration of the pulses
  - $n(t)$ is a sample function of a *white noise* with zero mean and PSD $\frac{N_0}{2}$

- Filtered signal and noise components: $z(t) = a(t) + n_0(t)$
  - $a(t) = h(t) \star s(t)$ or equivalently $A(f) = H(f)S(f)$
  - $a(t) = \mathcal{F}^{-1}[A(f)] = \int_{-\infty}^{\infty} H(f)S(f)e^{j2\pi ft}df$
  - $\mathbb{E}[n^2(t)] = \frac{N_0}{2}\int_{-\infty}^{\infty} |H(f)|^2 df$ (average power of output noise)

Critical Questions:

– At what time should we sample the filter's output?

– What $H(f)$ maximizes the SNR?

- peak pulse SNR is defined as

$$\text{SNR} = \frac{|a(t)|^2}{\mathbb{E}[n_0^2(t)]}$$

where $|a(t)|^2$ is the instantaneous power of the signal

Baseband Demodulation/Detection

- when the filter output is sampled at $t = T$ we have

$$(\text{SNR})_T = \frac{|a(T)|^2}{\mathbb{E}[n^2(t)]} = \frac{\left| \int_{-\infty}^{\infty} H(f)S(f)e^{j2\pi fT} df \right|^2}{\frac{N_0}{2}\int_{-\infty}^{\infty} |H(f)|^2 df}$$

$$\leq \frac{\int_{-\infty}^{\infty} |H(f)|^2 df \times \int_{-\infty}^{\infty} |S(f)e^{j2\pi fT}|^2 df}{\frac{N_0}{2}\int_{-\infty}^{\infty} |H(f)|^2 df}$$

$$= \frac{2}{N_0}\int_{-\infty}^{\infty} |S(f)e^{j2\pi fT}|^2 df$$

$$= \frac{2}{N_0}\int_{-\infty}^{\infty} |S(f)|^2 df$$

- The inequality is due to Cauchy-Schwartz inequality[1]
- The equality holds if and only if $\phi_2(x) = k\phi_1^*(x)$.
- With $\phi_1(x) = H(f)$ and $\phi_2(x) = S(f)e^{j2\pi fT}$, the "$\leq$" will be "$=$" when

$$H(f) = S^*(f)e^{-j2\pi fT}.$$

[1] $\left| \int_{-\infty}^{\infty} \phi_1(x)\phi_2^*(x)dx \right|^2 \leq \int_{-\infty}^{\infty} |\phi_1(x)|^2 dx \cdot \int_{-\infty}^{\infty} |\phi_2(x)|^2 dx$ for any bounded complex functions $\phi_1(x)$ and $\phi_2(x)$.

- Thus,

$$\boxed{\text{SNR}_{\max} = \frac{2}{N_0}\int_{-\infty}^{\infty}|S(f)|^2 df = \frac{2E}{N_0}}$$

and, it is achieved by the following $H(f)$:

$$H_{\text{opt}}(f) = kS^*(f)e^{-j2\pi fT}$$

– The time domain representation of the filter

◇ in general, $h_{\text{opt}}(t) = \mathcal{F}^{-1}[H_{\text{opt}}(f)] = \int_{-\infty}^{\infty} kS^*(f)e^{-j2\pi f(T-t)}df$

◇ for a *real signal* $s(t)$, we have $S^*(f) = S(-f)$ and we obtain

$$h_{\text{opt}}(t) = ks(T-t)$$

◇ We often let $k = 1$ and get

$$\boxed{h_{\text{opt}}(t) = s(T-t)}$$

∗ This filter is called matched filter as it is matched to the signal

2.1 Correlation Realization of the Matched Filter

- Substituting $h(t) = s(T - t)$ in the convolution equation we get

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

- When $t = T$, we get

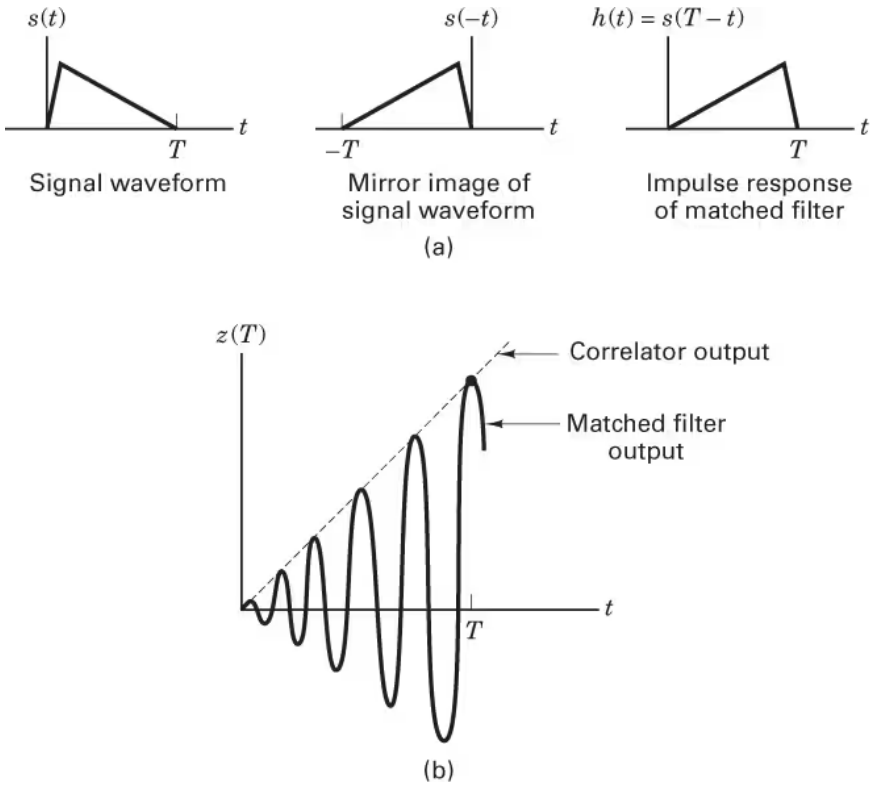
$$z(T) = \int_0^T r(\tau)s(\tau)\,d\tau$$

  which is the correlation of $r(t)$ and $s(t)$.

- Consider that a received signal $r(t)$ is correlated with each prototype signal $s_1(t), \ldots, s_M(t)$, using a bank of M correlators.
- The signal $s_i(t)$, whose correlation with $r(t)$ yields the maximum output, is the signal that matches $r(t)$ better than all the others.

- We will the use this correlation for the optimum detection of signals.

It is important to note that the correlator output and the matched filter output are the same only at time $t = T$.



Correlator and matched filter. **(a)** Matched filter characteristic. **(b)** Comparison of correlator and matched filter outputs for a sine wave input.

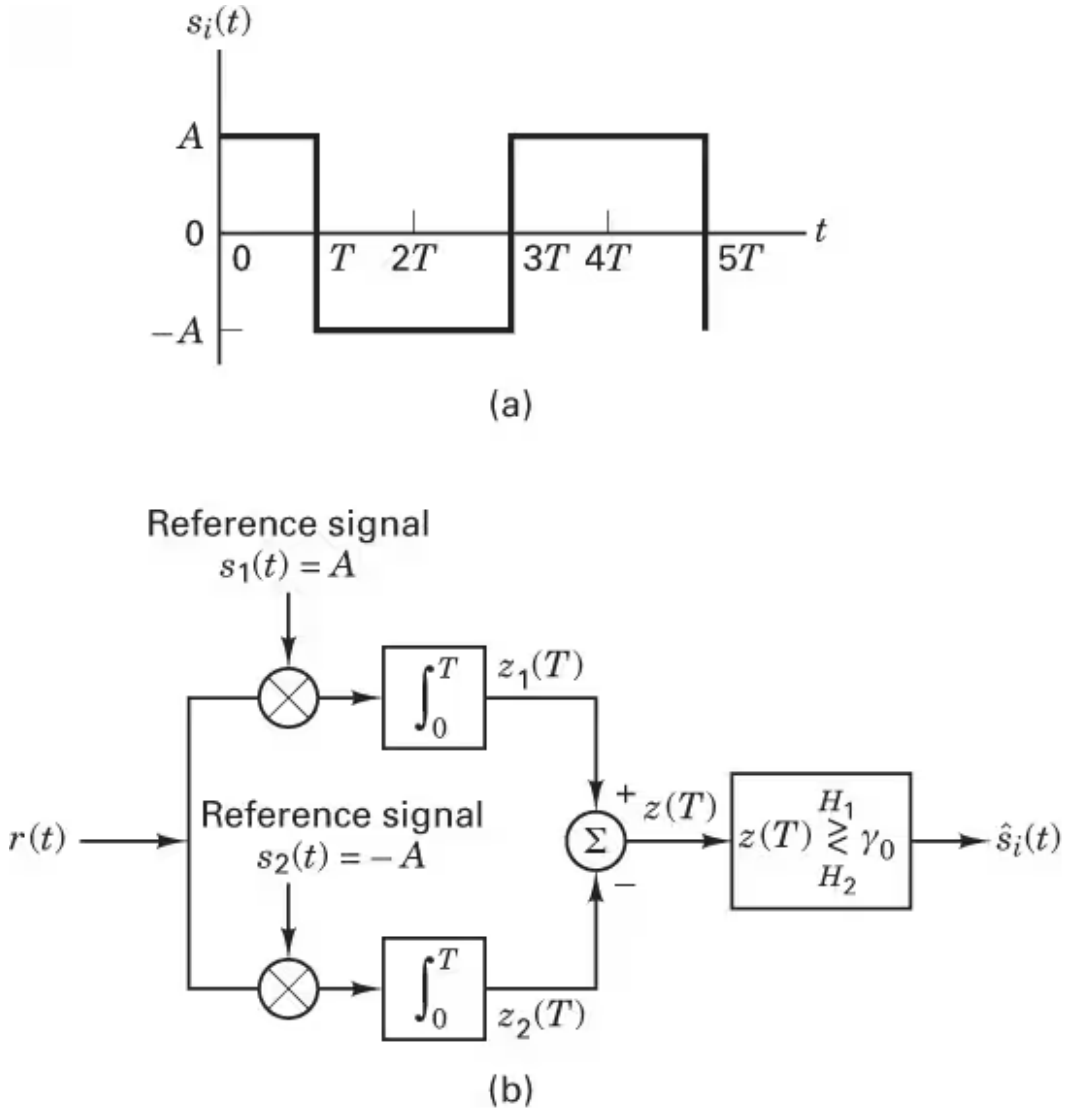
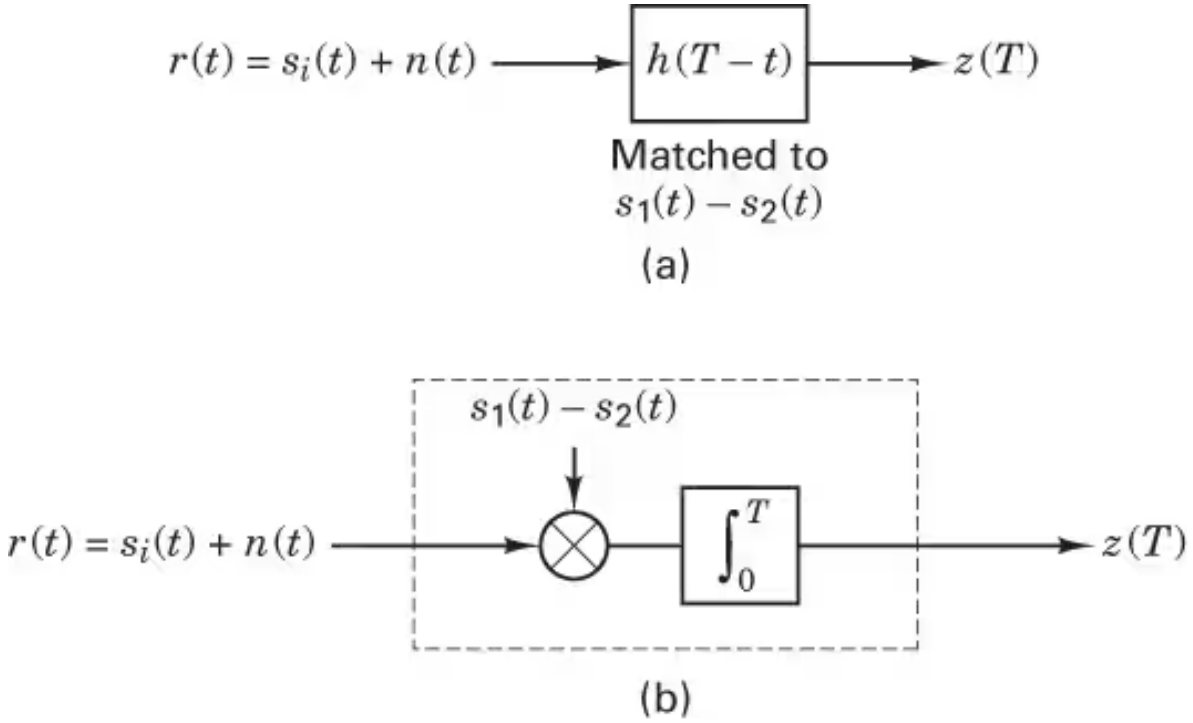
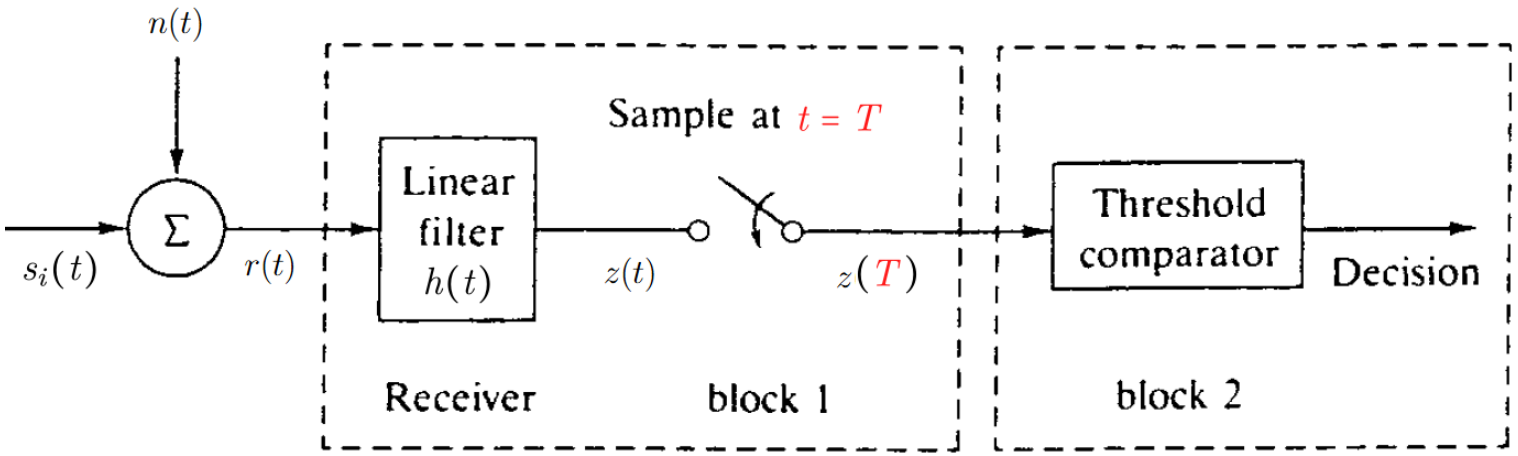


Figure 1: Detection of bipolar baseband signaling. (a) Bipolar signaling example. (b) Correlator detector.



Equivalence of matched filter and correlator. (a) Matched filter. (b) Correlator.



3 Binary Signal Detection

Two steps are involved in signal detection

- **Step 1:** reducing the received signal $r(t)$ into a number $z(T)$
  1. find a linear filter that *maximizes* output SNR [Answer: *matched filter*]
  2. sample the output of filter at certain time ($t = T$) to get a number $z(T)$
- **Step 2:** comparing $z(T)$ with a certain threshold and deciding which signal (i.e., $s_1(t)$ or $s_2(t)$) has been transmitted

We are going to discuss the last block (Step 2), in the following.

3.1 Error Analysis for AWGN Channel

- Consider binary transmission with

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

- The received signal $x(t)$ can be represented as

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

  where $n(t)$ is a zero mean AWGN.

- After filtering the signal $r(t)$ we can write

$$z(t) = a_i(t) + n_0(t), \quad 0 \le t \le T, \quad i = 1, 2$$

- After sampling at $t = T$, we get

$$z(T) = a_i(T) + n_0(T), \quad i = 1, 2$$

- For notational convenience, we write this as

$$z = a_i + n_0, \quad i = 1, 2$$

- The pdf of the Gaussian random noise n_0 can be expressed as

$$p_{n_0}(n) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{n^2}{2\sigma_0^2}}$$

  where σ_0^2 is the noise variance.

- Thus, it follows that the conditional pdfs $p(z|s_1)$ and $p(z|s_2)$ can be expressed as

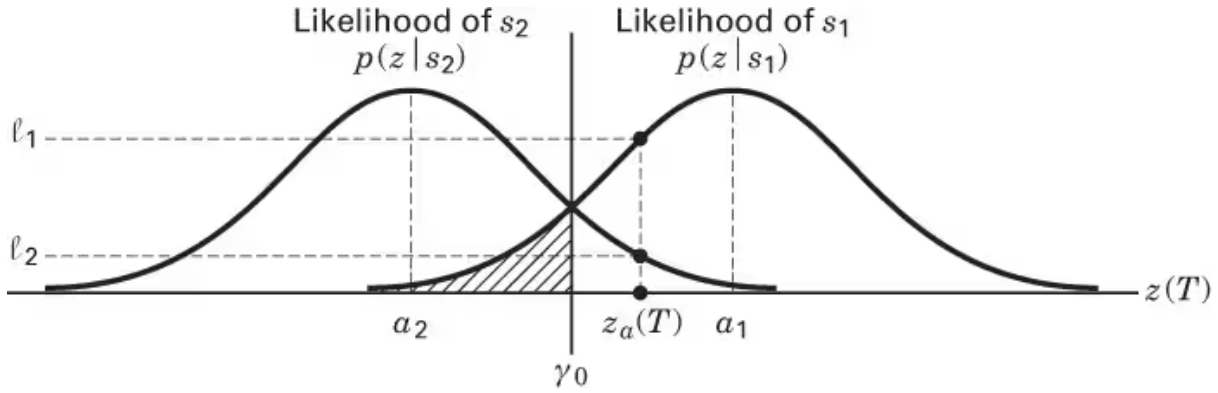
$$p(z|s_1 \text{ was sent}) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(z-a_1)^2}{2\sigma_0^2}} \tag{2}$$

  and

$$p(z|s_2 \text{ was sent}) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(z-a_2)^2}{2\sigma_0^2}} \tag{3}$$

  These conditional pdfs are illustrated in figure below.

- The rightmost conditional pdf, $p(z|s_1)$, called the likelihood of s_1, illustrates the probability density function of the random variable $z(T)$, given that symbol s_1 was transmitted.



Likelihoods of s_1 and s_2.

- Similarly, the leftmost conditional pdf, $p(z|s_2)$, called the likelihood of s_2, illustrates the pdf of $z(T)$, given that symbol s_2 was transmitted.
- At this point, by comparing $z(T)$ with a threshold γ, we can decide whether 0 or 1 was sent. Specifically, $z(T) \underset{s_2}{\overset{s_1}{\gtrless}} \gamma$.
- A popular criterion for choosing the threshold level γ for the binary decision is based on minimizing the probability of error.
- The computation for this minimum error value of γ starts with forming an inequality expression between the ratio of conditional probability density functions and the signal a priori probabilities.

- Since the conditional density function $p(z|s)$ is also called the likelihood of s, the formulation

$$\frac{p(z|s_1)}{p(z|s_2)} \underset{}{\overset{}{\gtrless}} \frac{p(s_2)}{p(s_1)} \tag{4}$$

  is called the likelihood ratio test.

- When $p(s_1) = p(s_2) = \frac{1}{2}$, the optimum threshold is $\gamma = \frac{a_1+a_2}{2} \triangleq \gamma_0$. That is, the decision is made based on the distance of z from a_1 and a_2.
  - If z is closer to a_1, we decide 1 (s_1) has been sent.
  - Otherwise, we decide 0 (s_2) has been sent.

3.2 Evaluating the Probability of Errors

- Now, we can find the decision error; there are two possible error events
  - 1 (s_1) is chosen when 0 (s_2) was actually sent $\triangleq p_{e_2}$
  - 0 (s_2) is chosen when 1 (s_1) was actually sent $\triangleq p_{e_1}$
- Mathematically,

$$p_{e_2} = \int_{\gamma_0}^{+\infty} p(z|s_2)dy$$

$$p_{e_1} = \int_{-\infty}^{\gamma_0} p(z|s_1)dy$$

- Finally, $P_e = p(s_1)p_{e_1} + p(s_2)p_{e_2}$

- Clearly, $p_{e_1} = p_{e_2}$. Thus, when $p(s_1) = p(s_2) = \frac{1}{2}$ we get $P_e = p_{e_1} = p_{e_2}$

$$\begin{aligned} P_e = p_{e_2} &= \int_{\gamma_0 = \frac{a_1+a_2}{2}}^{+\infty} p(z|s_2)dy \\ &= \int_{\gamma_0}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(z-a_2)^2}{2\sigma_0^2}} dy \\ &= Q\left(\frac{a_1 - a_2}{2\sigma_0}\right) \end{aligned}$$

  where the last step is obtained by changing variable $x = \frac{z-a_2}{\sigma_0}$ and using the definition of Q-function.

- Next, for minimizing P_e, it is necessary to choose the filter (matched filter) that maximizes the argument of $Q(x)$.

- Thus, we need to determine the linear filter that maximizes $\frac{a_1-a_2}{2\sigma_0}$ or, equivalently, that maximizes

$$\frac{(a_1 - a_2)^2}{\sigma_0^2}$$

Binary Signal Detection

- Consider that the filter is matched to the input difference signal $[s_1(t) - s_2(t)]$; thus, we can write the output SNR at time $t = T$ as

$$(\text{SNR})_T = \frac{|a_1 - a_2|^2}{\sigma_0^2} = \frac{2E_d}{N_0}$$

  - $\frac{N_0}{2}$ is the two-sided PSD of the noise at the filter input, and
  - $E_d = \int_0^T [s_1(t) - s_2(t)]^2 dt$ is the energy of $s_1(t) - s_2(t)$.

- Finally, $P_e = Q\left(\frac{a_1 - a_2}{2\sigma_0}\right) = Q\left(\sqrt{\frac{E_d}{2N_0}}\right)$

Summary: In binary signal detection,

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

where $E_d = \int_0^T [s_1(t) - s_2(t)]^2 dt$ It is the energy of the difference between the two pulses.

- This represents both `symbol error rate (SER)` and `bit error rate (BER)` since the signaling is binary.

- It is also very common to write P_e in terms of average energy per bit (E_b)
- $\boxed{E_b \triangleq \frac{E_1+E_2}{2}} = \frac{1}{2}[\int_0^T s_1^2(t)dt + \int_0^T s_2^2(t)dt]$
- Then, we just need to find the relationship between E_b and E_d

Example 1. *Probability of Error for Unipolar Signaling*

- $s_i(t) = \begin{cases} s_1(t) = A, & 0 \le t \le T, \text{ for 1} \\ s_2(t) = 0, & 0 \le t \le T, \text{ for 0} \end{cases}$
- $E_d = \int_0^T [s_1(t) - s_2(t)]^2 dt = \int_0^T A^2 dt = A^2T = 2E_b$ *(why?)*
- $P_e = Q(\sqrt{\frac{E_d}{2N_0}}) = Q(\sqrt{\frac{A^2T}{2N_0}}) = Q(\sqrt{\frac{E_b}{N_0}})$
- *Note that* $E_1 = \int_0^T s_1^2(t)dt = A^2T$ *and* $E_2 = 0$*; thus,* $E_b = \frac{A^2T}{2}$

Example 2. *Probability of Error for Bipolar Signaling (Figure 1)*

- $s_i(t) = \begin{cases} s_1(t) = A, & 0 \le t \le T, \quad \textit{for 1} \\ s_2(t) = -A, & 0 \le t \le T, \ \textit{for 0} \end{cases}$
- $E_d = \int_0^T [s_1(t) - s_2(t)]^2 dt = \int_0^T [2A]^2 dt = 4A^2T$
- $P_e = Q\left(\sqrt{\frac{E_d}{2N_0}}\right) = Q\left(\sqrt{\frac{2A^2T}{N_0}}\right)$
- *Since* $E_b = \frac{1}{2}[\int_0^T s_1^2(t)dt + \int_0^T s_2^2(t)dt] = A^2T$, *we can also write*

$$P_e = Q\left(\sqrt{\frac{E_d}{2N_0}}\right) = Q\left(\sqrt{\frac{2A^2T}{N_0}}\right) = Q\left(\sqrt{\frac{2E_b}{N_0}}\right)$$

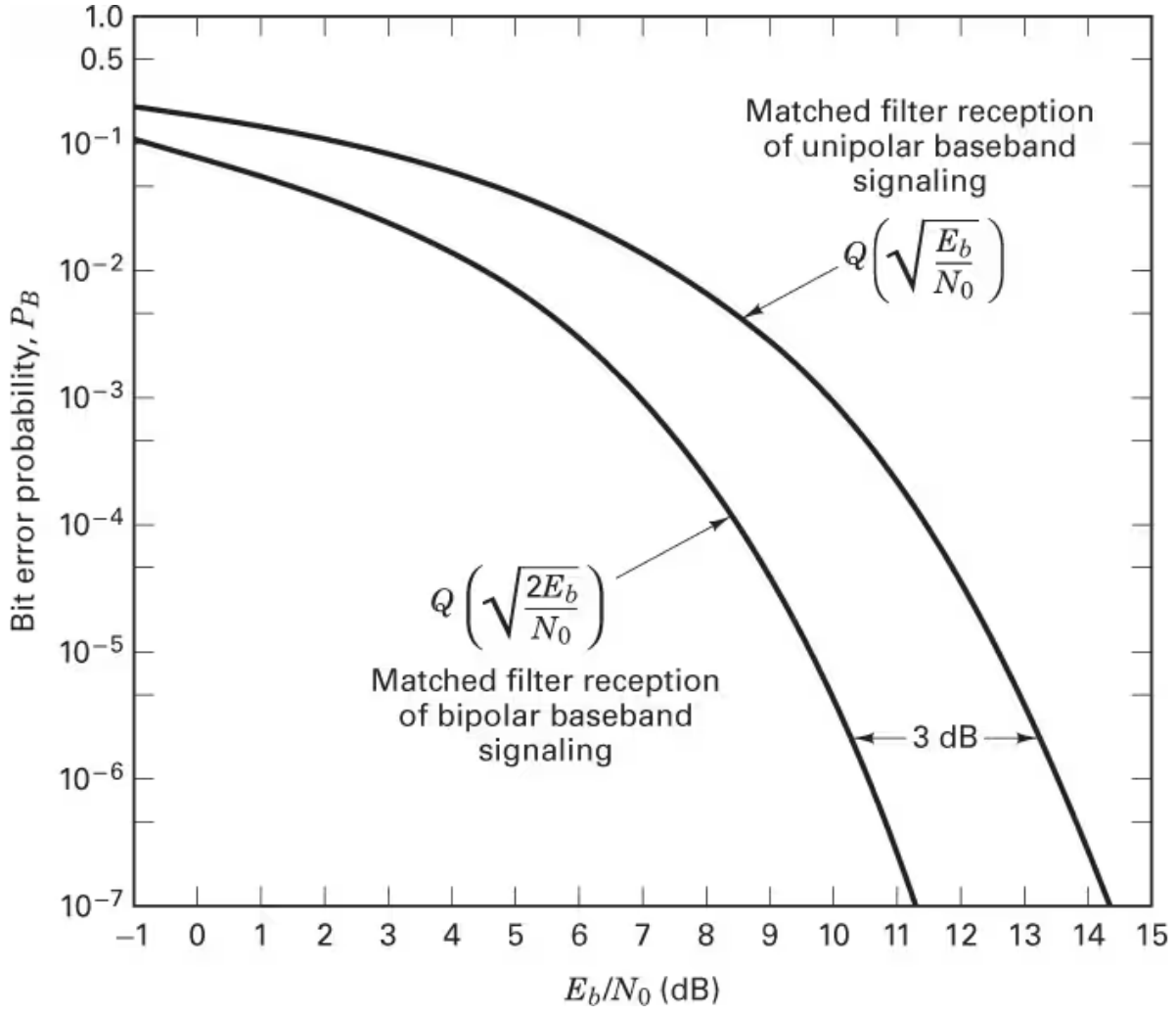
Example 3. *Assuming that $A = 5\ mV$, bit rate is 10^3 bits/s, and noise power spectral density is $\frac{N_0}{2} = 10^{-9}$ Watts/Hz find the probability of error for bipolar and unipolar signalings.*

- *Unipolar:*

$$P_e = Q\left(\sqrt{\frac{A^2T}{2N_0}}\right) = Q\left(\sqrt{\frac{0.005^2 \times 10^{-3}}{4 \times 10^{-9}}}\right) = Q(2.5) = 6.2 \times 10^{-3}$$

- *bipolar:*

$$P_e = Q\left(\sqrt{\frac{2A^2T}{N_0}}\right) = Q\left(\sqrt{\frac{2 \times 0.005^2 \times 10^{-3}}{2 \times 10^{-9}}}\right) = Q(5) = 2.8665 \times 10^{-7}$$



;

Figure 2: BER versus $\frac{E_b}{N_0}$ for unipolar and bipolar signalings.

3.3 Energy and Correlation

- $P_e = Q\left(\sqrt{\frac{E_d}{2N_0}}\right)$ is an interim result for the **energy of the difference signals** at the matched filter's input.
- From this a more general relation based on **bit energy** can be derived.
- A time cross-correlation coefficient ρ measures similarity between two signals:
  - Classical mathematical expression: $\rho = \frac{\int s_1(t)s_2(t)\,dt}{\sqrt{E_1 E_2}}$
  - Vector representation: $\rho = \cos\theta$
- θ is the angle between signal vectors:
  - Small θ: High correlation (similar signals).
  - Large θ: Low correlation (dissimilar signals).
- Expanding $E_d = \int_0^T [s_1(t) - s_2(t)]^2 dt$ yields

$$E_d = E_1 + E_2 - 2\rho\sqrt{E_1 E_2}$$

- When $E_1 = E_2 = E_b$, we get

$$E_d = 2E_b(1-\rho) \qquad \Longrightarrow \qquad P_e = Q\left(\sqrt{\frac{E_b(1-\rho)}{N_0}}\right)$$

- Special Cases of Correlation:
  - Perfect correlation ($\rho = 1$):
    - ◇ Signals are identical (not useful for digital signaling).
  - Anticorrelation ($\rho = -1$):
    - ◇ Signals are antipodal (Vectors $\mathbf{s}_1$ and $\mathbf{s}_2$ are 180° apart)
    - ◇ **Example:** Binary antipodal signals.
  - Zero correlation ($\rho = 0$):
    - ◇ Signals are orthogonal (Vectors $\mathbf{s}_1$ and $\mathbf{s}_2$ are 90° apart).
    - ◇ **Example:** Binary orthogonal signals.

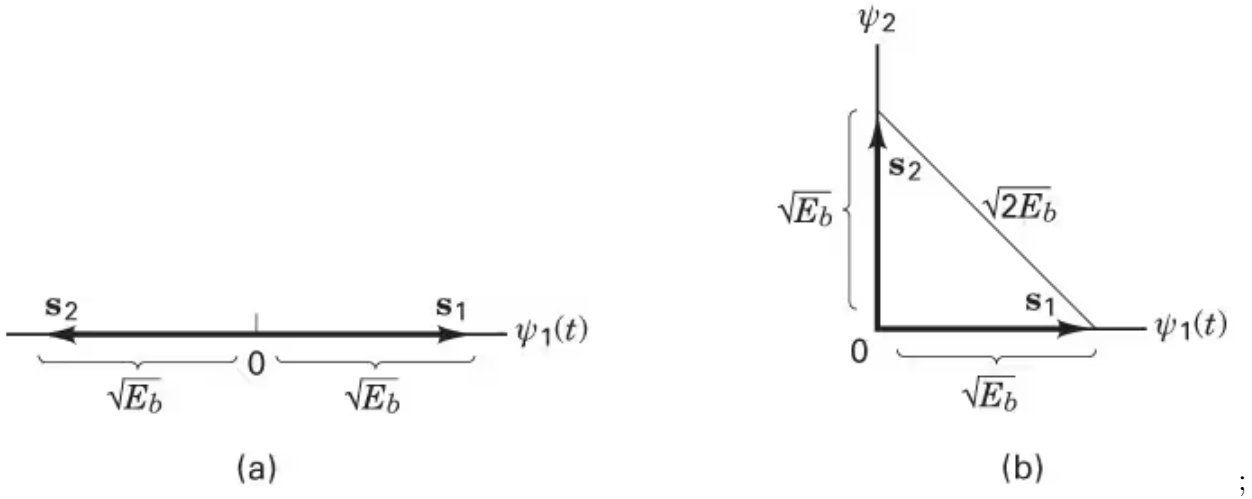


Figure 3: Binary signal vectors. (a) antipodal, (b) orthogonal

- Visualization and Matched Filter Detection
  - For antipodal signals ($\rho = -1$): Distance $= 2\sqrt{E_b}$, Energy $= 4E_b$.
  - For orthogonal signals ($\rho = 0$): Distance $= \sqrt{2E_b}$, Energy $= 2E_b$.
  - Error performance depends on the distance between signals:

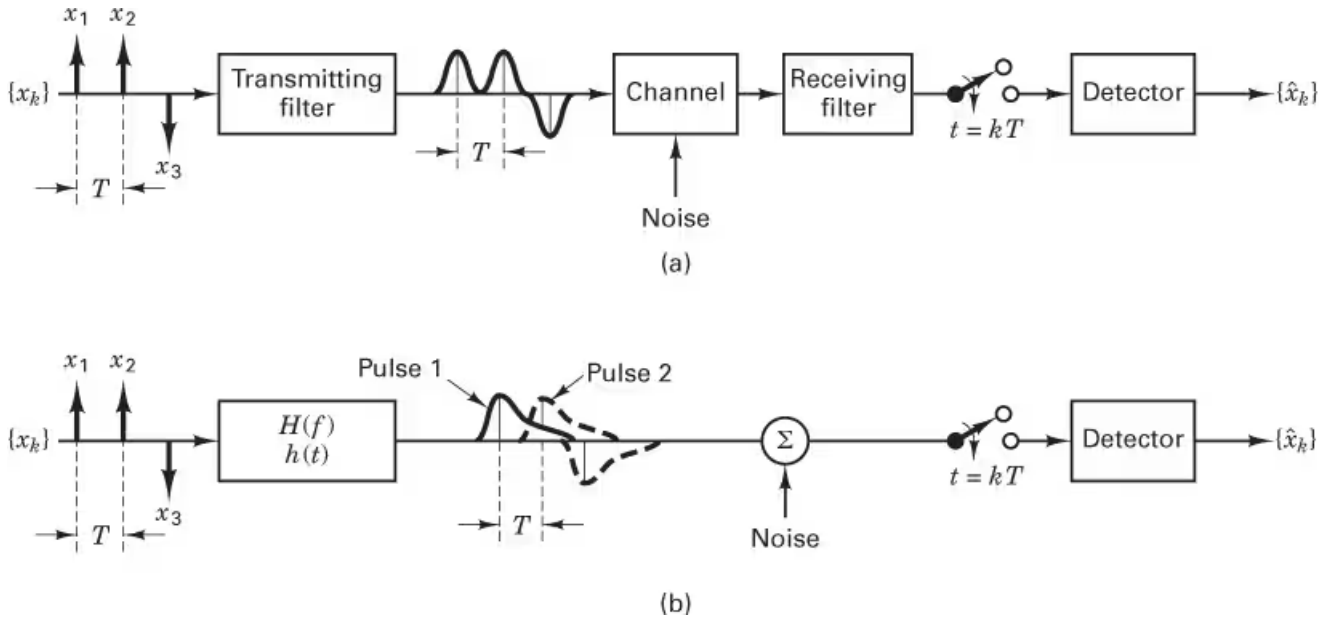


Figure 4: ISI in detection process. (a) Typical baseband systems, (b) Equivalent model.

4 Filtering without Intersymbol Interference (ISI)

Filtering & ISI

- Filtering aspects in a typical digital communication system.
  - **Transmitter:** Modulates pulses to comply with bandwidth constraints.
  - **Channel:** Distorts pulses due to distributed reactances (e.g., cables) or fading.

  - **Receiver:** Compensates for distortion using equalizing filters.

- The receiving filter is often referred to as an equalizing filter.
- An equivalent model for the system: all filtering effects are combined into one overall system transfer function:

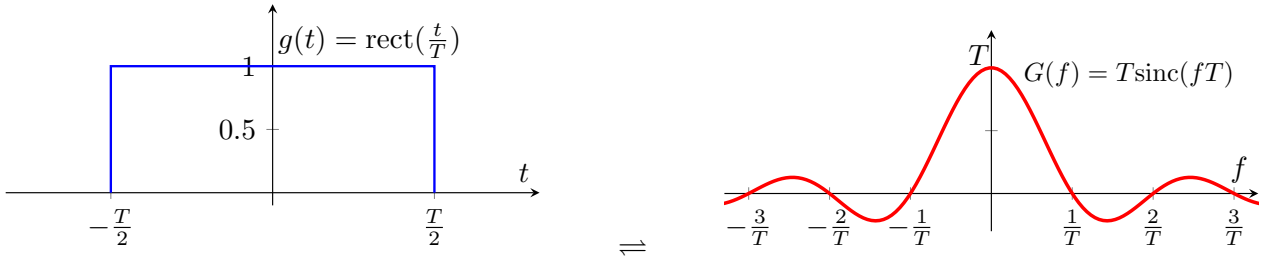
$$H(f) = H_t(f) \cdot H_c(f) \cdot H_r(f)$$

  - $H_t(f)$: Transmitting filter.
  - $H_c(f)$: Channel filtering.
  - $H_r(f)$: Receiving/equalizing filter.

- $H(f)$ represents the composite system transfer function.
- Filtering causes pulse overlap, leading to ISI:
  - Tails of pulses "smear" into adjacent symbol intervals.
  - Degrades error performance even in the absence of noise.

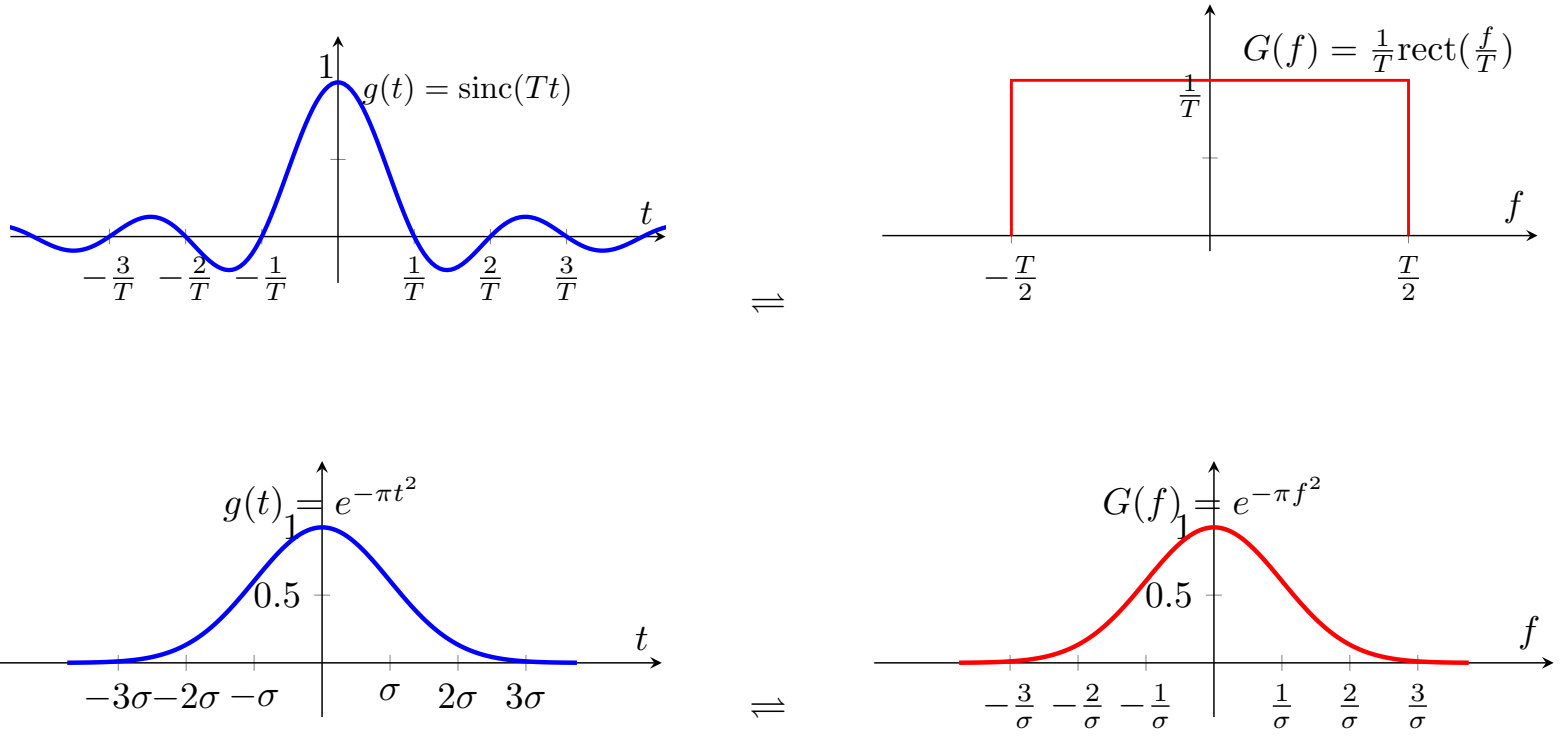
4.1 Why Pulse Shaping?

- Communication channels are **bandlimited**. Thus, we cannot transmit a rectangular pulse (it needs infinite bandwidth)



- We recall from Fourier transform properties that: time and frequency domains are **inversely related**
  - a band-limited signal extends infinitely in time, and vice versa.
  - stretching in one domain compresses the other by the same factor.
  - a smoother pulse results in a smoother Fourier transform, e.g. $e^{-\pi t^2} \rightleftharpoons e^{-\pi f^2}$

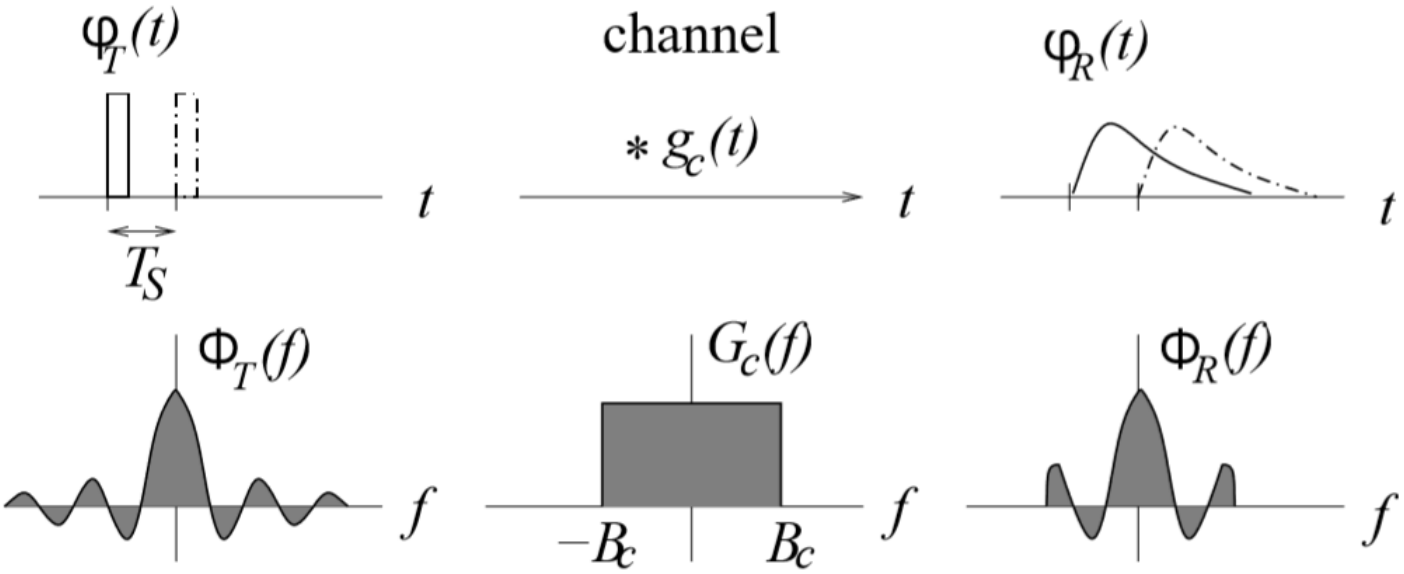
Pulse Shaping/Filtering without ISI



In general, we cannot use impulse, narrow pulse, or rectangular pulse to transmit data symbols. Therefore, discrete samples must be **pulse shaped**.

Pulse Shaping/Filtering without ISI

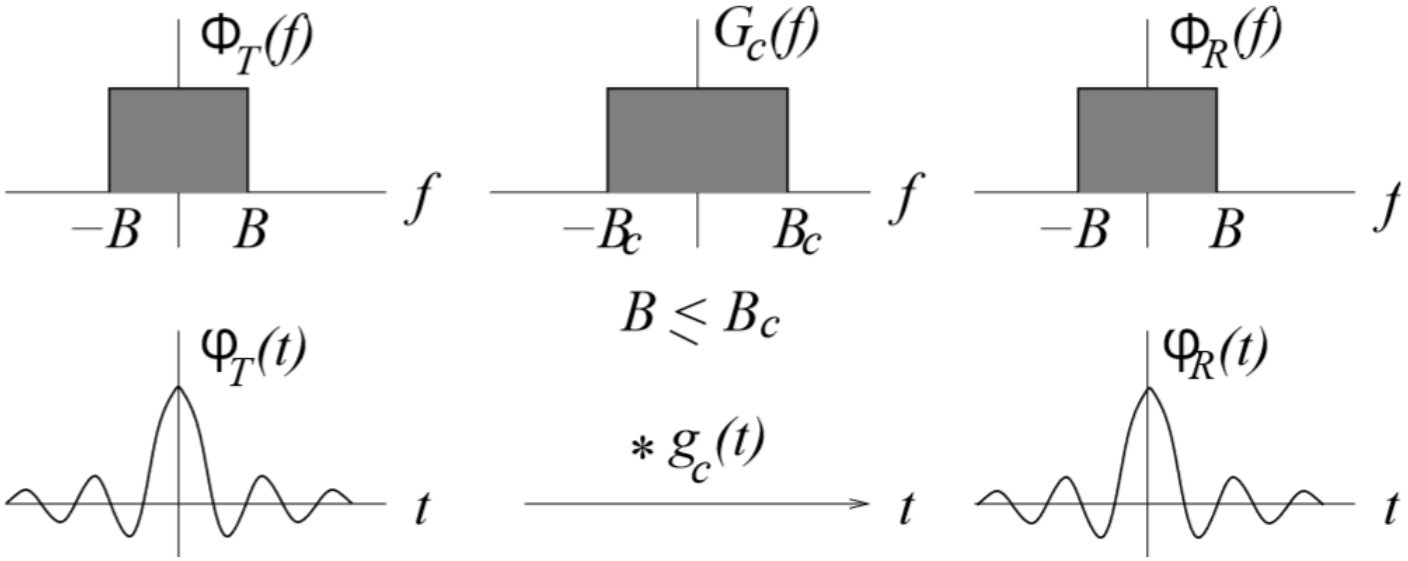
- If we transmit a narrow-width (or rectangular) pulses, and use a filter to limit the bandwidth then



- The pulses will spread out as their high-frequency components are suppressed, causing interference with neighboring pulses (symbols).
- This interference is called inter symbol interference (ISI).
  - Tails of pulses "smear" into adjacent symbol intervals.
  - Degrades error performance even in the absence of noise.

Pulse Shaping/Filtering without ISI

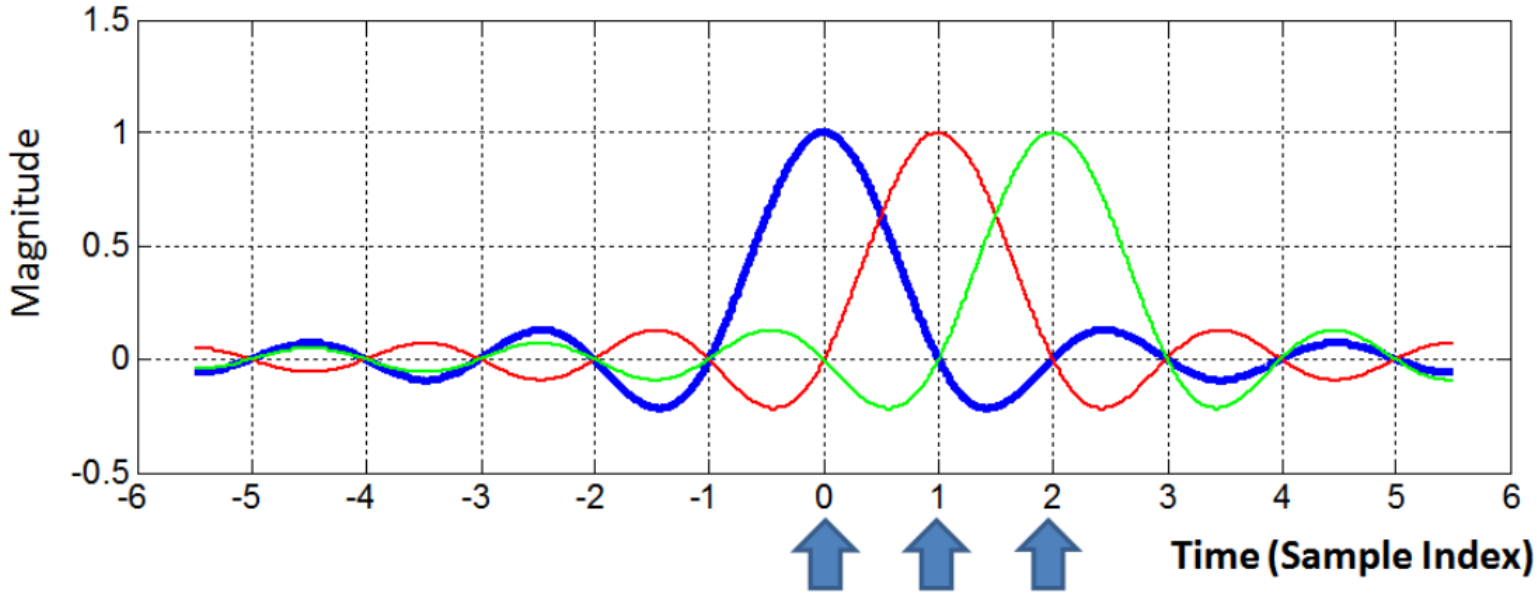
- Transmitted pulses should have a bandwidth $\leq$ channel bandwidth



- Then, a pulse (symbol) will be passed through the channel without distortion.
- However, a finite bandwidth means the waveform extends infinitely in the time domain, which would interfere in the time domain with neighboring pulses, **unless specific conditions are met**.
- The pulse shaping deals with this challenge!
- Nyquist investigated pulse shapes to eliminate ISI.

Pulse Shaping/Filtering without ISI

- ***Main idea:*** Pulses can spread more than one symbol time (T) as long as they do not interfere at sampling time.



- **Example:** the above pulses (sinc pulse) are band limited and at the sampling time no pulse interfere the other one.

Nyquist's Criteria for Zero ISI

Summary!

- Theoretical minimum bandwidth to detect R_s symbols/s without ISI is $\frac{R_s}{2}$ Hz. (recall that $R_s = 1/T$, and T is the pulse (symbol) duration).
- In other words, a system with a bandwidth of W Hz can support maximum transmission rate of $R_s = 2W$ symbols/s.

  Thus, the maximum symbol-rate packing is **2 symbols/s/Hz**.
- This occurs when overall $H(f)$ is made rectangular.

- Ideal Nyquist filter
  - Rectangular transfer function $H(f)$
  - Impulse response: $h(t) = \text{sinc}(t/T)$
- Key property of the ideal Nyquist pulse:
  - Tails pass through zero at sampling instants, eliminating ISI.
  - Assuming that the sample timing is perfect, there will be no ISI.

4.2 Practical Nyquist Filters

- The names "Nyquist filter" and "Nyquist pulse" are often used to describe the general class of filtering and pulse-shaping that satisfies zero ISI at the sampling points. Note that **Nyquist pulse $\leftrightharpoons$ Nyquist filter.**
- A Nyquist filter is a filter whose transfer function can be represented by a rectangular function convolved with any real even-symmetric frequency function.
- A Nyquist pulse is a pulse whose shape can be represented by a $\text{sinc}(\frac{t}{T})$ function multiplied by another time function.
  - Hence, there are a countless number of Nyquist filters and corresponding pulse shapes.
  - Among the class of Nyquist filters, the most popular ones are the **raised cosine** and the **root-raised cosine**.

- Ideal Nyquist filters are not realizable due to infinite pulse tails. Popular approximations:
  - Raised cosine filters.

– Root-raised cosine filters.

- The most popular form for $R_p(f)$ is the *spectral* **raised-cosine**, defined by

$$R_p(f) = \begin{cases} T, & 0 \le |f| \le \frac{1-\alpha}{2T} \\ \frac{T}{2}\left\{1 + \cos\left[\frac{\pi T}{\alpha}(|f| - \frac{1-\alpha}{2T})\right]\right\}, & \frac{1-\alpha}{2T} \le |f| \le \frac{1+\alpha}{2T} \\ 0, & |f| > \frac{1+\alpha}{2T} \end{cases}$$

  where $0 \le \alpha \le 1$ is the roll-off factor

- It's called the raised-cosine filter because, for $\alpha = 1$ its frequency response forms a half-cycle of a cosine wave shifted above the x-axis.

- The bandwidth of $R_p(f)$ is

$$W = \frac{1+\alpha}{2T},$$

  which is grater than the minimum bandwidth $W_{\text{Nyquist}} = \frac{1}{2T}$.

- The parameter α is also called *roll-off factor* because it controls the transition band. It is also called the *excess bandwidth*.

Pulse Shaping/Filtering without ISI

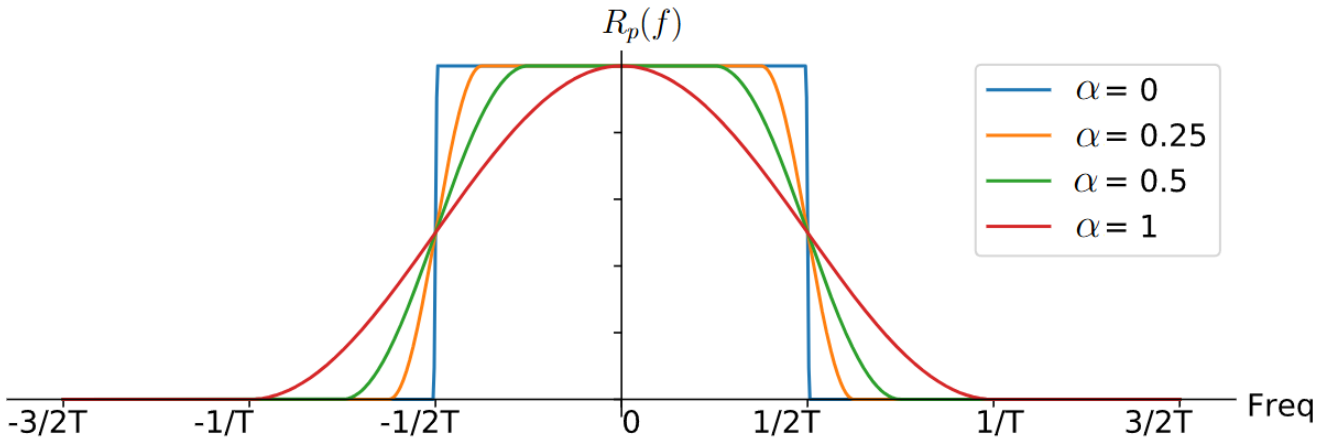


Figure 5: Transfer function of a raised cosine filter/pulse.

- The inverse Fourier transform of the raised cosine spectrum is given by:

$$r_p(t) = T\text{sinc}(t/T) \cdot \frac{\cos(\pi\alpha t/T)}{1 - (2\alpha t/T)^2}$$

  where:

  – The cosine term introduces a smooth roll-off controlled by α.

  – The denominator ensures proper decay and prevents singularities.

  Bandwidth slightly larger than $\frac{1}{2T}$; decay in time $|r_p(t)| \sim \frac{1}{|t|^3} \implies$ much narrower pulse.

Pulse Shaping/Filtering without ISI

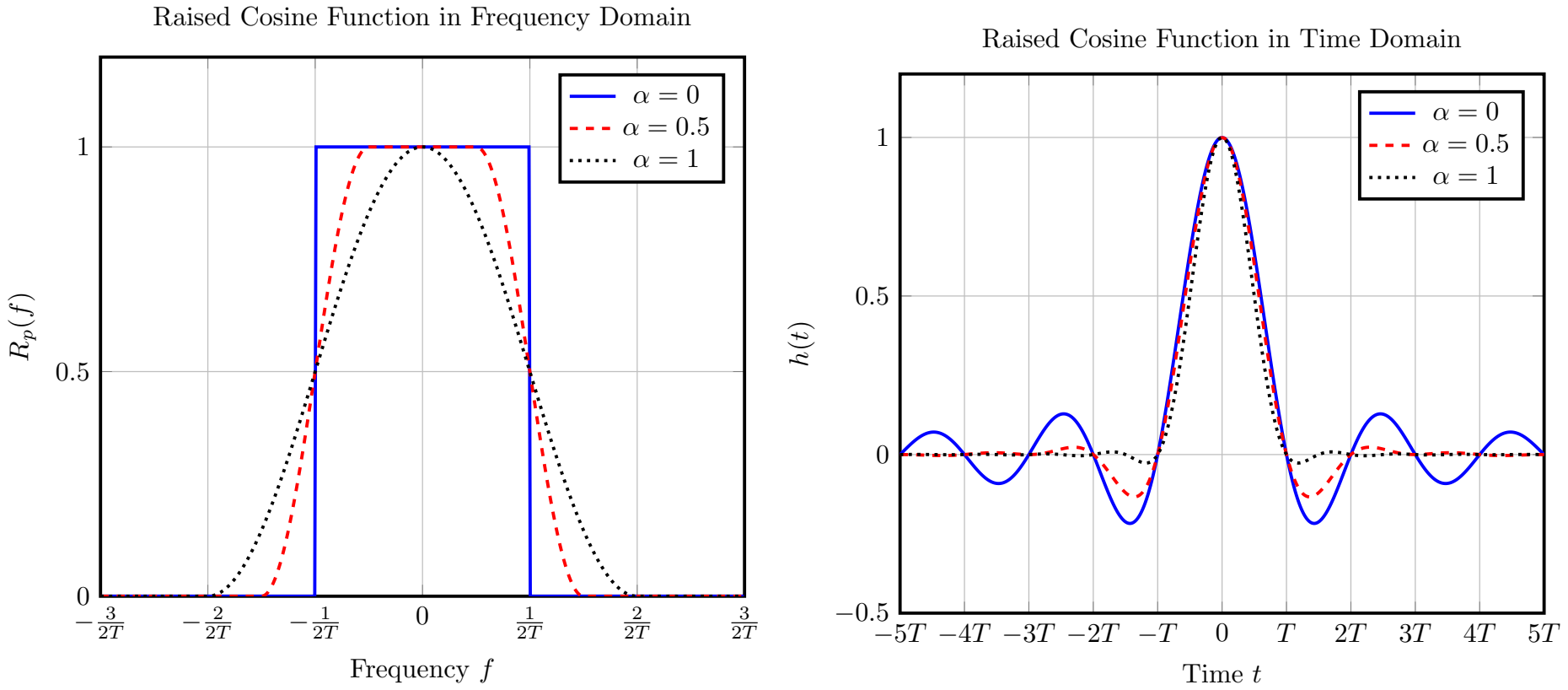


Figure 6: Raised-cosine spectrum (left) and corresponding time-domain waveform (right) for excess bandwidths of 0%, 50%, and 100%.

Pulse Shaping/Filtering without ISI

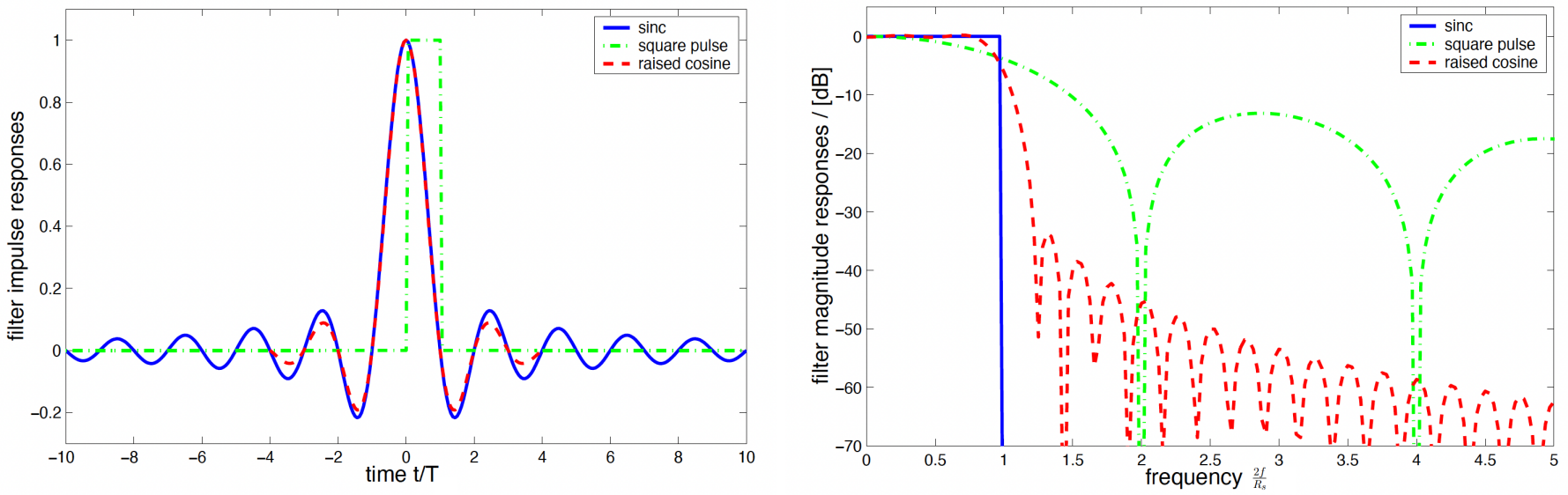
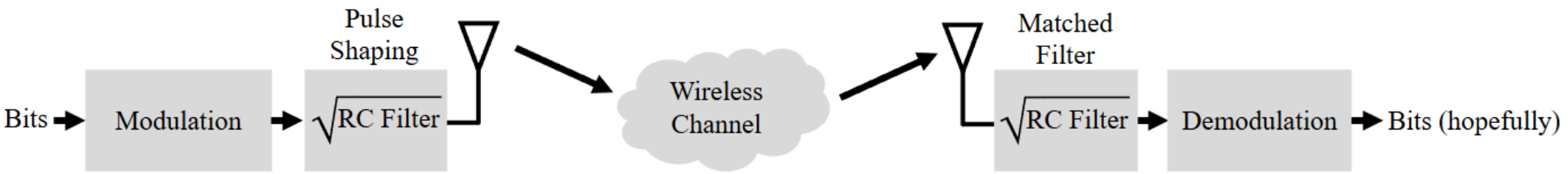


Figure 7: Square, sinc, and truncated raised cosine filters and the magnitude of their transfer function (in dB).

- Comparing sinc, raised cosine, and square (rectangle) pulses
  - Square pulse produces considerable excess bandwidth beyond the symbol rate $\frac{1}{2T} = R_s/2$
  - sinc impractical to realize
  - truncated raised cosine easy to realize



;

Figure 8: A raised cosine (RC) filter being split into two root raised cosine filters.

Square Root Raised Cosine Pulse Shape

- The $R_p(f)$ for the `square root raised cosine (RRC)` pulse shape is

$$R_p(f) = \begin{cases} \sqrt{T}, & 0 \leq |f| \leq \frac{1-\alpha}{2T} \\ \sqrt{\frac{T}{2}}\sqrt{1+\cos\left[\frac{\pi T}{\alpha}(|f| - \frac{1-\alpha}{2T})\right]}, & \frac{1-\alpha}{2T} \leq |f| \leq \frac{1+\alpha}{2T} \\ 0, & |f| > \frac{1+\alpha}{2T} \end{cases}$$

Then,

$$r_p(t) = \frac{1}{\sqrt{T}} \frac{\sin\left[\pi\frac{t}{T}(1-\alpha)\right] + 4\alpha\frac{t}{T}\cos\left[\pi\frac{t}{T}(1+\alpha)\right]}{\pi\frac{t}{T}\left[1-\left(4\alpha\frac{t}{T}\right)^2\right]}$$

5 The Nyquist Pulse Shaping (Filtering) Criterion

Theorem 1. ***(Nyquist no-ISI Pulse-Shaping Criterion):***
A necessary and sufficient condition for $r_p(t)$ *to satisfy*

$$r_p(nT) = \begin{cases} 1, & n = 0 \\ 0, & \text{other integer } n \end{cases}$$

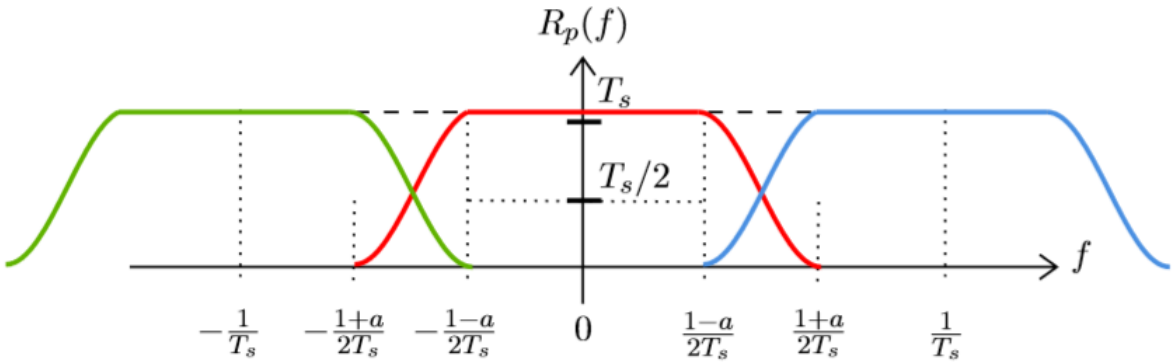
is that its Fourier transform $R_p(f)$ *satisfies*

$$\sum_{m=-\infty}^{\infty} R_p\left(f + \frac{m}{T}\right) = T.$$

That is, $1/T$*-frequency-shifted copies of* $R_p(f)$ *must add up to a constant.*

Proof: See Chapter 9.

- This ensures that there is no ISI in the transmitted signal.
- A pulse $r_p(t)$ satisfying the above condition is called "Nyquist pulse", and its Fourier transform is called "Nyquist filter".



;

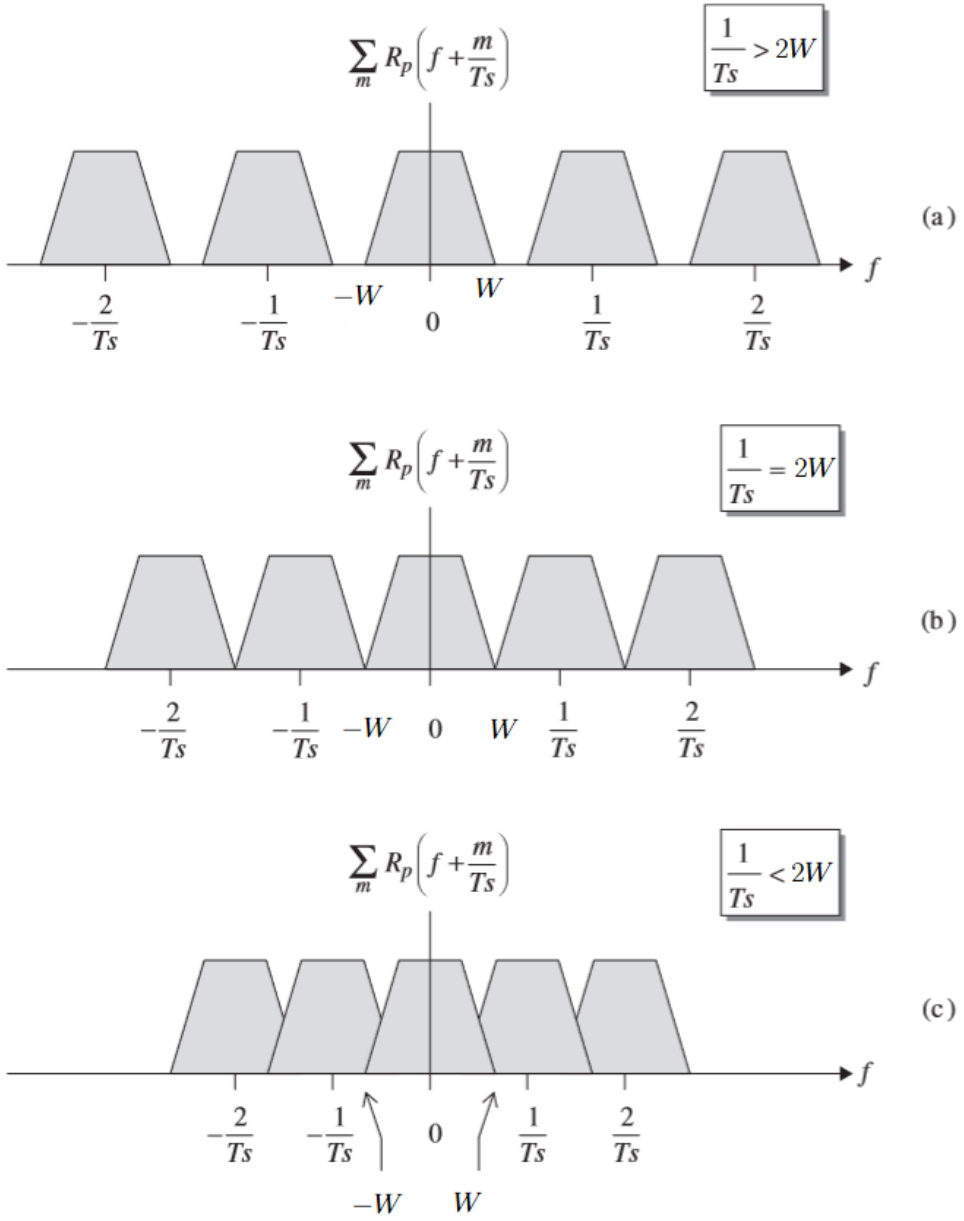
Figure 9: The Nyquist filtering criterion: $1/T$-frequency-shifted copies of $R_p(f)$ must add up to a constant (T). This plot shows $R_p(f)$ for the "square root raised cosine" pulse shape.

- **Examples of Nyquist Filter**
  - $R_p(f) = \text{rect}(fT)$ which is exactly constant within $-\frac{1}{2T} < f < \frac{1}{2T}$ and zero outside is a Nyquist filter.
  - Nyquist filter may bleed into the next *channel*, but the sum of

$$R_p\left(f - \frac{1}{T}\right) + R_p(f) + R_p\left(f + \frac{1}{T}\right) + \ldots$$

    must remain constant across all f.

Pulse Shaping/Filtering without ISI



;

Pulse Shaping/Filtering without ISI

- Now suppose that the channel has a bandwidth of W. We distinguish three cases:

  1. When $T < \frac{1}{2W}$, or, equivalently, $\frac{1}{T} > 2W$, since

$$B(f) = \sum_{n=-\infty}^{+\infty} R_p(f + n/T)$$

  consists of non-overlapping replicas of $R_p(f)$, as shown in previous figure, there is no choice for $R_p(f)$ to ensure $B(f) \equiv T$ in this case, and there is no way that we can design a system with no ISI.

  This means that the smallest value of T for which transmission with zero ISI is possible is $T = \frac{1}{2W}$.

  2. When $T = \frac{1}{2W}$, or, equivalently, $\frac{1}{T} = 2W$ (the Nyquist rate).
  There is only one $R_p(f)$ that results in $B(f) \equiv T$, and that is a rectangle for which, $r_p(t)$ has to be a sinc function.
     – The difficulty with this choice of $r_p(t)$ is that it is noncausal and, therefore, nonrealizable.

– To make it realizable, usually a delayed version of it, i.e.,

$$\text{sinc}\left(\frac{\pi(t-t_0)}{T}\right)$$

is used, and t_0 is chosen such that for $t < 0$, we have $\text{sinc}\left(\frac{\pi(t-t_0)}{T}\right) \approx 0$.

– With this choice of $r_p(t)$, the sampling time must also be shifted to $mT + t_0$.

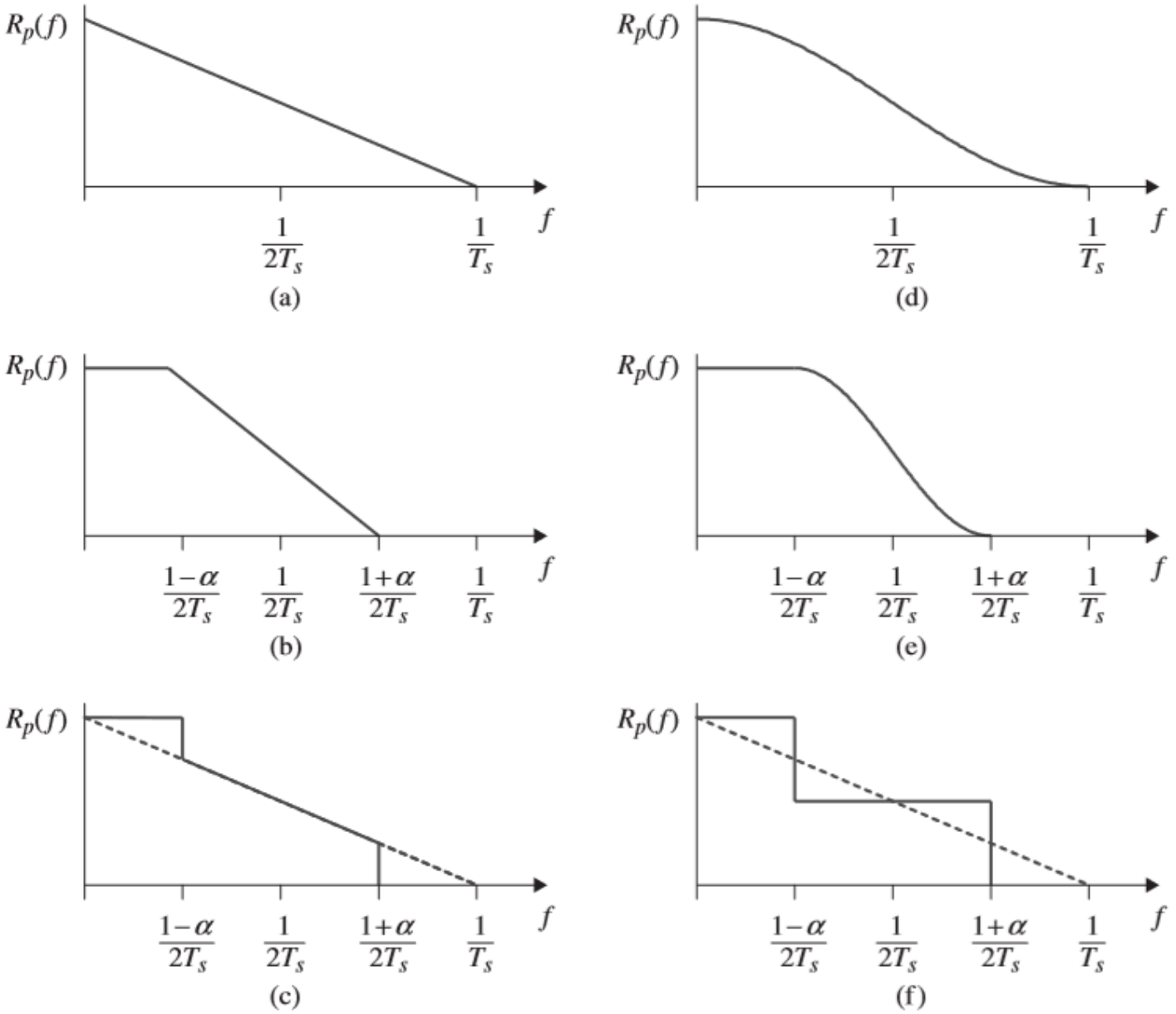
– A second difficulty with this pulse shape is that its rate of convergence to zero is slow. The tails of $r_p(t)$ decay as $1/t$; consequently, a small mistiming error in sampling the output of the matched filter at the demodulator results in an infinite series of ISI components.

– Such a series is not absolutely summable because of the $1/t$ rate of decay of the pulse, and, hence, the sum of the resulting ISI does not converge.

3. When $T > \frac{1}{2W}$,
$B(f)$ consists of overlapping replications are shown in the figure. In this case, there exist numerous choices for $R_p(f)$ such that $B(f) \equiv T$.

- A few of such choices are shown in figure below.

Pulse Shaping/Filtering without ISI



;

Pulse Shaping/Filtering without ISI

For zero ISI, it is necessary that the symbol rate $\frac{1}{T}$ be below the Nyquist rate of $2W$ symbols per second to enable the practical implementation of transmitting and receiving filters.

- According to Nyquist's criterion, the minimum required bandwidth for no-ISI is given by:

$$W_{\text{Nyquist}} = \frac{R_s}{2}$$

  where R_s is the symbol rate.

- With pulse-shaping filter with a roll-off factor α, the Nyquist's criterion becomes

$$W = \frac{1}{2}(1+\alpha)R_s.$$

- The above cases are for baseband signals. For bandpass signal, no-ISI requires

$$W = (1+\alpha)R_s.$$

Why are the above statements correct?

Pulse Shaping/Filtering without ISI

- In practical systems, *raised-cosine filtering* is applied, introducing an **excess bandwidth** controlled by the roll-off factor α. The total occupied bandwidth consists of:

$$\begin{aligned} W &= W_{\text{Nyquist}} + W_{\text{excess}} \\ &= \frac{R_s}{2} + \alpha \cdot \frac{R_s}{2} = \frac{1}{2}(1+\alpha)R_s \end{aligned}$$

- A bandpass system requires twice the bandwidth of a baseband system. Thus, for bandpass modulation schemes, the required bandwidth is given by:

$$W = (1+\alpha)R_s$$

5.1 Bandwidth Efficiency

- Bandwidth efficiency (R_b/W): Bits/s/Hz.
- Theoretical maximum symbol-rate packing: 2 symbols/s/Hz.

With $W = (1+\alpha)R_s$, for a banspass M-ary modulation schelem like QPSK, QAM, we can rewrite (1) as

$$\frac{E_b}{N_0} = \frac{S}{N} \cdot \frac{W}{R_b} = \frac{S}{N} \cdot \frac{1+\alpha}{\log_2 M} \tag{5}$$

- $\frac{R_b}{W} = \frac{\log_2 M}{1+\alpha}$ is called bandwidth efficiency, and is measured in bits per second per Hz (bits/s/Hz).
- The above $\frac{R_b}{W}$ is for M-QAM signaling. This value is different for other signalings like PAM, and we will study this in chapter 4.

Example 4. ***16-level PAM System without/with ISI.***

1. *For a 10 Mbps data using a baseband 16-level PAM without ISI, determine symbol rate and the theoretical minimum system bandwidth needed.*
2. *If ISI is present and the system bandwidth is fixed as previously described, what bit rate is required for a raised cosine filter with a 25% roll-off factor to avoid ISI?*

Solution:

1. We know that the symbol rate is

$$R_s = \frac{R_b}{\log_2 L}$$

where R_b is the bit rate of data and L is the number of levels in the PAM signal. For specified 16-level PAM and given $R_b = 10$ Mbps, we have

$$R_s = \frac{10 \text{ Mbps}}{\log_2(16)} = 2.5 \text{ Msymbols/s}$$

Theoretical minimum system bandwidth is given as

$$\text{BW} = \frac{R_s}{2} = \frac{2.5 \text{ Msymbols/s}}{2} = 1.25 \text{ MHz}$$

2. With ISI, $W = \frac{1}{2}(1+\alpha)R_s$. But, we have $W = 1.25$ MHz and $\alpha = 0.25$. Thus,

$$R_s = \frac{2W}{1+\alpha} = 2 \text{ Msymbols/s}$$

Thus, the bit rate (R_b) should be 8 Mbps at most.

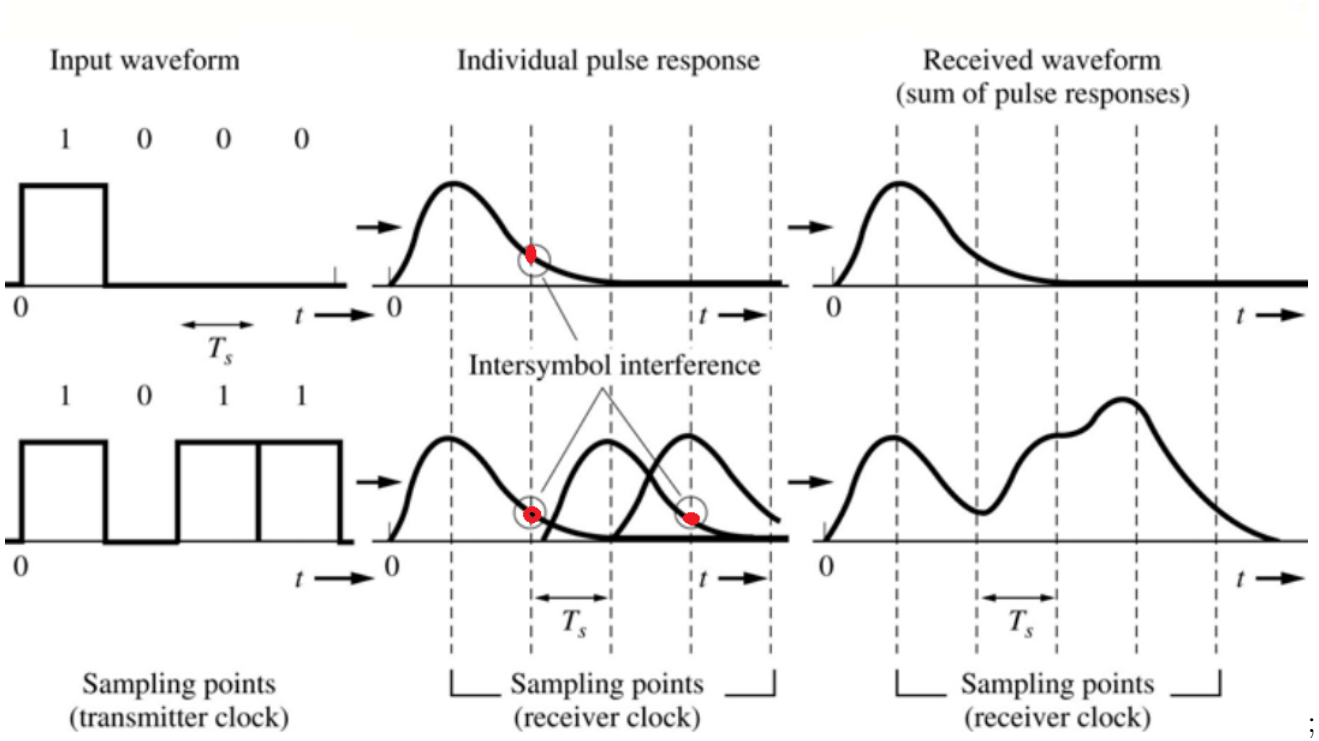
5.2 ISI and Pulse Shaping

- Pulse shaping is not performed to eliminate ISI that may already exist in a waveform; it does not act as an equalizer.
- The primary purpose of pulse shaping is to restrict the occupied bandwidth.
- Pulse shapes that satisfy the Nyquist ISI criterion can achieve the desired bandwidth restriction while ensuring that no additional ISI is introduced.

What is Inter-Symbol Interference (ISI)?

- ISI is a form of signal distortion in digital communication systems where multiple transmitted symbols overlap in time, causing the receiver to misinterpret them.
- This occurs when the transmission pulse for one symbol spreads into adjacent time slots.
- ISI occurs due to:
  - **Limited Bandwidth**: When a signal is band-limited, its time-domain representation (e.g., sinc pulses) extends indefinitely, causing overlap.

Pulse Shaping/Filtering without ISI



– **Multipath Propagation**: In wireless channels, signals take multiple paths due to reflection, refraction, or scattering, leading to delayed versions of the same signal arriving at different times.

– **Filtering Effects**: Imperfect filtering in the transmitter or receiver can distort the shape of pulses.

How Can We Remedy ISI?

Several techniques can mitigate or eliminate ISI in digital communication systems:

1. Zero ISI Pulse Shaping to Limit the Bandwidth
2. Equalization **(Chapter 9)**
   - **Linear Equalization**: Uses a filter to reverse the channel's distortion effects.
   - **Decision Feedback Equalization**: Uses past detected symbols to cancel out ISI.
   - **Adaptive Equalizers**: Algorithms such as the *least mean squares* and *recursive least squares* adaptively adjust filter coefficients to mitigate ISI.
3. Orthogonal Frequency Division Multiplexing (OFDM)

   OFDM converts the signal into multiple orthogonal subcarriers, reducing ISI since each subcarrier experiences less time dispersion. It introduces a **cyclic prefix** to absorb ISI due to multipath effects.
4. Equalization at the Receiver

- **Maximum likelihood sequence estimation (MLSE)**: Uses a Viterbi algorithm to detect the most probable transmitted symbol sequence.
- **Zero forcing (ZF) and minimum mean square error (MMSE) equalizers**: Remove ISI by inverting the channel response while minimizing noise amplification.

The best approach depends on the communication system, bandwidth constraints, and complexity trade-offs.