Base-Band Digital Data Transmission

- Components of Data Communication
- The Need for Digital Transmission
- Basic concepts of Line Coding
- The important characteristics of line coding
- Various line coding techniques
    - Unipolar
    - Polar
    - Bipolar

Components of Data Communication

- **Data**
  - Analog: Continuous value data (sound, light, temperature)
  - Digital: Discrete value (text, integers, symbols)
- **Signal**
  - Analog: Continuously varying electromagnetic wave
  - Digital: Series of voltage pulses (square wave)

Analog Data-->Signal Options

- **Analog data to analog signal**
  - Inexpensive, easy conversion (eg telephone)
  - Used in traditional analog telephony
- **Analog data to digital signal**
  - Requires a codec (encoder/decoder)
  - Allows use of digital telephony, voice Mail

Digital Data-->Signal Options

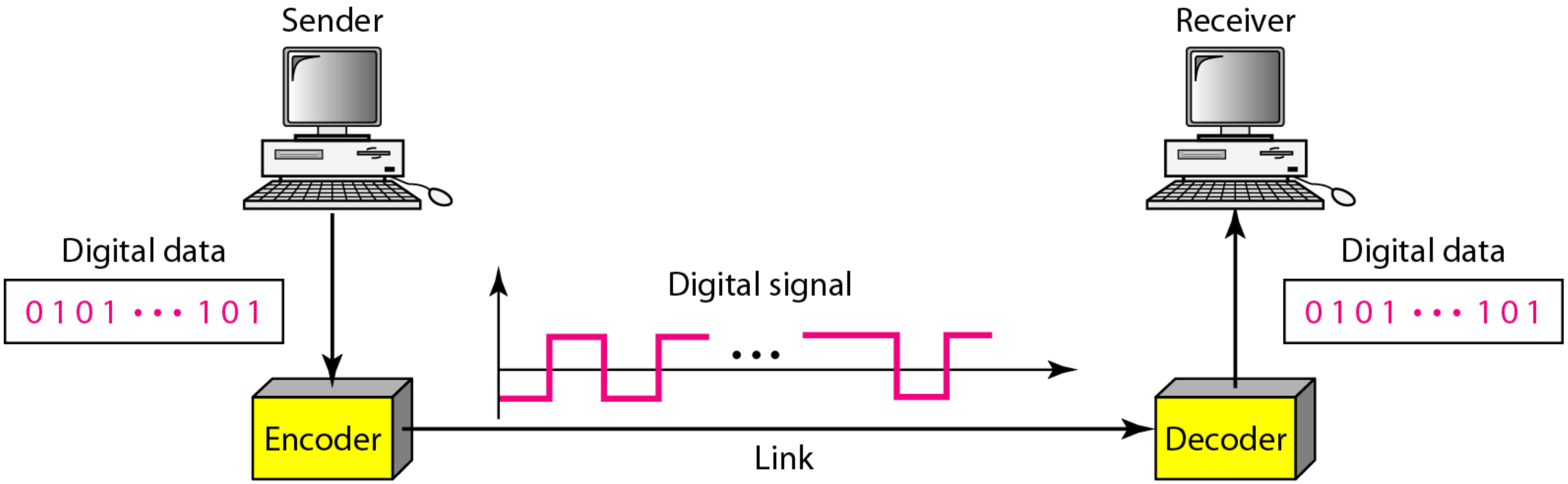
- **Digital data to analog signal**
  - Requires modem (modulator/demodulator)
  - Necessary when analog transmission is used

- **Digital data to digital signal**
  - Less expensive when large amounts of data are involved
  - More reliable because no conversion is involved

Digital to Digital Conversion

In this section, we see how we can represent digital data by using digital signals. The conversion involves three techniques: line coding, block coding, and scrambling.
Line coding is always needed; block coding and scrambling may or may not be needed.

Line Coding and Decoding



Introduction

- Digital vs. analog signals
- Analog → sampling → quantization → digital
- Baseband vs. passband
- Channel distortion – ISI

  (Channel) bandwidth limitation

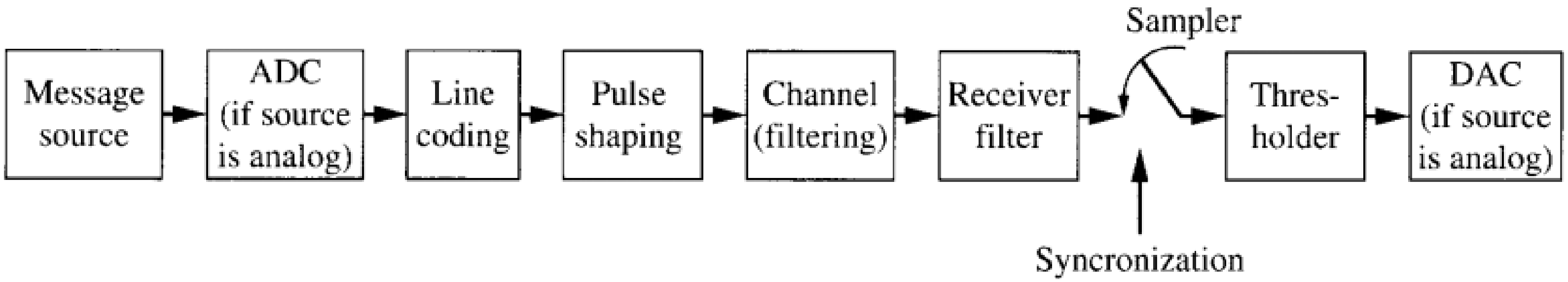


Figure 4.1
Block diagram of a baseband digital data transmission system.

Line Codes

- Baseband *data format* used to represent digital data (for transmission purpose).
- Examples are given on the next page
- <u>Operation</u>: time or frequency *shaping*
- <u>Purposes</u>: usually to cope with the channel limitations (or provide extra function such as synchronization)
- Needed for certain applications

Purposes of Line Codes

- **Self synchronization**
- **Proper power spectrum**
- **Transmission bandwidth**
- Transparency
- Error detection capability
- **Good error probability** performance

- Non-return-to-zero (NRZ) change
- NRZ mark (data1-> change in level; data0 -> no change)
- Unipolar return-to-zero (URZ) <1/2 width pulses>
- Polar RZ
- Bipolar RZ ("0"-> 0 level; "1"-> alternate sign
- Split phase (Manchester) ("1" -> level A to level –A at 1/2 interval; "0" -> level –A to level A at 1/2 interval)

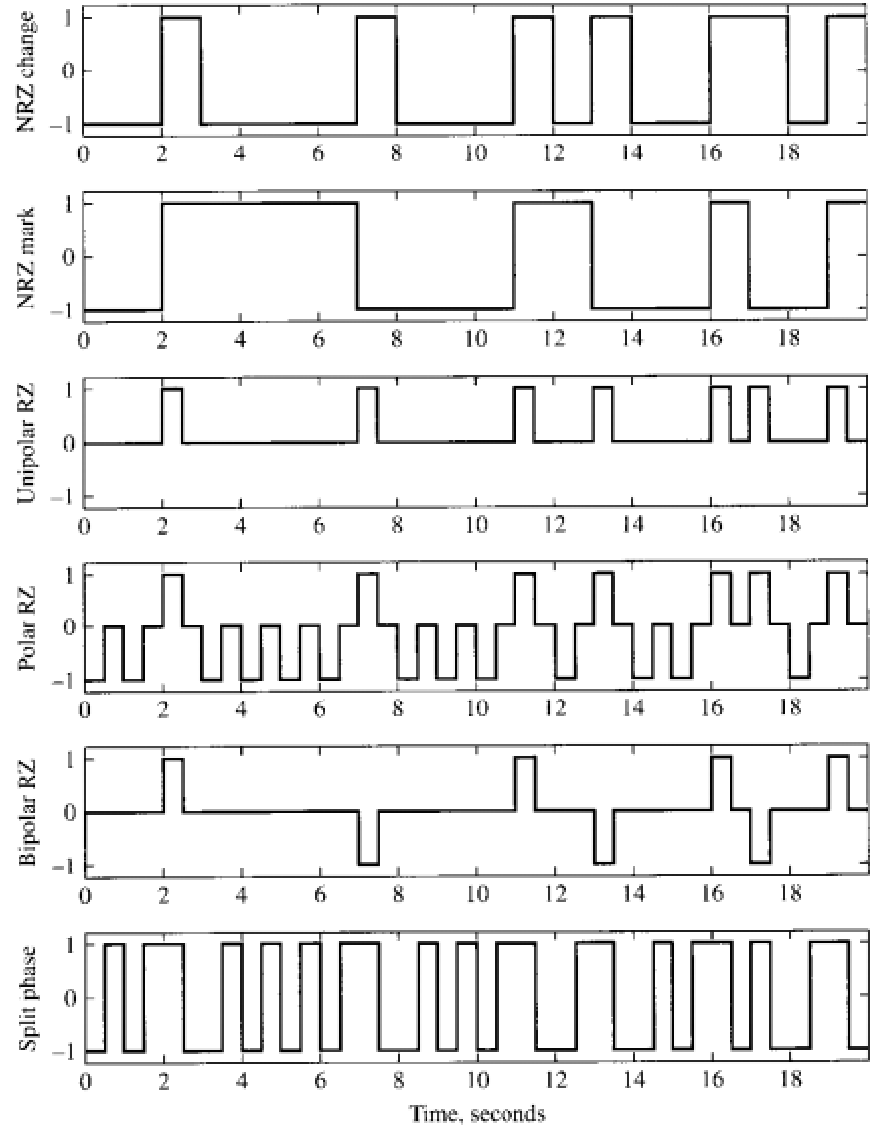
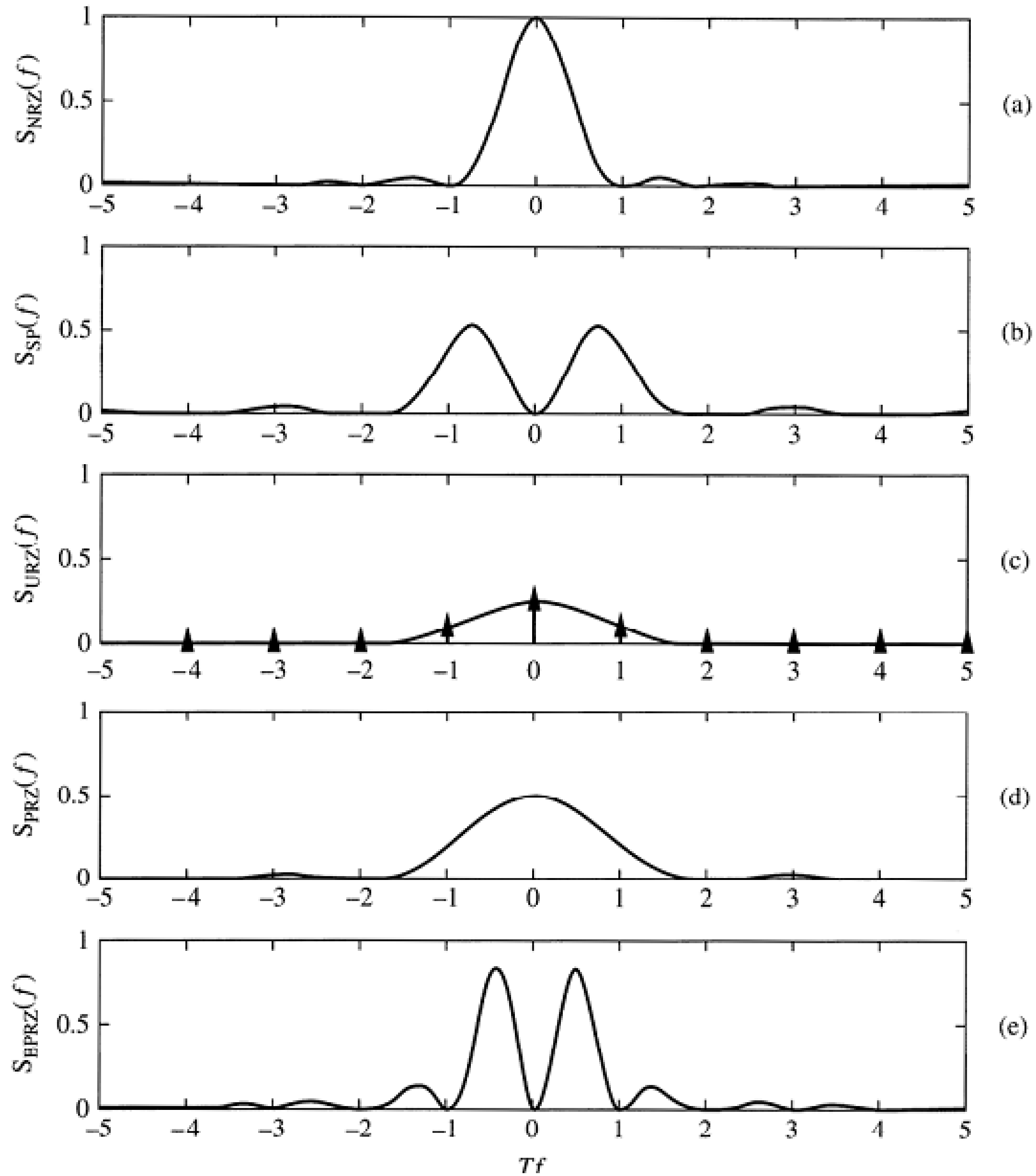


Figure 4.2
Abbreviated list of binary data formats.

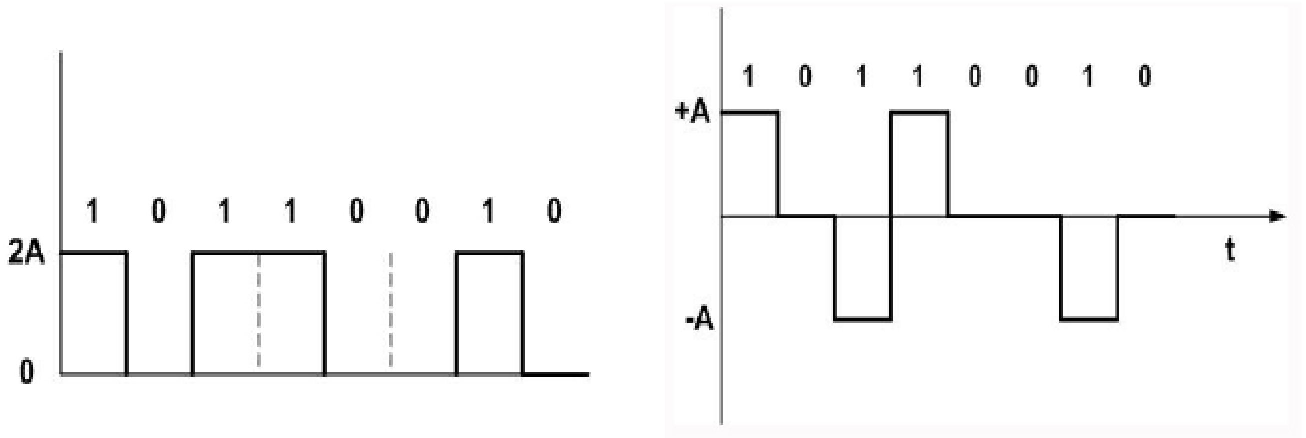


$S_{NRZ}(f)$
(a)
$S_{SP}(f)$
(b)
$S_{URZ}(f)$
(c)
$S_{PRZ}(f)$
(d)
$S_{EPRZ}(f)$
(e)
Tf

Important Characteristic of Line Coding

- No. Of Levels

This refers to the number values allowed in a signal, known as s**ignal levels**, to represent data.



(a) (b)

(a) Signal with two voltage levels, (b) Signal with three voltage levels

Important Characteristic of Line Coding

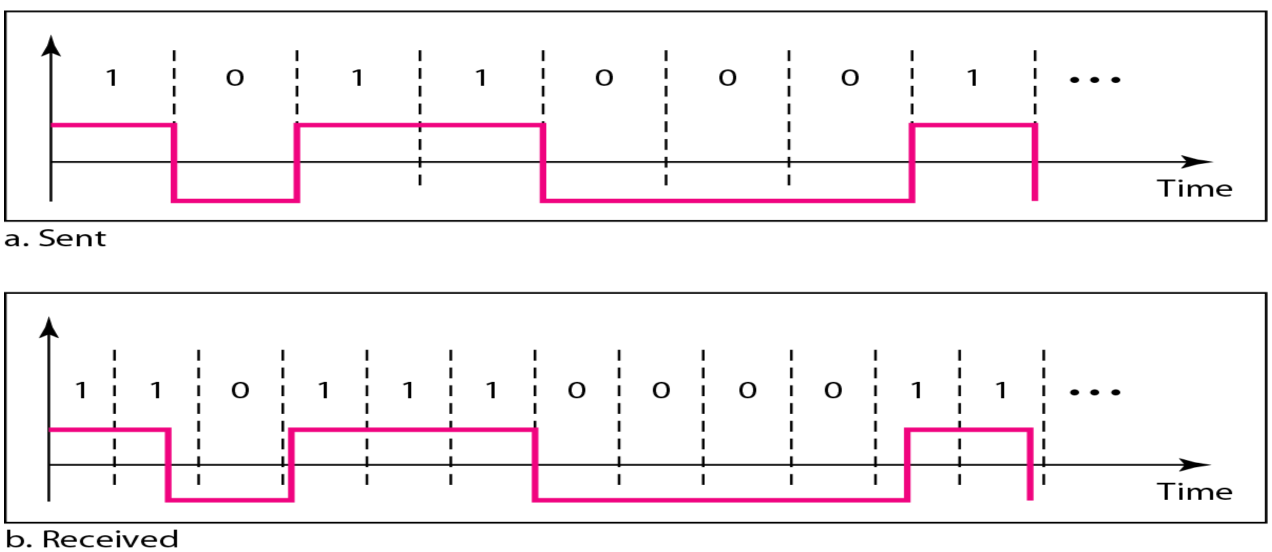
- Bit rate Versus Baud rate
  - The **bit rate** represents the number of bits sent per second,
  - whereas the **baud rate** defines the number of signal elements per second in the signal.
  - Depending on the encoding technique used, baud rate may be more than or less than the data rate.
- DC components
  - After line coding, the signal may have zero frequency component in the spectrum of the signal, which is known as the direct-current **(DC) component**.
  - Not Desirable. e.g. Transformer, Capacitive Coupling
  - Unwanted Energy Loss

Important Characteristic of Line Coding

- Signal Spectrum
  - Different encoding of data leads to different spectrum of the signal.
  - It is necessary to use suitable encoding technique to match with the medium so that the signal suffers minimum attenuation and distortion as it is transmitted through a medium.

Important Characteristic of Line Coding

- Synchronization
  - To interpret the received signal correctly, the bit interval of the receiver should be exactly same or within certain limit of that of the transmitter.
  - Any mismatch between the two may lead wrong interpretation of the received signal.



Effect of lack of synchronization

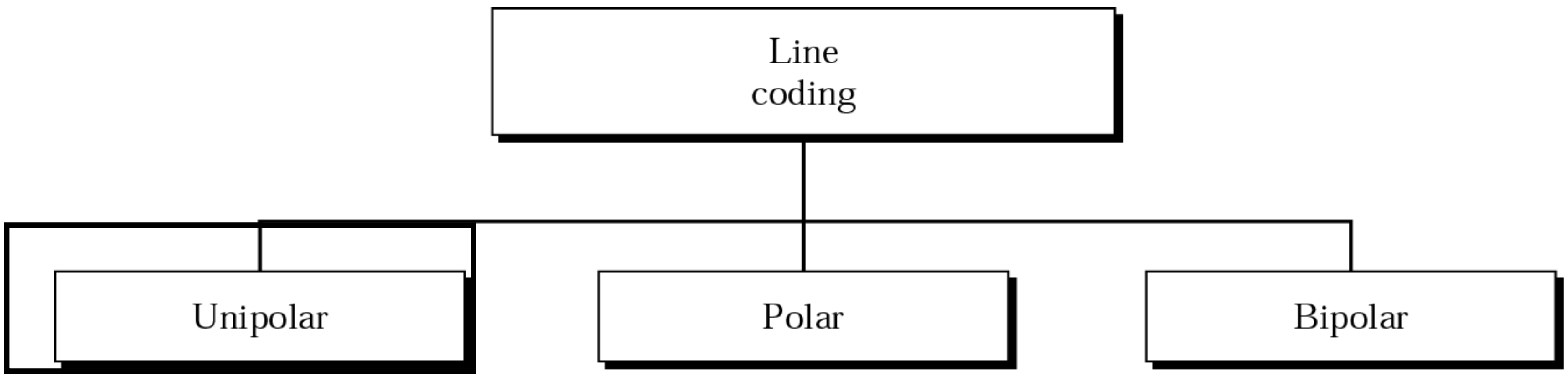
Self-Synchronization

- To correctly interpret the signals received from the sender, the receiver’s bit intervals must correspond exactly to the sender’s bit interval.
- A self-synchronizing digital signal includes timing information in the data being transmitted.
- This can be achieved if there are transitions in the signal that alert the receiver to the beginning, middle, or end of the pulse.
- If the receiver’s clock is out of synchronization, these alerting points can reset the clock.

Important Characteristic of Line Coding

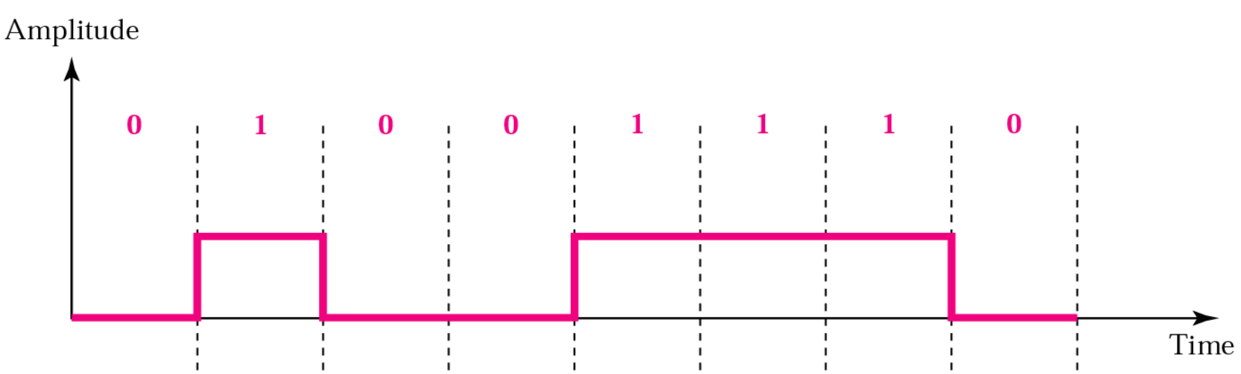
- Cost of Implementation
  - It is desirable to keep the encoding technique simple enough such that it does not incur high cost of implementation.

Line Coding Techniques



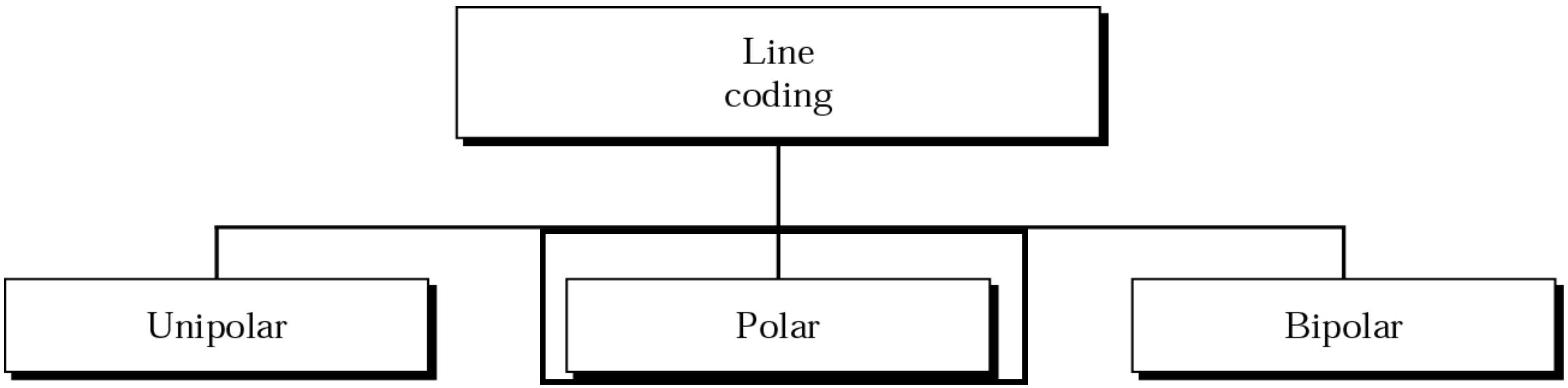
Unipolar Encoding

- Digital transmission systems work by sending voltage pulses along a cable.
- In unipolar encoding, the polarity is assigned to ‘1’ bit, while the ‘0’ bit is represented by zero voltage.
- Unipolar encoding has two problems: dc component and lack of synchronization
- The average amplitude of a unipolar encoded signal is nonzero, which creates a dc component.
- If data contains a long sequence of 1s or 0s, synchronization is lost.



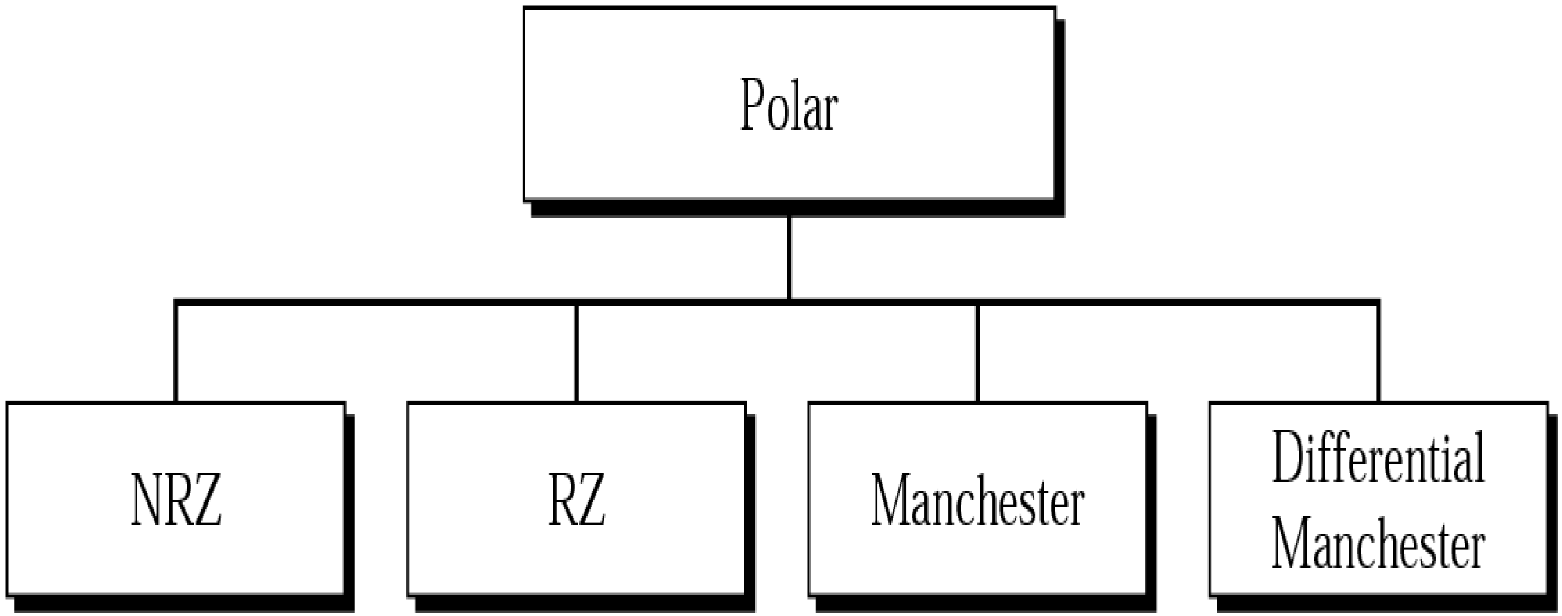
Unipolar Encoding

Line Coding Techniques



Polar Encoding

- Polar encoding uses two voltage levels (positive and negative).



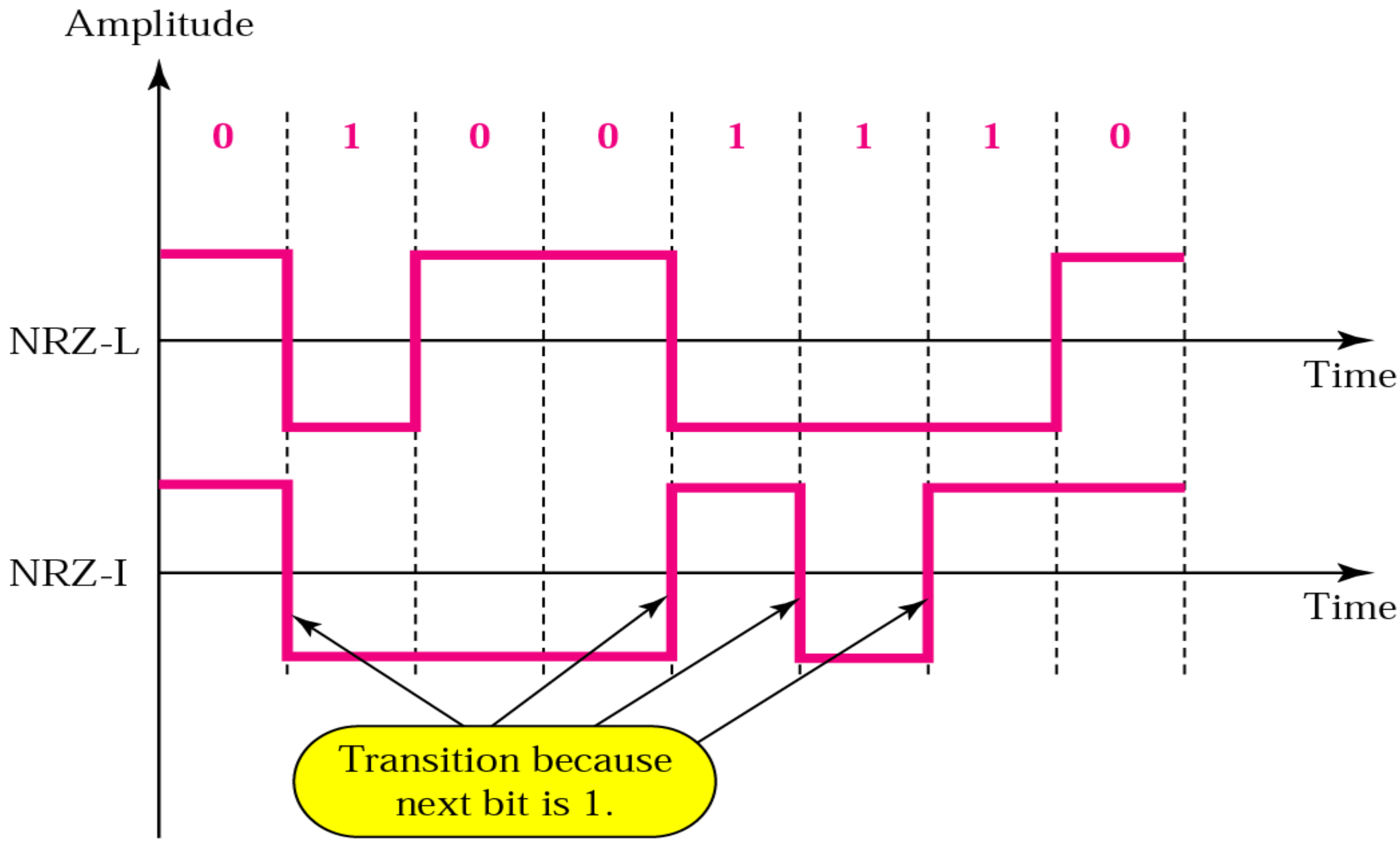
NRZ Encoding

- In Nonreturn to Zero (NRZ) encoding, the value of the signal is always either positive or negative.
- There are two popular forms of NRZ: NRZ-L and NRZ-I
- In NRZ-Level (NRZ-L), a positive voltage means ‘0’ bit, while a negative voltage means ‘1’ bit.
- The advantages of NRZ coding are:
  - Detecting a transition in presence of noise is more reliable than to compare a value to a threshold.
  - NRZ codes are easy to engineer and it makes efficient use of bandwidth.
- Disadvantage
  - A problem can arise when the data contains a long stream of 0s or 1s that can cause synchronization lost.

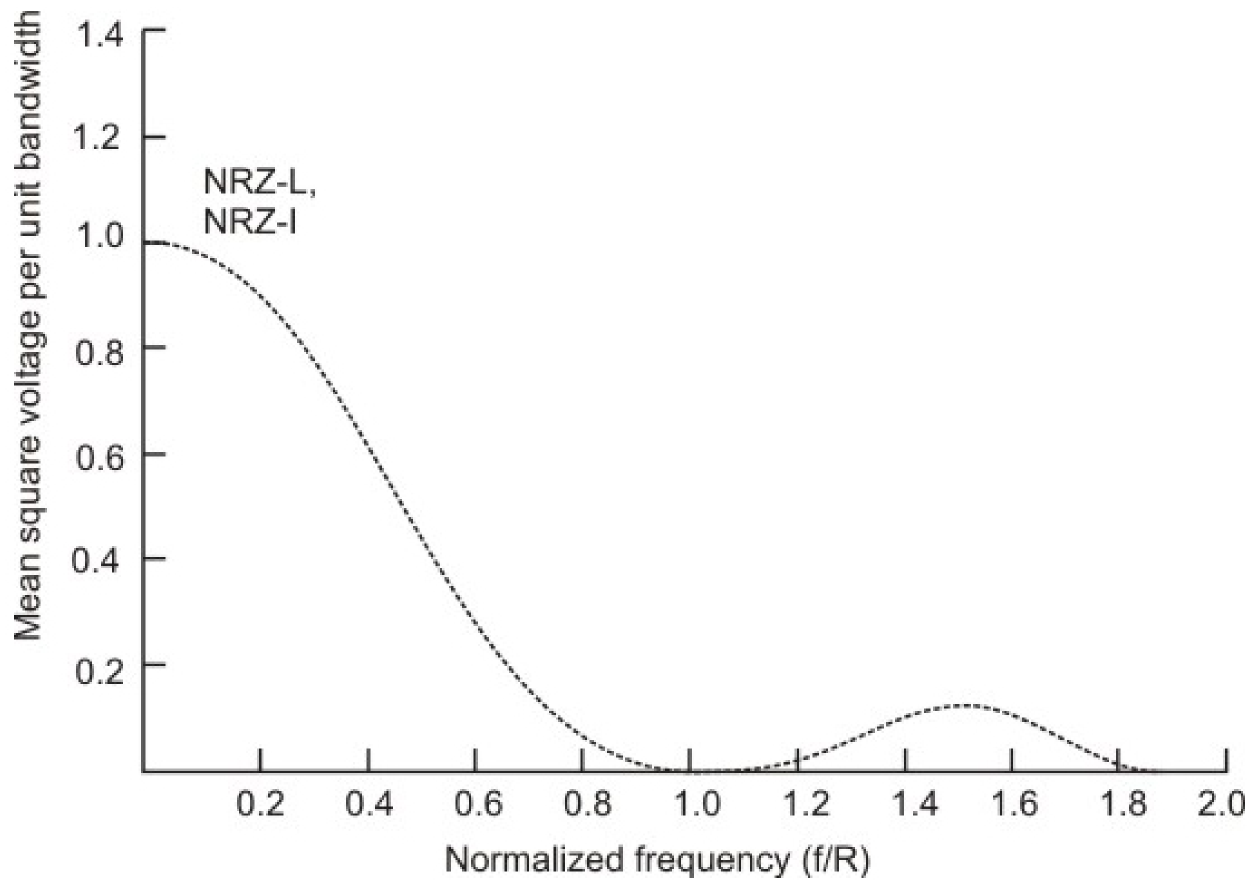
NRZ Encoding

- In NRZ-Invert (NRZ-I), it is the transition between positive and negative voltage, not the voltage itself, that represents a ‘1’ bit. A ‘0’ bit is represented by no change.
- NRZ-I is better than NRZ-L due to the synchronization provided by the signal change each time a 1 bit is encountered.
- Both NRZ-L and NRZ-I have synchronization problem.

NRZ-L and NRZ-I encoding



Spectrum of NRZ-L and NRZ-I



Return to Zero (RZ) Encoding

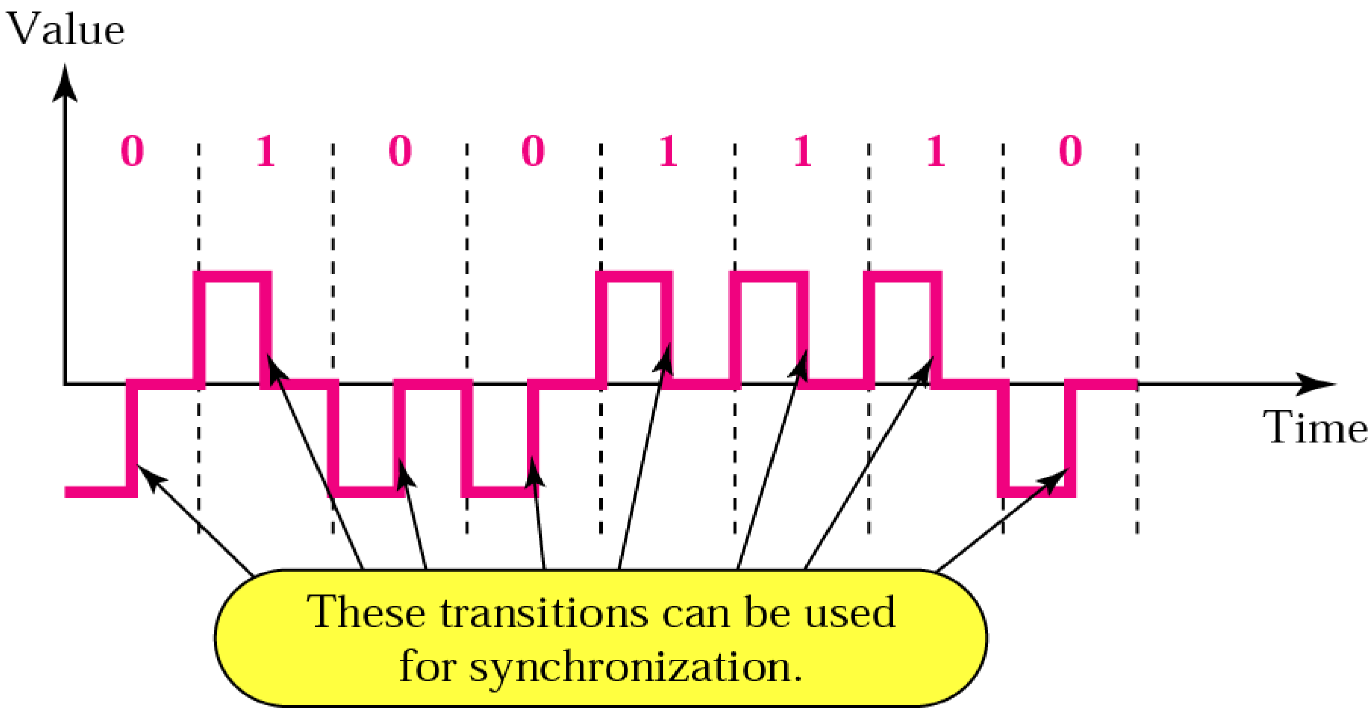
- To ensure synchronization, there must be a signal change for each bit. The receiver can use these changes to build up, update and synchronize its clock.

- Return to Zero (RZ) uses three values: positive, negative and zero. In RZ, the signal changes not between bits but during (middle of) each bit.

- A 1 bit is represented by positive-to-zero and a 0 bit by negative-to-zero.

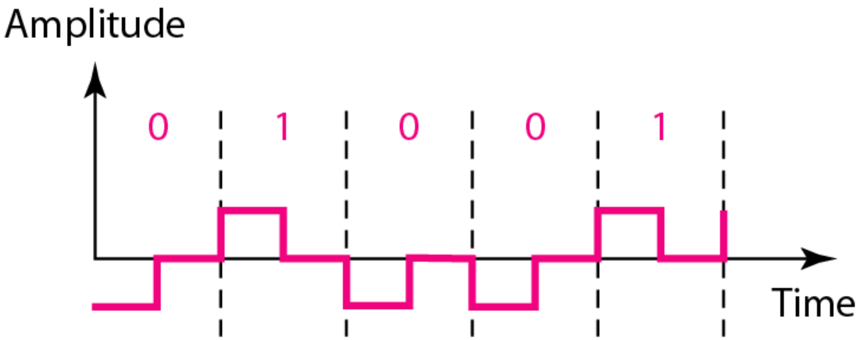
- The main disadvantage of RZ is that it requires two signal changes to encode one bit and therefore occupies more bandwidth.

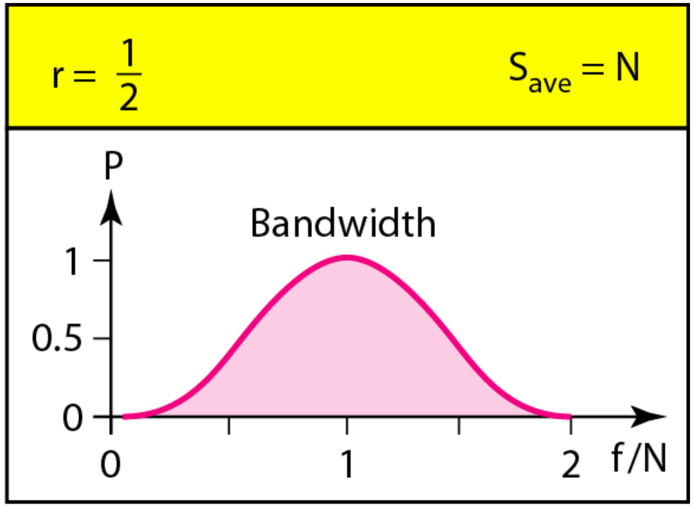
RZ Encoding



Value
0
1
0
0
1
1
1
0
Time
These transitions can be used for synchronization.

Spectrum of RZ





Biphase Coding

- To overcome the limitations of NRZ encoding, biphase encoding techniques can be adopted.
- *Manchester* and *differential Manchester Coding* are the two common Biphase techniques
- Bit changes at the Middle (mid bit Transition) serves as a clocking mechanism and also as data.

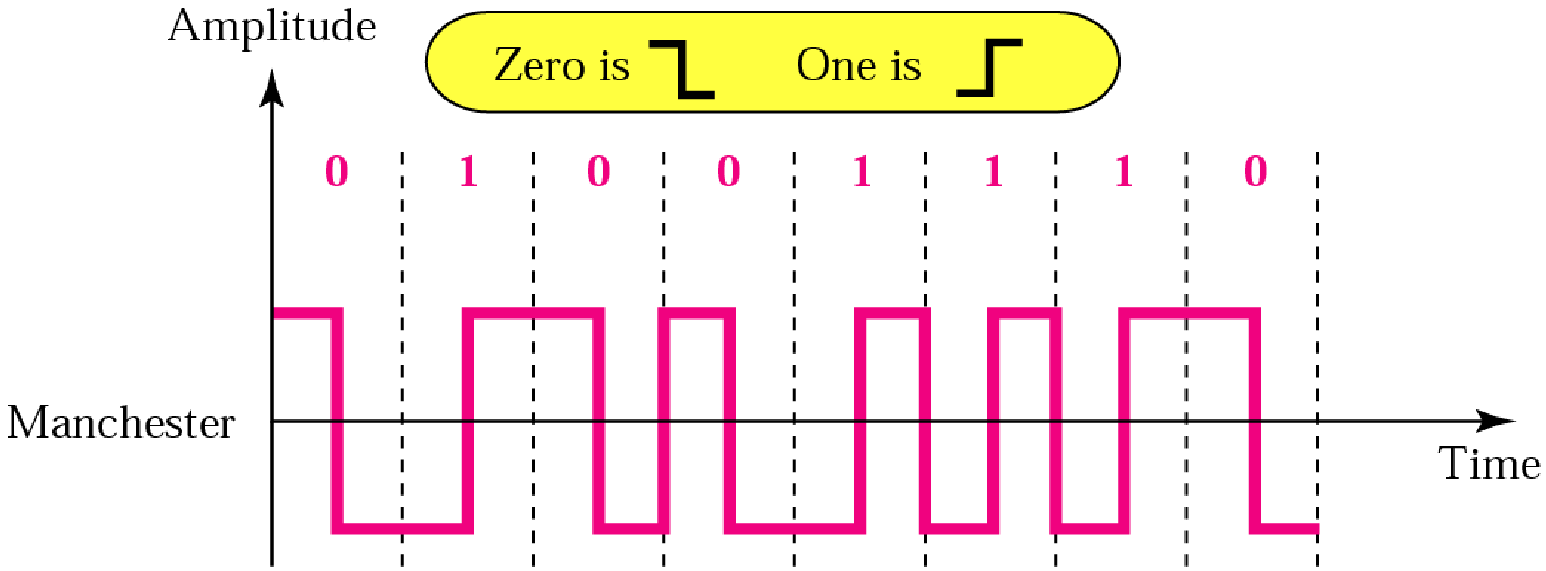
Manchester Encoding

- Manchester encoding uses an inversion at the middle of each bit interval for both synchronization and bit representation.

- A negative-to-positive transition represents 1 and a positive-to-negative transition represents 0.

- Manchester encoding achieves the same level of synchronization as RZ but with only two levels of amplitude.

Manchester Encoding



Note:

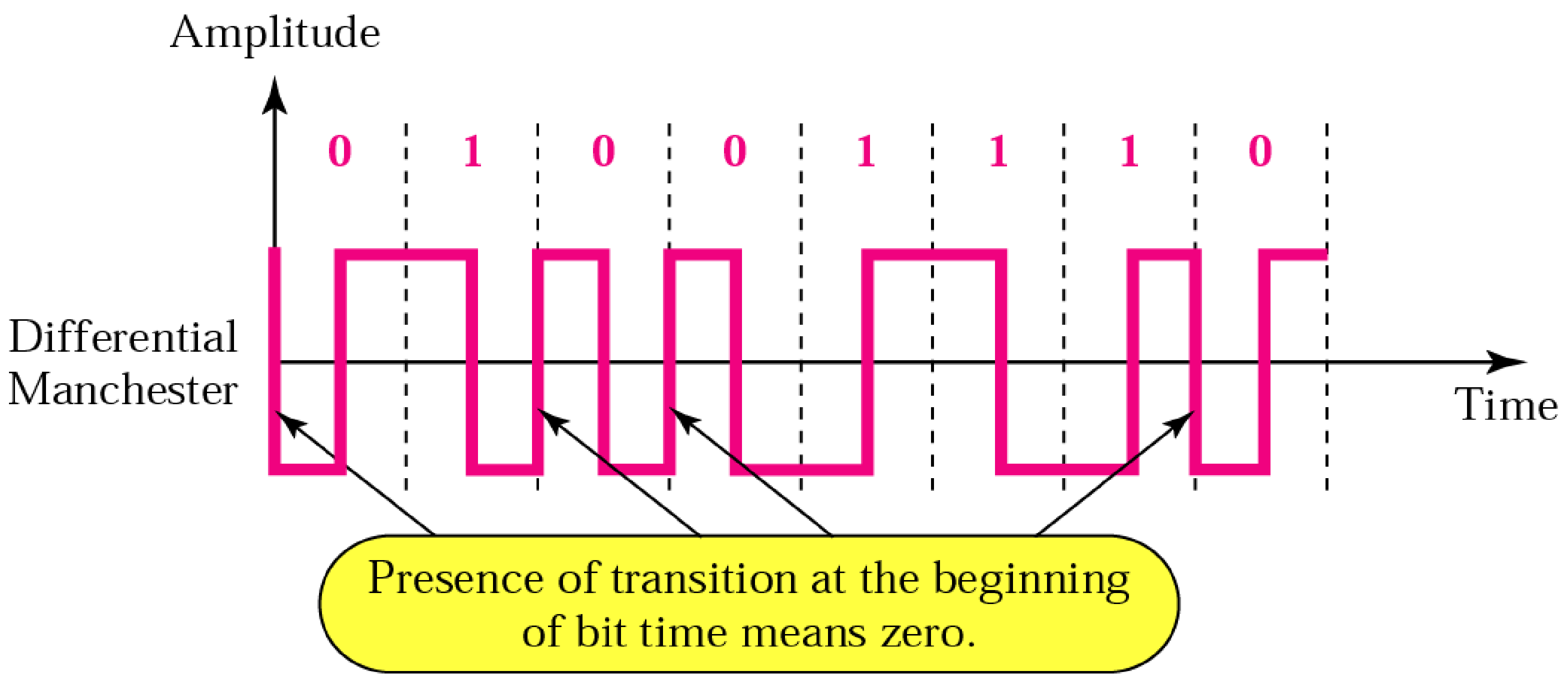
In Manchester encoding, the transition at the middle of the bit is used for both synchronization and bit representation.

Differential Manchester Encoding

- In differential Manchester encoding, the inversion at the middle of the bit interval is used for synchronization, but the presence (0) or absence (1) of an additional transition at the beginning of the interval is used to identify the bit.

- Differential Manchester encoding requires two signal changes to represent binary 0 but only one to represent binary 1.

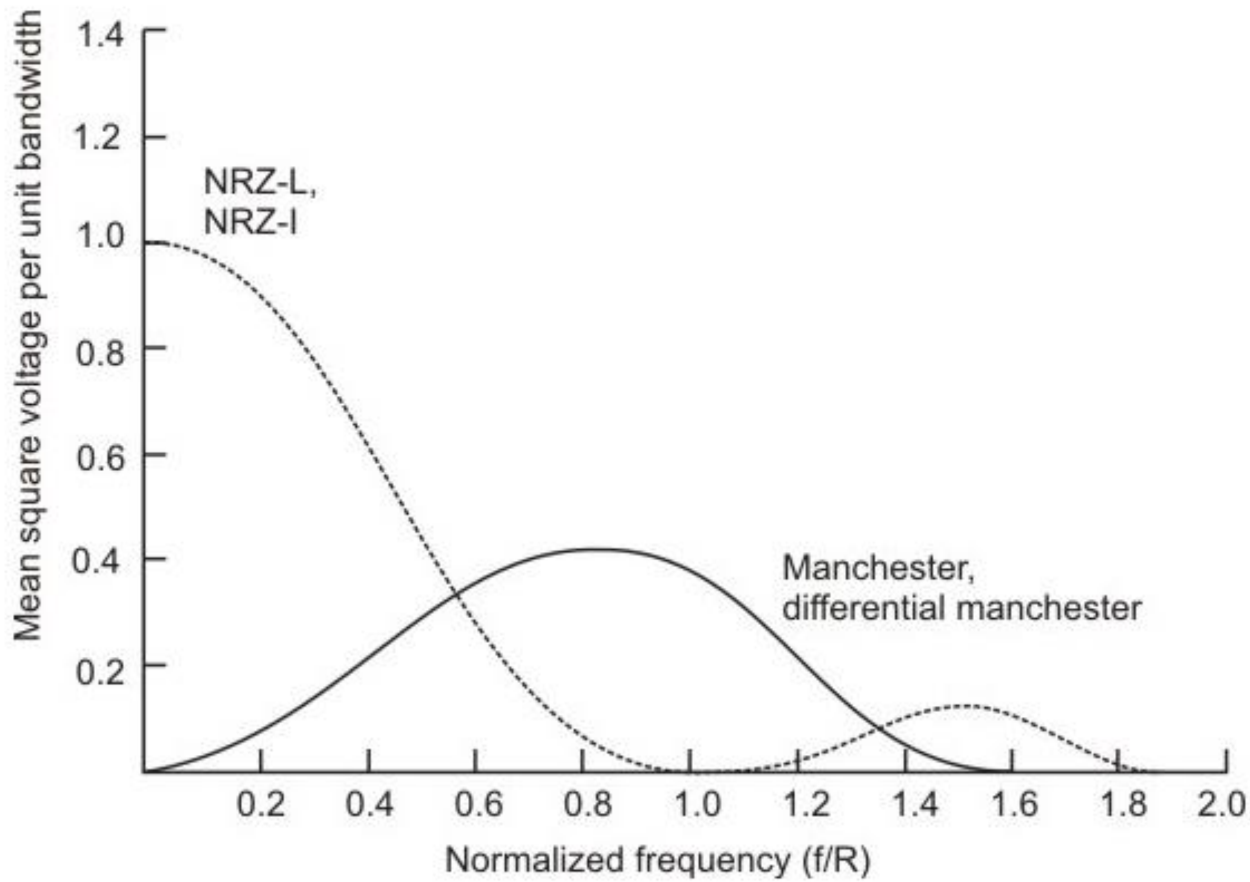
Differential Manchester Encoding



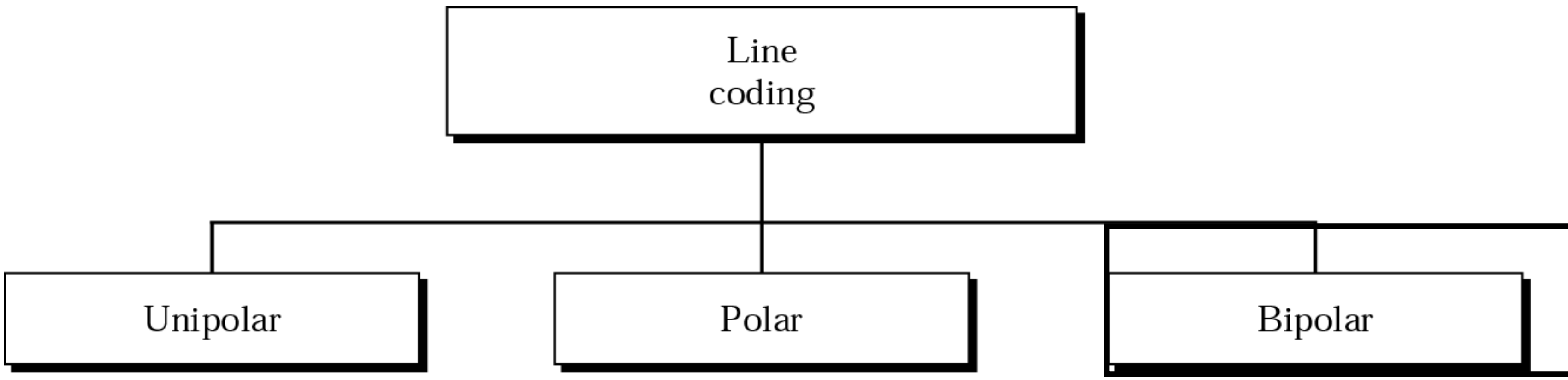
Note:

In differential Manchester encoding, the transition at the middle of the bit is used only for synchronization.
The bit representation is defined by the inversion or noninversion at the beginning of the bit.

Spectrum of Manchester Coding



Line Coding Techniques



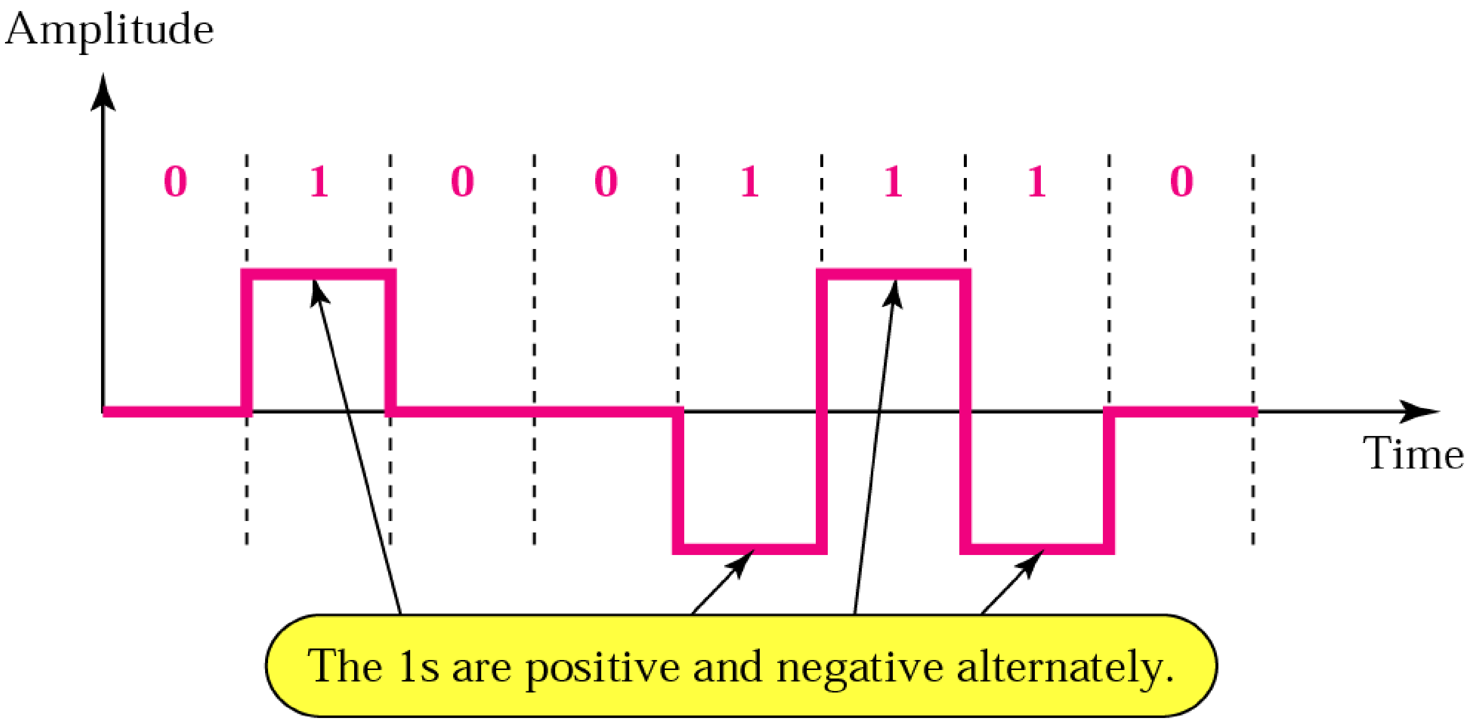
Bipolar Encoding

- In bipolar encoding, we use three levels: positive, zero, and negative.

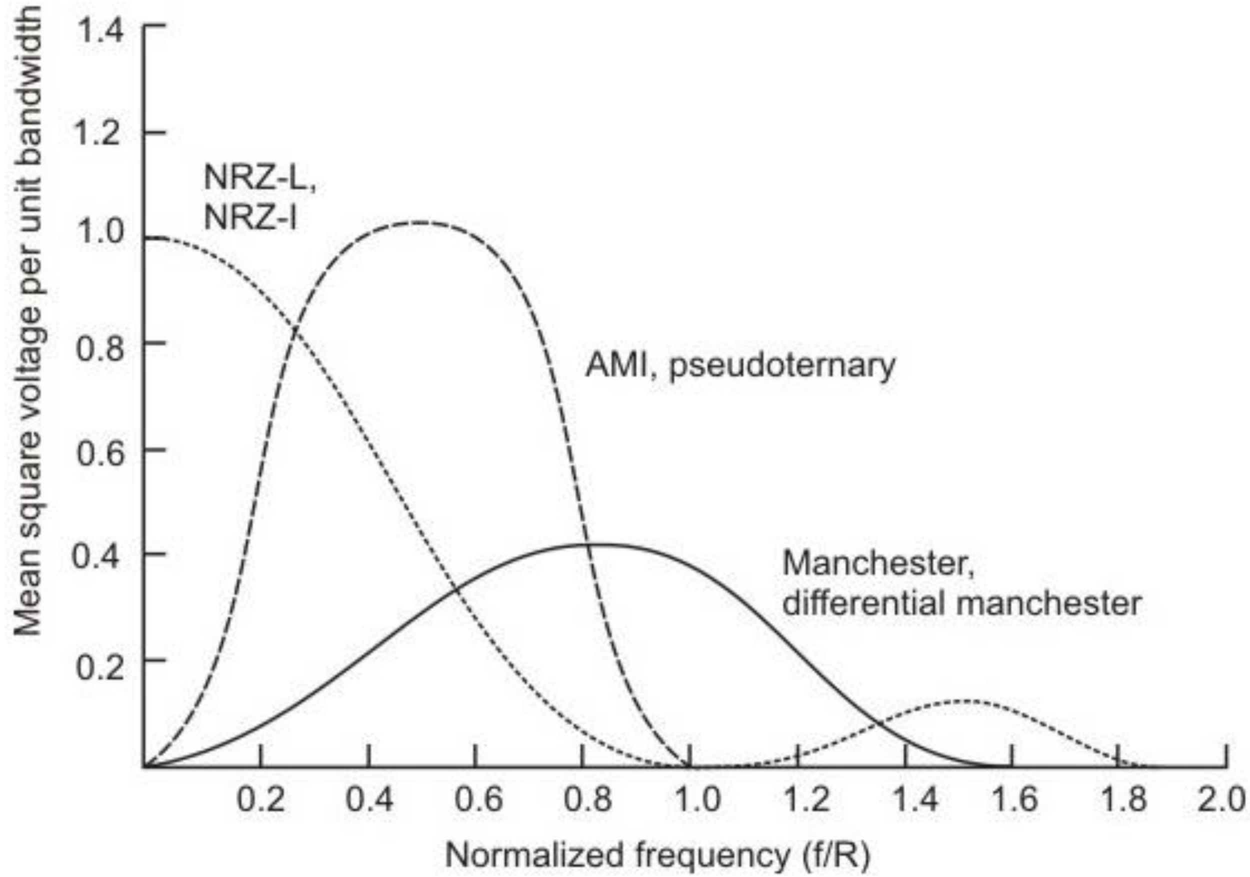
- Unlike RZ, the zero level in bipolar encoding is used to represent binary 0. The 1s are represented by alternating positive and negative voltages (even if the 1 bits are not consecutive).

- A common bipolar encoding scheme is called Alternate Mark Inversion (AMI). The word mark comes from telegraphy and it means 1.

Bipolar AMI Encoding



Frequency Spectrum of Various Modulation Scheme



Pseudometry

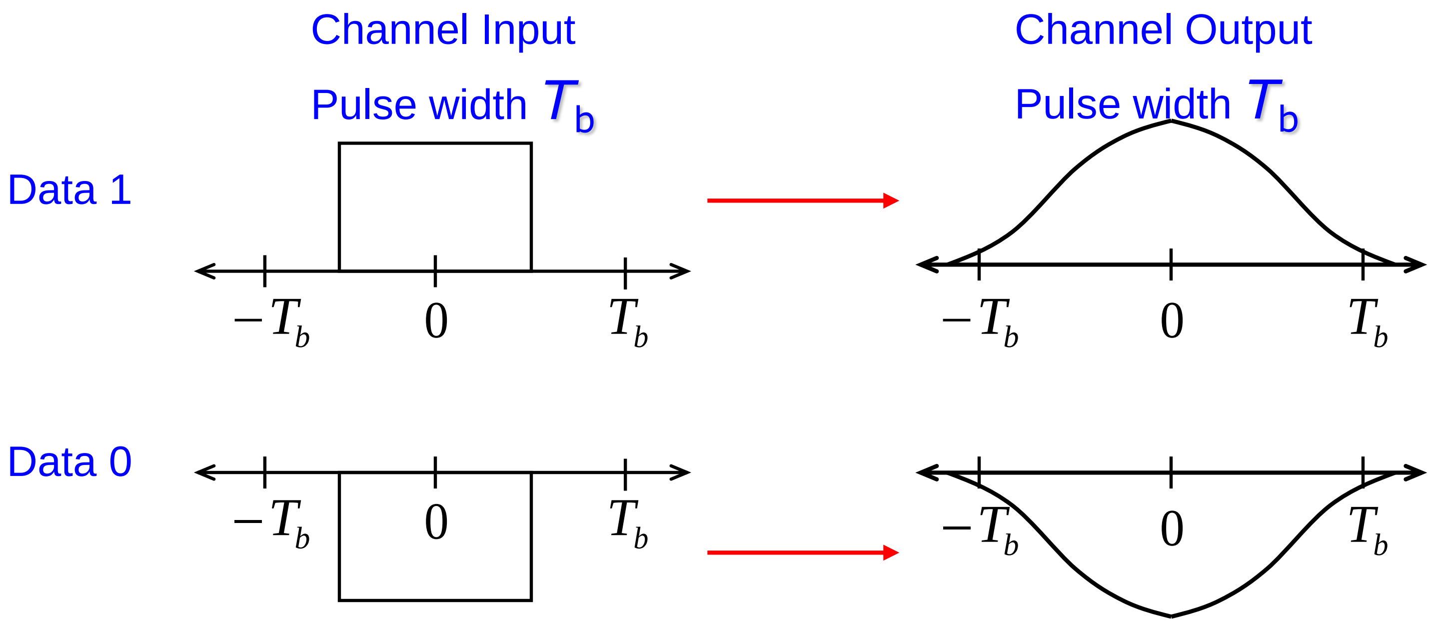
- Same as AMI.
- But alternating positive and negative pulse occur for binary 0 instead of binary 1.
- Three levels
- No DC comp
- Loss of synchronization for long seq. of 0.
- Lesser bandwidth

INTERSYMBOL INTERFERENCE (ISI)

- Intersymbol Interference
- ISI on Eye Patterns
- Combatting ISI
- Nyquist's First Method for zero ISI
- Raised Cosine-Rolloff Pulse Shape

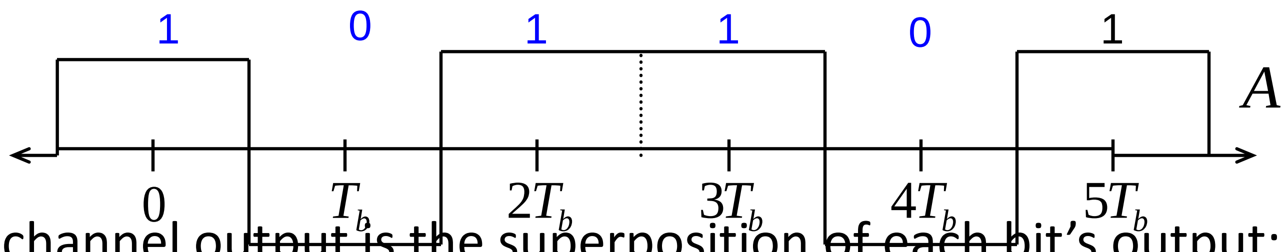
Intersymbol Interference

- **Intersymbol interference (ISI)** occurs when a pulse spreads out in such a way that it interferes with adjacent pulses *at the sample instant.*
- Example: assume polar NRZ line code. The channel outputs are shown as spreaded (width T_b becomes $2T_b$) pulses shown (Spreading due to bandlimited channel characteristics).

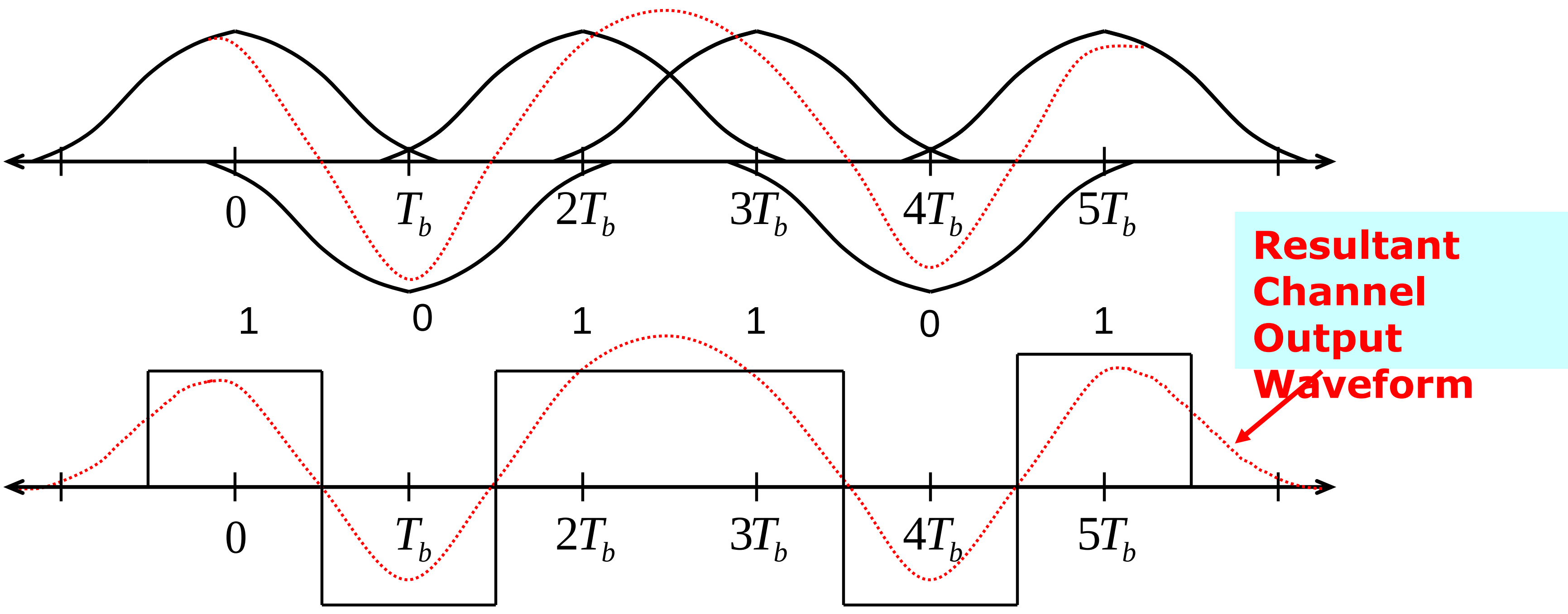


Intersymbol Interference

- For the input data stream:

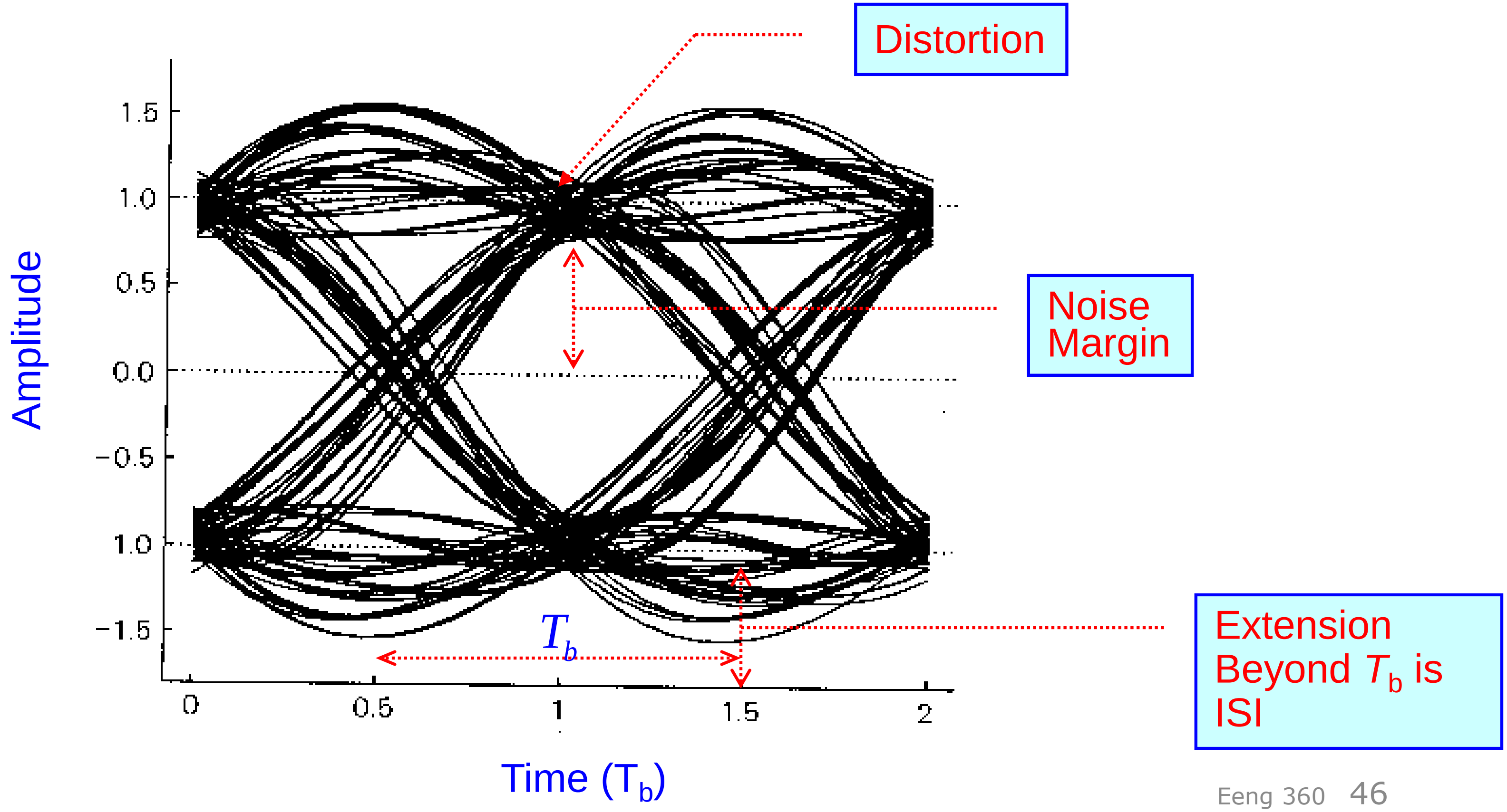


- The channel output is the superposition of each bit's output:



ISI on Eye Patterns

- The amount of ISI can be seen on an oscilloscope using an ***Eye Diagram*** or ***Eye pattern***.



Intersymbol Interference

➢ If the rectangular multilevel pulses are filtered improperly as they pass through a communications system, they will spread in time, and the pulse for each symbol may be smeared into adjacent time slots and cause ***Intersymbol Interference.***

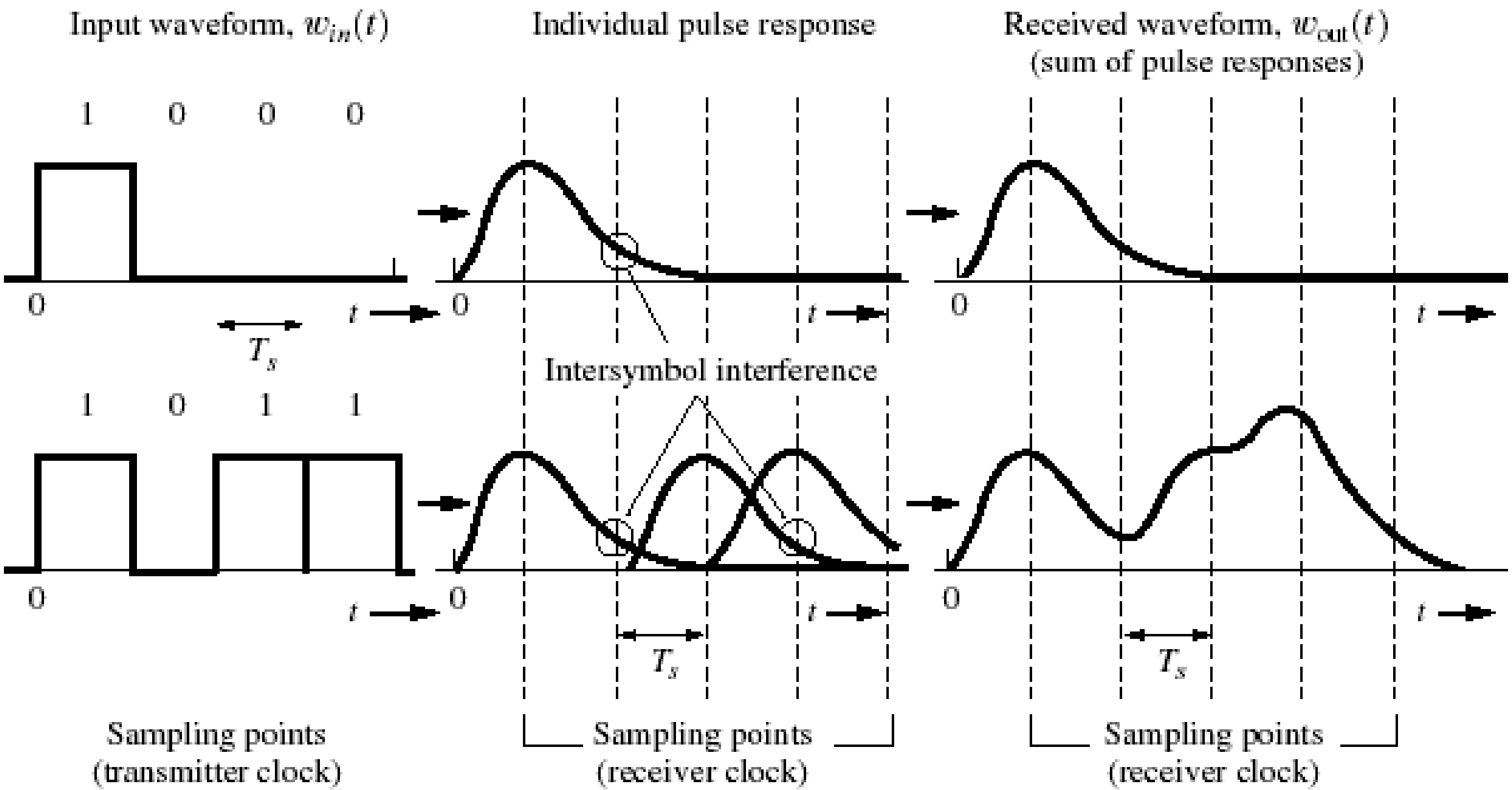


Figure 3–23 Examples of ISI on received pulses in a binary communication system.

➢ **How can we restrict BW and at the same time not introduce ISI?.**

Intersymbol Interference

- Flat-topped multilevel input signal having pulse shape $h(t)$ and values a_k:

$$w_{\text{in}}(t) = \sum a_n h(t - nT_s) = \sum_n a_n h(t) * \delta(t - nT_s) = \left[\sum_n a_n \delta(t - nT_s)\right] * h(t)$$

$$\text{Where } h(t) = \prod\left(\frac{t}{T_s}\right) \quad \text{Where} \quad D = \frac{1}{T_s} \text{ pulses/s}$$

$$w_{\text{out}}(t) = \left[\sum_n a_n \delta(t - nT_s)\right] * h_e(t) = \sum_n a_n h_e(t - nT_s)$$

$$\boxed{\text{Equivalent impulse response: } h_e(t) = h(t) * h_T(t) * h_C(t) * h_R(t)}$$

- $h_e(t)$ is the pulse shape that will appear at the output of the receiver filter.

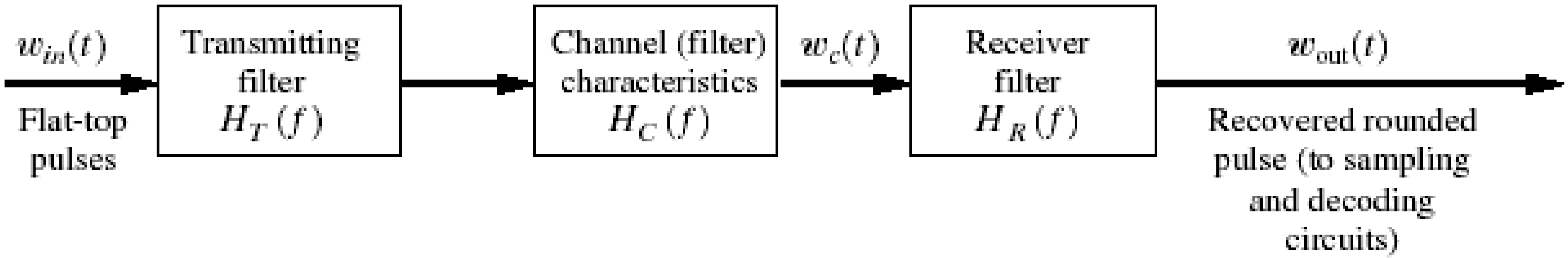


Figure 3–24 Baseband pulse-transmission system.

Intersymbol Interference

- Equivalent Impulse Response $h_e(t)$:

$$h_e(t) = h(t) * h_T(t) * h_C(t) * h_R(t)$$

- Equivalent transfer function:

$$H_e(f) = H(f) H_T(f) H_C(f) H_R(f) \quad \text{Where} \quad H(f) = F\left[\Pi\left(\frac{t}{T_s}\right)\right] = T_s\left(\frac{\sin \pi T_s f}{\pi T_s f}\right)$$

- Receiving filter can be designed to produce a needed $H_e(f)$ in terms of $H_T(f)$ and $H_C(f)$:

$$H_R(f) = \frac{H_e(f)}{H(f) H_T(f) H_C(f)}$$

- Output signal can be rewritten as:

$$w_{out}(t) = \sum_n a_n h_e(t - nT_s)$$

- $H_e(f)$, chosen such to minimize ISI is called EQUALIZING FILTER)

Combating ISI

- Three strategies for eliminating ISI:
  - Use a line code that is absolutely bandlimited.
    - Would require Sinc pulse shape.
    - Can’t actually do this (but can approximate).
  - Use a line code that is zero during adjacent sample instants.
    - It’s okay for pulses to overlap somewhat, as long as there is no overlap at the sample instants.
    - Can come up with pulse shapes that don’t overlap during adjacent sample instants.
      - Raised-Cosine Rolloff pulse shaping
  - Use a filter at the receiver to “undo” the distortion introduced by the channel.
    - Equalizer.

Nyquist's First Method for Zero ISI

- ISI can be eliminated by using an equivalent transfer function, $H_e(f)$, such that the impulse response satisfies the condition:

$$h_e(kT_s+\tau)=\begin{cases} C, & k=0 \\ 0, & k\neq 0 \end{cases}$$

k is an integer, T_s is the symbol (sample) period

τ is the offset in the receiver sampling clock times

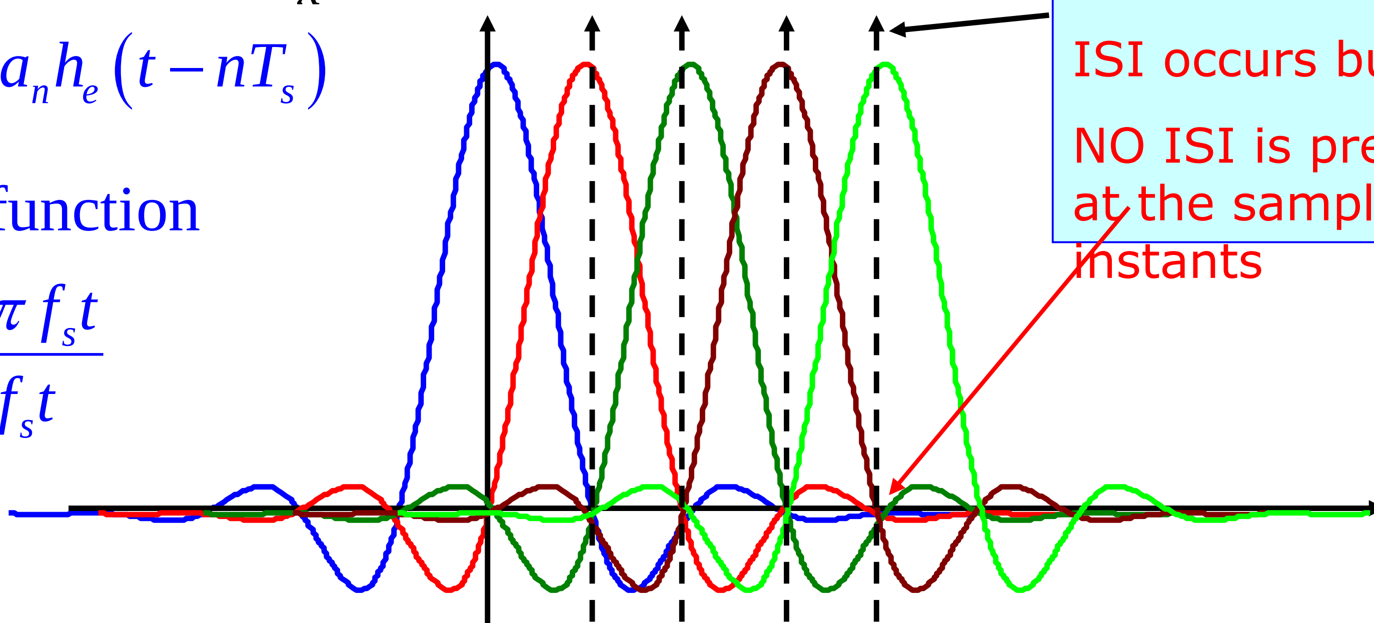
C is a nonzero constant

Now choose the $\dfrac{\sin x}{x}$ function for $h_e(t)$

$$w_{out}(t)=\sum_n a_n h_e(t-nT_s)$$

h_e is a Sa function

$$h_e(t)=\frac{\sin \pi f_s t}{\pi f_s t}$$



Sampling Instants

ISI occurs but,

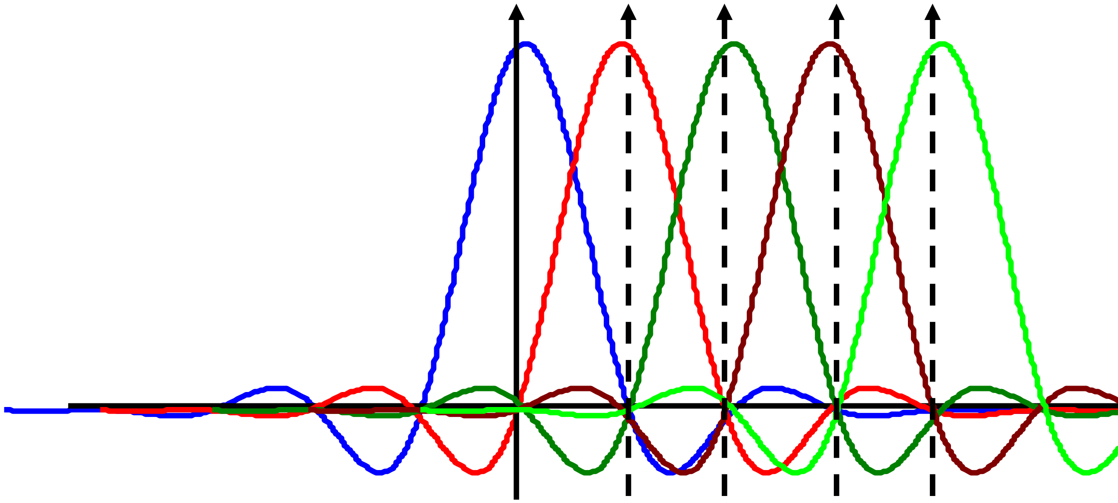
NO ISI is present at the sampling instants

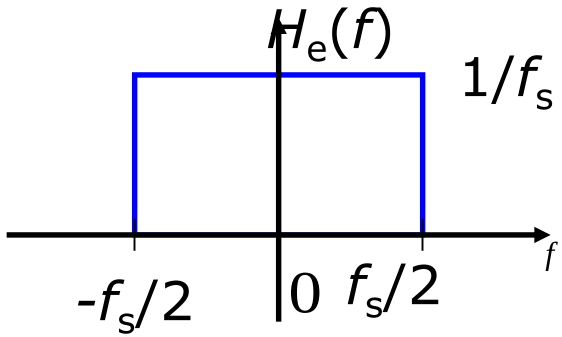
Nyquist's First Method for Zero ISI

- There will be **NO ISI and the bandwidth requirement will be minimum (Optimum Filtering)** if the transmit and receive filters are designed so that the overall transfer function $H_e(f)$ is:

$$H_e(f) = \frac{1}{f_s}\prod\left(\frac{f}{f_s}\right) \quad h_e(t) = \frac{\sin \pi f_s t}{\pi f_s t} \quad \text{Where} \quad f_s = \frac{1}{T_s}$$

- This type of pulse will allow signalling at a baud rate of $D=1/T_s=2B$ (for Binary $R=1/T_s=2B$) where B is the absolute bandwidth of the system.

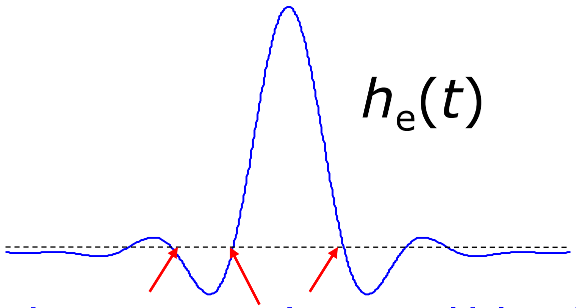




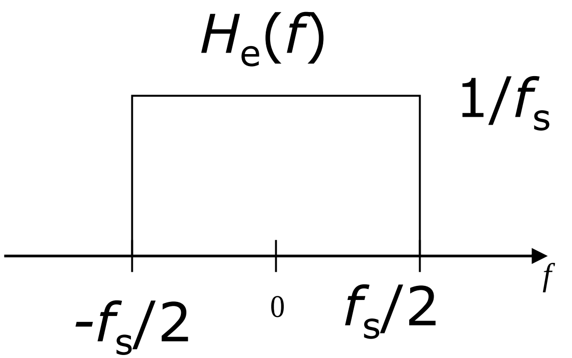
$$\text{Absolute bandwidth is: } B = \frac{f_s}{2} \quad \text{MINIMUM BANDWIDTH}$$

$$\text{Signalling Rate is: } D=1/T_s = 2B \text{ Pulses/sec}$$

Nyquist's First Method for Zero ISI



Zero crossings at non-zero integer multiples of the bit period



- Since pulses are not possible to create due to:
  - Infinite time duration.
  - Sharp transition band in the frequency domain.

- The Sinc pulse shape can cause significant ISI in the presence of timing errors.
  - If the received signal is not sampled at *exactly* the bit instant (Synchronization Errors), then ISI will occur.

- We seek a pulse shape that:
  - Has a more gradual transition in the frequency domain.
  - Is more robust to timing errors.
  - Yet still satisfies Nyquist's first method for zero ISI.

Raised Cosine-Rolloff Nyquist Filtering

- **Because of the difficulties caused by the Sinc type pulse shape, consider other pulse shapes which require more bandwidth such as the Raised Cosine-rolloff Nyquist filter but they are less affected by synchrfonization errors.**
- **The Raised Cosine Nyquist filter is defined by its rollof factor number *r*=f_Δ/f_o.**

$$H_e(f)=\begin{cases}1, & |f|<f_1\\ \frac{1}{2}\left\{1+\cos\left[\frac{\pi\left(|f|-f_1\right)}{2f_\Delta}\right]\right\}, & f_1<|f|<B\\ 0, & |f|>B\end{cases}\qquad \text{B is the Absolute Bandwidth}$$

$f_\Delta = B - f_0 \quad f_1 \equiv f_0 - f_\Delta$ Where f_o is the 6-dB bandwidth of the filter

Rolloff factor: $r=\frac{f_\Delta}{f_0}$ Bandwidth: $\mathrm{B}=\frac{R_b}{2}(1+r)$

$$h_e(t)=F^{-1}\left[H_e(f)\right]=2f_0\left(\frac{\sin 2\pi f_0 t}{2\pi f_0 t}\right)\left[\frac{\cos 2\pi f_\Delta t}{1-\left(4f_\Delta t\right)^2}\right]$$

Rolloff factor: $r=\frac{f_\Delta}{f_0}$ Bandwidth: $\mathrm{B}=\frac{R_b}{2}(1+r)$

Raised Cosine-Rolloff Nyquist Filtering

- Now filtering requirements are relaxed because absolute bandwidth is increased.
- Clock timing requirements are also relaxed.
- The r=0 case corresponds to the previous Minimum bandwidth case.

$$H_e(f) = \frac{1}{2}\left\{1 + \cos\left[\frac{\pi\left(|f| - f_1\right)}{2f_\Delta}\right]\right\}$$

$$\boxed{B = f_o + f_\Delta}$$

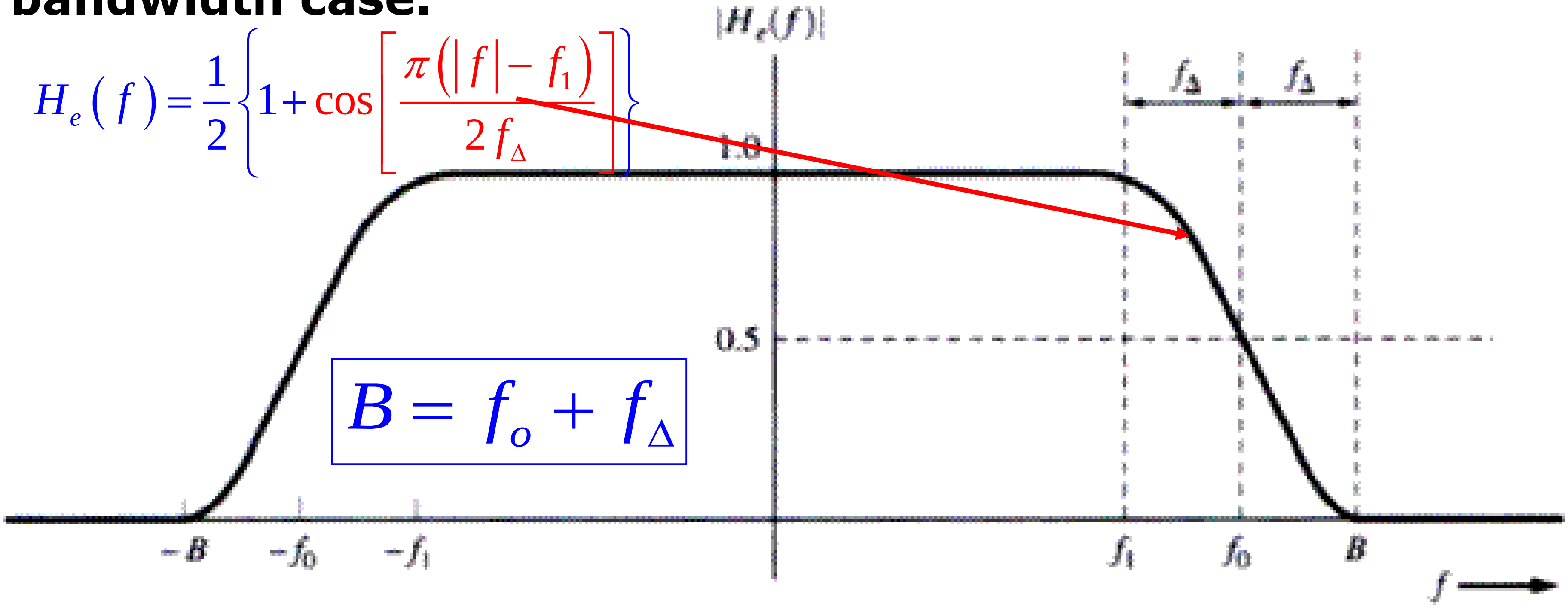


Figure 3–25 Raised cosine-rolloff Nyquist filter characteristics.

Rolloff factor: $r = \frac{f_\Delta}{f_0}$ Bandwidth: $B = \frac{R}{2}(1+r) = \frac{D}{2}(1+r)$

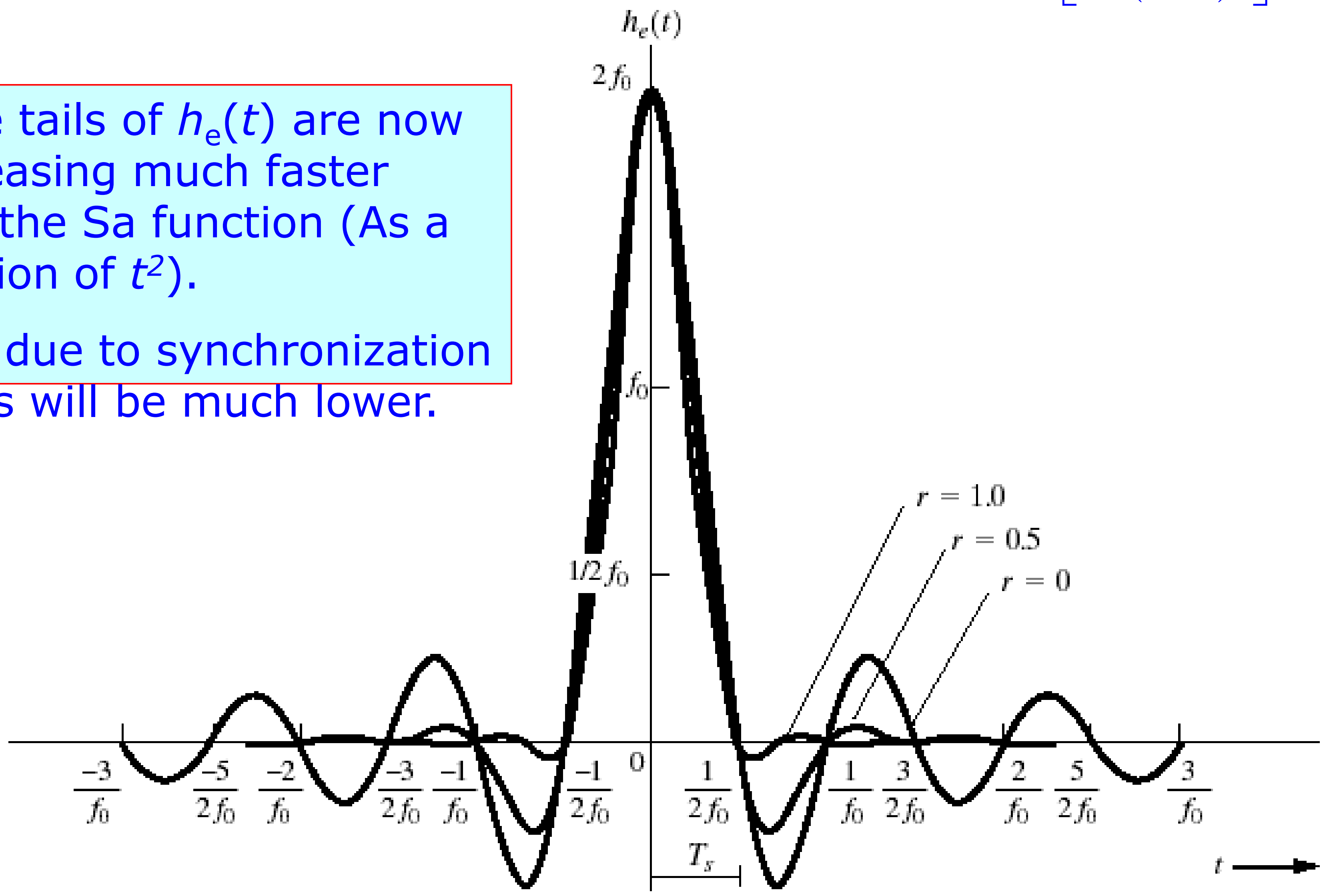
Raised Cosine-Rolloff Nyquist Filtering

- Impulse response is given by:

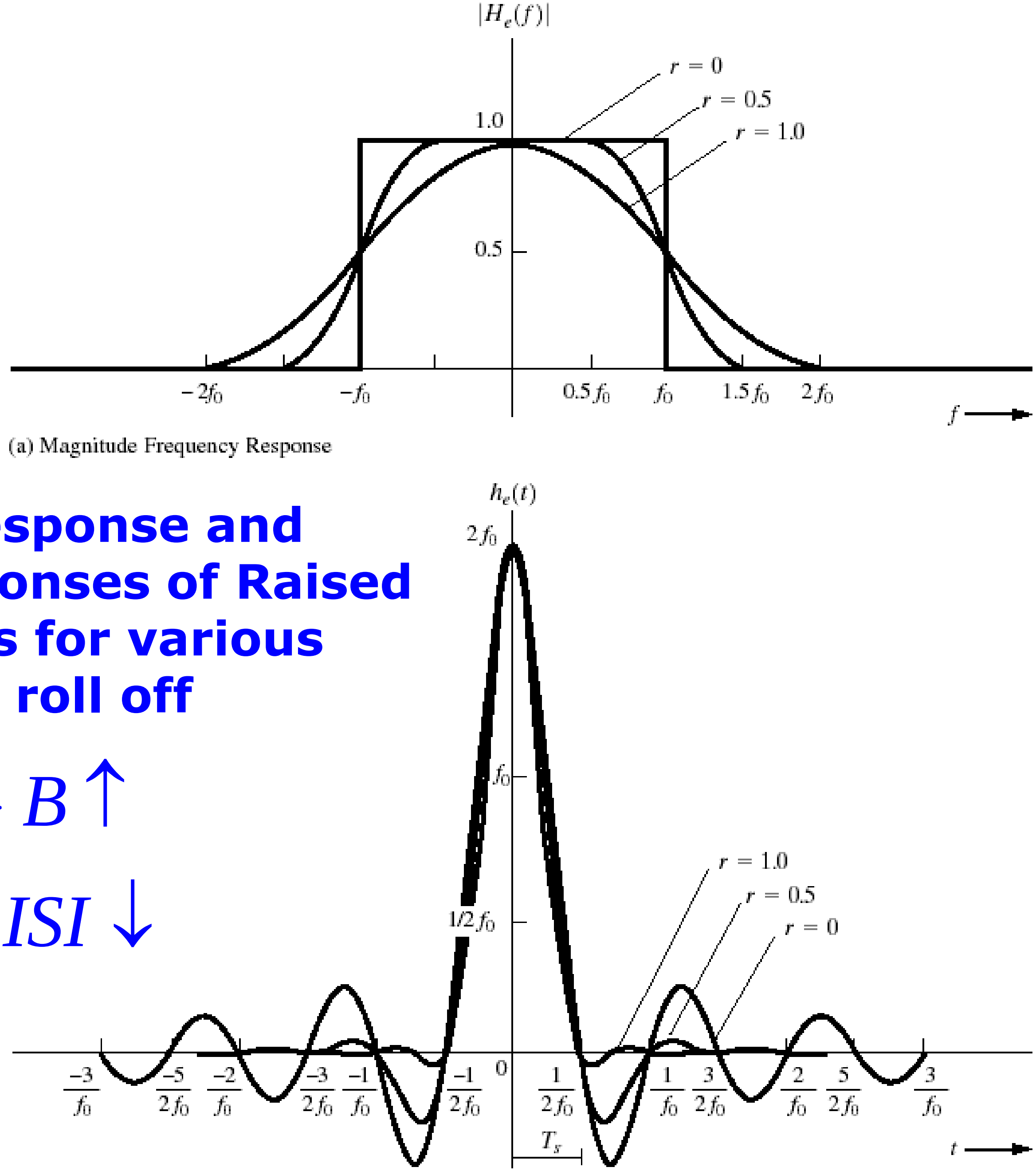
$$h_e(t) = F^{-1}\left[H_e(f)\right] = 2f_0\left(\frac{\sin 2\pi f_0 t}{2\pi f_0 t}\right)\left[\frac{\cos 2\pi f_\Delta t}{1-(4f_\Delta t)^2}\right]$$

- The tails of $h_e(t)$ are now decreasing much faster than the Sa function (As a function of t^2).

- ISI due to synchronization errors will be much lower.



Raised Cosine-Rolloff Nyquist Filtering



Frequency response and impulse responses of Raised Cosine pulses for various values of the roll off parameter.

$$r \uparrow \rightarrow B \uparrow$$

$$r \uparrow \rightarrow ISI \downarrow$$

Raised Cosine-Rolloff Nyquist Filtering

- **Illustrating the received bit stream of Raised Cosine pulse shaped transmission corresponding to the binary stream of 1 0 0 1 0 for 3 different values of *r*=0, 0.5, 1.**

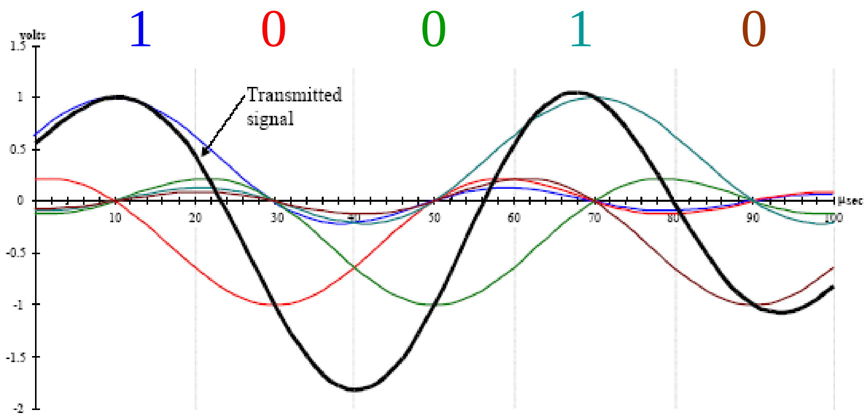


Figure 3.28 Transmitted binary PAM waveform for the data sequence "10010" using sinc-shaped pulses (raised cosine pulse shaping, $\alpha = 0$) at a transmission speed of 50,000 bits/sec. Note that this is the same plot as Figure 3.20a.

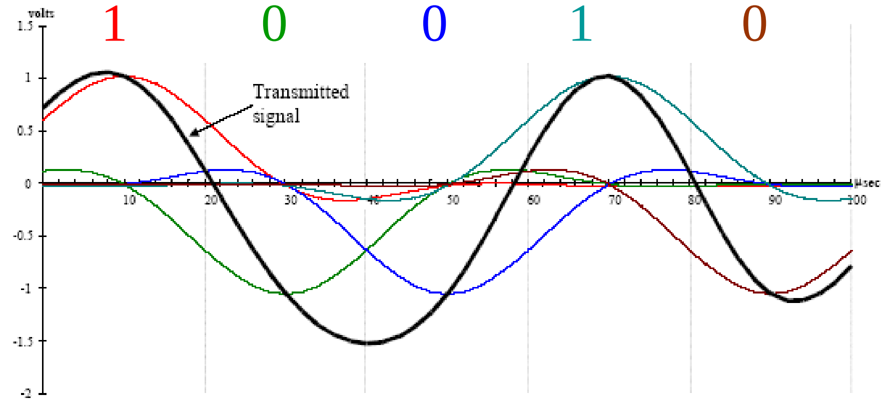


Figure 3.30 "10010" with raised cosine pulse shaping, $\alpha = 0.5$

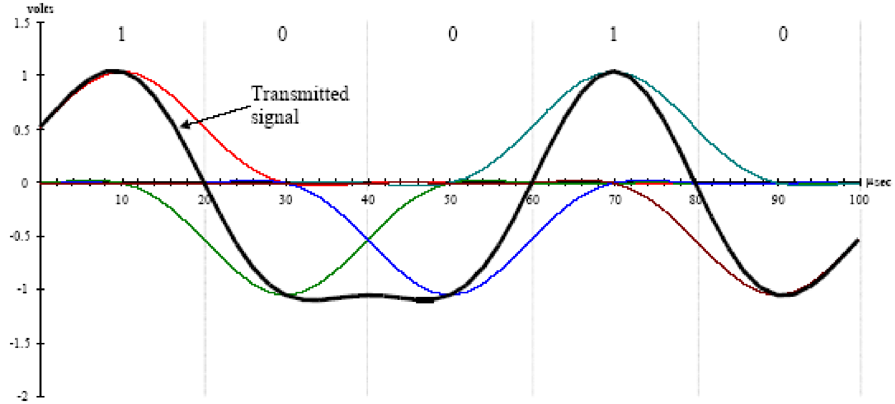


Figure 3.32 "10010" with raised cosine pulse shaping, $\alpha = 1$

Bandwidth for Raised Cosine Nyquist Filtering

- The bandwidth of a Raised-cosine (RC) rolloff pulse shape is a function of the bit rate and the rolloff factor:

$$B = f_o + f_\Delta = f_o\left(1+\frac{f_\Delta}{f_o}\right) = f_o(1+r)$$

$$B = \frac{R}{2}(1+r)$$

$$B = \frac{D}{2}(1+r) \qquad \text{Multilevel Signalling}$$

- Or solving for bit rate yields the expression:

$$R = \frac{2B}{1+r}$$

  - This is the maximum transmitted bit rate when a RC-rolloff pulse shape with Rolloff factor *r* is transmitted over a baseband channel with bandwidth *B*.

Optional

Power Spectra (I)

The transmitted signal is a pulse train:

$$X(t) = \sum_{k=-\infty}^{\infty} a_k p(t - kT - \Delta).$$

The amplitudes can be viewed as random variables with

$$R_m = \langle a_k a_{k+m} \rangle \quad m = 0,\ \pm 1,\ \pm 2,\ \ldots$$

The autocorrelation function of the waveform is

$$R_X(\tau) = \sum_{m=-\infty}^{\infty} R_m r(\tau - mT),$$

in which $r(\tau) = \dfrac{1}{T}\int_{-\infty}^{\infty} p(t+\tau)p(t)dt.$

Power Spectra (II)

The power spectral density is the Fourier transofrm of $R_X(\tau)$:

$$S_X(f) = \mathbf{FT}[R_X(\tau)] = \mathbf{FT}[\sum_{m=-\infty}^{\infty} R_m r(\tau - mT)]$$

$$= \sum_{m=-\infty}^{\infty} R_m \mathbf{FT}[r(\tau - mT)] = \sum_{m=-\infty}^{\infty} R_m S_r(f) e^{-j2\pi mTf}$$

$$= S_r(f) \sum_{m=-\infty}^{\infty} R_m e^{-j2\pi mTf}.$$

Note that $S_r(f) = \mathbf{FT}[r(\tau)] = \mathbf{FT}[\frac{1}{T} p(-t) * p(t)] = \frac{|P(f)|^2}{T}.$

Example 1: NRZ

- Assume the message $m[n]$ is random (white noise) with equally "0" and "1" values.
- Step 1: Compute R_m based on the above assumption and "format"
- Step 2: Compute $r(t)$ based on the pulse shape.

1. NRZ, 50% "0" and 50% "1".

$$R_m = \tfrac{1}{2}A^2 + \tfrac{1}{2}(-A)^2 = A^2, m = 0;$$

$$R_m = \tfrac{1}{4}A^2 + \tfrac{1}{4}A(-A) + \tfrac{1}{4}(-A)A + \tfrac{1}{2}(-A)^2 = 0 \text{ , } m \neq 0.$$

$$p(t) = \Pi(t/T) \rightarrow P(f) = T\text{sinc}(Tf)$$

Therefore, $S_{NRZ}(f) = A^2 S_r(f) = A^2 T\text{sinc}^2(Tf)$.

Example 2: Unipolar RZ

2. Unipolar RZ, 50% 1-level and 50% 0-level.

$$R_m = \begin{cases} \frac{1}{2}A^2 + \frac{1}{2}(0)^2 = \frac{1}{2}A^2, & m = 0 \\ \frac{1}{4}A \cdot A + \frac{1}{4}A \cdot 0 + \frac{1}{4}0 \cdot A + \frac{1}{4}0 \cdot 0 = \frac{1}{4}A^2, & m \neq 0 \end{cases}$$

$$p(t) = \Pi(2t/T) \rightarrow P(f) = \tfrac{T}{2}\mathrm{sinc}(\tfrac{T}{2}f)$$

$$S_{URZ}(f) = \frac{T}{4}\mathrm{sinc}^2(\frac{T}{2}f)\left[\frac{1}{2}A^2 + \frac{1}{4}A^2 \sum_{m=-\infty, m\neq 0}^{m=\infty} e^{-j2\pi mTf}\right]$$

$$= \frac{T}{4}\mathrm{sinc}^2(\frac{T}{2}f)\left[\frac{1}{4}A^2 + \frac{1}{4}A^2 \sum_{m=-\infty}^{m=\infty} e^{-j2\pi mTf}\right]$$

$$= \frac{T}{4}\mathrm{sinc}^2(\frac{T}{2}f)\left[\frac{1}{4}A^2 + \frac{1}{4}\frac{A^2}{T} \sum_{n=-\infty}^{n=\infty} \delta(f - \frac{n}{T})\right] . \because \sum_{m=-\infty}^{m=\infty} e^{-j2\pi mTf} = \frac{1}{T}\sum_{n=-\infty}^{n=\infty} \delta(f - \frac{n}{T})$$