Digital Communication

Asst. Lec. Ghufran Saady

Digital Communication Systems

The term digital communication covers a broad area of communications techniques, including digital transmission and digital radio.

Digital transmission, is the transmitted of digital pulses between two or more points in a communication system. Digital radio, is the transmitted of digital modulated analog carriers between two or more points in a communication system.

- **Why Digital**

There are many reasons

- The primary advantage is the ease with which digital signals, compared to analog signal, are regenerative.

The shape of the waveform is affected by two mechanisms:

(1) As all the transmission lines and circuits have some nonideal transfer function, there is a distorting effect on the ideal pulse.

(2) Unwanted electrical noise or other interference further distorts the pulse waveform.

Both of these mechanisms cause the pulse shape to degrade as a function of distance.

During the time that the transmitted pulse can still be reliably identified, the pulse is thus regenerated. The

- Digital circuits are less subject to distortion and interference than analog circuits.
- Digital circuits are more reliable and can be produced at lower cost than analog circuits. Also, digital hardware lends itself to more flexible implementation than analog hardware.
- Digital techniques lend themselves naturally to signal processing functions that protect against interference and jamming.
- Much data communication is computer to computer, or digital instrument or terminal to computer. Such digital

Communication System Models

Generally, there are two types for communication system models, ***base-band*** model and ***pass-band*** model.

In base-band model, the spectrum of signal from zero to some frequency (i.e. carrier frequency=0). For transmission of base-band signal by a digital communication system, the information is formatted so that it is represented by digital symbols. Then, pulse waveforms are assigned that represented these symbols. This step referred to as ***pulse modulation*** or ***base-band modulation***. These waveforms can be transmitted over a cable. Base-band signal also called ***low-pass signal***.

In pass-band (or band-pass) signal, the signal has a spectral magnitude that is nonzero for frequency in some band concentrate about a frequency $f = \pm f_c$ and negligible elsewhere, where f_c is the carrier frequency need to be much greater than zero. For radio

Multiplexing

Multiplexing is the transmission of information (either voice or data) from more than one source to more than one destination on the same transmission medium.

Two most common methods are used, ***frequency division multiplexing (FDM) and time division multiplexing (TDM).***

 FDM

In FDM multiple sources that originally occupied the same frequency spectrum are each converted to a different frequency band and transmitted simultaneously over a single transmission medium. FDM is an analog multiplexing

If two input signals to a mixer are sinusoids with frequencies fA and fB, the mixing or multiplication will yield new sum and difference frequencies at fA+B and fA-B. Equation below desc

$$\cos A \cos B = \frac{1}{2}[\cos(A+B) + \cos(A-B)]$$

A simple FDM example with three translated voice channels is shown

in figure below.

TDM

With TDM system, transmission from multiple sources occurs on the same transmission medium but not at the same time. Transmission from various sources is interleaved in time domain. Figure below shows the time-frequency plan in TDM system, the

same communication resources is shared by assigning each of N symbols or users the full spectral occupancy of the system for a short duration of time called ***time slot***. The unused time regions between slot assignments, called

The multiplexing operation consists of providing each source with an opportunity to occupy one or more slots. The demultiplexing operation consists of desloting the information and delivering the data to the intended sink.

The communication switches (S1…SM) have synchronized so that the massage corresponding to signal(1), for example, appears on the channel (1) output, and so on.

Time is segmented in to intervals called ***frames***. Each frame is further partitioned in to assignable user time slots.

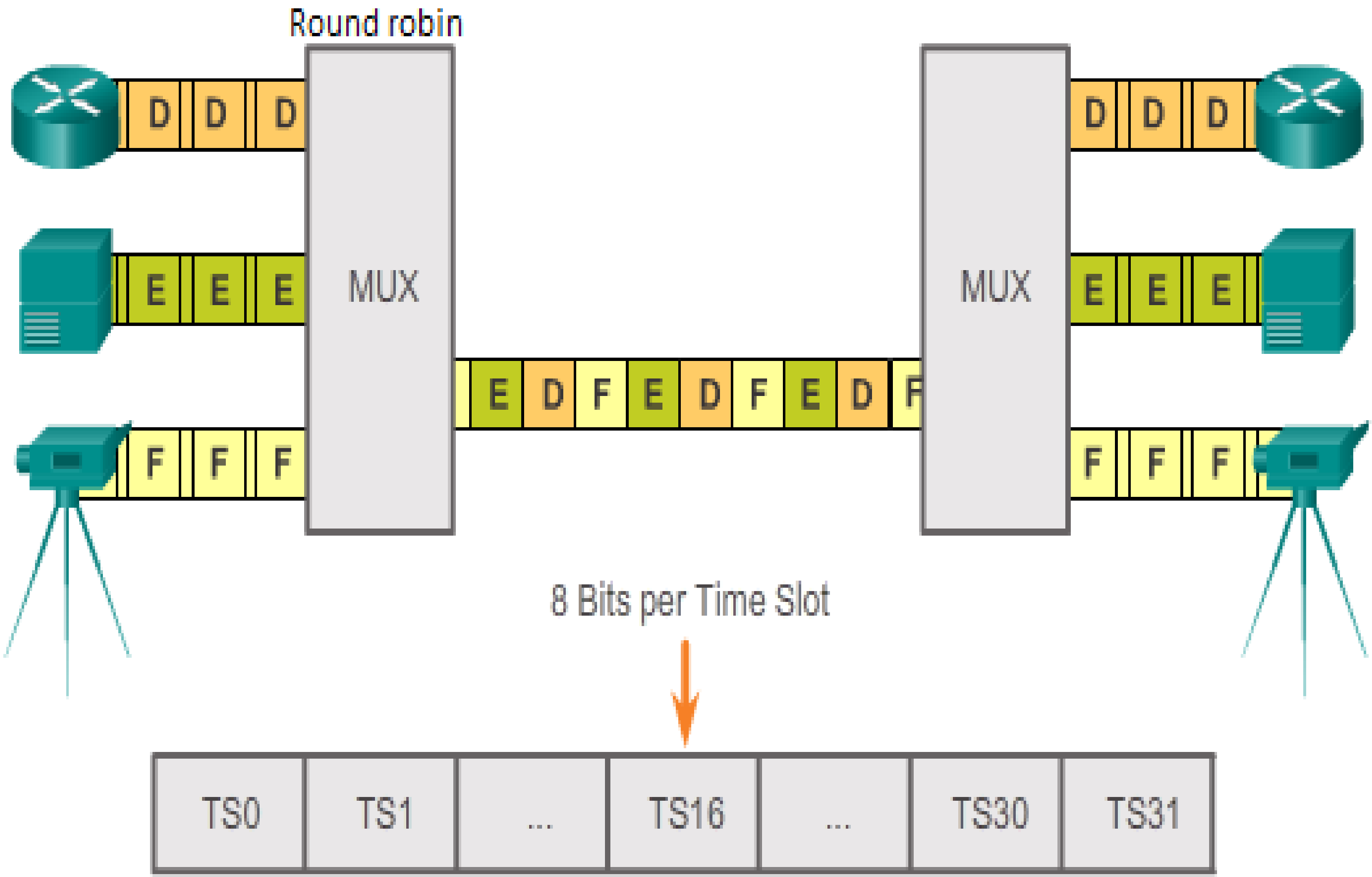
The simplest TDM scheme called ***fixed-assignment TDM***. In fixed assignment TDM scheme, all of the slot has no data to sent during a particular frame, that slot is wasted.

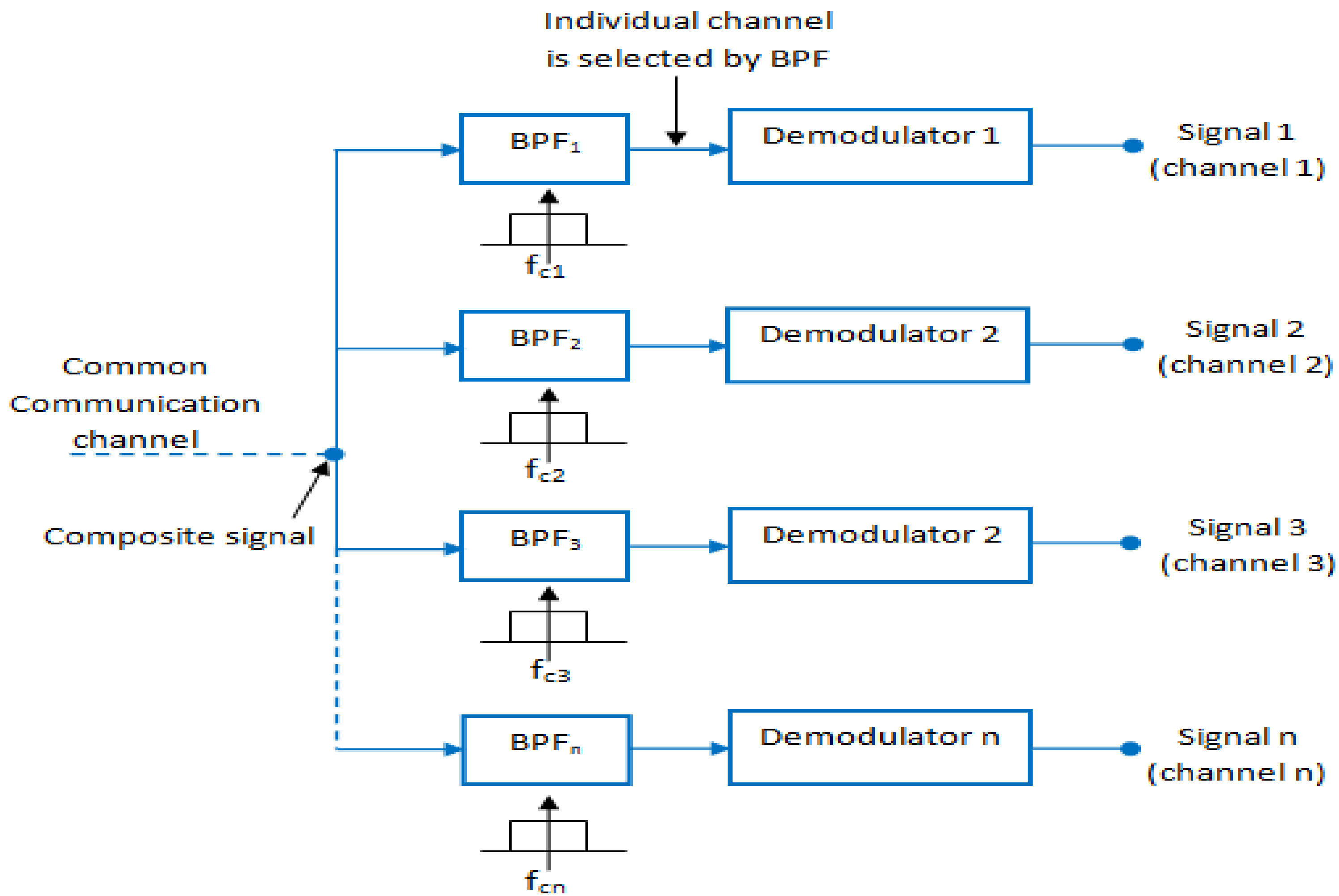
Another more efficient scheme, involving the ***dynamic assignment*** of the slots rather than fixed assignment.

Time Division Multiplexing (TDM)

- The best utilization if everyone always has something to send.
- Wastes time if this is not the case.
- Slots can be unevenly assigned.

Time-Division Multiplexing



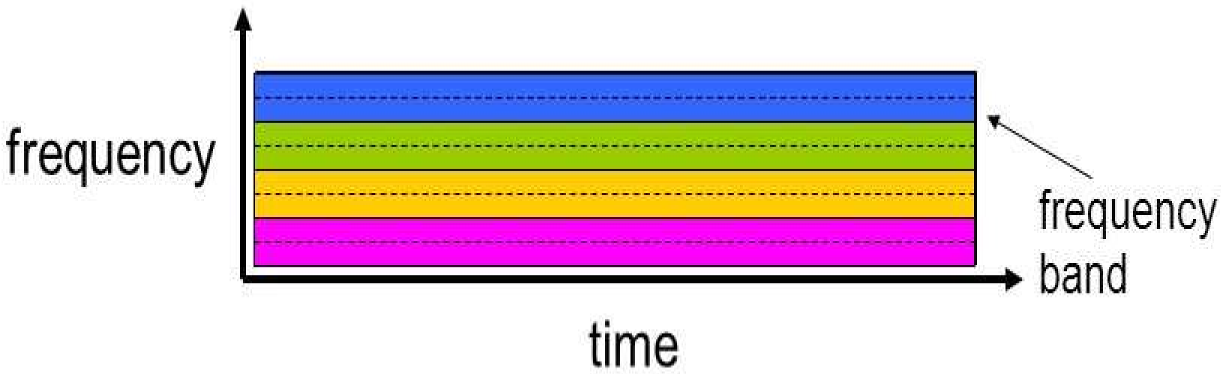

Individual channel
is selected by BPF
BPF1
Demodulator 1
Signal 1
(channel 1)
fc1
BPF2
Demodulator 2
Signal 2
(channel 2)
fc2
Common
Communication
channel
Composite signal
BPF3
Demodulator 2
Signal 3
(channel 3)
fc3
BPFn
Demodulator n
Signal n
(channel n)
fcn

Circuit switching: FDM and TDM

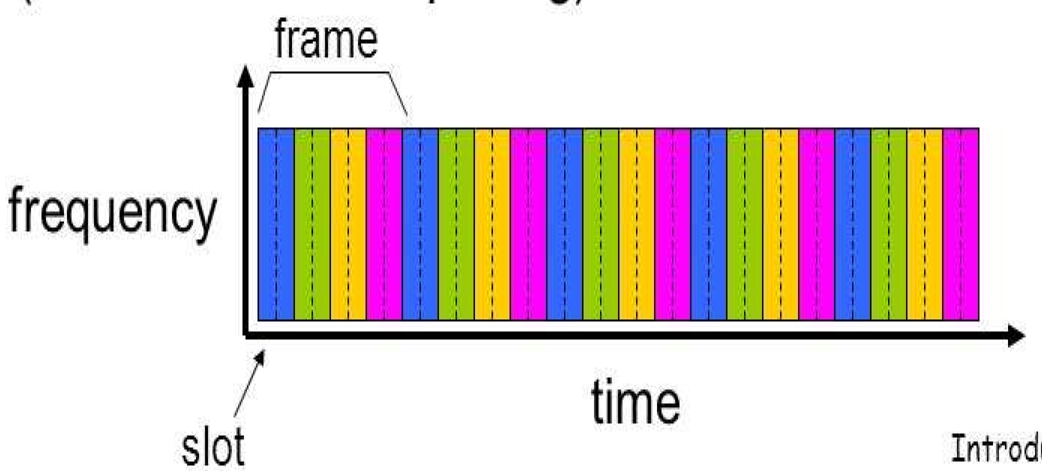
Example:

4 users

FDM (Frequency division multiplexing)



TDM (Time division multiplexing)



Sampling Theorem

The link between an analog waveform and its sampled version is provided by what is known as the sampling process.

A band limited signal having no spectral components above (fm Hz) can be determined uniquely by values sampled at

$$T_s \leq \frac{1}{2 f_m}$$

Stated another way, the upper limit on T_s can be expressed in terms of the ***sampling rate***, denoted $f_s = \frac{1}{T_s}$

The restriction, stated in terms of sampling rate, is known as the Nyquist criterion. The statement is

$$f_s \geq 2 f_m$$

The sampling rate ($f_s = 2 f_m$) also called ***Nyquist rate***.

The Nyquist criterion is a theoretically sufficient condition to allow an analog signal to be reconstructed completely from a set of uniformly spaced discrete time samples.

❖ Impulse Sampling

Assume an analog waveform x(t), as shown in Fig. (a), with a Fourier transform, X(f), which is zero outside the interval ($-f_m<f<f_m$), as shown in Fig. (b). The sampling of x(t) can be viewed as the product of x(t) with a train of unit impulses functions, $x_s(t)$, shown in Fig. (c), and defined as follows:

$$x_\delta(t)=\sum_{n=-\infty}^{\infty}\delta(t-nT_s)$$

Let us choose $T_s=\frac{1}{2f_m}$, so that Nyquist rate is just satisfied.

Using shifting property of the impulse function the $x_s(t)$, shown in Fig. (e), can be given by

$$x_s(t) = x(t)x_\delta(t) = \sum_{n=-\infty}^{\infty} x(t)\delta(t - nT_s)$$

$$= \sum_{n=-\infty}^{\infty} x(nT_s)\delta(t - nT_s)$$

Using frequency convolution property of Fourier transform, the time product $x(t)x_\delta(t)$ transforms to the frequency domain convolution $X(f) \otimes X_\delta(f)$, where $X_\delta(f)$ is the Fourier transform of $x_\delta(t)$ and given by

$$X_\delta(f) = \frac{1}{T_s} \sum_{n=-\infty}^{\infty} \delta(f - nf_s)$$

The convolution with an impulse function simply shifts the original function, as follows:

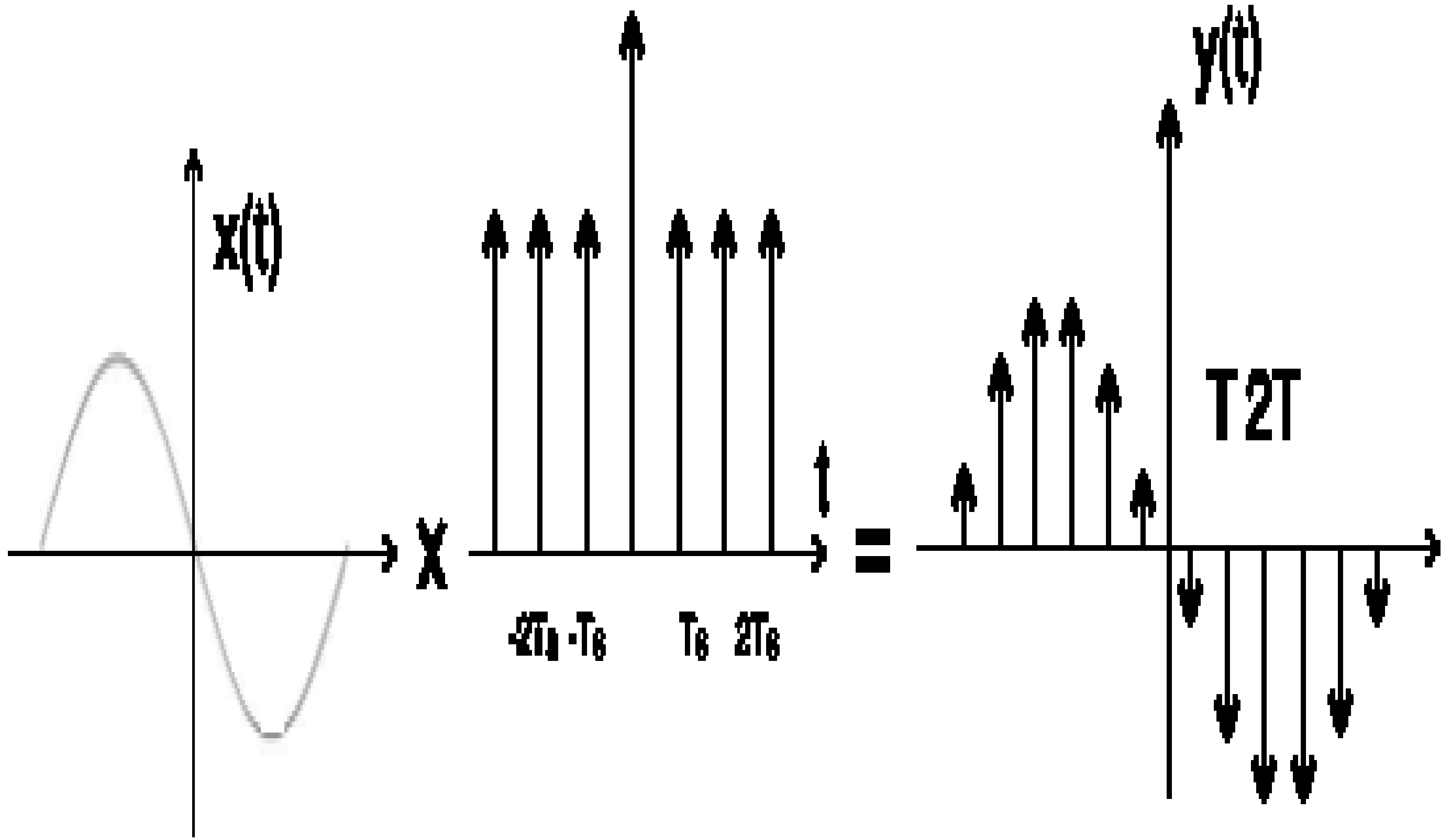
$$X(f) \otimes \delta(f - nf_s) = X(f - nf_s)$$

The Fourier transform of the sampled waveform, $X_s(f)$, can be given by:

$$X_s(f) = X(f) \otimes X_\delta(f) = X(f) \otimes [\frac{1}{T_s}\sum_{n=-\infty}^{\infty}\delta(f - nf_s)]$$

$$= \frac{1}{T_s}\sum_{n=-\infty}^{\infty} X(f - nf_s)$$

Figure below shows the sampling theorem using the frequency convolution property of the Fourier transform (Impulse sampling).


x(t)
X
-2Ts -Ts
Ts 2Ts
t
=
y(t)
T2T

❖ Natural Sampling

In this way the band limited analog signal x(t), shown in Fig. (a_1), is multiplied by the pulse train or switching waveform $x_p(t)$, shown in Fig. (c_1). Each pulse in $x_p(t)$ has width T and amplitude 1/T.

The resulting sampled data sequence, $x_s(t)$, is shown in Fig. (e_1) and is expressed as

$$x_s(t) = x(t)x_p(t)$$

The periodic pulse train, $x_p(t)$, can be expressed as a Fourier series in the form

$$x_p(t) = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi n f_s t}$$

and $$c_n = \frac{1}{T_s}\operatorname{sinc}\left(\frac{nT}{T_s}\right)$$

where $f_s=2f_m$, T is the pulse width, and 1/T is the pulse amplitude.

The envelope of magnitude spectrum of the pulse train, seen as a dashed line in Fig. (d_1), has characteristic ***sinc*** shape.

$$\therefore x_s(t) = x(t)\sum_{n=-\infty}^{\infty} c_n e^{j2\pi n f_s t}$$

The Fourier transform of $x_s(t)$ is found as follows

$$X_s(f) = F[x(t)\sum_{n=-\infty}^{\infty} c_n e^{j2\pi n f_s t}]$$

For linear system the operation of summation and Fourier transformation can be interchanged. Therefore,

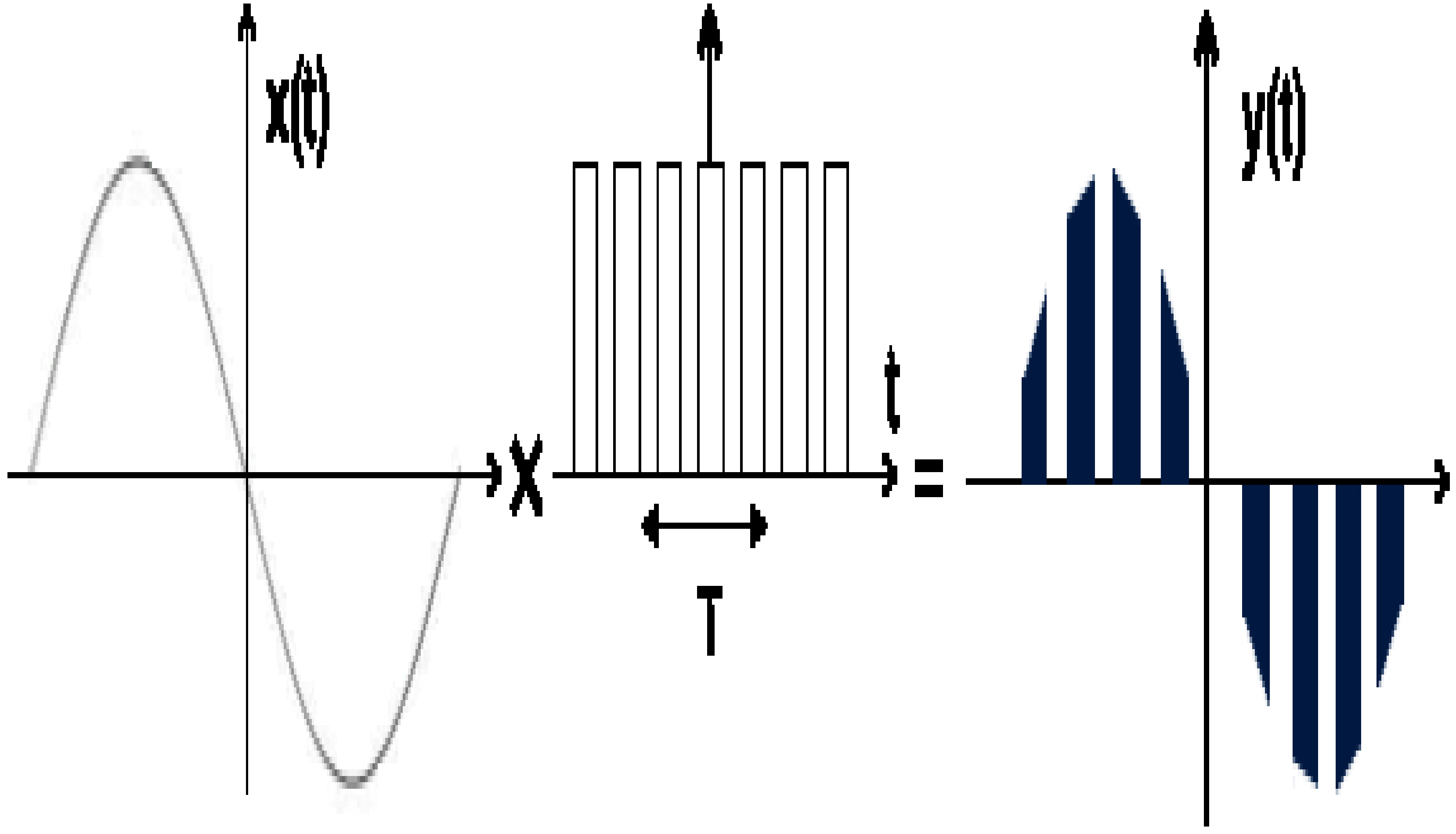
$$X_s(f) = \sum_{n=-\infty}^{\infty} c_n F[x(t)e^{j2\pi n f_s t}]$$

Using frequency translation property of Fourier transform,

$$X_s(f) = \sum_{n=-\infty}^{\infty} c_n X(f - nf_s)$$

Note

The sampling here is termed natural sampling, since the top of each pulse in the $x_s(t)$ sequence retains the shape of its corresponding analog segment during the pulse interval.



x(t)
X
t
T
=
y(t)

Pulse Modulation

In pulse modulation some parameter of a pulse train is varied in accordance with the massage signal.

Two families of pulse modulation may be distinguished: ***analog pulse modulation and digital pulse modulation***. In analog pulse modulation, a periodic pulse train is used as the carrier wave, and some characteristics features of each pulse (e.g. Amplitude, Position, and Width) is varied in a continuous manner in accordance with the corresponding sample value of the message signal. Thus in analog pulse modulation, information is transmitted basically in analog form, but the transmission takes place at discrete times.

In digital pulse modulation, on the other hand, the massage signal is represented in a form that is discrete in both time and amplitude; thereby permitting its transmission in digital form as a sequence of coded pulses.

(1) Pulse Amplitude Modulation (PAM)

PAM is the simplest and most basic form of analog pulse modulation. In PAM the amplitude of regularly spaced pulses are varied in proportion to the corresponding sample values of a continuous message signal, the pulses can be of a rectangular form or other appropriate shape.

PAM as defined here is somewhat similar to natural sampling where the message signal is multiplied by a periodic train of rectangular pulses. However, in natural sampling the top of each modulated rectangular pulse varies with the message signal, whereas in PAM it is maintained flat.

The waveform of PAM signal is shown in figure below.

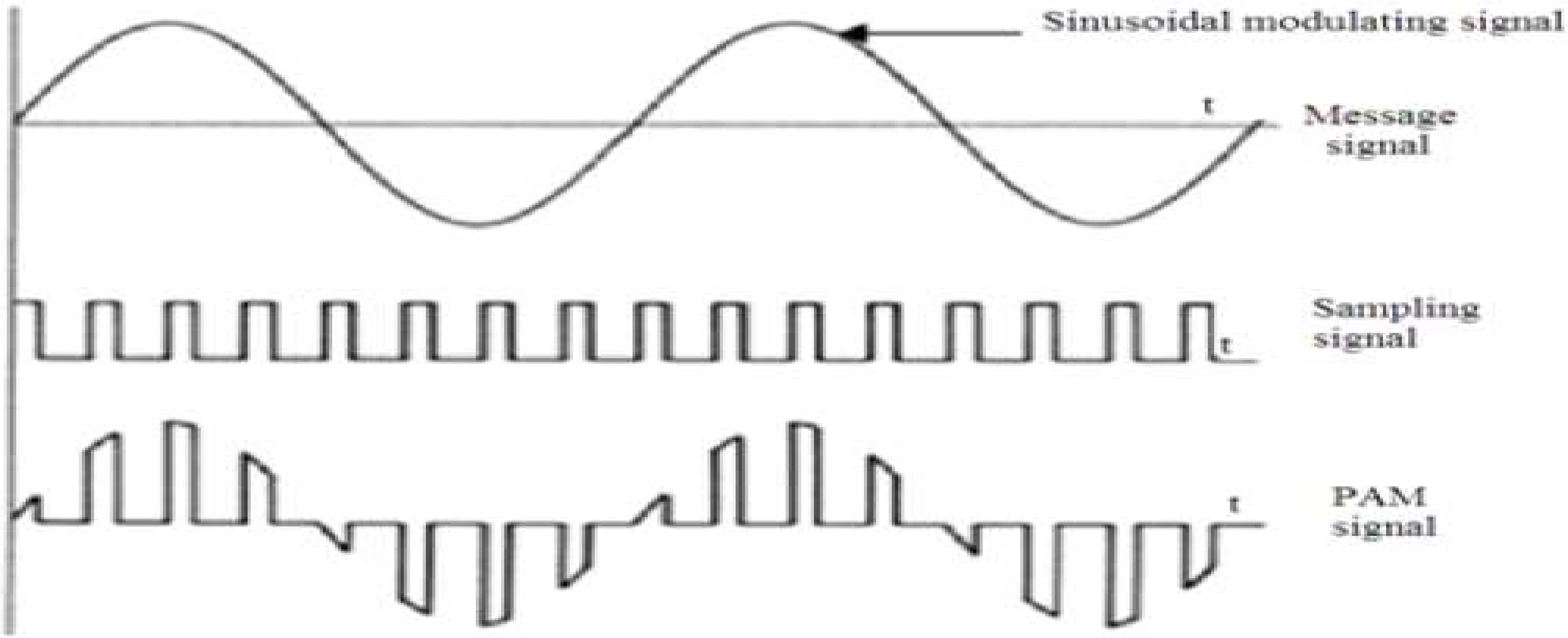


Fig2. Spectrum of PAM signal

There are two operations involved in the generation of the PAM signal:-

(i) Instantaneous sampling of the message signal every Ts second, where the sampling rate $f_s=1/T_s$ is chosen in accordance with the sampling theorem.

(ii) Lengthening the duration of each sample so obtained to some constant value (τ).

➤ *PAM/TDM System*

Suppose we wish to time multiplexed two signals using PAM. Let us assume that both input signal $f_1(t)$ and $f_2(t)$ are low pass, and band limited to 3KHz. The sampling theorem states that each must be sampled at a rate no less than 6KHz. This requires a 12KHz minimum clock rate for the two channel system. Figure below shows the block diagram of PAM/TDM system.

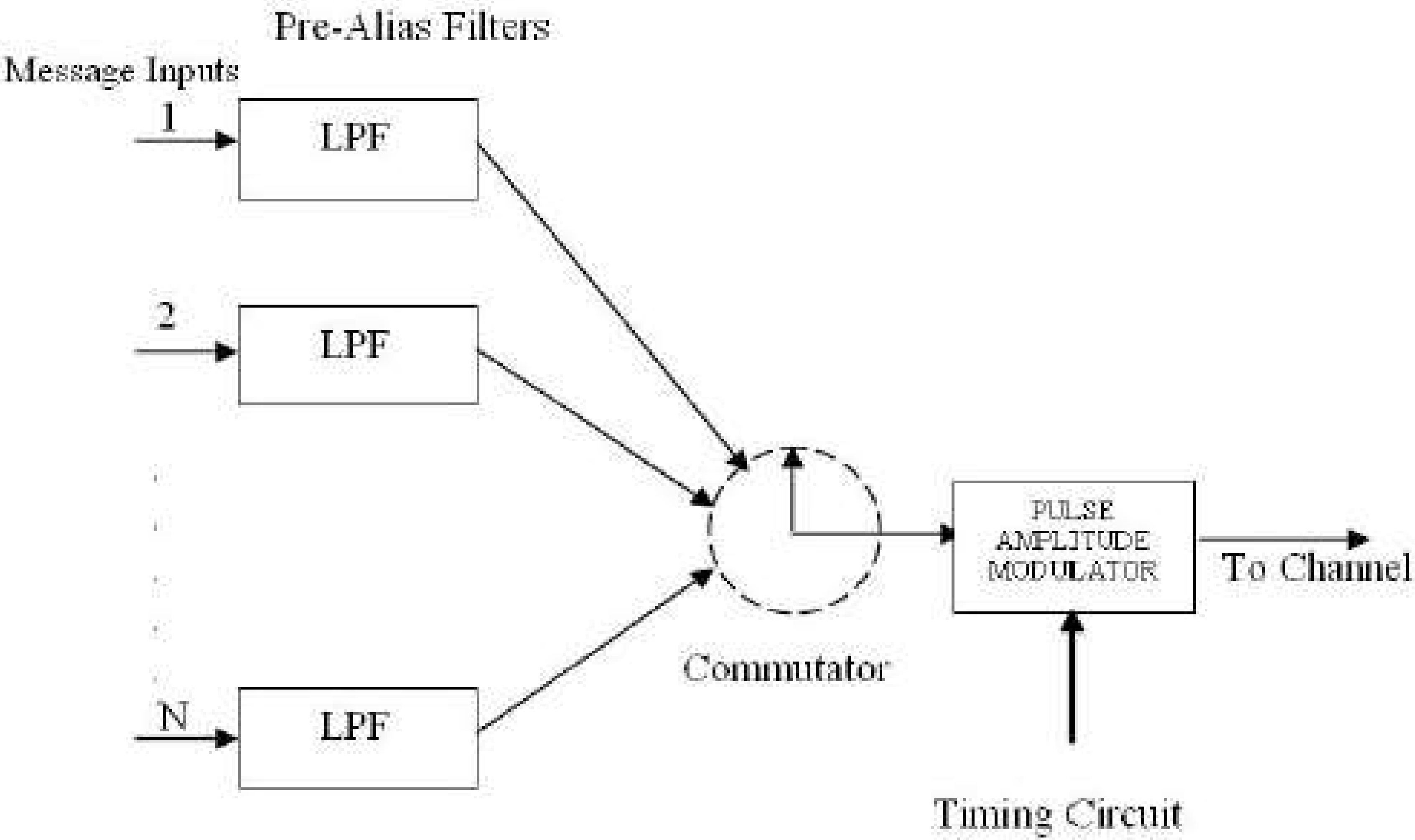


Fig-2.21 : TDM-PAM: Transmitter

The time spacing between adjacent samples in the time multiplex signal waveform (T_x), can be defined as

$$T_x = \frac{T_s}{n}$$

where T_s =sampling rate, and n=number of input signals.

To prevent any irretrievable loss of information in the composite waveform then requires that bandwidth B_x of LPF must satisfy the criterion

$$B_x \geq \frac{1}{2T_x}$$

At the receiver the composite time multiplexed and filtered waveform must be resampled and separated into the appropriate channel. One the pulses are separated, the normal sampling considerations applies and the analog reconstruction of signals can be obtained by LPF. The block diagram of PAM/TDM receiver is shown below.

Example:

Channel 1 of two channels PAM system handles 8KHz signal. Channel 2 handles 10 KHz signals. The two channels are sampled at equal intervals of time using very narrow pulses at the lowest frequency that is theoretical adequate. The sampled signals are time multiplexed and passed through a LPF before transmission.

(1) What is the minimum clock frequency of the PAM system?

(2) What is the minimum cut off frequency of LPF used before transmission that will preserve the amplitude information on the output pulses?

(3) What would be the minimum bandwidth if these channels were frequency multiplexed, using AM technique and SSB technique?

Solution

(1)

$$f_{s1} = 2 * f_{m1}$$

$$f_{s1} = 2 * 8 = 16KHz$$

$$f_{s2} = 2 * f_{m2}$$

$$f_{s2} = 2 * 10 = 20KHz$$

In order to sample channel 2 adequately

$$f_s = f_{s2} = 20KHz$$

$\therefore$ The minimum clock rate $= n * f_s$

$$= 2 * 20 = 40KHz$$

(2)

$$T_s = \frac{1}{f_s} = \frac{1}{20KHz} = 50\,\mu\sec$$

$$\because n = 2$$

$$T_x = \frac{T_s}{n} = \frac{50}{2} = 25\,\mu\sec$$

$$\because B_x \geq \frac{1}{2T_x}$$

$$\therefore B_x = 20KHz$$

(4) For AM

$$\min.BW. = 2(f_{m1} + f_{m2}) = 2(8+10) = 36KHz$$

For SSB

$$\min.BW. = f_{m1} + f_{m2} = 8+10 = 18KHz$$

H.W

Two low pass signals, each band limited 4KHz, are to be time multiplexed into a single channel using PAM. Each signal is impulse sampled at a rate 10KHz. The time multiplexed signal waveform is filtered by an ideal LPF before transmission.

(a) What is minimum clock frequency of the system?

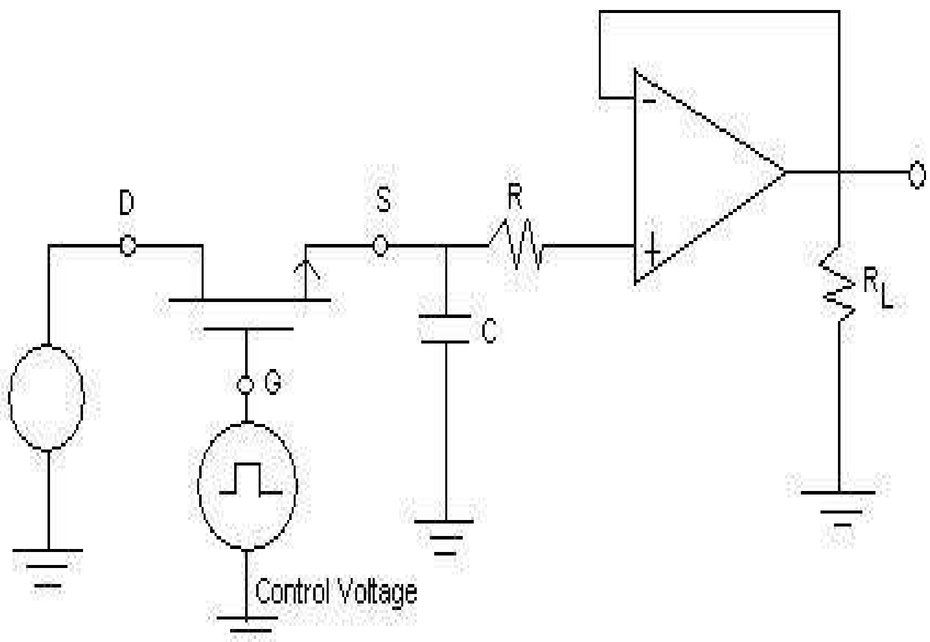
(b) What is the minimum cut off frequency of the LPF?

(c) In the receiver side, determine the minimum and maximum acceptable bandwidth of the LPF used in retrieving the analog signal?

Ans. (a) 20KHz (b) 10KHz (c) 4KHz, 6KHz.

❖ ***Sample and Hold Circuit***

Figure below shows the sample and hold circuit



The switch closes only when that particular channel is to be sampled. If the source impedance **r** is small, the capacitor voltage changes to the input voltage within the time τ that switch is closed.

The load impedance **R** is arranged to be high so that the capacitor retains the voltage level until the switch is closed again. Therefore the sample and hold circuit accepts only those values of the input which occur at the sampling times and then holds them until the next sampling time.

(2) Other Types of Analog Pulse Modulation (PWM&PPM)

One type of pulse timing modulation uses constant amplitude pulses whose width is proportional to the value of message signal at the sampling instants. This type is designated as ***pulse width modulation (PWM) or pulse duration modulation (PDM)*** is also called.

Another possibility is to keep both the amplitude and the width of the pulses constant but vary the pulse position in proportion to the value of message signal at sampling instant. This is designated as ***pulse position modulation (PPM)***.

PAM, PWM and PPM waveforms for a given message signal are shown below: -

Figure shows PWM signal generation using *natural sampling*.

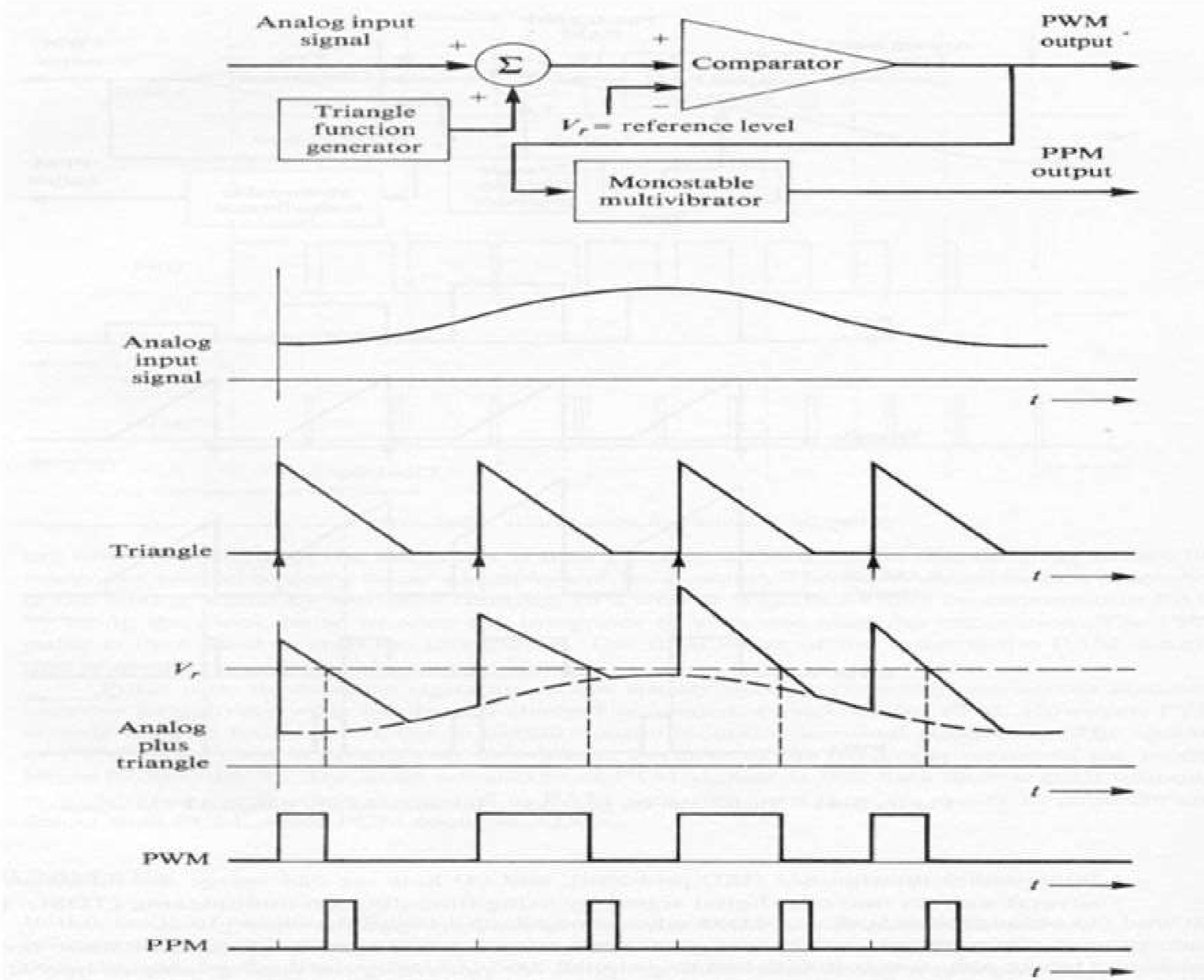


Figure 3-45 Technique for generating naturally sampled PTM signals.

Fig:Direct method of generation of PTM signals

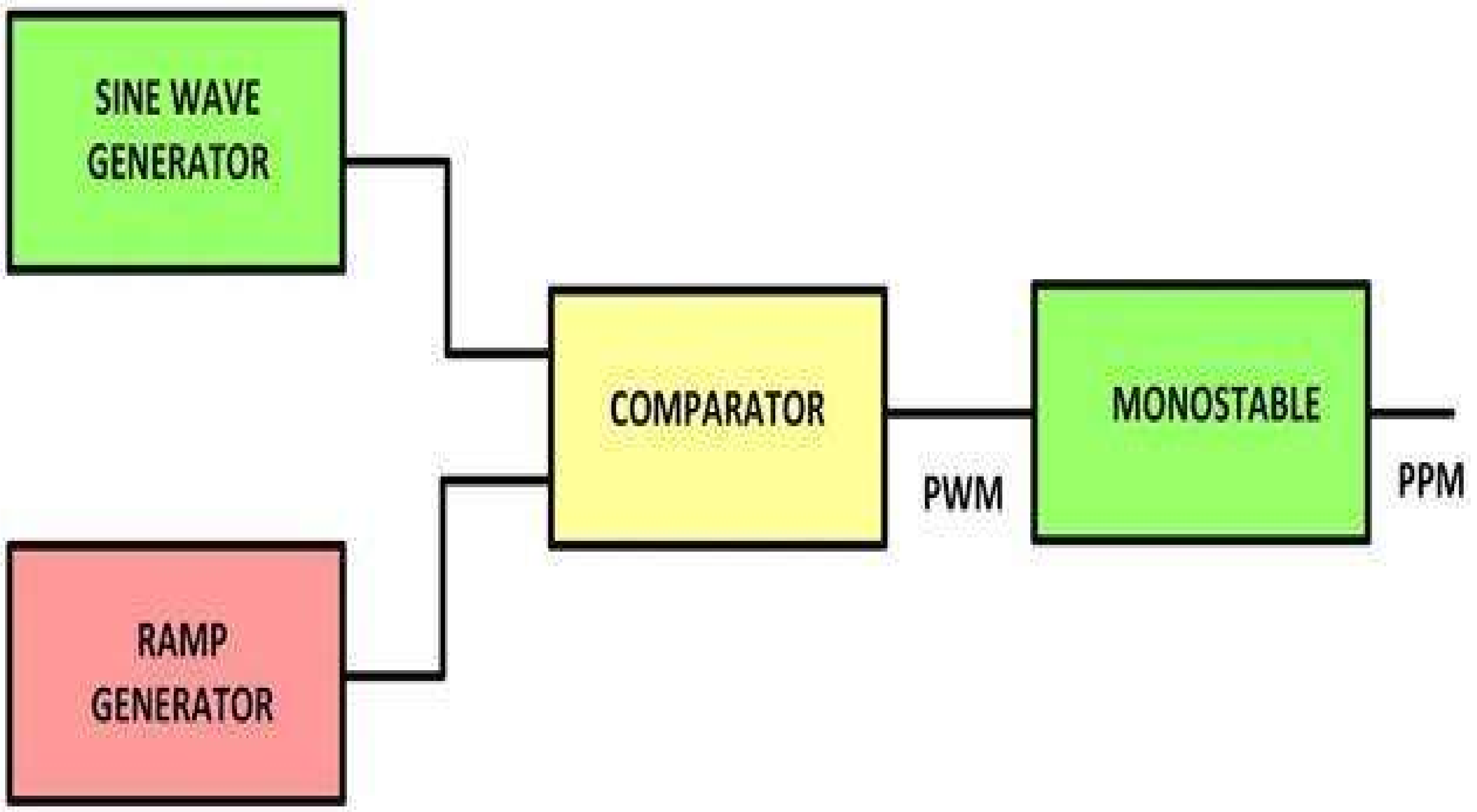
In PWM, the signal f(t) is sampled periodically at a rate fast enough to satisfy the requirements of the sampling theorem. At each sampling instant a pulse is generated with fixed amplitude and a width

that is proportional to the sample value of f(t). A minimum pulse width is assigned to the minimum value of f(t).

In PPM, these are sent as constant width, constant amplitude pulses. The minimum pulse delay is used to designate the minimum value of f(t) and the change in delay is proportional to the modulating signal. The constant of proportionality is the modulation constant.

Generation of PWM & PPM

Generation of PWM and PPM commonly employs various combinations of a sample and holed circuit, a precision ramp voltage generator and a comparator. The block diagram of a typical circuit for generation PWM and PPM is shown in figure below: -


SINE WAVE GENERATOR
RAMP GENERATOR
COMPARATOR
PWM
MONOSTABLE
PPM

The ramp generator produces a precision ramp voltage which has peak to peak amplitude slightly larger than the maximum amplitude range of the input signals. This ramp voltage is the basis for the amplitude to timing conversion and therefore must be accurately known.

The comparator is a high gain amplifier intended for two stated operation. If input signal is higher than a preset reference level, the output is held in one state (i.e. a given voltage level). Whenever the input signal level is less than the reference level, the output is held in the other state. Which output state is present, then, depends upon whether the input is above and below the threshold (reference level) of the comparator.

The voltage reference level of the comparator is adjust so that there is always an intersection with the sum of the sample and hold circuit and ramp voltage. In this system, the first crossing of the reference level indicates the clock timing and the second crossing generates the variable trailing edge.

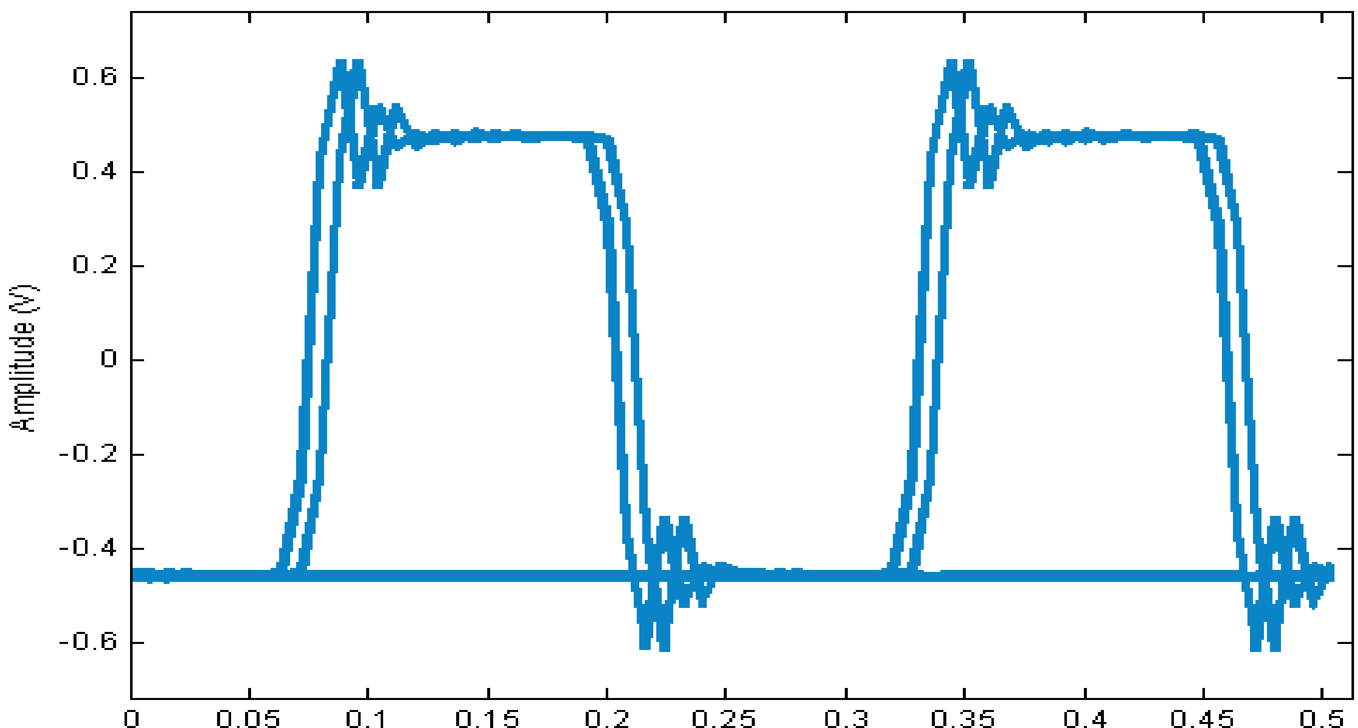
A convenient way to generate PPM is to use PWM waveform generated above and then trigger a constant width pulse generation those edge of the PWM waveform with a negative slope.

➢ Signal to noise ratio in analog pulse modulation

The performance of analog pulse modulation system in the presence of additive noise is investigated here.

- **PAM**

Noise is added in the transmission of the PAM signal as illustration in figure below.



The noise occurring between pulses adds noise power to the transmission without any increase in signal power. To avoid this, a synchronized gating circuit is used in the receiver to accept samples only when the signal is known to be present.

We shall assume that the signal and the additive noise present in the input to the PAM receiver are band limited and that the conditions of the sampling theorem are satisfied. Because the PAM receiver is linear, we can apply the signal and the noise separately measures their power, and then combine. The sampling and low pass filtering at the receiver reproduce the band limited signal and noise spectra within a

Thus $$\overline{S_0^2(t)} = K\overline{S_i^2(t)}$$

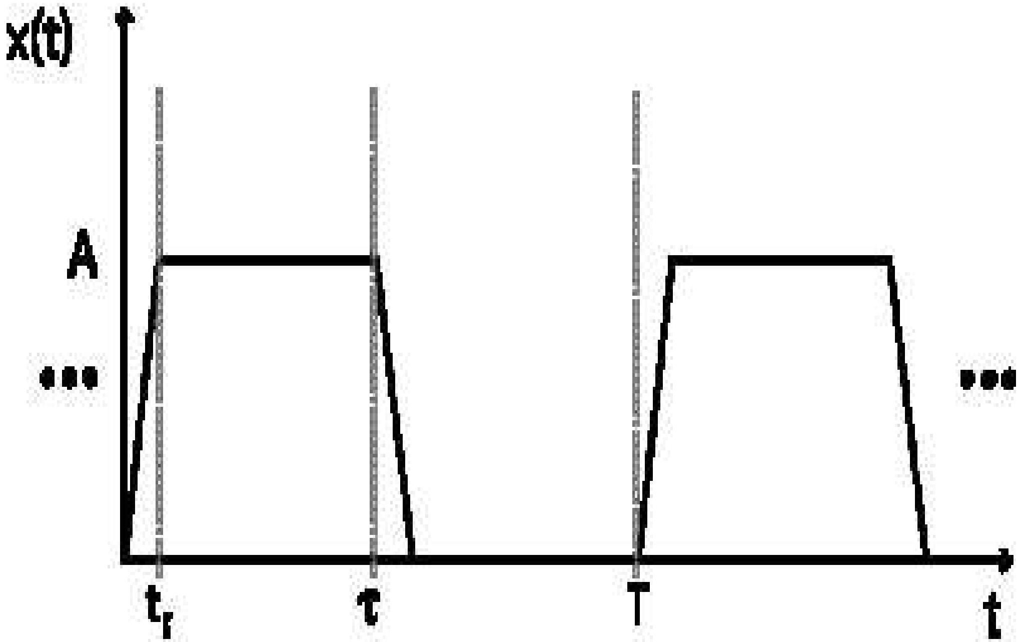
$$\overline{n_0^2(t)} = K\overline{n_i^2(t)}$$

So that

$$\frac{S_0}{N_0} = \frac{S_i}{N_i}$$

- **Pulse timing modulation**

Although the pulses to convey the information may be generated with extremely short (fast) rise time, after passing through a band limited system they have rise time which are governed by the bandwidth of the system. This rise time can be approximated by a linear ramp. As shown in figure below, so that the pulse assumes a trapezoidal shape.



x(t)
A
•••
•••
t_r
τ
T
t

The position of the trapezoidal pulse is sensitive to additive noise. If the noise voltage is assumed to vary slowly compared to the rise time of the pulse, the variation in the pulse amplitude, n, may be

represented by a shift, ε, in the pulse position as shown in above figure.

From the geometry of above figure, we have

$$\frac{\varepsilon}{t_r} = \frac{n}{A}$$

or $$\overline{\varepsilon^2} = \left(\frac{t_r}{A}\right)^2 \overline{n^2} \qquad \text{.....................(1)}$$

The output signal amplitude is proportional to the modulating signal f(t) through a modulation constant k.

$$\therefore S_0(t) = kf(t)$$

or

$$\overline{S_0^2(t)} = k^2 \overline{f^2(t)} \qquad \text{..................(2)}$$

The output noise use Eq.(1)

$$\overline{n_0^2} = \overline{\varepsilon^2} = \left(\frac{t_r}{A}\right)^2 \overline{n_i^2(t)} \qquad \text{..................(3)}$$

Also we have

$$\overline{S_i^2} = A^2\left(\frac{\tau}{T_s}\right) \qquad \ldots\ldots\ldots\ldots\ldots(4)$$

and for ideal LPF (giving a nearly linear rise time)

$$B = \frac{1}{t_r} \qquad \ldots\ldots\ldots\ldots\ldots(5)$$

Combining Eqs. (2) and (5) we have

$$\frac{S_0}{N_0} = \frac{K^2 \overline{f^2(t)}}{\tau / T} B^2 \frac{S_i}{N_i} \longrightarrow \frac{S_0}{N_0} \propto B^2 \frac{S_i}{N_i}$$

∴ the S/N improvement in a PPM is proportional to the square of the bandwidth.

Example 1

In the pulse timing modulation receiver the ratio of peak amplitude to additive r.m.s noise=10, the pulse duration=1 μsec., the guard time= 1 μsec., the minimum bandwidth=3.3 MHz, and the SNR before demodulation=10.

(a) Find SNR after demodulation.

(b) If the amplitude rang ± 1volt, calculate the signal resolution.

Solution:-

$$\frac{S_i}{N_i} = \frac{A^2(\frac{\tau}{T_s})}{\overline{n^2(t)}}$$

$$10 = \frac{100 * 10^{-6}}{T} \Rightarrow T = 10\mu\,\text{sec}.$$

$$T_{\text{mod}} = T - \tau - \tau_g = 10 - 2 = 8\mu\,\text{sec}.$$

$$\frac{S_0}{N_0} = \frac{k^2 \overline{f^2(t)} B^2}{\frac{\tau}{T}} \frac{S_i}{N_i}$$

$$= \frac{k^2 \overline{f^2(t)} (3.3 * 10^6)^2}{\frac{1}{10}} \frac{S_i}{N_i}$$

$$\because k = \frac{T_{\text{mod}}}{2V_m}$$ k=modulation constant.

$$k * 2V_m = T_{\text{mod}}$$

$$k^2 * 4V_m^2 = T_{\text{mod}}^2 \Rightarrow k^2 * 2V_m^2 = \frac{T_{\text{mod}}^2}{2}$$

for sin wave $\overline{f^2(t)} = 2V_m^2$

$$\therefore k^2 * \overline{f^2(t)} = \frac{T_{\text{mod}}^2}{2} = \frac{(8*10^{-6})^2}{2} = 32*10^{-12}\,\text{sec.}$$

$$\frac{S_0}{N_0} = 32*10^{-12}*10.89*10^{12}*100 = 348.48*10^2$$

(b) $$k = \frac{8*10^{-6}}{2} = 4*10^{-6}\,\text{sec.}$$

$$\Delta\tau = t_r\left(\frac{\overline{n^2}(t)}{A^2}\right)^{\frac{1}{2}}$$

$$B = \frac{1}{2t_r} \Rightarrow \therefore t_r = 0.15\,\mu\,\text{sec.}$$

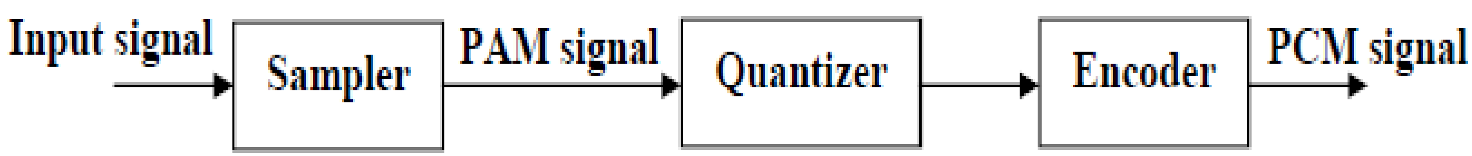
$$\therefore \Delta\tau = 0.15 * 10^{-6} * (0.01)^{\frac{1}{2}} = 0.015 * 10^{-6}\ \sec.$$

$$\therefore \Delta\tau = k * resolution$$

$$\therefore resolution = \frac{\Delta\tau}{k} = \frac{0.015 * 10^{-6}\ \sec.}{4 * 10^{-6}\ \sec./V} = 3.75V$$

Pulse code modulation (PCM)

Pulse code modulation (PCM) is the name given to the class of baseband signals obtained from the quantized PAM signals by encoding each quantized sample into a digital word. Figure below shows the steps required in PCM communication.



The source of information is sampled and quantized to one of **L**-levels, then each quantized sample is digitally encoded into a **k**-bits code word.

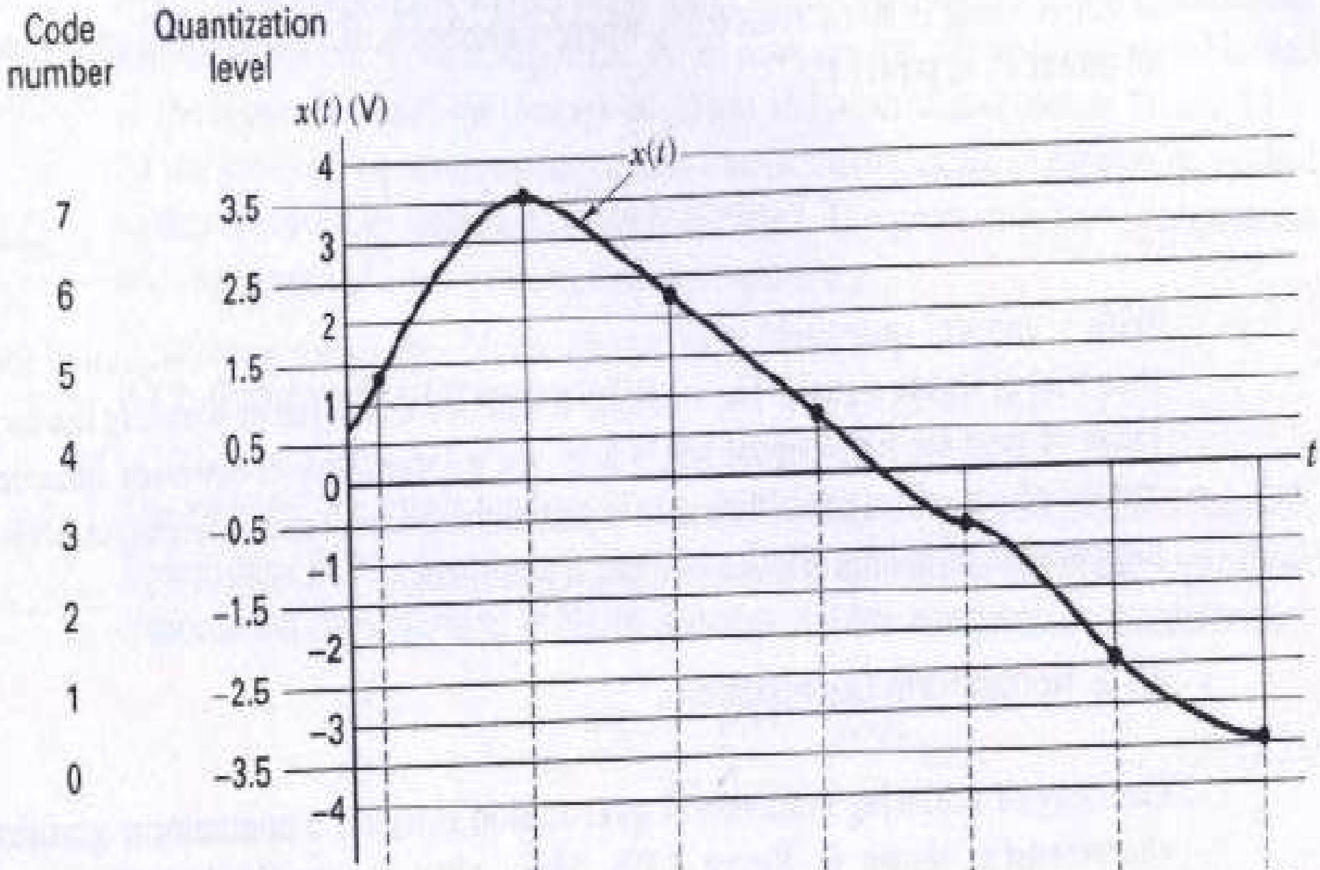
Where $k = \log_2 L$

$$L = 2^k$$

The essential features of binary PCM are shown in figure below. Assume that an analog signal, x(t), is limited in its excursions to the range (-4V to +4V). The step size between quantization levels has been set at 1V. Thus eight quantization level are employed, these located at -3.5V, -2.5V, ………., +3.5V.

The code number 0 may be assigned to the level at -3.5V; the code number 1 may be assigned to the level at -2.5V, and so on until the level at 3.5V, which is assigned the code number 7.

Each code number has its representation in binary arithmetic, ranging from 000 for code number 0 to 111 for code number 7.



Natural sample value	1.3	3.6	2.3	0.7	−0.7	−2.4	−3.4
Quantized sample value	1.5	3.5	2.5	0.5	−0.5	−2.5	−3.5
Code number	5	7	6	4	3	1	0
PCM sequence	101	111	110	100	011	001	000

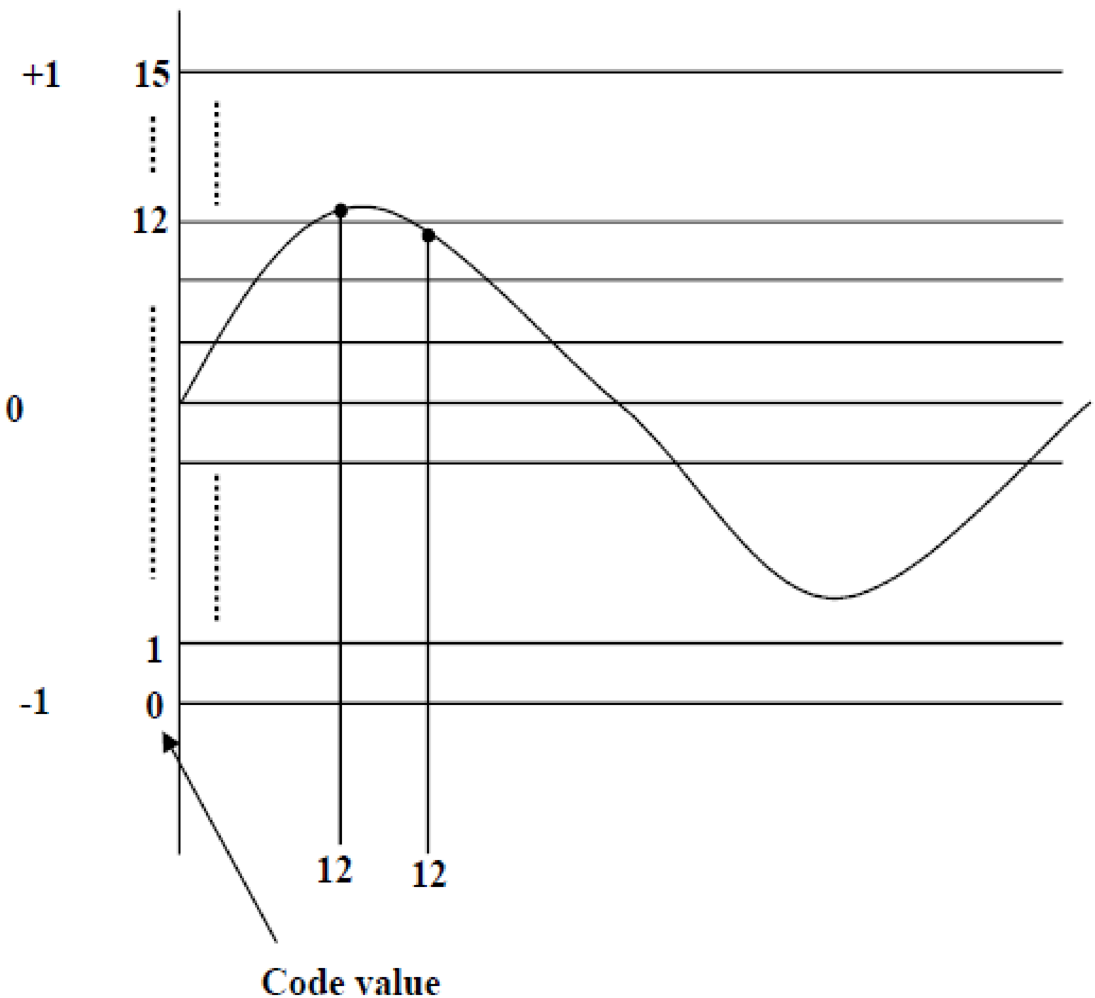
From the above figure each sample of analog signal is assigned to the quantization level closest to the value of the sample. Beneath the analog waveform, x(t), are seen four representations of x(t) as follows:- the natural sample value, the quantized sample value, the code numbers, and the PCM sequence.

- ***Quantization***

The objective of the quantization step in PCM process is to represent each sample by a fixed number of bits.

For example, if the amplitude of PAM resulting from sampling process ranges between (-1V and +1V), there can be infinite values of voltage between (-1 and +1). For instance, one value can be -0.27689V. To assign a different binary sequence to each voltage value, we would have to construct a code of infinite length. Therefore, we can take a limit number of voltage values between (-1V and +1V) to represent the original signal and these values must be discrete.

Assume that the quantization steps were in 0.1V increment, and the voltage measurement for one sample is 0,58V. That would have to be rounded off to 0.6V, the nearest discrete value. Note that there is a 0.02V error, the difference between 0.58V and 0.6V. See figure below. Take step 12 in the curve, for example, the curve is passing through a maximum and is given tow values of 12. For the first value, the actual curve is above 12 and for second value below 12. That error from the true value to the quantum value is called ***quantization distortion***. This distortion is the major source of imperfection in PCM system.


+1
15
12
0
1
-1
0
12
12
Code value

The more quantization level, the better quality the system will deliver. However, increasing the number of quantization level has two major costs:-

1) The cost of designing a system with large binary code size needed.
2) The time it takes to process this large number of quantizing steps by the coder.

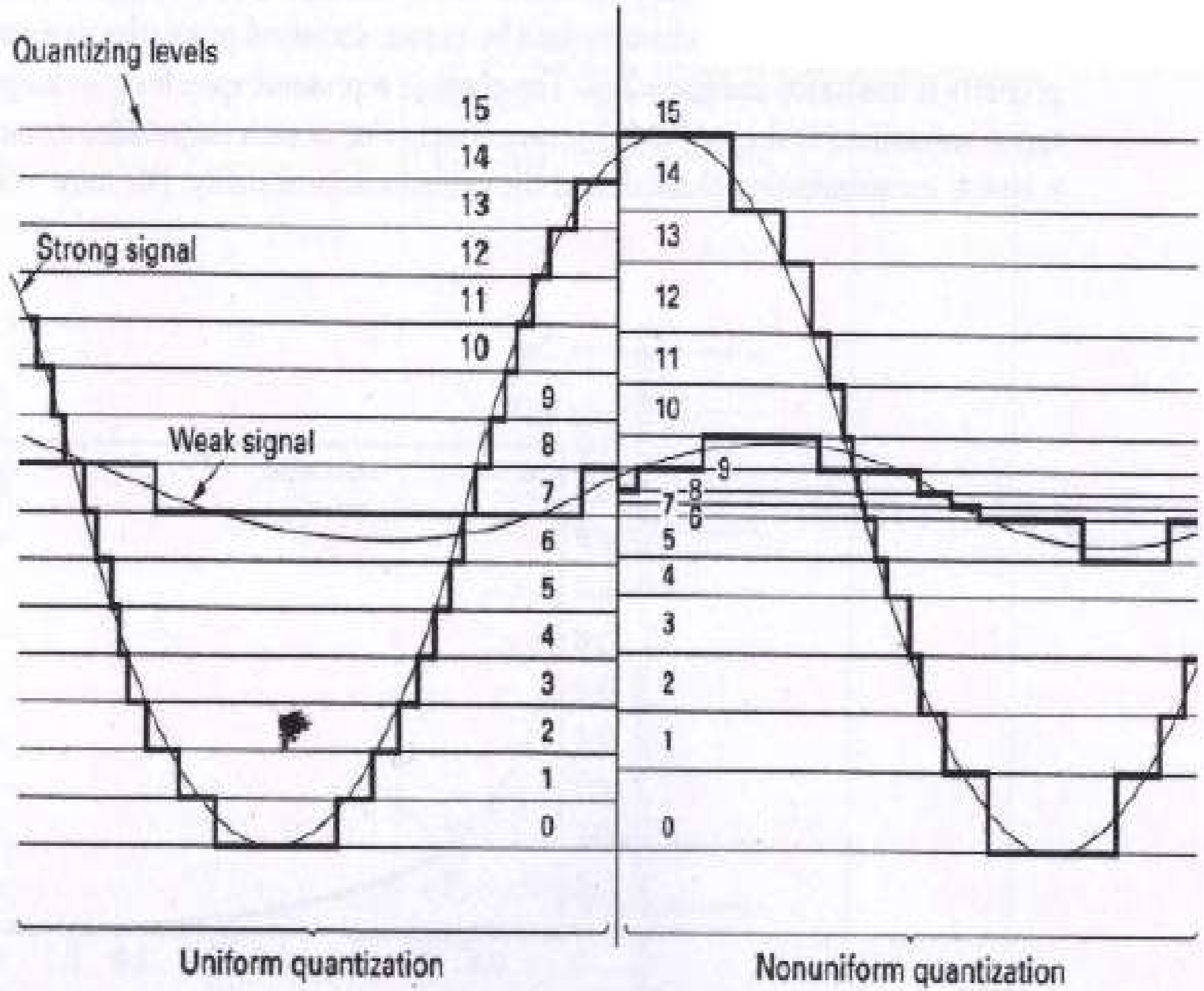
Therefore, a very large number of quantizing levels may induce unwanted delays in the system.

Uniform and Nonuniform Quantization

Form the above discussion it cam be seen that the quantization noise depends on the step size. When the steps have uniform size the quantization called as ***uniform quantization***.

For uniform quantization, the quantization noise is the same for all signal magnitudes. Therefore, with uniform quantization the signal to noise ratio (SNR) is worse for low level signals than for high level signals.

Nonuniform quantization can provide fine quantization of the weak signal and coarse quantization of the strong signal. Thus in the case of nonuniform quantization, quantization noise can be made proportional to signal size. The effect is to improve the overall SNR by reducing the noise for the predominant weak signals, at the expense of an increase in noise for the rarely occurring strong signals. Figure below compares the quantization of strong signal versus a weak signal for uniform and nonuniform quantization.



Quantizing levels
Strong signal
Weak signal
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
Uniform quantization
Nonuniform quantization

Example 2

The information in an analog waveform, with maximum frequency f_m=3 KHz, is to be transmitted over M level PCM system, where number of pulse level M=16. The quantization distortion is specified not to exceed $\pm 1\%$ of the peak-to-peak analog signal.

(a) What is the minimum number of bits/sample, or bits/PCM word that should be used in this PCM system.

(b) What is the minimum required sampling rate, and what is the resulting bit transmission rate.

(c) What is the PCM pulse or symbol transmission rate.

Solution:-

Note:- in this example we are considered with two types of levels, the number of quantization levels (L), and the 16 level of the multilevel PCM pulses (M).

(a) By using

$$L \geq \frac{1}{2p} \text{ levels}$$

$$\therefore k \geq \log_2 \frac{1}{2p} \text{ bits}$$

where L=number of quantization level, k=number of bits, and p=fraction of peak-to-peak analog voltage.

$$\therefore k \geq \log_2 \frac{1}{2*0.01} = \log_2 50 = 5.6$$

$$\therefore k = 6 \Rightarrow L = 2^k = 64$$

The number of bit/samples =k =6

(b) $f_s=2f_m=6000$ sample/second

$\therefore$ bit transmission rate $R_b \geq kf_s$

$\therefore R_b = 6*6000 = 36000$ bit/sec.

(c) since multilevel pulses are to be used with $M=2^m=16$

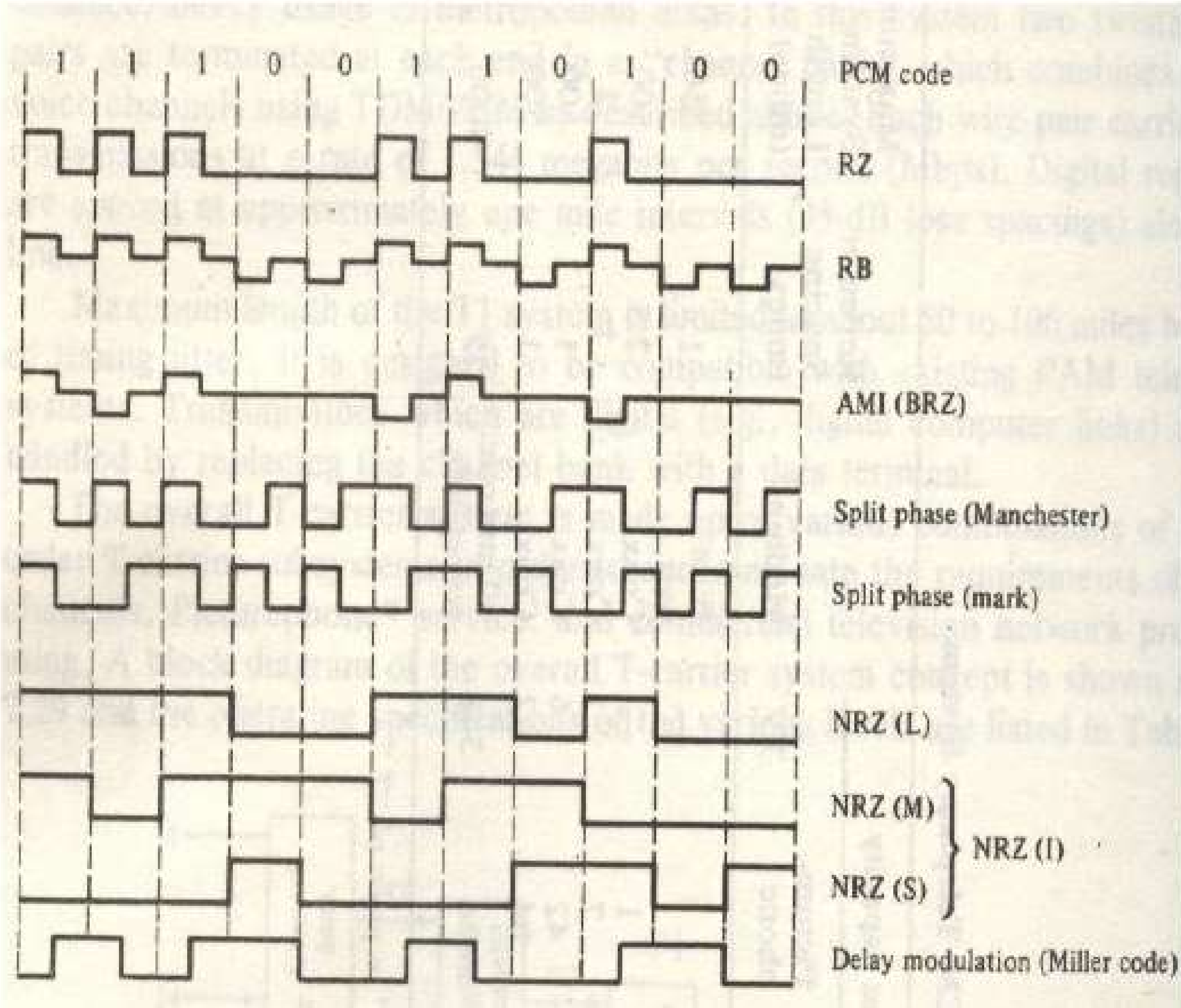
$\therefore$ m=4 bit/symbol

$\therefore$ The bit stream will be partitioned into groups of 4-bits to form a new 16-level PCM digit.

$\therefore$ Symbol transmission rate (R_s)

$$R_s = \frac{R_b}{m} = \frac{36000 bit/\sec.}{4 bit/symbol} = 9000 \frac{symbol}{\sec.}$$

Figure below some of the more commonly used PCM representation.



1) Return-to-zero (RZ) method represents a 1 by a change to the level for one-half the bit interval, after which the signal returns to the reference level for the remaining half-bit interval. A 0 is indicated in this method by no change, the signal remaining at the reference level.

2) Return-to-bias (RB) method, in this method three levels are used 0, 1, and a bias level. The bias level may be chosen either below or between the other two levels. The waveform returns to the bias level during the last half of each bit interval.

3) Alternate Mark Inversion (AMI), in this method the first binary one is represented by +1, the second by -1, the third by +1, etc. The AMI representation is easily derived from an RZ binary code (and vice versa) by alternately inverting the 1's. It has zero average value and is widely used in telephone PCM systems. This is also referred to as a bipolar return-to-zero (BRZ) representation.

4) Split-phase representations eliminate the variation in average value using symmetry. In the Manchester split-phase method, a 1 is represented by a 1 level during the first half-bit interval, then shifted to the 0 level for the latter half-bit interval; a 0 is indicated by the reverse representation. In the split-phase (mark) method, a similar symmetric representation is used expect that a phase reversal relative to the previous phase indicates a 1 (i.e. mark) and no change in phase is used to indicate a 0.

5) Nonreturn-to-zero (NRZ) representations reduce the bandwidth needed to send PCM code. In NRZ (L) representations a bit pulse remains in one of its two levels for the entire bit interval. In NRZ (M) method a level change is used to indicate a mark (i.e a 1) and no level change for a 0; the NRZ (S) method uses the same scheme except that a level change is used to indicate a space (i.e. a 0). Both of these are examples of the more general classification NRZ (I) in which a level change (inversion) is used to indicate one kind of binary digit and no level change indicates the other digit. Note that use of split-phase and NRZ representations require some added receiver complexity to determine the clock frequency.

6) Delay modulation (Miller code), in this method a 1 is represented by a signal transition at the midpoint of a bit interval. A 0 is represented by no transition unless it is followed by another 0, in which case the signal transition occurs at the end of the bit interval. In this method, a succession of 1's and a succession of 0's each are represented by a square wave at the bit rate, but one is delayed a half-bit interval from the other.

➢ ***Noise consideration in PCM system***

The performance of a PCM system is influenced by two major sources of noise.

1) ***Channel noise***, which is introduced anywhere between the transmitter output and the receiver input, channel noise is always present, once the equipment is switched on.

2) ***Quantization noise***, which is introduced in the transmitter and is carried all the way along to the receiver output.

Quantization Noise

The peak signal to r.m.s noise power ratio is given by

$$\frac{S_0}{N_0} = 3L^2$$

$$\frac{S_0}{N_0})_{dB} = 4.8 + 20\log_{10} L$$

where L=number of quantizer level.

S_0= peak signal power.

N_0= r.m.s noise power.

Increasing L increases the number of code pulses and hence the bandwidth. We can thus relate SNR to bandwidth. This is easily done by noting that

$$L = n^m$$

where m=the number pulses in code group.

n=the number of code levels.

$$\therefore \frac{S_0}{N_0} = 3n^{2m}$$

and
$$\frac{S_0}{N_0})_{dB} = 4.8 + 20m\log_{10} n$$

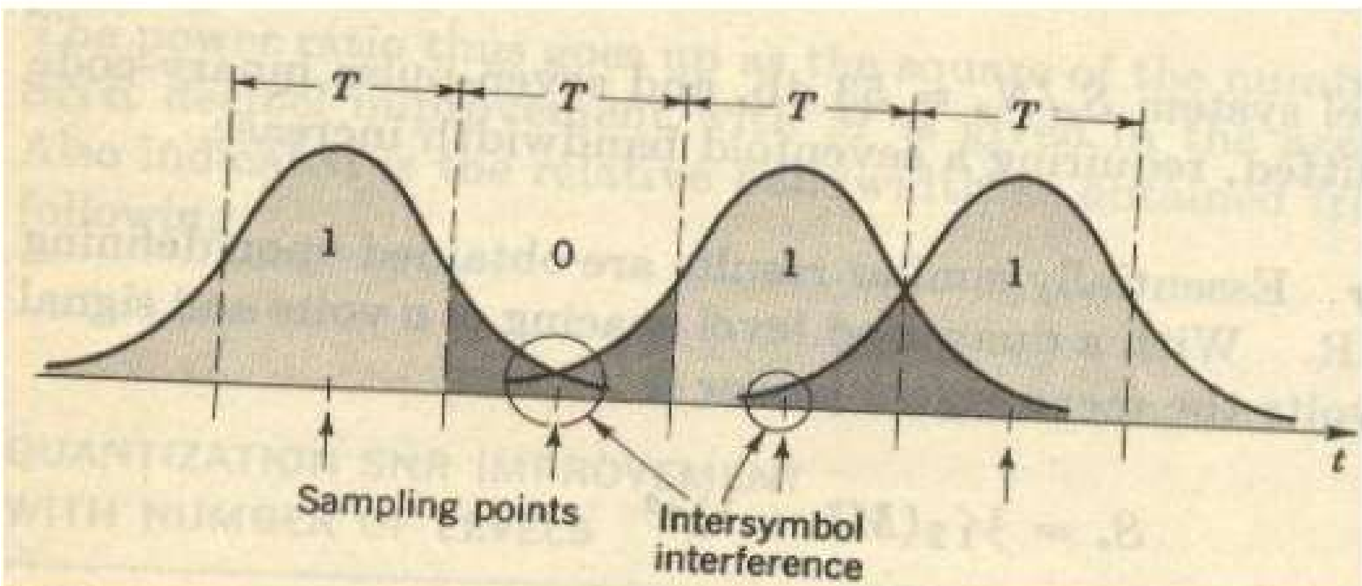
In particular, for binary code n=2.

$$\frac{S_0}{N_0})_{dB} = 4.8 + 6m$$

Since the bandwidth is proportional to m, the output SNR increases exponentially with bandwidth.

Intersymbol interference (ISI) and pulse shaping to reduce ISI

Consider the sequence of pulses shown in figure below. Although these are shown as binary pulses, they could well be pulses of identical shape, but of arbitrary height. They are shown recurring at T_s second, where T_s is the sampling interval.

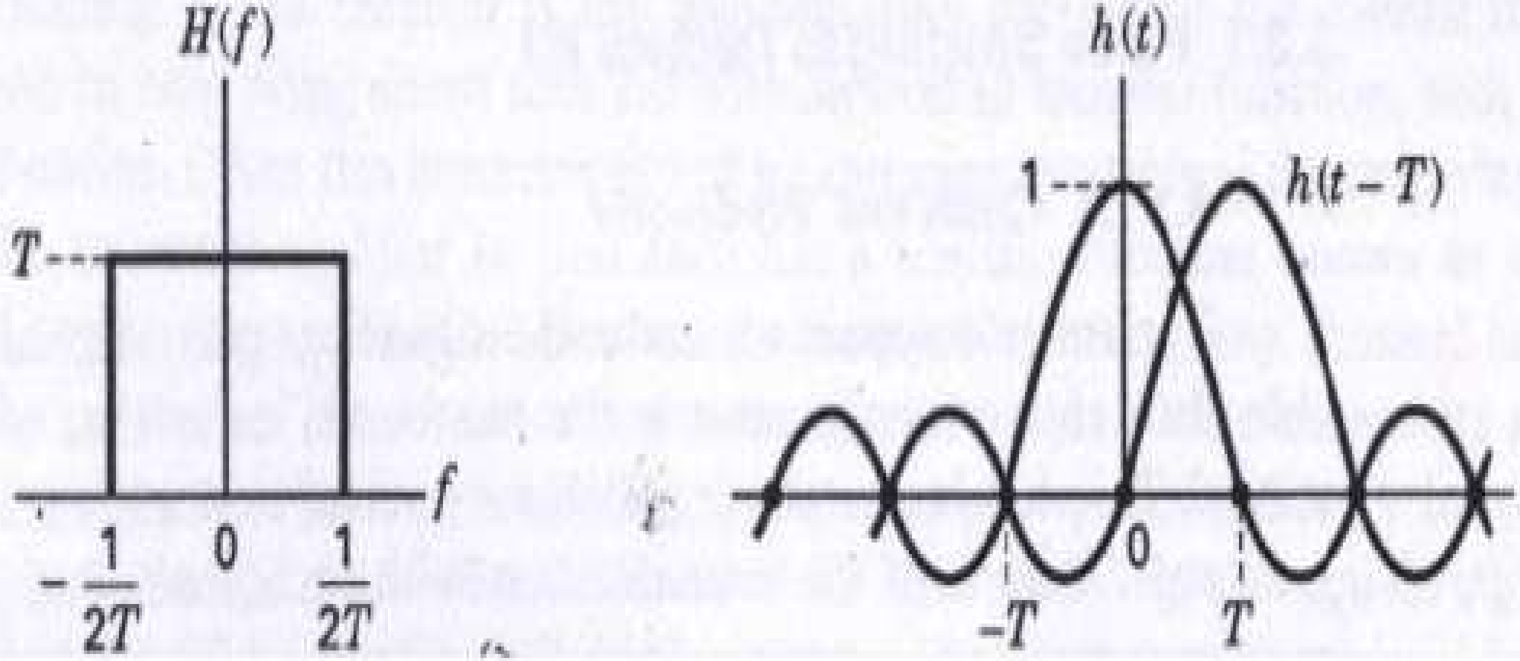


System filtering causes these pulses to spread out as they traverse the system. And they overlap into adjacent time slot as shown. At the receiver the original pulse message may be derived by sampling at the center of each time slot, and then basing a decision on the amplitude of the signal measured at that point.

The signal overlapped into adjacent time slots may, if too strong, results in an erroneous decision. Thus, as an example, in the case of above figure the 0 transmitted may appear as a 1 if tails of adjacent pulses add up to too high a value.

This phenomenon of pulse overlap and the resultant difficulty with receiver decision is termed intersymbol interference (ISI).

Nyquist investigated the problem of specifying a received pulse shape so that no ISI occurs at the detector. He showed that the theoretical minimum system bandwidth needed to detect R_s symbol/second without ISI is $R_s/2$ Hz. This occurs when the system transfer function, H(f), is made rectangular as shown in figure below.



H(f)
T
f
−1/2T
0
1/2T
h(t)
1
h(t − T)
t
0
−T
T

Note that, when H(f) is such an ideal filter with bandwidth 1/2T, its impulse response

$$h(t) = \sin c(\frac{t}{T})$$

Therefore, the bandwidth required to detect 1/T symbol/sec. is 1/2T Hz.

The Nyquist pulse shape is not physically realizable since it dictates a rectangular bandwidth characteristic. Also, with such a characteristic, the detection process would be very sensitive to small timing error.

One frequently used system transfer function H(f) is called the ***raised cosine*** filter. It can be expressed as

$$H(f)=\begin{cases}1 & 0\le |f|\le \dfrac{1-r}{2T}\\ \dfrac{1}{2}[1+\cos(\dfrac{\pi[(2T|f|)-1+r]}{2r})] & \dfrac{1-r}{2T}<|f|\le \dfrac{1+r}{2T}\\ 0 & |f|>\dfrac{1+r}{2T}\end{cases}$$

where $r=\dfrac{\omega-\omega_0}{\omega_0}$ is called roll-off factor.

ω = absolute bandwidth.

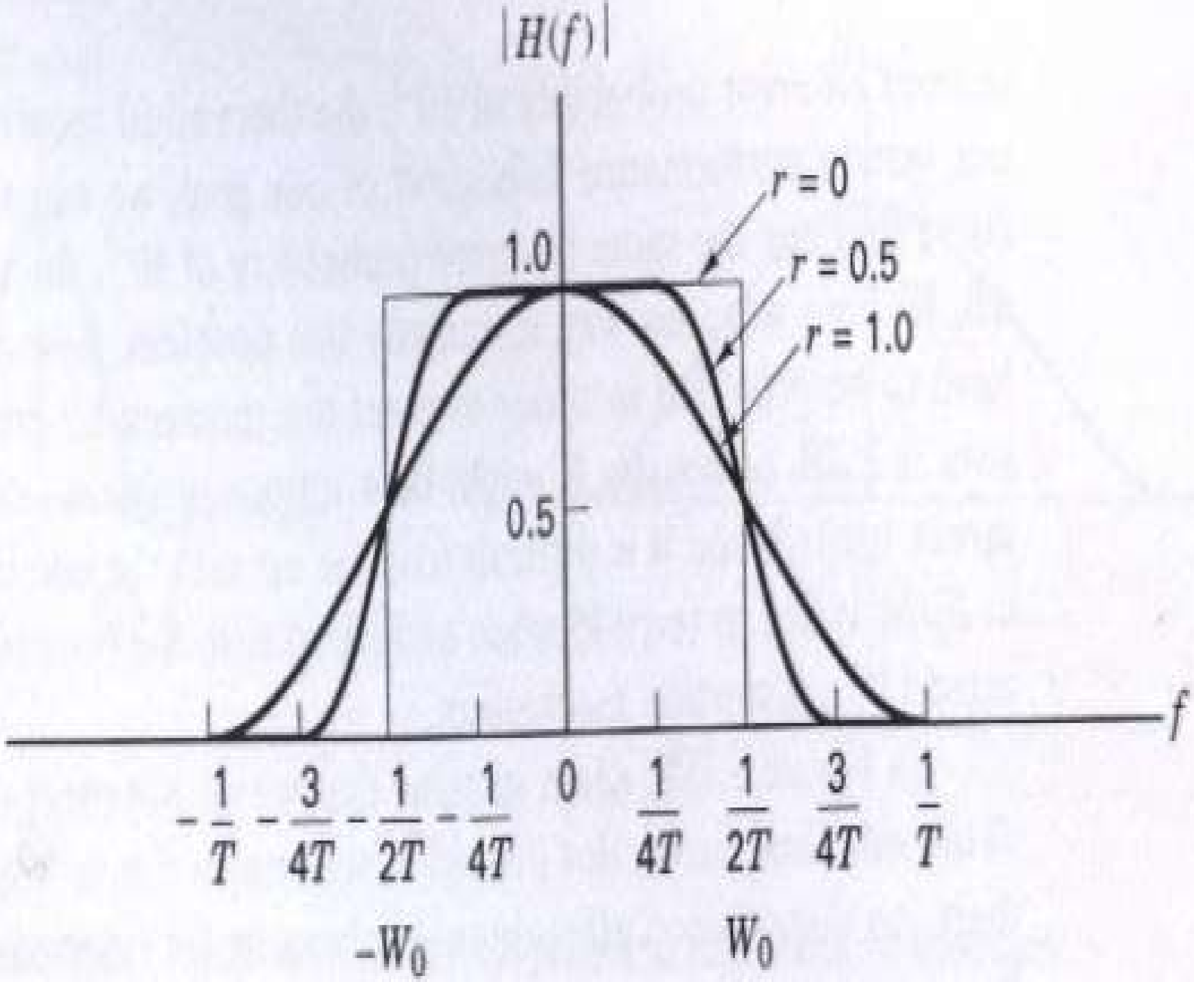
$\omega_0=\dfrac{1}{2T}$ =minimum Nyquist bandwidth.

Therefore, the bandwidth may be given by:-

$$B.W = \frac{(1+r)}{2T}$$

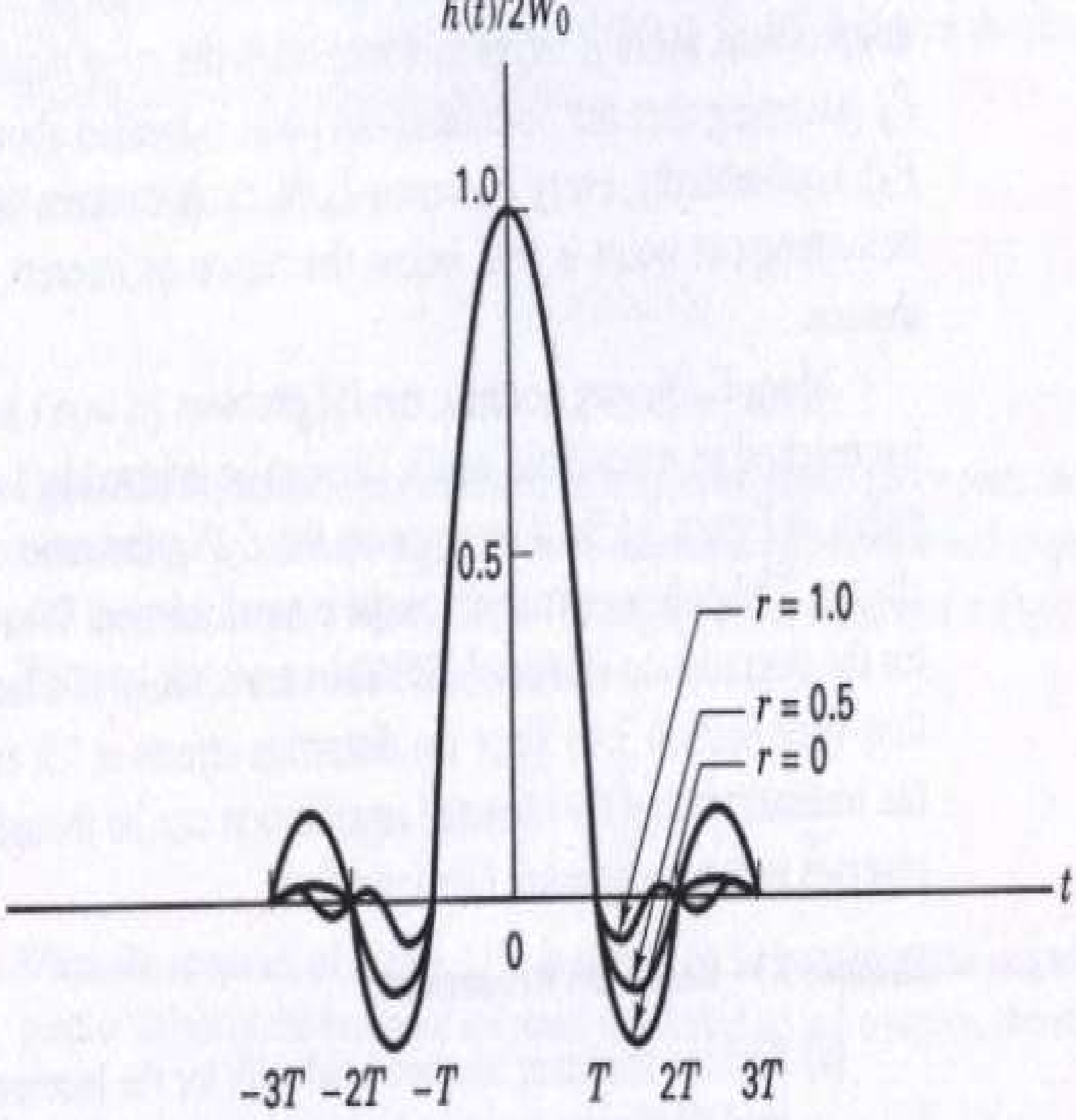
❖ ***Note:-***

- If $r = 0 \rightarrow \omega = \omega_0$, is the Nyquist minimum bandwidth.
- If $r = 0 \rightarrow \omega = 2\omega_0 = \frac{1}{T}$.



|H(f)|
r = 0
r = 0.5
r = 1.0
1.0
0.5
f
$-\frac{1}{T}$
$-\frac{3}{4T}$
$-\frac{1}{2T}$
$-\frac{1}{4T}$
0
$\frac{1}{4T}$
$\frac{1}{2T}$
$\frac{3}{4T}$
$\frac{1}{T}$
$-W_0$
W_0

(a)



h(t)/2W0
1.0
0.5
r = 1.0
r = 0.5
r = 0
t
0
−3T −2T −T
T 2T 3T

- ***Channel Capacity***

The maximum rate of transmission was found by Shannon to be given by:-

$$C = W \log_2 (1 + \frac{S}{N})$$

Shannon's maximum capacity expression provides an upper bound on the rate at which one can communicate over a channel of bandwidth W, and signal to noise ratio SNR ($\frac{S}{N}$).

Delta modulation (DM)

In the basic form, DM provides a stair case approximation to the over sampled version of the message signal, as shown in figure below, the difference between the input and the approximation is quantized into only two levels, namely, $\pm\Delta$, corresponding to positive and negative differences, respectively. Thus, if the approximation below the signal at any sampling epoch, it is increased by Δ, on the other hand, the approximation lies above the signal, it is diminished by Δ.

Denoting the input signal as m(t), and its stair case approximation as $m_q(t)$, the basic principle of DM may be formalized in the following set of discrete time relations:-

$$e(nT_s) = m(nT_s) - m_q(nT_s - T_s) \qquad \text{............(a)}$$

$$e_q(nT_s) = \Delta Sgn[e(nT_s)] \qquad \text{............(b)}$$

$$m_q(nT_s) = m_q(nT_s - T_s) + e_q(nT_s) \qquad \text{............(c)}$$

where

T_s=sampling time, $e(nT_s)$=error signal representing difference between $m(nT_s)$ and the latest approximation to it $m(nT_s)-m_q(nT_s-T_s)$, $e_q(nT_s)$= the quantized version of $e(nT_s)$.

Finally, the quantizer output $e_q(nT_s)$ coded to produce the DM signal.

DM modulator and demodulator are shown below.

The comparator computes the difference between two inputs. The quantizer consists of hard limiter with an input/output relation that is scaled version of the signum function. The accumulator increments the approximation by a step Δ in positive or negative direction, depending on the algebraic sign of the error signal $e(nT_s)$.

Demodulation is subjected to two types of error:-

(1) Slop over load direction.

(2) Granular noise.

Equation (c) may be observed as a digital equivalent of integration in the sense that it represents the accumulation of positive and negative increments of magnitude Δ, also, denoting the quantization error by $\boldsymbol{q(nT_s)}$, as shown by

$$m_q(nT_s) = m(nT_s) + q(nT_s)$$

From equation (a) may be observed

$$e(nT_s) = m(nT_s) - m(nT_s - T_s) - q(nT_s - T_s)$$

Thus, except for the quantization error $q(nT_s - T_s)$, the quantizer input is a first backward difference of the input signal, which may be viewed as a digital approximation to the derivative of the input signal or, equivalently, as the inverse of the digital integration process.

Consider the maximum slope of input m(t), it is clear that in order for the sequence of samples $m_q(nT_s)$ to increase as fast as the input sequence of samples $m(nT_s)$ in a region of maximum slope of m(t), the condition

$$\frac{\Delta}{T_s} \geq \max\left|\frac{dm(t)}{dt}\right|$$

must be satisfied.

The step size (Δ) is too small for the staircase approximation $m_q(t)$ to follow a steep segment of the input signal m(t), with the result that $m_q(t)$ falls behind m(t), as shown in figure below. This condition called ***slope- overload***, and the resulting quantization noise called ***slope- overload distortion***.

Delta modulation using a fixed step size (Δ) is often referred to as a ***linear delta modulator***.

In contrast to slop-overload distortion, ***granular noise*** occurs when the step size (Δ) is too large relative to the local slope characteristics of the input waveforms m(t), there by causing the staircase approximation $m_q(t)$ to hunt around a relatively flat segment of the input waveform. This noise also illustrated in above figure.

Example 5

A Delta modulator is used to encode speech signal band-limited to 3KHz with sampling frequency 100 KHz. For ± 1 volt peak signal voltage, find

(a) Minimum step size to avoid slope overloading.

(b) Signal to quantization noise ratio if speech is assumed to have nonuniform probability density function (PDF).

Solution:-

For DM system, if input signal $f(t) = b\cos\omega_m t$.

(a) $\therefore \left|\frac{df(t)}{dt}\right|_{\max} = b2\pi f_m$

if step size used in DM system=a

$$\therefore f_s \geq \frac{2\pi f_m}{a/b}$$

$$\therefore a \geq \frac{2\pi f_m b}{f_s}$$

$a \geq \frac{2\pi * 3 * 10^3 * 1}{100 * 10^3} = 0.188V$ minimum step size.

(b) $$\frac{S}{N} = \frac{3}{8\pi^2}\left(\frac{f_s}{f_m}\right)^3$$

$$= \frac{3}{8\pi^2}\left(\frac{100}{3}\right)^3$$