Chapter 4

BASEBAND DATA COMMUNICTION SYSTEMS

Introduction

Till now we have studied communication primarily in terms of "electrical signal", responses of the systems to the signal, the effect of noise on the quality of the signal and so on. However, the signal theory does not providea concrete view of the fundamental communication process of information transfer. To represent the digital communication system in terms of information, Claude Shannon in 1948 came up with a theory known as 'Mathematical Theory of Communication' in which he concentrated on the information content of the message rather than on the signal. His approach was soon renamed as Information theory, and deals with mathematical modeling and analysis of a communication system rather than with physical sources and physical channel.

There two fundamental issues related to information theory pertaining to thedigital communication system are:

a. What is the rate at which the given source is emitting "information"?

b. What is the maximum rate of information transmission over a noisy channel?

The answer to the first question is the information theory, which deals with the definition, measurement and modeling of information. The answer to the second question is given by channel capacity theorem.

4.1 Information

A message is a sequence of symbols intended to reduce uncertainty of the receiver. If the sequence of symbols does not change the uncertainty level of the receiver then the message does not contain any information.

Example 4.1:

Suppose you are planning to go to Biratnagar during summer vacation. You called your friend in Biratnagar to know the weather conditions of Biratnagar. Assume that you could have received the following message regarding the weather condition in Biratnagar in summer

(a) It is sunny and hot.

(b) It is cold

(c) Snow is falling

The Information content among these three messages is different. From the knowledge about the weather condition of Biratnagar, in the past, you know that in summer Biratnagar is sunny and hot. Therefore, the reader will hardly notice the first headline (a) as it contains virtually no information.

The information content in message (b) is relatively high than that in (a) because you do not expect, in general, a cold weather during summer in Biratnagar, but may happen rarely.

Message (c) contains highest level of information (of course among the three possible messages you could have received) as snow during summer and in Biratnagar is almost next to impossible.(but still could happen).

Therefore, the information content of any message is closely related to the past knowledge of the occurrence of event and the level of uncertainty it contains with respect to the recipient of message. The same message, if given to a person living in Argentina and have never heard of Biratnagar and have no plan to go there, would have contained almost zero information.

Thus in general "on an intuitive basis, the amount of information received from the knowledge of occurrence of an event is related to the probability or likelihood of occurrence of the event".

In other words the message related to an event least likely to occur (with the knowledge of occurring of such event in the past) contains more information.

The communication system can never be described in the deterministic sense, it can be considered to be statistical in nature. To describe a communication system completely we have to use its unpredictable or uncertain behaviour.

It can be easily understood by example that each transmitter transmit the information randomly, we cannot predict which message will be transmitted the next moment. But we know the probability of transmitting a particular message.

So to define a system completely, we need statistical study of system and statistical study of system is performed with the help of concept of probability.

4.1.1 Information content of a symbol

The information source can be classified as having memory or being memoryless. A source with memory is one for which the occurrence of current symbol depends on previous symbols. A memoryless source is one for which each symbol produced is independent of the previous symbols.

For our study we assume that the symbols emitted by the source during successive signaling intervals is statistically independent (i.e., the source is memoryless).

Let us consider a discrete memoryless source (DMS) with "q" possible message represented by $\{x_1, x_2, \ldots, x_q\}$ with probabilities of occurrence $\{p_1, p_2, \ldots, p_q\}$ such that

$$p_1 + p_2 + \ldots + p_q = 1 \tag{4.1}$$

Let $I(x_i)$ be the amount of information content in the i^{th} message. Based on intuition, for $I(x_i)$ to represent information content of x_i message, the following condition of Eq.(4.2), (4.3), (4.4), (4.5) and (4.6) should be met.

$$I(x_i) > I(x_j), \text{ If } p_i < p_j \tag{4.2}$$

Eq.(4.2) means that the information content of an event with less probability is higher than that of with high probability of occurrence.

$$I(x_i) \rightarrow 0 \text{ if } p_i \rightarrow 1 \tag{4.3}$$

In Eq.(4.3) the information content in an event with the probabilty of occurrence near to unity (also called "almost certain" event) is near to zero. The example is the message to a person living in Earth that "Sun rises from the east".

$$I(x_i) \rightarrow 1 \text{ if } p_i \rightarrow 0 \tag{4.4}$$

The example for Eq.(4.4) is the information content in the message about an event which is "almost impossible" e.g. an event like "Sun rises from the west".

$$I(x_i) \geq 0 \text{ when } 0 \leq p_i \leq 1 \tag{4.5}$$

Eq.(4.5) means that any message should contain some information i.e., $I(x_i)$ should be non-negative (i.e. there should not be any information in a message that would increase uncertainityin the receipentafter receiving message).

$$I(x_i \text{ and } x_j) = I(x_i\, x_j) = I(x_i) + I(x_j) \tag{4.6}$$

Eq.(4.6) means, if x_i and x_j are two statistically independent messages coming from the same source, then the total information received from two message is equal to the sum of information contents in each message. For example, in weather forecast from Radio Nepal you hear "it will be rainy today and sunny tomorrow". There are two messages in it and if weather of today does not affect the weather of tomorrow, then the total information in this message is equal to sum of the two individual informations.

The following relationship between $I(x_i)$ and p_i would satisfy all the above conditions for $I(x_i)$ to represent the information contained by message x_i with the probability of occurrence p_i.

$$I(x_i) = \log_b \frac{1}{p_i} = \log_b p_i \tag{4.7}$$

The unit of $I(x_i)$ depends upon the base b assigned to log.

- base b is 'e', i.e., logarithm is natural (ln). The unit of information is 'nat(natural unit)'
- base b is 10, the unit is Hartley or decit(decimal unit).
- base b is 2, the unit is bit(binary unit)

If two bit (binary digits) 1 and 0 occur with equal probability and are correctly detected at the receiving end, then the information content in each digit is 1 bit.

$$I(0 \text{ or } 1) = -\log_2 \frac{1}{2} = 1\ bit \tag{4.8}$$

4.1.2 Entropy or Average Information Content

The information source is usually transmitting a long sequence of symbols as in most of the real cases a single symbol virtually carry no information. Thus, instead of define the information content of individual message symbol, it is more practical to determine the average information content of a long sequence of symbols.

Let the a source emit one of k possible symbols (long independent sequence of symbols i.e., the length of the symbols N in the sequence of much more greater than the possible symbols "k") x_1, x_2, x_k in statistically independent sequence (i.e., probability of occurrence of x_i, does not depend upon previous occurrence of x_h or future occur of x_j) with probabilities $p_1, p_2, \ldots p_k$ respectively. Now, the information content of individual symbol is discrete and random in nature (i.e., cannot be determined), that takes on the value $I(x_1), I(x_2), \ldots, I(x_k)$. The mean value of information content $I(x_k)$ is called Entropy and is denoted by H(X). Where, X is the discrete random variable.

$$H(X) = E[I(x_k)] \tag{4.9}$$

$$H(X) = \sum_{k=1}^{k} p_k\, I(x_k) \tag{4.10}$$

$$H(X) = \sum_{k=1}^{k} p_k \log_2 \frac{1}{p_i} \tag{4.11}$$

Entropy is the measure of the average information content per source symbol. If the information is in bits then the unit of entropy is bits/symbol

4.1.3 Information Rate

If the rate at which source emits symbols is r (sumbols/sec), than the information rate R of the source is given by

$$R = rH(X)\ bit\,/\,sec \tag{4.12}$$

Here R is the information rate, H(X) is source entropy or average information and r is the rate at which symbols are generated.

4.2 Shannon-Hartley channel capacity theorem

Channel capacity is concerned with the information handling capacity of a given channel. It is affected by numerous factors among which the major are:

- The attenuation of a channel which varies with frequency as well as channel length.
- The noise induced into the channel which increases with distance.
- Non-linear effects such as clipping on the signal.

Some of the effects may change with time e.g. the frequency response of a copper cable changes with temperature and age. Obviously, there is a need to model a channel in order to estimate how much information can be passed through it safely. The non linear effects and attenuation can be taken care of but it is extremely difficult to remove noise.

The highest rate of information that can be transmitted through a channel is called the channel capacity, C.

Shannon's Channel Capacity Theorem (or the Shannon-Hartley Theorem) states that:

$$C = B \log_2 \left(1 + \frac{S}{N}\right) bits/sec \tag{4.13}$$

Where C is the channel capacity, B is the channel bandwidth in hertz, S is the signal power and N is the noise power (i.e., N_0B with $N_0/2$ being the two sided noise PSD).

Note: *In the Eq.(4.13), the S/N is expressed in power ratio watt/watt and not in dB.*

Implication of Theorem

1. Indicates the upper limit of data transmission for reliable communication. That is if the data rate input to the channel exceeds its capacity then reliable communication cannot be ensured.

   A designer can estimate C for given SNR and B for reliable communication. Shannon's Channel Coding Theorem states that, if the information rate, R (r×H bits/s) is equal to or less than the channel capacity, C, (i.e. R < C) then there is, in principle, a coding technique which enables transmission over the noisy channel with no errors.

   The inverse of this is that if R > C, the probability of error is close to 1 for every symbol received and decoded.

2. **Trade off between B and SNR for given C**

   The channel capacity is limited by various extraneous factors that the designer has no control over. For example the maximum frequency that can be transmitted over a pair of cable is limited by its construction. But the other two parameters B (the signal bandwidth) and SNR are in hand of the designer. Bandwidth of the signal can be compressed and the SNR can be increased by increasing signal power or by introducing low noise devices. Therefore, the designer can trade-off between B and SNR in optimal way for given C.

Example:

A signal with the data rate R=10,000 bits/sec is required to transmit over a channel with limited bandwidth of B=3000 Hz. Theoretically, the absolute minimum bandwidth required for transmission of the above data is 10,000/2 = 5000 Hz (i.e., B=R/2). Now for transmission of R bits/sec, the minimum channel capacity should also be equal to this value, i.e., C_{min}=R. In this case the required SNR will be.

$$SNR = 2^{(C/B)} - 1 = 2^{(10000/3000)} - 1 = 9$$

This shows that the signal power must be 9 times higher than the noise power if a signal with the data rate 10 kbps is to be transmitted error free over a channel with 3 kHz bandwidth. But now if we consider a channel with B=10000 Hz, then the required SNR is

$$SNR = 2^{(C/B)} - 1 = 2^{(10000/10000)} - 1 = 1$$

It means that, if the channel bandwidth B is higher, same quality of signal transmission can be achieved with less power. In other words channel bandwidth compression from 10,000 to 3,000 Hz is possible but at the cost of increasing signal power by the factor of 9.

3. **Signal Bandwidth Compression**

Shannon channel capacity theorem indicates that it is possible to transmit signal with upper frequency f_{max} through a channel having bandwidth less than f_{max}:

Example:

Let a signal x(t) have upper frequency limit of f_{max}. Let us sample x(t) at $3f_{max}$. Then the data rate will be

$$R = nf_s = 3nf_{max} \tag{4.14}$$

where n is the number of bits per sample

Now if channel bandwidth is

$$B = f_{max}/2$$

then for the channel capacity C ≥ R and for n=6, the required SNR would be:

$$SNR = 2^{(C/B)} - 1 = 2^{(3\times 6\times f_{max}/(f_{max}/2))} - 1\ 2^{36} - 1 \approx 7\times 10^{10}$$

If we increase signal power by 7×10^{10} times in comparison to the noise power we can transmit a signal through a channel having bandwidth equal to half of the signal bandwidth. This is although possible, but very impracticable.

Theoretical limits of Shannon's channel capacity theorem

As the noise in the channel tends to zero, the value of SNR will tend to infinity (SNR→∞). Subsequently, the channel capacity C will tend to infinity (C→∞). This means that the noiseless channel has an infinite capacity. This type of channel is referred to as ideal channel.

At first sight from the Eq.(4.13), it seems that if channel bandwidth B=∞, channel capacity C=∞. However, this is not true because noise power is proportional to the bandwidth (i.e., the noise signal considered is a white noise with a uniform power density spectrum over the entire frequency range) and as the bandwidth B increases, noise N also increases and hence the channel capacity will converge even if B=∞.

If $N_0/2$ is the noise power density, then we have noise power within the bandwidth B i.e., $N=N_0B$, and then

$$C = B\log_2\left(1+\frac{S}{N_oB}\right) \text{ bits/sec} \tag{4.15}$$

$$\lim_{B\to\infty} C = \lim_{B\to\infty} \frac{S}{N_o}\frac{N_oB}{S}\log_2\left(1+\frac{S}{N_oB}\right) \text{ bis/sec} \tag{4.16}$$

$$\lim_{B\to\infty} C = \lim_{B\to\infty} \frac{S}{N_o}\log_2\left(1+\frac{S}{N_oB}\right)^{\frac{N_oB}{S}} \text{ bits/sec} \tag{4.17}$$

The above limit may be found with the help of the following standard expression:

$$\lim_{B\to 0} \log_2(1+x)^{1/x} = \log_2 e = 1.44 \tag{4.18}$$

Therefore, replacing $\frac{S}{N_oB}$ by x we have

$$\lim_{B\to\infty} C\ \lim_{B\to\infty} \frac{S}{N_o}\log_2\left(1+\frac{S}{N_oB}\right)^{\frac{N_oB}{S}} \tag{4.19}$$

$$\lim_{B\to\infty} C = \frac{S}{N_o}\lim_{x\to 0}(1+x)^{1/x} \tag{4.20}$$

$$\lim_{B\to\infty} C = \frac{S}{N_o}\log_2 e \tag{4.21}$$

$$\lim_{B\to\infty} C = C_{max} = 1.44\frac{S}{N_o} \tag{4.22}$$

So as the bandwidth goes to infinity the capacity goes to $1.44S/N_0$, i.e., it goes to a finite value.

4.3 Baseband Data Communication

This topic is dedicated to the study of the transmission of digital data over a baseband channel. Baseband digital communication system refers to a system in which transmission and reception of digital signals over band-limited channel is accomplished without employing carrier modulation (band pass) technique

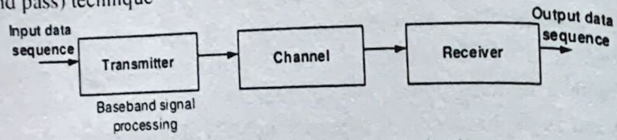


Fig. 4.1: Baseband digital communication

Digital data have a broad spectrum with significant low-frequency content. Baseband transmission of digital data therefore requires the use of low-pass channel with bandwidth large enough to accommodate the essential frequency content of the data stream. Typically, however, the channel is dispersive i.e., its frequency response deviates from that of an ideal low-pass filter. Thus, the received pulse is affected somewhat by adjacent pulses, thereby giving rise to a common form of interference called intersymbol interference (ISI).

Intersymbol Interference is a major source of bit errors in the reconstructed data stream at the receiver output. To avoid ISI certain control measures have to be implemented. One of the method is the use of discrete pulse modulation, in which the amplitude, duration, or position of the transmitted pulses is varied in discrete manner in accordance with the given data stream. However for the baseband transmission of digital data, the use of discrete pulse-amplitude modulation (PAM) is one of the most efficient schemes in terms of power and bandwidth utilization. We begin the study by considering the case of binary data.

Baseband data communication system using PAM have the following functional blocks shown in Fig.4.2.

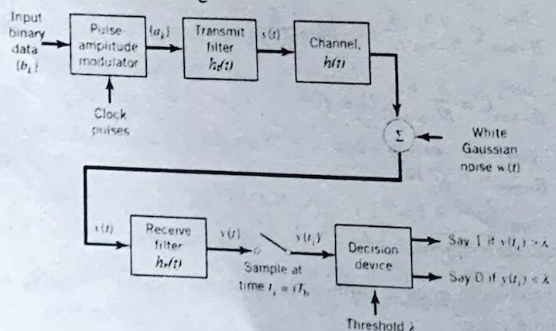


***Fig.4.2:** Block diagram of Baseband binary data transmission system*

The input binary sequence $\{b_k\}$ consists of symbols 1 and 0, each of duration T_b. This sequence is applied to a pulse generator, producing the discrete PAM signal

$$x(t) \sum_{k=\infty}^{\infty} a_k p_g(t - kT_b) \tag{4.23}$$

Where $p_g(t)$ denotes the basic pulse whose amplitude a_k depends upon the input data sequence as

$$a_k = \begin{cases} +1 & \text{if symbol } b_k \text{ is } 1 \\ -1 & \text{if symbol } b_k \text{ is } 0 \end{cases} \tag{4.24}$$

And is normalized such that

$$p_g(0) = 1 \tag{4.25}$$

Signal $x(t)$ is passed through a transmission filter of impulse response $h_t(t)$. The output further modified as it passes through the channel of impulse response $h_c(t)$. In addition, the channel adds random noise to the signal at the receiver input. The channel output is then passed through a receive filter of impulse response $h_r(t)$. The receive filter output is written as

$$y(t) = \mu \sum_k a_k p_r(t - kT_b) - \tau_d) + n(t) \tag{4.29}$$

Where μ is the scaling factor, τ_d is delay introduced in the system and $n(t)$ is noise.

The pulse $p_r(t)$ is normalized such that

$$p_r(0) = 1 \tag{4.27}$$

The resulting filter output y(t) is sampled synchronously with the transmitter, with the sampling instants being determined by a clock or timing signal that is usually extracted from the receive filter output. Finally, the sequence of samples thus obtained is used to reconstruct the original data sequence by means of a decision device. Specifically, the amplitude of each sample is compared to a threshold λ. If the threshold λ is exceeded, a decision is made in favor of symbol 1. If the threshold λ is not exceeded, a decision is made in favor of symbol 0.

For simplicity of further analysis we assume that $\tau_d = 0$ and that the channel is noiseless, i.e., $n(t) = 0$.

Then in frequency domain the received pulse can be expressed as the response of the cascaded connection of the transmitting filter, the channel, and the receiving filter, which is produced by the pulse $p_g(t)$ applied to the input of this cascade connection. Therefore, we may relate $p_g(t)$ and $p_r(t)$ in the frequency domain by

$$\mu P_r(f) = P_g(f)\, H_T(f) H_C(f)\, H_R(f) \tag{4.28}$$

Where $P_g(f)$ and $P_r(f)$ are the Fourier transforms of $p_g(t)$ and $p_r(t)$, respectively. $H_T(f)$, $H_C(f)$ and $H_R(f)$ are the transfer functions of transmitting filter, channel and receiving filter respectively.

The receiving filter output y(t) is sampled at time $t=mT_b$ is

$$y(t = mT_b) = \mu a_m + \mu \sum_{\substack{k=-\infty \\ k \neq m}}^{\infty} a_k p_r(mT_b - kT_b) \tag{4.29}$$

$$y(t = mT_b) = \mu a_m + \mu \sum_{\substack{k=-\infty \\ k \neq m}}^{\infty} a_k p_r[(m-k)T_b] \tag{4.30}$$

In Eq.(4.30), the first term μa_m is the m^{th} decoded bit and the second term represents the residual effect of all other transmitted bit on the decoding of the m^{th} bit; this residual effect is called intersymbol interference (ISI).

ISI arises due to dispersion of pulse shape by the filters and channel. Therefore, one of the major task of the system designer is to optimally design transmitting and receiving filters and the shape of the basic pulse to minimize ISI.

4.3.1 Nyquist's criterion for distortionless baseband binary transmission

Generally, the properties of the channel and the shape of the transmitted pulse are known to the receiver. However, as stated above, the problem is to determine the transfer function of the transmitter and receiving filters so as to reconstruct the transmitted data sequence $\{b_k\}$ without any error. The receiver does this by extracting and then decoding the corresponding sequence of weights, $\{a_k\}$, from the output y(t) at the decisiontime of m-th bit, $t_m = mT_b$. The decoding requires that the weighted pulse $a_k p_r(mT_b - kT_b)$ for $k=m$ be free from ISI due to the overlapping tails of all other weighted pulse represented by $k \neq m$. This, in turn, requires that we control the received pulse $p_r(t)$, as shown by Eq.(4.31)

$$p_r(mT_b - kT_b) = \begin{cases} 1 & m = k \\ 0 & m \neq k \end{cases} \tag{4.31}$$

Or in general,

$$p_r(iT_b) = \begin{cases} 1 & i = 0 \\ 0 & i \neq 0 \end{cases} \tag{4.32}$$

Where, integer $i=m-k$ and $p_r(0)=1$. If $p_r(t)$ satisfies the condition of Eq.(4.31), the receiver output, given by Eq.(4.30), simplifies to

$$y(t = mT_b) = \mu a_m \tag{4.33}$$

which implies zero ISI. Hence, the condition of Eq.(4.32) assures perfect reception in the absence of noise. The shaping of the pulse according to the Eq.(4.32) is also referred to as Nyquist Pulse Shaping Criteria for zero ISI.

Frequency Domain Analysis

From a design point of view, it is informative to transform the condition of Eq.(4.32) into the frequency domain. Consider then the sequence of samples $\{p_r(nT_b)\}$, where n = 0, ±1, ±2, ...From chapter 2 on the sampling process for a low-pass function, we know that sampling in the time domain produces periodicity in the frequency domain. Thus, from Eq.(2.3) in chapter 2 we may write.

$$P_\delta(f) = R_b \sum_{n=-\infty}^{\infty} P_r(f - nR_b) \tag{4.34}$$

Where $R_b=1/T_b$ is the bit rate; $P_\delta(f)$ is the Fourier transform of an infinite periodic sequence of delta function of period T_b, and whose strengths are weighted by the receptive sample values of $p(t)$. That is, $P_\delta(f)$ is given by

$$P_\delta(f) = \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} [p_r(iT_b)\, \delta(t - iT_b)] \exp(-j2\pi f t)\, dt \tag{4.35}$$

Let the integer $i = m - k$. Then, $m = k$ corresponds to $i = 0$, and likewise $m \neq k$ corresponds to $i \neq 0$. Accordingly, imposing the condition of Eq.(4.32). On the sample values of $p(t)$ in the integral of Eq.(4.35)., we get

$$P_\delta(f) = \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} p_r(0)\, \delta(t) \exp(-j2\pi f t)\, dt \tag{4.36}$$

$$P_\delta(f) = p_r(0) \tag{4.37}$$

Where, we have made use of the shifting property of the delta function. Since $p_r(0)=1$, by normalization, we thus see from Eq.(4.34).and Eq.(4.37). that the condition for zero ISI is satisfied if

$$\sum_{n=-\infty}^{\infty} P_r(f - nR_b) = T_b \tag{4.38}$$

Eq.(4.31) in terms of the time function $p_r(t)$, or equivalently, Eq.(4.38) in terms of the corresponding frequency function P(f), is called the Nyquist criterion for distortionless baseband transmission in the absence of noise.

Ideal Solution

From Eq.(4.38) we see that the function P(f) represents the series of shifted spectrums. P(f) is obtained by permitting only one non-zero component in the series (i.e., for n=0). The range of frequencies for P(f) extend from $-B_o$ to B_o where B_o denotes half the bit rate:

Hence, $B_o = \dfrac{R_b}{2}$

That is, P(f) is specify as a rectangular spectrum given by

$$P(f) = \frac{1}{2B_o} \operatorname{rect}\left(\frac{f}{2B_o}\right) \tag{4.39}$$

This is the spectrum of a signal which produces zero ISI which is shown in Fig.4.3(a).

The time domain representation of the signal is nothing but a sinc function obtained by taking inverse Fourier Transform of Eq.(4.39) and is shown below

$$p(t) = F^{-1}[P(f)] = F^{-1}\left[\frac{1}{2B_o} rect\left(\frac{f}{2B_o}\right)\right] \tag{4.40}$$

$$p(t) = \frac{\sin(2\pi B_o t)}{2\pi B_o t} \tag{4.41}$$

$$\therefore \quad p(t) = \text{sinc}\,(2B_o t) \tag{4.42}$$

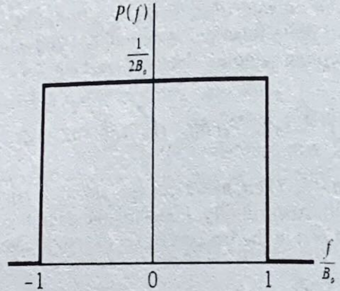
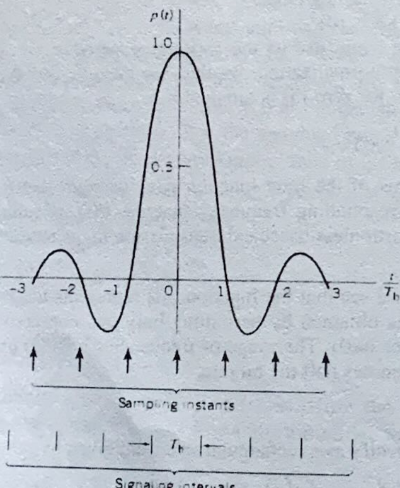


Fig.4.3: (a) Graphical representation of P(f)



(b)

Figure 4.3(b): Time-domain response

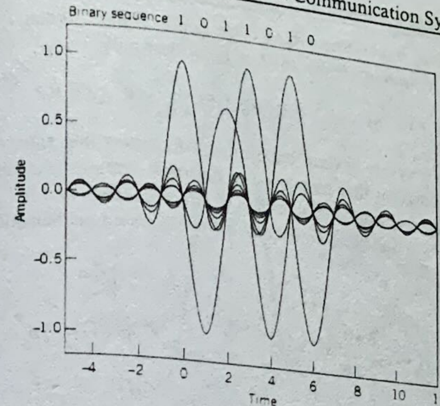


Fig.4.4: The series of sinc pulses corresponding to the sequence 1011010

Fig.4.3 (b) show plot of p(t). In Fig.4.3(b), the signaling intervals and the corresponding centered sampling instants is also shown. The function p(t) can be regarded as the impulse response of an ideal low-pass filter with pass-band amplitude response $1/(2B_o)$ and bandwidth B_o. The function p(t) has its peak value at the origin and goes through zero at integer multiples of the bit duration T_b. It is apparent that if the received waveform y(t) is sampled at the instants of time $t = 0, \pm T_b, \pm 2T_b, \ldots$, then the pulses defined by $\mu p(t - iT_b)$ with arbitrary amplitude μ and index $i = 0, \pm 1, \pm 2, \ldots$, will not interfere with each other.

This choice of pulse shape for $p(t)$ solves the problem of ISI with the minimum bandwidth possible. However, there are two practical difficulties that make it an undesirable for system design:

1. The amplitude characteristic of P(f) needs to be flat from $-B_0$ to B_0, and zero elsewhere, which is physically unrealizable because of the abrupt transitions at $\pm B_0$.
2. Due to discontinuity of P(f) at $\pm B_0$, there is practically no margin of error in sampling times in the receiver.

4.3.2 Raised Cosine Spectrum (Practical Consideration)

The practical difficulties faced in the ideal Nyquist channel can be overcome by extending the bandwidth from $B_o = R_b/2$ to an adjustable value between B_o and $2B_o$. In doing so, we expand the series on the left side of Eq.(4.38) i.e.,

$$\sum_{n=-\infty}^{\infty} P_r(f - nR_b) = T_b \tag{4.43}$$

and retain only three terms which corresponds to n=-1, 0 and 1 and restrict the frequency band of interest to $|f| \le B_o$ as shown by

$$P(f)+P(f-2B_o)+P(f+2B_o)=\frac{1}{2B_o} \quad -B_o \le f \le B_o \tag{4.44}$$

Now, we may devise several band-limited functions that satisfy Eq.(4.44). One of the function that match many desirable features is a raised cosine spectrum. The spectrum characteristic of which consists of a flat portion and a roll-off portion that has a sinusoidal form, expressed mathematically as

$$P(f)=\begin{cases} \dfrac{1}{2B_o} & |f|<f_1 \\ \dfrac{1}{4B_o}\left\{1+\cos\left[\dfrac{\pi(|f|-f_1)}{2B_o-2f_1}\right]\right\} & f_1 \le |f| < 2B_o-f_1 \\ 0 & |f| \ge 2B_o-f_1 \end{cases} \tag{4.45}$$

The frequency f_1 and bandwidth B_o are related by

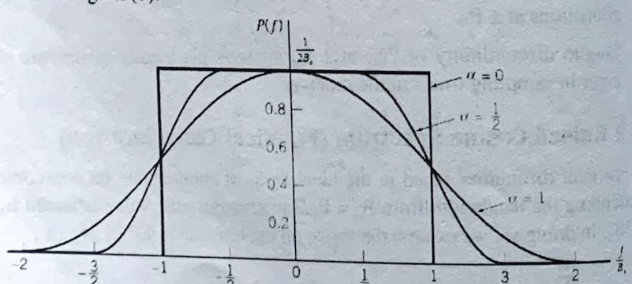
$$\alpha = 1-\frac{f_1}{B_o} \tag{4.46}$$

The parameter α is called the rolloff factor, which indicates the excess bandwidth over the ideal solution B_o. For $\alpha = 0$, that is, $f_1=B_o$, we get the minimum bandwidth solution as described in ideal case.

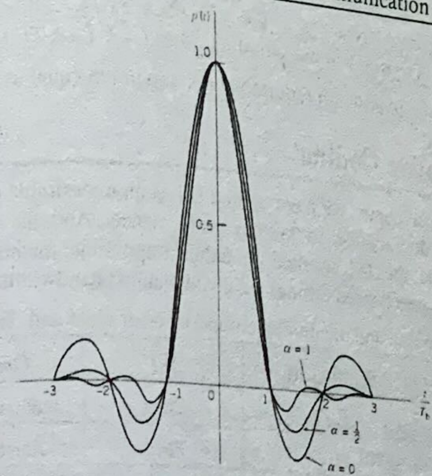
The time response p(t), that is, the inverse Fourier transform of P(f), is defined by

$$p(t)=\text{sinc}(2B_o t)\frac{\cos(2\pi\alpha B_o t)}{1-16\alpha^2 B_o^2 t^2} \tag{4.47}$$

The normalized frequency response of the raised cosine function is obtained by multiplying P(f) by $2B_o$, and is shown plotted in Fig. 4.5(a) for three values of a namely, 0, 0.5, and 1. The corresponding time response p(t) is plotted in Fig.4.5(b).



(a)



(b)

Fig.4.5: Responses for different roll-off factors (a) Frequency response. (b) Time response

This function consists of the product of two factors: the factor $\text{sinc}(2B_o t)$ associated with the ideal filter, and a second factor that decreases as $1/|t|^2$ for large |t|. The first factor ensures zero crossings of p(t) at the desired sampling instants of time $t = iT_b$ with i an integer (positive and negative).

The second factor reduces the tails of the pulse considerably below that obtained from the ideal low-pass filter, so that the transmission of binary waves using such pulses is relatively insensitive to sampling time errors. In fact, the amount of ISI resulting from this timing error decreases as the roll-off factor α is increased from zero to unity.

For the special case of $\alpha = 1$ (i.e., $f_1=0$) is known as the full-cosine roll-off characteristics, for which the frequency response of Eq.(4.45) simplifies to

$$P(f)=\begin{cases} \dfrac{1}{4B_o}\left[1+\cos\left(\dfrac{\pi f}{2B_o}\right)\right], & 0<|f|<2B_o \\ 0, & |f| \ge 2B_o \end{cases} \tag{4.48}$$

the function p(t) simplifies to

$$p(t)=\frac{\text{sinc}(4B_o t)}{1-16B_o^2 t^2} \tag{4.49}$$

General Observation

1. For α=0.5 and 1, the characteristics of P(f) changes gradually with respect to frequency. Hence, is practically realizable.

2. There are zero crossings at $t = \pm 3T_b/2, \pm 5T_b/2....$ in addition to the usual zero crossings at the sampling times $t = \pm T_b, \pm 2T_b)......$
3. For $\alpha=1$, the bandwidth requirement is maximum equal to $2B_o$.

4.4 Correlative Coding

In the previous schemes, we have treated ISI as an undesirable phenomenon that produces degradation in system performance. And the measures to reduce/eliminate the ISI resulted in either impossible method (ideal) or reduce the data rate (raised cosine) for given channel bandwidth.

Table.4.1: Data rate and bandwidth comparison for ideal and raised cosine.

Data rate	Bandwidth	ISI	Condition
R_b	$R_b/2$	Zero	Ideal (unrealizable)
R_b	R_b	Zero	Raised cosine

It is possible to achieve a bit rate of R_b bits per second in a channel of bandwidth $R_b/2$ hertz by adding ISI to the transmitted signal in a controlled manner. Such schemes are called correlative coding or partial-response signaling schemes. As ISI introduced into the transmitted signal is known, its effect can be compensated at the receiver. Thus, correlative coding is a practical means of achieving the theoretical maximum signaling rate of R_b bits per sec in a bandwidth of $R_b/2$ hertz, but at the cost of increased transmitted power.

4.4.1 Duobinary Signaling

Duobinary signaling describes the basic idea behind correlative coding, where "duo" implies doubling of the transmission capacity of a binary system.

Consider a binary input sequence $\{b_k\}$ consisting of uncorrelated binary digits each having duration T_b seconds, with symbol 1 represented by a pulse of amplitude +1 volt, and symbol 0 by a pulse of amplitude -1 volt. When this sequence is applied to a duo binary encoder, it is converted into a three-level output, namely, -2, 0, and +2 volts. This transformation is produced according to scheme shown in Fig. 4.6.

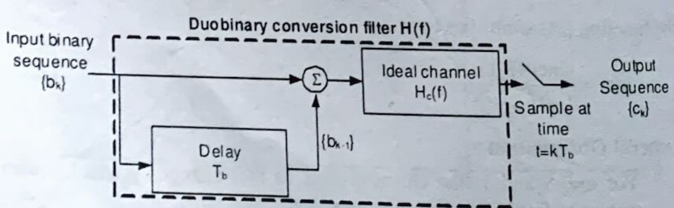


Fig. 4.6: Duobinary signaling scheme

The binary sequence $\{b_k\}$ is first passed through a simple filter involving a single delay element. The digit c_k at the duobinary coder output is the sum of the present binary digit b_k and its previous value b_{k-1}, as shown by

$$c_k = b_k + b_{k-1} \tag{4.50}$$

Such that

$$c_k = \begin{cases} +2 & \text{if } b_k \text{ and } b_{k-1} \text{ are both } 1 \\ 0 & \text{if } b_k \text{ and } b_{k-1} \text{ are different} \\ -2 & \text{if } b_k \text{ and } b_{k-1} \text{ are both } 0 \end{cases} \tag{4.51}$$

The effects of the transformation described by Eq.(4.50) is to change the input sequence $\{b_k\}$ of uncorrelated binary digits into a sequence $\{c_k\}$ of correlated digits. This correlation may be viewed as imposing intersymbol interference into the transmitted signal in an artificial manner. However, this intersymbol interference will be known to the receiver and it is the basis of correlative coding.

An ideal delay element, producing a delay of T_b seconds, has the transfer function $\exp(-j2\pi f T_b)$, so that the transfer function of the simple filter shown in Fig.4.6 is $1 + \exp(-j2\pi f T_b)$. Hence, the overall transfer function of this filter connected in cascade with the ideal channel $H_c(f)$ is

$$H(f) = H_c(f)[1 + \exp(-j2\pi f T_b)] \tag{4.52}$$

$$H(f) = H_c(f)[\exp(j\pi f T_b) + \exp(-j\pi f T_b)]\exp(-j\pi f T_b) \tag{4.53}$$

$$H(f) = 2H_c(f)\cos(\pi f T_b)\exp(-j\pi f T_b) \tag{4.54}$$

For an ideal channel of bandwidth $B_o=R_b/2$, we have

$$H_c(f) = \begin{cases} 1 & |f| \le R_b/2 \\ 0 & \text{otherwise} \end{cases} \tag{4.55}$$

Thus the overall frequency response has the form of a half-cycle cosine function, as shown by

$$H(f) = \begin{cases} 2\cos(\pi f T_b)\exp(-j\pi f T_b) & |f| \le R_b/2 \\ 0 & \text{otherwise} \end{cases} \tag{4.56}$$

for which the amplitude and phase response are as shown in Fig.4.7(a) and Fig.4.7(b), respectively. An advantage of this frequency response is that it can be easily approximated in practice.

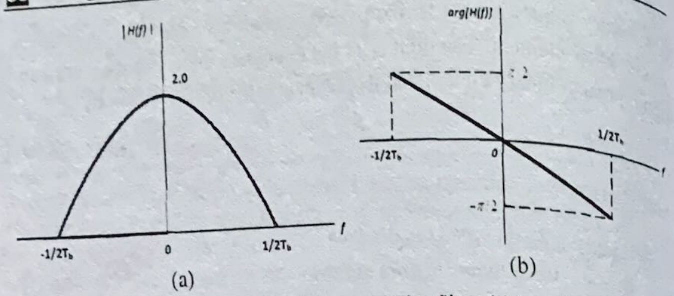


Fig.4.7: Frequency response of duobinary conversion filter. (a) Amplitude response. (b) Phase response.

The corresponding value of the impulse response consists of sinc pulse, time-displaced by T_b seconds, as shown by

$$h(t) = \frac{\sin(\pi t/T_b)}{\pi t/T_b} - \frac{\sin[\pi(t-T_b)/T_b]}{\pi(t-T_b)/T_b} \tag{4.57}$$

$$h(t) = \frac{\sin(\pi t/T_b)}{\pi t/T_b} - \frac{\sin(\pi t/T_b)}{\pi(t-T_b)/T_b} \tag{4.58}$$

$$h(t) = \frac{T_b^2 \sin(\pi t/T_b)}{\pi t(T_b - t)} \tag{4.59}$$

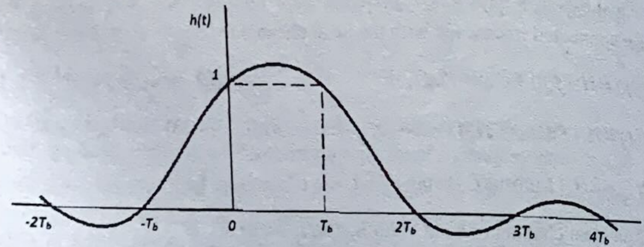


Fig. 4.8: Impulse response of duobinary conversion filter

The above Fig.4.8 shown that h(t) has two distinguishable values at the sampling instants $+T_b$ and $-T_b$.

Detection

The original data (b_k) may be detected from the duobinary-coded sequence (c_k) by subtracting the previous decoded binary digit from the currently received digit c_k in accordance with Eq.(4.50). Suppose, $\hat{b}_k$ represent the estimate of the original binary digit b_k, we have

$$\hat{b}_k = c_k - \hat{b}_{k-1} \tag{4.60}$$

Eq.(4.60) will yield correct $\hat{b}_k$ only if the previous bit $\hat{b}_{k-1}$ was correctly decoded at sampling instance $t = (k - 1)T_b$. But if the previous bit was wrongly decoded then all the remaining bits will be wrongly decoded. This phenomenon is referred to as error propagation.

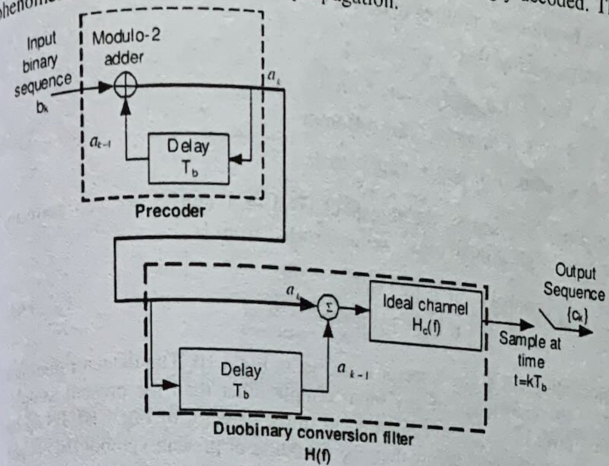


Fig.4.9: A precoded duobinary scheme

A practical means of avoiding this error propagation is the use of precoding before the duobinary coding (Fig.4.6), as shown in Fig. 4.9. The precoding operation performed on the input binary sequence (b_k) converts it into another binary sequence (a_k) defined by

$$a_k = b_k \oplus a_{k-1} \tag{4.61}$$

The '$\oplus$' sign in Eq.(4.61) represents module-2 addition which is equivalent to the EXCLUSIVE-OR operation. The resulting precoder output (a_k) is applied to the duobinary coder, thereby producing sequence (c_k) that is related to (a_k) as follows.

$$c_k = a_k + a_{k-1} \tag{4.62}$$

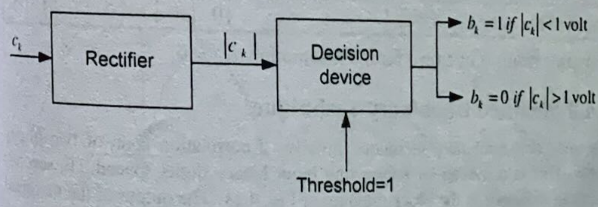


Fig.4.10: Detector for recovering original binary sequence from the precoded duobinary coder output.

It is assumed that symbol 1 at the precoder output in Fig.4.9 is represented by +1 volt and symbol 0 by -1 volt. Then if m^{th} bit b_m is 0 then $a_m = a_{m-1}$ from Eq.(4.61). It is evident that for $b_m = 0$, c_m will be either 2 or -2. Similarly, if $b_m = 1$, then a_m is the complement of a_{m-1} i.e., a_m and a_{m-1} will always be different resulting in $c_m = 0$.

Therefore, we find that

$$c_k = \begin{cases} 2 \text{ volts, if } a_k \text{ and } a_{k-1} \text{ are both } 1 \\ 0 \text{ volts, if } a_k \text{ and } a_{k-1} \text{ are different} \\ -2 \text{ volts, if } a_k \text{ and } a_{k-1} \text{ are both } 0 \end{cases} \quad (4.63)$$

From Eq.(4.63), setting the threshold level at 1 and -1 we can correctly detect the original input binary sequence $\{b_k\}$ from $\{c_k\}$.

$$b_k = \begin{cases} \text{symbol 0} & \text{if } |c_k| > 1 \text{ volt} \\ \text{symbol 1} & \text{if } |c_k| < 1 \text{ volt} \end{cases} \quad (4.64)$$

A block diagram of the detector is shown in Fig.4.10. The detector does not require the knowledge of any input sample other than the present sample. Hence, error propagation cannot occur in the detector of Fig.4.10. It can be shown from example below that, the decoding of present symbol (b_k) is not affected by the error in its past value (b_{k-1}).

Example: Assumption: $a_{k-1} = 1$

Consider an input binary sequence 0010110. Now the process of duobinary encoding and decoding is explained in the table below.

Count	(k – 1)	k						
Input Seq. (b_k)		0	0	1	0	1	1	0
precoder output ($a_k = b_k \oplus a_{k-1}$)	1 (assumed value)	1	1	0	0	1	0	0
Polar representation of a_k	+1	+1	+1	–1	–1	+1	–1	–1
DB encoder output ($c_k = a_k + a_{k-1}$)		+2	+2	0	–2	0	0	–2
Decoded bit using Eq.(4.64)		0	0	1	0	1	1	0

The same result is obtained for the assumption $a_{k-1} = 0$.

4.4.2 Modified Duobinary Technique

The modified duobinary technique involves a correlation span of two binary digits. This is achieved by subtracting input binary digits spaced $2T_b$ seconds apart, as indicated in the block diagram of Fig. 4.11. The output of the modified duobinary conversion filter is related to the sequence at its input as follows

$$c_k = a_k - a_{k-2} \quad (4.65)$$

Here, three-level signal is generated. If $a_k = \pm 1$ volt assumed previously, c_k takes on one of three values i.e., 2, 0, and -2 volts.

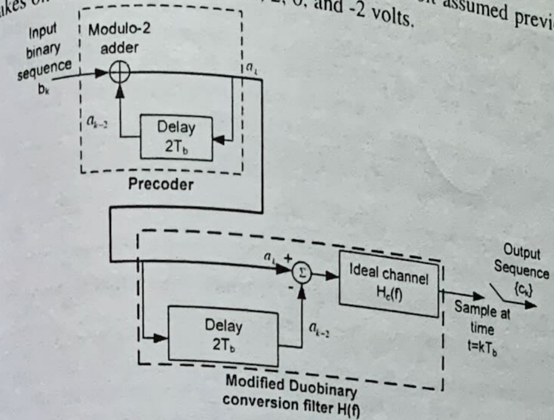


Fig.4.11: Modified duobinary signaling scheme

The overall transfer function of the tapped-delay-line filter connected in cascade with the ideal channel, as in Fig. 4.11, is given by

$$H(f) = H_c(f)\,[1 - \exp(-j4\pi f T_b)]$$

$$H(f) = 2jH_c(f)\sin(2\pi f T_b)\exp(-2jfT_b) \quad (4.66)$$

$$(4.67)$$

Here, $H_c(f)$ is as defined in Eq.(4.55). Therefore, the overall frequency response is in the form of a half-cycle sine function, as

$$H(f) = \begin{cases} 2j\sin(2\pi f T_b)\exp(-j2\pi f T_b) & |f| \le R_b/2 \\ 0 & \textit{elsewhere} \end{cases} \quad (4.68)$$

The corresponding amplitude response and phase response of the modified duobinary-coder are shown in Fig.4.12 (a) and (b), respectively. The phase response in Fig.4.12(b) does not include the constant 90° phase shift due to the multiplying factor j. A useful feature of the modified duobinary coder is that its output has no dc component. This property is important, as in practice many communication channels cannot transmit a dc component.

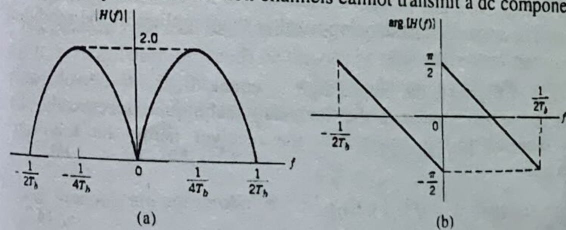


Fig.4.12: Frequency response of the modified duobinary conversion filter (a) Magnitude response.(b) Phase response.

The impulse response of the modified duobinary coder consists of sinc pulse that is time-displaced by $2T_b$ seconds, as shown

$$h(t) = \frac{\sin(\pi t/T_b)}{\pi t/T_b} - \frac{\sin[\pi(t-2T_b)/T_b]}{\pi(t-2T_b)/T_b} \tag{4.69}$$

$$h(t) = \frac{\sin(\pi t/T_b)}{\pi t/T_b} + \frac{\sin(\pi t/T_b)}{\pi(t-2T_b)/T_b} \tag{4.70}$$

$$h(t) = \frac{2T_b^2 \sin(\pi t/T_b)}{\pi t(2T_b - t)} \tag{4.71}$$

This impulse response is plotted in Fig. 4.13, which shows that it has three distinguishable levels at the sampling instants.

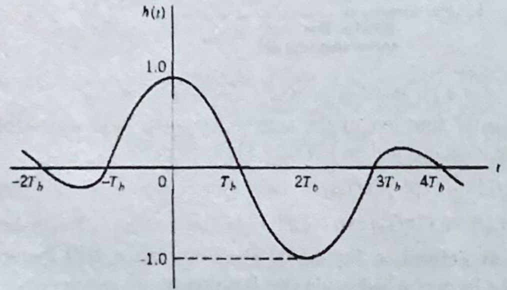


Fig.4.13: Impulse response of the modified duobinary conversion filter.

In order to eliminate the possibility of error propagation in the modified duobinary system, we use a precoding procedure similar to that used for the duobinary case. Prior to the generation of the modified duobinary signal, a modulo-2 logical addition is used on signals $2T_b$ seconds apart, as shown in Fig. 4.11.

$$a_k = b_k \oplus a_{k-2} \tag{4.72}$$

where (b_k) is the input binary sequence and (a_k) is the sequence at the precoder output. Note that modulo-2 addition and modulo-2 subtraction are the same. The sequence (a_k) thus produced is then applied to the modified duobinary conversion filter.

In the case of Fig.4.11, the output digit c_k equals 0,+2, or -2 volt. c_k is obtained in the way similar to the duobinary technique. Specifically, the original sequence (b_k) is obtained at the receiver using the following decision rule:

$$b_k = \begin{cases} symbol\ 0 & if\ |c_k| < 1volt \\ symbol\ 1 & if\ |c_k| > 1volt \end{cases} \tag{4.73}$$

4.5 M-ary Signaling

In binary signaling, the output of the pulse generator can have one of two possible levels. In M-ary signaling, the output can have one of M possible levels. In M-ary system the input source emits one of M distinct symbols and each symbol is assigned a distinct level out of M possible levels. For example, for M=4, the signal representation can be defined as

Input symbol	Representation Level	Bit representation
B	–3A	00
C	–A	01
D	+A	10
E	+3A	11

Combination of two bits can be represented by one level in 4-ary signaling system. For M=4 (i.e., 2^2), the combination of input bit stream in binary form can be represented by slicing the stream into four groups of two bits (00, 01, 10, 11). Each group is then assigned a fixed level.

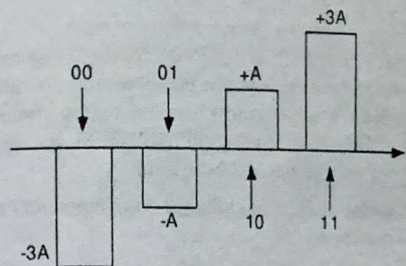


Fig.4.14: 4-ary Signaling scheme.

This signaling allows us to transmit each pair of binary digits by one 4-ary pulse. Hence to transmit n binary digits, we need only (n/2) 4-ary pulse.

At receiving end, the decoded output is compared with present threshold values (slicing levels) and decision is made.

In general, the information I_M transmitted by an M-ary symbol is

$$I_M = \log_2 M \text{ bits} \tag{4.74}$$

The signaling rate and bandwidth requirement of M-ary

Let R_s (symbols/sec or Baud) be the rate of symbols emitted by the source. Now if the M symbols emitted are equiprobable and statistically independent, then the source entropy will be.

$$H(X) = \sum_{i=1}^{M} p_i \log_2(1/p_i) \tag{4.75}$$

As all symbols are equiprobable

$$p_1 = p_2 = \dots = p_M = \frac{1}{M}$$

Thus,

$$H(X) = \sum_{i=1}^{M} \frac{1}{M} \log_2(M) \tag{4.76}$$

$$H(X) = \frac{M}{M} \log_2(M) \tag{4.77}$$

$$H(X) = \log_2(M) \tag{4.78}$$

If the rate at which source emits symbols is r, than the information rate R of the source is given by

$$R = r\,H(X)\ bit/\sec \tag{4.79}$$

$$R = R_s \log_2 M\ bit/\sec \tag{4.80}$$

The signaling interval duration $T_s = 1/R_s$, is same for both binary and M-ary systems. Therefore, the absolute minimum bandwidth required to transmit $R_s \log_2 M$ bits/sec of information is $R_s/2$ Hz.

Under similar conditions (i.e.. $T_b = T_s$), the signaling rate for binary system is

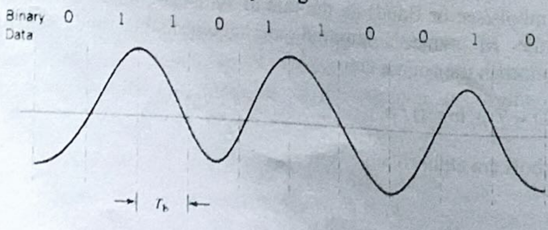
$$R_b = R_s \log_2 2 = R_s \tag{4.81}$$

And the bandwidth is also $R_s/2$. It means M-ary signaling can transmit data $\log_2 M$ times faster than binary system under similar conditions. The price paid for higher speed (or subsequently less bandwidth compared to binary) in M-ary system is the power required to transmit M-ary signal. Thus there exists a trade-off between power and bandwidth.

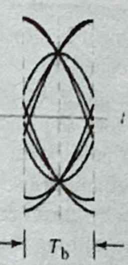
M-ary is more complex since it requires (M-1) comparators at the receiving end for threshold detection.

4.6 Eye Diagram

Eye diagram is a practical way to visualize and evaluate the ISI and its effect on PCM. Eye pattern is obtained on cathode ray oscillator (CRO) by applying received signal to vertical input and sawtooth wave at the transmission symbol rate (i.e., $R=1/T_b$) to horizontal input. This resulting oscilloscope display is called an eye pattern because of its resemblance to the human eye for binary waves as shown in Fig.4.15. The interior region of the eye pattern is called the eye opening.



(a)



(b)

Fig.4.15: (a) Distorted binary wave. (b) Eye pattern

An eye pattern provides a great deal of information about the performance of the pertinent system, as shown in Fig.4.16.

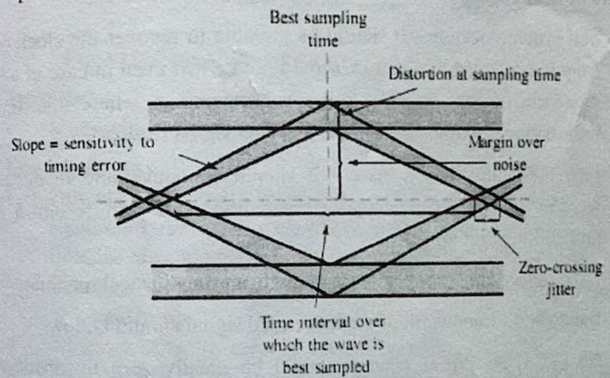


Fig.4.16: Interpretation of eye pattern.

The information provided by the eye pattern is

1. The width of the eye opening defines the time interval over which the received wave can be sampled without error from intersymbol interference. It is apparent that the preferred time for sampling is the instant of time at which the eye is open widest.
2. The sensitivity of the system to timing error is determined by the rate of closure of the eye as the sampling time is varied.
3. The height of the eye opening, at a specified sampling time, defines the margin over noise.

When the effect of intersymbol interference is severe, traces from the upper portion of the eye pattern cross traces from the lower portion, with the result that the eye is completely closed. In such a situation, it is impossible to avoid errors due to the combined presence of intersymbol interference and noise in the system, and a solution has to be found to correct for them.

4.7 Line Coding

Binary 1's and 0's, in PCM signaling, may be represented in various serial-bit electrical signaling formats called line codes. The simplest is to represent '1' by a square pulse of A volt and '0' by 0 volt. But the simplest way may not always be good enough. A long sequence of '0' may appear as a loss of transmission as nothing except noise level is transmitted. In order to take care of these and many other requirements, the symbols are transformed in to various different wave shapes by the process named line coding. Different wave shapes result in different spectrum, suiting different needs. Hence in a way, line codes are also spectrum shaping codes. Thus line coding is mapping of binary information sequence into the electrical signal that enters the channel.

Necessary properties of line code

Types of Line coding

1. Self synchronization - It should be possible to recover the clock pulse from the received data. Clock should not be lost even in case of a long sequence of '0' or '1'. Self-synchronization, where the timing information is extracted from the received signal itself
2. Low probability of bit error - It should be convenient to design a receiver, which receives the specific line code and results a low probability of bit error.
3. PSD - Spectrum of the line code should suit the physical medium.
4. Bandwidth - Bandwidth of the line coded signal should be low.
5. No DC - DC power content should be ideally zero to enable AC coupling.
6. Low frequency power - Power at very low frequency should be as low as possible.
7. To suit channel coding - Line code should be such that subsequent coding for error detection/correction is easy.
8. Power efficiency - Required transmission power should be small.
9. Transparency –a line code is transparent if any bit pattern does not affect the accuracy of the timing. A transmitted signal would not be transparent if there are a long series of 0's which would cause an error in the timing information.

Unipolar Signaling

Unipolar signaling (also called on-off keying, OOK) is the type of line coding in which one binary symbol (representing a 0 for example) is

represented by the absence of pulse (i.e., a SPACE or OFF) and the other binary symbol (denoting a 1) is represented by the presence of a pulse (i.e., a MARK or ON). There are two common variations of unipolar signaling

i. **Unipolar Non-Return to Zero (NRZ) Signaling**

In this line code, symbol 1 is represented by transmitting a pulse of amplitude A for the duration of the symbol, and symbol 0 is represented by switching off the pulse, as in Fig.4.17. The unipolar NRZ line code is also referred to as ON-OFF signaling. Each ON pulse is equal to the duration of T_b of the symbol slot.

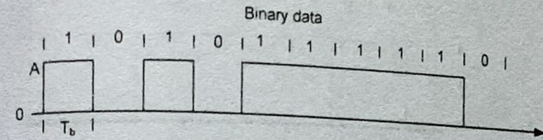


Fig.4.17: Unipolar Non-Return to Zero (NRZ) signaling.

Advantages

- Simplicity in implementation.

Disadvantages

- Presence of DC level (indicated by spectral line at 0 Hz in Fig.4.18)
- Requires large bandwidth (the first null bandwidth is R_b, i.e., 2 time the absolute minimum bandwidth of $R_b/2$), $Rb=1/T_b$.
- Contains large amount of low frequency components. Causes "Signal Droop".
- Does not have any error correction capability.
- Does not possess any clocking component for ease of synchronization.
- Is not transparent long string of zeros causes loss of synchronization.

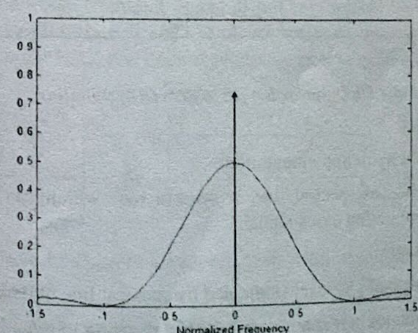


Fig.4.18: PSD of Unipolar NRZ.

When Unipolar NRZ signals are transmitted over links with either transformer or capacitor coupled (AC) repeaters, the DC level is removed converting them into a polar format.

The continuous part of the PSD is also non-zero at 0 Hz (i.e. contains low frequency components) as shown in Fig.4.18. This means that AC coupling will result in distortion of the transmitted pulse shapes. AC coupled transmission lines typically behave like high-pass RC filters and the distortion takes the form of an exponential decay of the signal amplitude after each transition. This effect is referred to as "Signal Droop" and is illustrated in Fig.4.19.

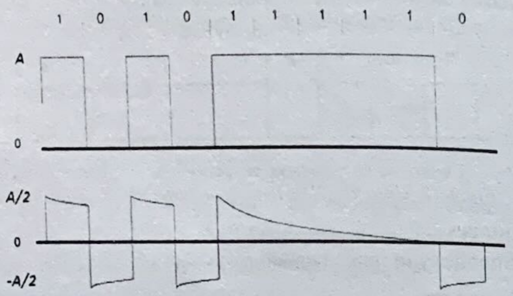


Fig.4.19: Distortion (Signal Droop) due to AC coupling of unipolar NRZ signal

ii. Unipolar Return to Zero (RZ) Signaling

In this line code, symbol 1 is represented by a rectangular pulse of amplitude A for duration half of symbol width (i.e.,$T_b/2$) and symbol 0 is represented by transmitting no pulse, as illustrated in Fig.4.20.

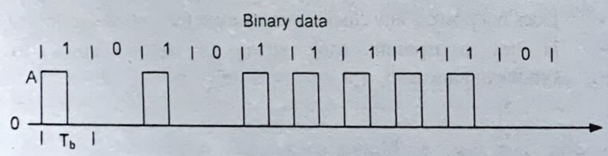


Fig.4.20: Unipolar Return to Zero (RZ) signaling.

Advantages

- Simplicity in implementation.
- Presence of spectral line at symbol rate which can be used as symbol timing clock signal.

Disadvantages

- Presence of DC level (indicated by spectral line at 0 Hz in Fig.)
- Continuous part is non-zero at 0 Hz. Causing "Signal Droop".
- Does not have any error correction capability.

- Occupies twice as much bandwidth as Unipolar NRZ (simply because the pulse width is reduced to half).
- Is not Transparent. Long string of zeros causes loss of synchronization.

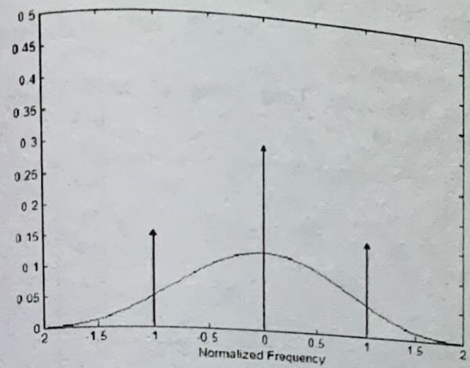


Fig.4.21: psd of Unipolar RZ.

Positive property of the unipolar RZ line code is the presence of delta functions at $f = 0, \pm 1/T_b$ in the power spectrum of the transmitted signal; the delta functions can be used for bit-timing recovery at the receiver. However, its disadvantage is that it requires 3 dB more power than polar RZ signaling for the same probability of symbol error.

Conclusion is that neither variety of unipolar signals is suitable for transmission over AC coupled lines.

Polar Signaling

In polar signaling a binary "1" is represented by a positive pulse of fixed height and a binary "0" by the opposite (negative) pulse.

i. Polar Non-Return to Zero (NRZ) Signaling

In this line code, symbols 1 and 0 are represented by transmitting pulses of amplitudes +A and –A respectively, as illustrated in Figure 4.22. The polar NRZ line code is relatively easy to generate.

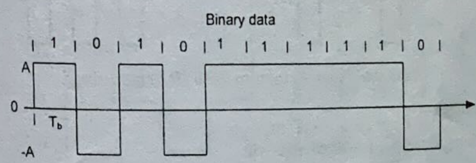


Fig.4.22: Polar Non Return to Zero (NRZ) signaling.

Advantages

- Simplicity in implementation.
- No DC component.

Disadvantages

- Continuous part is non-zero near 0 Hz. Causing "Signal Droop".
- The required channel bandwidth is R_b.
- Does not have any error correction capability.
- Does not possess any clocking component for ease of synchronization..
- Is not Transparent. Long string of zeros causes loss of synchronization.

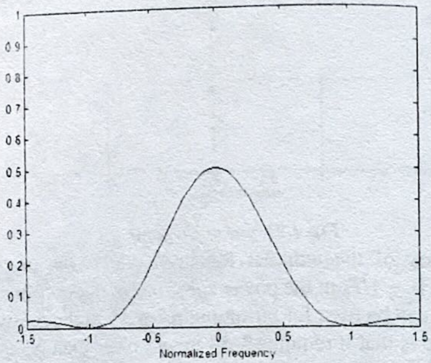


Fig.4.23: psd of Polar NRZ

ii. Polar Return to Zero (RZ) Signaling

In this line code, symbol 1 is represented by a rectangular pulse of amplitude A for duration half of symbol width (i.e.,$T_b/2$) and symbol 0 is represented by transmitting a rectangular pulse of amplitude -A pulse for duration half of symbol width (i.e.,$T_b/2$), as illustrated in Fig.4.24.

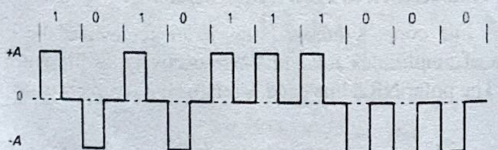


Fig.4.24: Polar Return to Zero (RZ) signaling.

Advantages

- Simplicity in implementation.
- No DC component.

Disadvantages

- Continuous part is non-zero near 0 Hz. Causing "Signal Droop".
- Does not have any error correction capability.

- Does not posses any clocking component for ease of synchronization. However, clock can be extracted by rectifying the received signal.
- Occupies twice as much bandwidth as Polar NRZ.

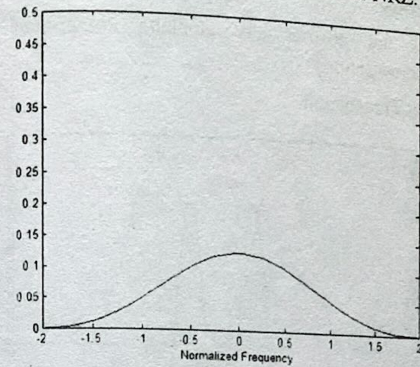


Fig.4.25: psd of Polar RZ

Bipolar RZ Signaling

Bipolar signaling is also called Alternate Mask Inversion (AMI) which uses three amplitude levels (A,0,-A). '0s' as in unipolar are represented by the absence of a pulse and '1s' (MARK or ON) are represented by alternating voltage levels of +A and –A. Alternating the MARK level voltage ensures that the bipolar spectrum has a null at DC and the signal droop on AC coupled line is avoided. The alternating mark voltage also gives bipolar signaling a single error detection capability. Like the Unipolar and Polar cases, Bipolar also has NRZ and RZ variations.

i. Bipolar/AMI NRZ

In this line code, '0s' are represented by the absence of a pulse and '1s' (MARK or ON) are represented by alternating voltage levels of +A and –A with each pulse having symbol width equal to T_b as shown in Fig.4.26.

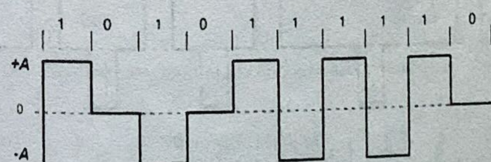


Fig.4.26: AMI NRZ

Advantages

- No DC component.
- Occupies less bandwidth than unipolar and polar NRZ schemes.

- Does not suffer from signal droop (suitable for transmission over AC coupled lines).
- Possesses single error detection capability.

Disadvantages

- Does not possess any clocking component for ease of synchronisation.
- Is not Transparent.

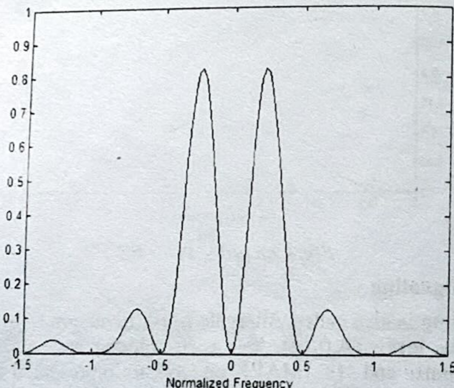


Fig.4.27: psd of Bipolar/AMI NRZ

ii. Bipolar/AMI RZ

In this line code, '0s' are represented by the absence of a pulse and '1s' (MARK or ON) are represented by alternating voltage levels of +A and –A with each pulse having half symbol width (i.e., $T_b/2$) as shown in Fig.4.28.

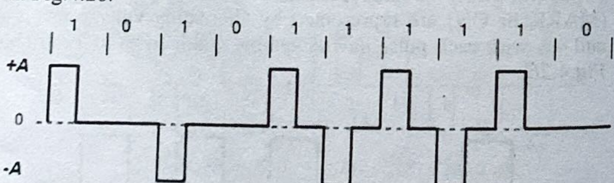


Fig.4.28: Bipolar RZ

Advantages

- No DC component.
- Occupies less bandwidth than unipolar and polar RZ schemes.
- Does not suffer from signal droop (suitable for transmission over AC coupled lines).

- Possesses single error detection capability.
- Clock can be extracted by rectifying (a copy of) the received signal.

Disadvantages

- Is not Transparent.

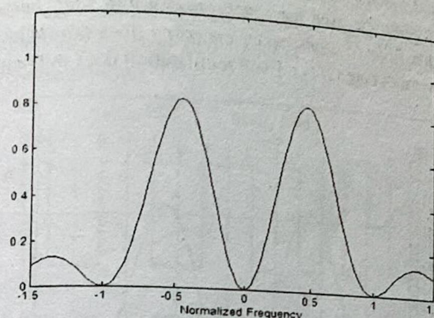


Fig.2.29: psd of Bipolar RZ

Split-Phase (Manchester Code) Signaling

In this method of signaling, illustrated in Fig.2.30, the voltage remains at one level during the first half and moves to the other level during the second half.

A '1' is represented by a positive pulse of half bit duration followed by a negative pulse of same magnitude and half bit duration. The '0' is represented by negative pulse of half bit duration followed by a positive pulse of same magnitude and half bit duration.

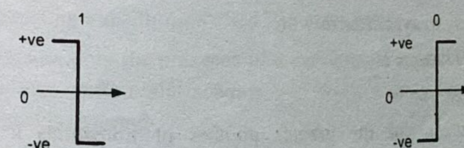


Fig.2.30: Representation of one and zero using Manchester code.

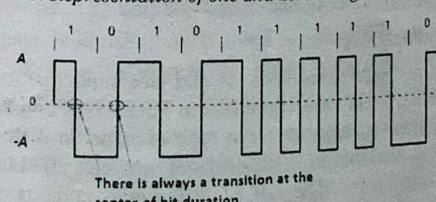


Fig.2.31: Split-Phase (Manchester Code) Signaling

The transition at the centre of every bit interval is used for synchronization at the receiver. Manchester encoding is called self-synchronizing. Synchronization at the receiving end can be achieved by locking on to the transitions, which indicate the middle of the bits.

It is worth highlighting that the traditional synchronization technique used for unipolar, polar and bipolar schemes, which employs a narrow BPF to extract the clock signal cannot be used for synchronization in Manchester encoding. This is because the PSD of Manchester encoding does not include a spectral line/ impulse at symbol rate ($1/T_b$). Even rectification does not help.

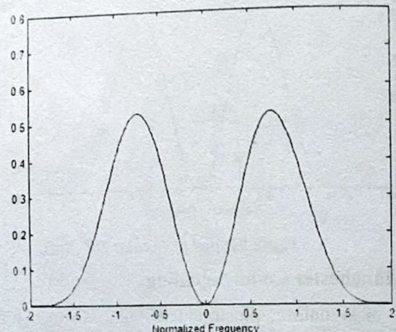


Fig.2.32: pdf of Manchester

Advantages

- No DC component.
- Does not suffer from signal droop (suitable for transmission over AC coupled lines).
- Easy to synchronize with.
- Is Transparent.

Disadvantages

- Because of the greater number of transitions it occupies a significantly large bandwidth.
- Does not have error detection capability.

Differential coding

Change in polarity i.e., +A becomes –A and vice versa due to error in some system results in error in subsequent bits in Polar NRZ coding. Differential line coding provides robustness to this type of error. In differential coding, '1' is mapped as a transition in signal level whereas, '0' is mapped as no transition in signal level. The pdf of differential code is same as NRZ scheme.

Differential Manchester coding

In differential Manchester encoding shown in Fig.2.33, the inversion at the middle of the bit interval is used for synchronization but the presence or absence of an additional transition at the beginning of the interval is used to identify the bit.

A transition means binary 0 and no transition means binary 1. Differential Manchester encoding requires two signal changes to represent binary 0 but only one to represent binary 1.

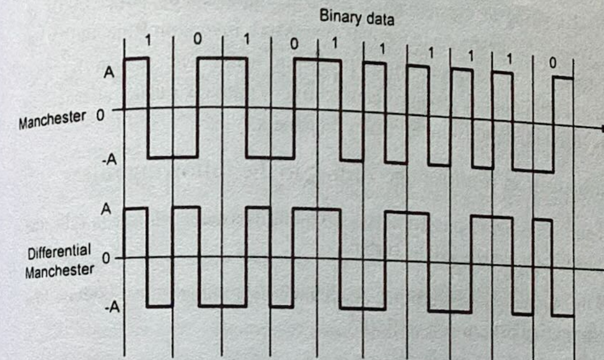


Fig.2.33: Manchester and Differential Manchester.

Thus, differential encoder uses the presence or absence of transitions to indicate logical value. This gives it several advantages over Manchester encoding

- Detecting transitions is often less error-prone
- Because only the presence of a transition is important rather than polarity, differential scheme will work exactly the same if the signal is inverted.

High Density Bipolar (HDB) Signaling

In case of unipolar RZ, unipolar NRZ, bipolar NRZ or AMI signal, the transmitted signal is equal to zero when a binary 0 is to be transmitted. Transmission of long sequence of binary 0 can cause problem in synchronization at the receiver. The solution is to add pulses when long strings of 0's exceeding a number n are being transmitted. This type of coding is called High Density Bipolar (HDB) coding. It is denoted by HDBn where n=1, 2, 3,.... The most widely used HDB format is with n=3 i.e., HDB3 as shown in Fig.2.34.

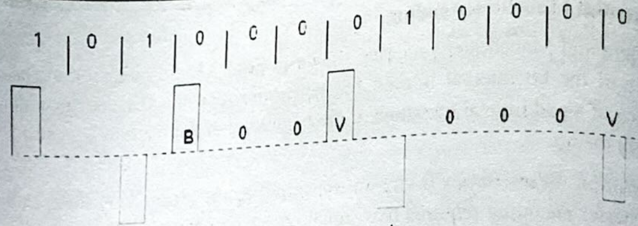


Fig.2.34: HDB3 code

In HDB-3 a string of 4 consecutive zeros are replaced by either 000V or B00V. Where, 'B' conforms to the Alternate Mark Inversion Rule and 'V' is a violation of the Alternate Mark Inversion Rule. The reason for two different substitutions is to make consecutive Violation pulses alternate in polarity to avoid introduction of a DC component.

The substitution is chosen according to the following rules:

1. If the number of nonzero pulses after the last substitution is odd, the substitution pattern will be 000V.
2. If the number of nonzero pulses after the last substitution is even, the substitution pattern will be B00V.

Advantages

- No DC component.
- Occupies less bandwidth than unipolar and polar RZ schemes.
- Does not suffer from signal droop (suitable for transmission over AC coupled lines).
- Possesses single error detection capability.
- Clock can be extracted by rectifying (a copy of) the received signal.
- Is Transparent.

Disadvantage

- As HDB3 is a bipolar signal, the receiver has to distinguish between three levels (+A, -A, and 0), instead of just two levels as in other signaling formats previously discussed.

These characteristic make this scheme ideal for use in Wide Area Networks

Previous Exam Questions

1. Define Inter Symbol Interference (ISI)? State Nyquist Pulse Shaping criteria for Zero ISI, Discuss Raised Cosine Pulse Shaping method of ISI reduction.
2. What is ISI? Explain two practical methods of minimizing ISI.
3. Derive the expression of the signal at the input of the receiver of a baseband DCS and based on that expression, define ISI. State and explain Nyquist Pulse Shaping criteria for zero ISI.
4. What do you mean by duo-binary encoding? What is its importance; Explain duo-binary encoding with example.
5. Define information and entropy. Relate message, entropy and information.

   Calculate the upper limit of the channel capacity as the bandwidth of the channel (B) tends to infinity.
6. What is the basic condition to be fulfilled if the parameter "i" is to represent the amount of information contained in an event (message) "mk".
7. Differentiate between message and information. Derive the expression for evaluating the entropy of source that emits M non-equiprobable symbols in statistically independent manner.
8. Differentiate between entropy and information. Define entropy and derive the expression for evaluating the entropy of source emitting symbols in statistically independent manner.
9. A discrete source emits one of six possible symbols per microsecond in statistically independent manner. The symbol probabilities are 1/4, 1/4. 1/4, 1/8, 1/16 and 1/16 respectively. Calculate: (a) Symbol Rate (b) Entropy (c) information Rate.
10. A discrete source emits one of 6 symbols per 10 μs in statistically independent manner. The symbol probabilities are 1/4,. 1/4, 1/4,.1/8, 1/16 and 1/16 respectively. Calculate (a)symbol Rate (b) Entropy (c) Information Rate.
11. A signal of bandwidth 4.5 kHz is sampled at the double rate given by Nyquist Rate and the signal is quantized in 8 levels with probabilities of occurrence of the level are 0.1, 0.15, 0.15, 0.05, 0.2, 0.05, 0.18, 0.12. Calculate (a) minimum no. of bits per sample (b) Information Rate.

12. A analog signal band-limited to 10 kHz is sampled at Nyquist Rate and quantized in 8 levels with probabilities of 1/4. 1/5. 1/5, 1/10,.1/10, 1/20, 1/20 and 1/20 respectively. Calculate: (a) Entropy (b) Information Rate.

13. State and discuss Shannon's Channel Capacity Theory, Based on this theory discuss how the bandwidth and SNR can be trade-off for given channel capacity. Discuss the implications and theoretical limits of this theory.

14. Represent binary sequence 1011001010 in Polar NRZ, Polar RZ, Manchester and AMI codes.

15. What is ISI? State Nyquist pulse shaping criteria for zero ISI. Explain duo-binary encoding with example.

16. Write short Notes on:

    Performance evaluation of DCS using Eye Diagram.

✡✡✡

Chapter

5 BANDPASS (MODUL COMMUNICATION S

Introduction

In baseband digital system, the information bearing without any changes in spectral components of baseband signal power is localized at low frequen impractically large size antennas to radiate the low- the signal efficiently. Hence, for wireless comm spectrum must be shifted to higher frequency modulation techniques.

In digital communication, the message signal (also cal consists of digital data .If we change amplitude, fre frequency sinusoidal carrier signal according to the m the modulation is called Amplitude Shift Keying (A Keying (FSK) and Phase Shift Keying (PSK) respecti ASK, FSK and PSK is shown in Fig. 5.1 respectively.

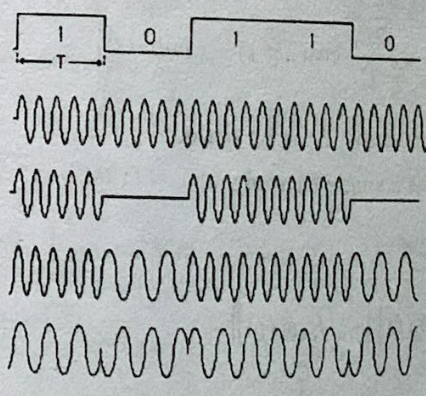


Fig. 5.1: Waveform of (a) Binary signal, (b) Carrier signal (e) PSK modulation

In M-ary signaling, the modulator produces one of (i.e., equal to 2^n) distinct signals in response to n b time. Binary modulation is a special case of M-ary equals to 2 (i.e., 2^1).

At receiver, demodulation can be performed by eit coherent detection methods as in case of analog modul